ABSTRACT

Lorraine L. Richards: Evidence-as-a-Service: State Recordkeeping in the Cloud
(Under the Direction of Christopher A. Lee)

The White House has engaged in recent years in efforts to ensure greater citizen access to government information and greater efficiency and effectiveness in managing that information. The Open Data policy and recent directives requiring that federal agencies create capacity to share scientific data have fallen on the heels of the Federal Government’s “Cloud First” policy, an initiative requiring Federal agencies to consider using cloud computing before making IT investments.

Still, much of the information accessed by the public resides in the hands of state and local records creators. Thus, this exploratory study sought to examine how cloud computing actually affects public information recordkeeping stewards. Specifically, it investigated whether recordkeeping stewards’ concerns about cloud computing risks are similar to published risks in newly implemented cloud computing environments, it examined their perceptions of how cross-occupational relationships affect their ability to perform recordkeeping responsibilities in the Cloud, and it compared how recordkeeping roles and responsibilities are distributed within their organizations. The distribution was compared to published reports of recordkeeping roles and responsibilities in archives and records management journals published over the past 42 years.
The study used an interpretive, constant comparative approach to data collection and an analytical framework from Structuration Theory. Findings were drawn from 29 interviews and their associated transcripts and from 682 published articles from six archives and records management journals dating from 1970 onwards.

It was found that the actual work environments reported by interview participants most resembled the recordkeeping environments published by archival continuum theorists. In addition, records managers reported greater worry about status and a lack of clearly demarcated lines of responsibility in their work than did the archivists. Records managers also reported less impact from the new technology as physical artifact than from political and inter-occupational power adjustments that altered their status after the cloud implementations. It was also found that current cloud computing environments exhibit a variety of disincentives for accurate and complete recordkeeping, some of which are primarily due to political changes and others from the distributed nature of information storage in the Cloud.
To Don Robert Richards, Sr. (1922-2000) and Donna Lorraine Richards (1924-1988)
ACKNOWLEDGEMENTS

So many people have helped me in so many ways along my path. It isn’t possible to thank them all here. Chances are good they know who they are, though, and will forgive any omissions here because they also know how absent-minded I can sometimes be. Obviously, I owe a tremendous debt to my advisor Cal (Christopher) Lee. He spent a great deal of time discussing and critiquing ideas, urging me onward, reading a few hideous first drafts, and pushing me to think (and therefore) write more clearly. I also owe a huge debt to my co-advisor Helen Tibbo, who has been a constant professional role model. I have benefited so much from her unrelenting honesty, her incisive and extensive knowledge of the profession, and her friendly help in every aspect of my degree program when I’ve needed it. I also want to thank the other members of my comprehensive exam committee and my dissertation committee, who have each provided helpful comments and critiques, and have offered professional insights along the way – Paul Jones, Richard Marciano, and Jeffrey Pomerantz from the School of Information and Library Science (SILS) at the University of North Carolina, Chapel Hill (UNC-CH) and Theresa A. Pardo from the Center for Technology and Government (CTG) at the University at Albany, the State University of New York (SUNY).

I want to thank the interview respondents who took the time to discuss their workplace environments, concerns, and work objectives. I hope you keep in mind that this paper is not attempting to report “objective truths” about your organizations but rather, to highlight the perceptions which you have discussed so openly.
A large number of friends and colleagues have also provided help to me while writing this. I want to thank Amber Cushing and Sarah Ramdeen for some very early comments and critiques that helped me fine tune my ideas and better present them. I also want to thank the past and current members of the Curation and Archives Research Group at SILS. I received a number of excellent comments and questions from the members of this group during the various presentations and discussions we had. In addition, I want to extend a huge thank you to the SILS administrative folks who have provided excellent services along the way – Lara Bailey, Tammy Cox, Kaitlyn Murphy, Susan Sylvester, Shaundria Williams, and others too numerous to mention! And of course, thank you to Barbara Wildemuth and Gary Marchionini, who taught my first class in the Ph.D. program at SILS and who now perform so many administrative and academic duties to help students succeed that it boggles my mind.

I want to thank two very good friends who have provided moral support, intellectual aid, and good humor along the way – Jewel Ward and Carolyn Hank. I couldn’t have done it without your advice, your jokes, and your willingness to listen to whining. Carolyn, I decided finally to write my own acknowledgments instead of just using yours. Your parents are undoubtedly great, but if I used your acknowledgments *you* wouldn’t have been in them! Jewel, thank you for keeping me going when things seemed dark and thank you for rushing to Hillsborough to be my witness on the morning Garnet and I decided to just go get married.

I want to thank my very patient, very loving (and did I mention very patient?) family – Garnet Bornn, Jennifer Ryden, Christopher Nayar, and Jay Eakin. They put up with the days I could do nothing but write and the days that I couldn’t write to defend myself. They offered support, made sarcastic comments when necessary (and funny), and have always
made not just my work, but my entire life, brighter for their being in it. Garnet - thank you for coming back into my life. You are my best friend and my true love. Luna, Chris, and Jay – I love you and am so proud of you! Thank you for putting up with me!

And of course, thank you Bootsie, Jimoa, Hobo, Saber, Matches, and Butters who gave me lots of furry love these past years.

Finally, I want to thank my dear, departed friend and boss Joe Fuhrig from my Golden Gate University days. I haven’t followed directly in your path but I have learned so many lessons from you about the ways the market can work when founded on ethical behavior, about the nature of dedication and the pursuit of knowledge, and about sticking to your fundamental beliefs even when no one else accepts them. You never gave up and you never gave in. I miss you, and think of you often.
TABLE OF CONTENTS

1. INTRODUCTION ........................................................................................................ 2
   1.1. Background ........................................................................................................... 2
       1.1.1. Why cloud computing? ................................................................................ 8
       1.1.2. Why state government? .............................................................................. 12
   1.2. Research Questions and Methods ....................................................................... 15

2. LITERATURE REVIEW ........................................................................................... 18
   2.1. The Work Environment ...................................................................................... 18
       2.1.1. Work in Complex Sociotechnical Systems .............................................. 18
       2.1.2. The Structuration of a Social System ....................................................... 22
       2.1.3. Changing Patterns of Work ...................................................................... 25
           2.1.3.1. Increasing Collaboration in State Government ................................. 25
           2.1.3.2. Accountability Concerns in Society and at Work ............................ 31
           2.1.3.3. Accountability in the ARM Literature ............................................. 37
   2.2. The Workers ...................................................................................................... 43
       2.2.1. ARM Occupational Identity in the Pre-Computerization Era ................... 43
       2.2.2. Occupational Identity of ARM Workers in the Computerized Era .......... 48
       2.2.3. Current Thoughts on ARM Identity ......................................................... 57
       2.2.4. Recordkeeping Roles and Responsibilities ............................................... 59
           2.2.4.1. ISO 15489-1 ...................................................................................... 62
           2.2.4.2. DoD 5015.02 ..................................................................................... 64
2.3. The Technology .......................................................................................................................... 65
  2.3.1. Cloud Computing as a New and Emerging Technology ..................................................... 65
  2.3.2. Defining Cloud Computing .................................................................................................. 69
    2.3.2.1. Cloud as Hardware, Architecture, or Market ............................................................... 69
      2.3.2.1.1. Delivery Models ....................................................................................................... 71
      2.3.2.1.2. Deployment Models .............................................................................................. 73
    2.3.2.2. Cloud as Structure ........................................................................................................ 74
    2.3.2.3. Comparing Cloud to Earlier Forms of Computing .................................................... 76
      2.3.2.3.1. The Evolution of Cloud Computing ........................................................................ 76
      2.3.2.3.2. Grid, Supercomputers, Cluster, Web 2.0 and P2P .................................................. 79
      2.3.2.3.3. Comparing Cloud and Its Relatives ........................................................................ 85
    2.3.2.1. Handling Recordkeeping Risks in the Cloud ................................................................. 88

3. METHODS ......................................................................................................................................... 98
  3.1. Methodological Strategy .......................................................................................................... 100
  3.2. Rationale for Case Selection .................................................................................................... 102
  3.3. Selection of Cases .................................................................................................................... 103
  3.4. Post-Interview Process ............................................................................................................. 112
  3.5. Unit of Analysis ........................................................................................................................ 114
  3.6. Data Sources .............................................................................................................................. 115
  3.7. Documentary Analysis of Site-Specific Materials ..................................................................... 116
  3.8. Content Analysis of Published ARM Literature ..................................................................... 116
  3.9. Literature Selection .................................................................................................................. 117

4. RESULTS ........................................................................................................................................ 122
  4.1. Case 1: Statewide Email and Communications Cloud ............................................................ 125
    4.1.1. Introduction ......................................................................................................................... 125
4.1.8. Synopsis of Case 1 Findings ............................................................................. 179

4.2. Case 2: Multi-Jurisdictional Cloud Implementation ............................................. 181

4.2.1. Introduction ....................................................................................................... 181

4.2.1.1. The Decision to Move to the Cloud ........................................................... 184

4.2.2. Recordkeeping Stewards ................................................................................... 188

4.2.2.1. United States Centers for Disease Control and Prevention (CDC) ............ 189

4.2.2.2. State Departments of Health ....................................................................... 190

4.2.2.3. Association of State and Territorial Health Officials (ASTHO) ............ 192

4.2.2.4. National Association of City and County Health Officials (NACCHO) ... 193

4.2.2.5. The International Society for Disease Surveillance (ISDS) ............... 193

4.2.2.6. Local Pharmacies ....................................................................................... 193

4.2.2.7. States and Local Jurisdictions .................................................................... 194

4.2.2.8. Department of Defense (DoD); Veterans’ Administration (VA) .......... 195

4.2.2.9. National Laboratories ............................................................................... 195

4.2.2.10. Research and Data Management Organizations .................................... 195

4.2.3. Legal Environment Affecting Stewards ........................................................ 196

4.2.4. Requirements and Actions .......................................................................... 198

4.2.4.1. Defining the Recordkeeping Requirements ............................................. 213

4.2.4.2. Acting on Requirements ........................................................................... 214

4.2.4.3. Vendor Selection .................................................................................... 216

4.2.5. Stewards’ Perceptions .................................................................................... 216

4.2.5.1. Perceptions of “Records” ........................................................................ 216

4.2.5.2. Perceptions of “Cloud Computing” ......................................................... 218

4.2.5.3. Perceptions of Roles and Responsibilities .............................................. 219

4.2.6. Concerns and Perceptions of Risk ................................................................. 220
4.2.7. Interactions between Recordkeeping Stewards ................................................ 222
4.2.7.1. Perceptions of Working (Together) in the Cloud ........................................ 222
4.2.7.2. Perceptions of Changes Brought by the Cloud ........................................... 225
4.2.8. Synopsis of Case 2 Findings ........................................................................... 227
4.3. Case 3: Agency-Specific Cloud Implementation.................................................. 228
4.3.1. Introduction ..................................................................................................... 228
4.3.1.1. The Decision to Move to the Cloud ........................................................... 230
4.3.2. Recordkeeping Stewards .................................................................................. 232
4.3.2.1. Office of Knowledge & Information and Data Services (KIDS) ............... 232
4.3.2.2. Records Officers ....................................................................................... 232
4.3.2.3. Kentucky Commonwealth Office of Technology (COT) ........................... 233
4.3.2.4. Kentucky State Archives and Records Commission .................................. 233
4.3.2.5. Kentucky Department for Libraries and Archives (KDLA) ....................... 234
4.3.2.6. Public School System Archives & Records Units ...................................... 234
4.3.3. Legal Environment Affecting Stewards ........................................................ 235
4.3.4. Requirements and Actions ............................................................................. 236
4.3.4.1. Defining the Recordkeeping Requirements ................................................ 236
4.3.4.1.1. Recordkeeping Requirements as Seen by COT ................................... 236
4.3.4.2. Recordkeeping Requirements as Seen by the Department of Education ... 237
4.3.4.3. Recordkeeping Requirements as Seen by Local Archivists/Records Managers ................................................................. 237
4.3.5. Stewards’ Perceptions .................................................................................... 238
4.3.5.1. Perceptions of “Records” ........................................................................... 238
4.3.5.2. Perceptions of “Cloud Computing”......................................................... 239
4.3.6. Concerns and Perceptions of Risk ................................................................. 240
4.3.7. Interaction between Recordkeeping Stewards .................................................. 240
4.3.7.1. Perceptions of Working (Together) in the Cloud ....................................... 240
4.3.7.2. Perceptions of Changes Brought by the Cloud ........................................... 248
4.3.8. Synopsis of Case 3 Findings ............................................................................. 251
4.4. The ARM Literature ............................................................................................. 252
4.4.1. The ARM Literature on Recordkeeping Roles and Responsibilities ............ 252
4.4.2. The ARM Literature on Cloud Computing ....................................................... 262
4.5. Analysis and Discussion ....................................................................................... 264
4.5.1. Introduction ....................................................................................................... 264
4.5.2. Reported Recordkeeping Stewards ................................................................ 266
4.5.3. Perceived Requirements in the Cloud ............................................................... 269
4.5.4. Interaction between Recordkeepers ................................................................. 272
4.5.5. Evidence of ARM Functions in the Cloud Environments ................................ 276
4.5.5.1. ARM Functions and Their Occupational Allocation .................................. 276
4.5.5.2. Stewards’ Perceptions of ARM Roles and Responsibilities ....................... 288
4.5.6. Cloud Recordkeeping Risks .............................................................................. 290
4.5.6.1. Risks Reported in the ARM Literature ....................................................... 290
4.5.6.2. Recordkeeping Risks as Reported by Recordkeeping Stewards ................ 291
4.6. Overall Findings ................................................................................................... 297
4.6.1. A variety of individuals from different occupational groups act as recordkeepers within their organization ............................................................ 298
4.6.2. Thinking of stewardship as an organizational (as opposed to individual) responsibility may create disincentives for engaging in appropriate levels of long-term preservation ............................................................... 301
4.6.3. Different occupational group members may have different notions of “stewardship.” .......................................................... 306
4.6.4. The distribution of information ownership and processing in a new cloud computing environment creates disincentives to engage in appropriate levels of recordkeeping. ........................................................... 313

4.6.5. Cloud computing is used to help justify and smooth the changing power dynamics within organizations when they implement new technology perceived to represent this social structure. ........................................ 329

4.6.6. The degree of worker mobility affects the participants’ reported levels of comfort with new cloud computing implementations......................... 332

4.6.7. Presence or lack of trust between participants is a key factor in participants’ comfort with the implementation of cloud computing. ......................... 334

4.7. Final Comments ........................................................................................................... 336

5. CONCLUSION ................................................................................................................ 340

5.1. Recap ........................................................................................................................ 340

5.2. Limitations ................................................................................................................. 343

5.3. Import of Findings .................................................................................................... 345

5.4. Further Research Requirements ............................................................................. 346

APPENDIX A - DEFINITIONS OF CLOUD COMPUTING, BY SOURCE .................. 349

APPENDIX B - DEFINITIONS OF GRID COMPUTING, BY SOURCE ................. 351

APPENDIX C – MINNESOTA ARCHIVAL REQUIREMENTS FOR THE MICROSOFT 365 IMPLEMENTATION, JULY 2009 ........................................ 353

APPENDIX D – DISSERTATION REQUEST TO PARTICIPATE LETTER .......... 358

APPENDIX E – CONSENT TO PARTICIPATE FORM .................................................... 359

APPENDIX F – MINNESOTA RECORDKEEPING LAWS ............................................. 362

APPENDIX G – LEGAL ENVIRONMENT AFFECTING BIOSENSE 2.0 .............. 364

APPENDIX H – KENTUCKY RECORDKEEPING LAWS .......................................... 367

APPENDIX I – ARTICLES REFERENCING CHANGES IN ARM OCCUPATIONS AS A RESULT OF COMPUTER TECHNOLOGIES ................................. 369

REFERENCES ..................................................................................................................... 380
# LIST OF TABLES

Table 1 - Comparing Cloud to Related Architectures, by Feature and Source.......................... 89
Table 2 - Yin's Criteria for Selecting Research Design............................................................... 100
Table 3 - States That Meet the Selection Criteria.................................................................... 106
Table 4 – Project Research Questions Mapped to Chapter 4 Section Headings ....................... 123
Table 5 - In-Interview Frequency of Conversational Use of the Stub “Record,” by Occupation.................................................................................................................... 155
Table 6 - Self-Described Responsibilities of MN Records Managers........................................ 160
Table 7 - Employee Concerns Regarding the New Cloud Computing Service....................... 166
Table 8 - Status of BioSense 2.0 Jurisdiction Onboarding, July 2013 ..................................... 187
Table 9 - BioSense 2.0 Minimum Set of Shared Data Elements (ISDS 2011, 56-65) ............. 199
Table 10 - Reported ARM Functions....................................................................................... 258
Table 11 - Cloud Recordkeeping Risks Reported in ARM Literature...................................... 263
Table 12 - Reported Occupations of Record Recordkeeping Stewards, with Interviewee Occupation and Case(s) ........................................................................................ 267
Table 13 - Steward Performance of Tasks within ARM Functional Area.................................. 278
Table 14 - Comparison of Interviewee Reports with ARM Literature Reported Risks .......... 294
Table 15 - Interviewee Concerns Missing from ARM Literature............................................. 295
Table 16 - Similarities and Differences between Cases............................................................ 341
LIST OF FIGURES

Figure 1 - Computation Systems Compared (Foster et al. 2008) ........................................... 82
Figure 2 – Organizational Location of Stewards in Minnesota’s Microsoft Cloud............. 169
Figure 3 - BioSense 2.0 Information Providers and Data Warehouse (Kass-Hout, Barr, and Alletto 2012) ...................................................................................................... 215
Figure 4 - Functions Attributed to ARM Personnel More than Five Times in ARM Journal Articles, 1970-2010............................................................................................... 255
Figure 5 - Respondents' Reported Risks and Concerns about Cloud Computing .............. 292
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A29 DPWP</td>
<td>Article 29 Data Protection Working Party</td>
</tr>
<tr>
<td>ACL</td>
<td>Access Control List</td>
</tr>
<tr>
<td>AICPA</td>
<td>American Institute of Certified Public Accountants</td>
</tr>
<tr>
<td>ARM</td>
<td>Archives and Records Management</td>
</tr>
<tr>
<td>ARPA</td>
<td>Advanced Research Projects Agency</td>
</tr>
<tr>
<td>ARMA</td>
<td>Association of Records Managers and Administrators</td>
</tr>
<tr>
<td>ARRA</td>
<td>American Recovery Reinvestment Act</td>
</tr>
<tr>
<td>AST</td>
<td>Adaptive Structuration Theory</td>
</tr>
<tr>
<td>ASTHO</td>
<td>Association of State and Territorial Health Officials</td>
</tr>
<tr>
<td>AWS</td>
<td>Amazon Web Services</td>
</tr>
<tr>
<td>BCS</td>
<td>Bureau of Computer Services</td>
</tr>
<tr>
<td>BT</td>
<td>Bioterrorism</td>
</tr>
<tr>
<td>CCUCDG</td>
<td>Cloud Computing Use Case Discussion Group</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control</td>
</tr>
<tr>
<td>CFR</td>
<td>Code of Federal Regulations</td>
</tr>
<tr>
<td>CIITS</td>
<td>Continuous Instructional Improvement Technology System</td>
</tr>
<tr>
<td>CIO</td>
<td>Chief Information Officer</td>
</tr>
<tr>
<td>CISO</td>
<td>Chief Information Security Officer</td>
</tr>
<tr>
<td>CJIS</td>
<td>Criminal Justice Information System</td>
</tr>
<tr>
<td>CO2</td>
<td>Carbon dioxide</td>
</tr>
<tr>
<td>CoSA</td>
<td>Council of State Archivists</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>COT</td>
<td>Commonwealth Office of Technology</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer Relationship Management</td>
</tr>
<tr>
<td>CSP</td>
<td>Cloud Service Provider</td>
</tr>
<tr>
<td>CSTE</td>
<td>Council of State and Health Officials</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-separated values</td>
</tr>
<tr>
<td>DCC</td>
<td>Digital Curation Centre</td>
</tr>
<tr>
<td>DFEH</td>
<td>Department of Fair Employment and Housing</td>
</tr>
<tr>
<td>DGS</td>
<td>Department of General Services</td>
</tr>
<tr>
<td>DHHS</td>
<td>Department of Health and Human Services</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defense</td>
</tr>
<tr>
<td>DoT</td>
<td>Department of Transportation</td>
</tr>
<tr>
<td>DPH</td>
<td>Department of Public Health</td>
</tr>
<tr>
<td>DTI</td>
<td>Department of Technology and Information</td>
</tr>
<tr>
<td>DTS</td>
<td>Department of Technology Services (Utah)</td>
</tr>
<tr>
<td>DUA</td>
<td>Data Use Agreement</td>
</tr>
<tr>
<td>EASC</td>
<td>Enterprise Architecture and Standards Committee</td>
</tr>
<tr>
<td>EC2</td>
<td>Elastic Compute Cloud</td>
</tr>
<tr>
<td>EOC</td>
<td>Emergency Operations Center</td>
</tr>
<tr>
<td>HER</td>
<td>Electronic health record</td>
</tr>
<tr>
<td>EUCC</td>
<td>Enterprise Unified Communications and Collaboration</td>
</tr>
<tr>
<td>FCCSET</td>
<td>Federal Coordinating Council for Science, Engineering, and Technology</td>
</tr>
<tr>
<td>FIPS</td>
<td>Federal Information Processing Standards</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>FISMA</td>
<td>Federal Information Security Management Act</td>
</tr>
<tr>
<td>FOIA</td>
<td>Freedom of Information Act</td>
</tr>
<tr>
<td>GARP</td>
<td>Generally Accepted Recordkeeping Principles</td>
</tr>
<tr>
<td>GigaPOP</td>
<td>Gigabit network points of presence</td>
</tr>
<tr>
<td>GIS</td>
<td>Geographic Information System</td>
</tr>
<tr>
<td>GSA</td>
<td>General Services Administration</td>
</tr>
<tr>
<td>HEAA</td>
<td>Higher Education Assistance Authority</td>
</tr>
<tr>
<td>HHS</td>
<td>Health and Human Services</td>
</tr>
<tr>
<td>HIPAA</td>
<td>Health Insurance Portability and Accountability Act</td>
</tr>
<tr>
<td>HITECH</td>
<td>Health Information Technology for Economic and Clinical Health</td>
</tr>
<tr>
<td>HL7</td>
<td>Health Level Seven</td>
</tr>
<tr>
<td>HPCC</td>
<td>High Performance Computing and Communications</td>
</tr>
<tr>
<td>H.R.</td>
<td>House of Representatives</td>
</tr>
<tr>
<td>IaaS</td>
<td>Infrastructure-as-a-Service</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machines</td>
</tr>
<tr>
<td>ICA</td>
<td>International Council on Archives</td>
</tr>
<tr>
<td>ICD-10</td>
<td>International Statistical Classification of Diseases and Related Health Problems, 10th Revision</td>
</tr>
<tr>
<td>ICT</td>
<td>Information and Communication Technology</td>
</tr>
<tr>
<td>ILP</td>
<td>Individual Learning Plan</td>
</tr>
<tr>
<td>InterPARES</td>
<td>International Research on Permanent Authentic Records in Electronic Systems</td>
</tr>
<tr>
<td>IPAD</td>
<td>Information Policy Analysis Division</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>Abbr.</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>ISDS</td>
<td>International Society for Disease Surveillance</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organization for Standardization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>JISC</td>
<td>Joint Information Systems Committee</td>
</tr>
<tr>
<td>KDE</td>
<td>Kentucky Department of Education</td>
</tr>
<tr>
<td>KDLA</td>
<td>Kentucky Department of Libraries and Archives</td>
</tr>
<tr>
<td>KETS</td>
<td>Kentucky Education Technology Services</td>
</tr>
<tr>
<td>KIDS</td>
<td>Kentucky Office of Knowledge, Information and Data Services</td>
</tr>
<tr>
<td>KSIS</td>
<td>Kentucky Student Information System</td>
</tr>
<tr>
<td>LARM</td>
<td>Libraries, Archives, and Records Management</td>
</tr>
<tr>
<td>LOC</td>
<td>Library of Congress</td>
</tr>
<tr>
<td>LOINC</td>
<td>Logical Observation Identifiers Names and Codes</td>
</tr>
<tr>
<td>MEEC</td>
<td>Maryland Education Enterprise Consortium</td>
</tr>
<tr>
<td>MHS</td>
<td>Minnesota Historical Society</td>
</tr>
<tr>
<td>MIS</td>
<td>Management Information Systems</td>
</tr>
<tr>
<td>MOU</td>
<td>Memo of Understanding</td>
</tr>
<tr>
<td>MPI</td>
<td>Master Patient Index</td>
</tr>
<tr>
<td>MREN</td>
<td>Metropolitan Research and Education Network</td>
</tr>
<tr>
<td>MS</td>
<td>Microsoft</td>
</tr>
<tr>
<td>NACCHO</td>
<td>National Association of City and County Health Officials</td>
</tr>
<tr>
<td>NARA</td>
<td>National Archives and Records Administration</td>
</tr>
<tr>
<td>NC DETECT</td>
<td>North Carolina Disease Event Tracking and Epidemiologic Collection Tool</td>
</tr>
<tr>
<td>NCSU</td>
<td>North Carolina State University</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>NHPRC</td>
<td>National Historical Publications and Records Commission</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NPACI</td>
<td>National Partnerships for Advanced Computational Infrastructure</td>
</tr>
<tr>
<td>NPI</td>
<td>National Provider Identifier</td>
</tr>
<tr>
<td>NUBC</td>
<td>National Uniform Billing Committee</td>
</tr>
<tr>
<td>OET</td>
<td>Office of Enterprise Technology (Minnesota)</td>
</tr>
<tr>
<td>OMB</td>
<td>Office of Management and Budget</td>
</tr>
<tr>
<td>OPRR</td>
<td>Office for Protection from Research Risks</td>
</tr>
<tr>
<td>P2P</td>
<td>Peer-to-Peer</td>
</tr>
<tr>
<td>PaaS</td>
<td>Platform-as-a-Service</td>
</tr>
<tr>
<td>PACI</td>
<td>Partnerships for Advanced Computational Infrastructure</td>
</tr>
<tr>
<td>PASS</td>
<td>Provenance-Aware Storage System</td>
</tr>
<tr>
<td>PCAOB</td>
<td>Public Company Accounting Oversight Board</td>
</tr>
<tr>
<td>PCI DSS</td>
<td>Payment Card Industry Data Security Standard</td>
</tr>
<tr>
<td>PHEP</td>
<td>Public Health Emergency Preparedness</td>
</tr>
<tr>
<td>PHSIP</td>
<td>Public Health Surveillance and Informatics Program</td>
</tr>
<tr>
<td>PII</td>
<td>Personally Identifiable Information</td>
</tr>
<tr>
<td>PKI</td>
<td>Public Key Infrastructure</td>
</tr>
<tr>
<td>POIT</td>
<td>Psychological Ownership of Information Theory</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>REST</td>
<td>Representational State Transfer</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposal</td>
</tr>
<tr>
<td>RIM</td>
<td>Records and Information Management</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>RMA</td>
<td>Records Management Application</td>
</tr>
<tr>
<td>S3</td>
<td>Simple Storage Service</td>
</tr>
<tr>
<td>SAA</td>
<td>Society of American Archivists</td>
</tr>
<tr>
<td>SaaS</td>
<td>Software-as-a-Service</td>
</tr>
<tr>
<td>SARC</td>
<td>State Archives and Records Commission</td>
</tr>
<tr>
<td>SAS-70</td>
<td>Statement of Auditing Standard 70</td>
</tr>
<tr>
<td>SEC</td>
<td>Securities and Exchange Commission</td>
</tr>
<tr>
<td>SERI</td>
<td>State Electronic Records Initiative</td>
</tr>
<tr>
<td>SLA</td>
<td>Service Level Agreement</td>
</tr>
<tr>
<td>SME</td>
<td>Subject matter expert</td>
</tr>
<tr>
<td>SNOMED</td>
<td>Systemized Nomenclature of Medicine</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>SOX</td>
<td>Sarbanes-Oxley</td>
</tr>
<tr>
<td>SSL</td>
<td>Secure Sockets Layer</td>
</tr>
<tr>
<td>SSRC</td>
<td>Southwood Shared Resource Center</td>
</tr>
<tr>
<td>UNC</td>
<td>University of North Carolina</td>
</tr>
<tr>
<td>US</td>
<td>United States</td>
</tr>
<tr>
<td>vBNS</td>
<td>Very high-speed backbone network service</td>
</tr>
<tr>
<td>VIA</td>
<td>Virtual Interface Architecture</td>
</tr>
<tr>
<td>VO</td>
<td>Virtual Organization</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

The motivation for undertaking this exploratory study is to gain increased understanding of how new technologies affect the theory and practice of archives and records management (ARM) in complex organizational settings. The specific goals of this study are to examine how recordkeeping stewards who work in state government or alongside other state government recordkeeping stewards in cloud computing environments perceive and act upon electronic recordkeeping requirements in the Cloud, to understand which of the functions of ARM work described by ARM academic literature occur in the recordkeeping environments examined, and to determine whether these functions are performed by ARM workers or by other recordkeeping stewards when they do occur.

1.1. Background

In the past forty years, archives and records management researchers have paid a great deal of attention to the effect of new computing technology on ARM functions, capabilities, and roles and responsibilities. As early as the late 1960s members of the Society of American Archivists engaged in conversations about the potential impacts of automatic data processing and information retrieval technologies (Brown 1984). Discussions of computing technologies from that point onward focus on a wide variety of technical, theoretical, and social dimensions influencing how technology affects the archives and records management profession. For example, in the 1970s, Evans and Gustafson admitted to the “possibility” that computers may someday have a role in archival operations (1975),
while Hickerson, Winters, and Beale spoke of the actual role of automated indexing in archival practice (1976). By the 1980s, a number of computer-based records management systems and automated indexes had been built in the United States and other countries; one can see descriptions of many of these projects in each issue of the journal *The American Archivist* throughout the 1980s, in its regular section called “The International Scene: News and Abstracts.” Although the articles specifically referencing computer technology in the 1970s were for the most part practical and although this pragmatic trend continued into the 1980s, a number of researchers began to question the ways in which technology would change the nature of archives, the roles and responsibilities of the archivist, and archival theory and concepts (Bearman 1989a; Ham 1981, 1984; Cook 1983; Lytle and Dürr 1980; Burke 1981; Weldon 1983; Jimerson 1989; Peterson 1984) during the latter decade.

During the 1990s the ongoing presence of computerized technology had largely become accepted as a lasting component of archival practice, albeit still somewhat problematic from technical (Martin 1994; Weissman 1994; Curtin 1990; Pederson 1990; Greenberg 1998; Cox 1990), managerial (Tibbo 1995; Curtin 1990; Hedstrom 1991; Bearman 1992; Gilliland-Swetland and Hughes 1992), educational (Gilliland-Swetland 1993; Henry 1993; Kesner 1993; Ruller 1993; Walch 1994), political (Elliott 1990; Lyman 1994; O'Toole 1994; Lubar 1999), legal (Brown 1995; Piasecki 1995; Kahin 1988), social (Lockwood 1990), and theoretical (Hedstrom; Stielow 1991; O'Toole 1993; Bearman 1995) perspectives. In 1993, Victoria Irons Walch conducted a study which combined Everett Roger’s theory of diffusion of innovation (1983) with Houle’s analysis of professional work settings (1970) to conclude that the archival profession had by then reached Roger’s “early majority” stage in relation to the automation of their activities, a point at which members of
the occupation could be expected rapidly to accept new technologies and to begin demanding assistance in using them.

During the 1990s and 2000s, some researchers began to place more emphasis on the need to build theoretical structures to understand better the relationship between archival theory and practice and changing technologies. For example, Margaret Hedstrom (1991) argued early in the decade that archivists “need a framework for research on electronic records issues” (1991, 334) and that “research on electronic records issues spans all archival functions and may challenge basic archival theory and practice” (335). She recommended a more interdisciplinary and theoretical research agenda for understanding the role of technology in ARM theory, urging researchers to pay close attention to research methodology and pointing out that the term “technology” is very ambiguous. Frederick Stielow argued that the nature of electronic records requires archivists to “anticipate a restructuring of descriptive theory” (1992, 335). In 1996, Elizabeth Yakel made a strong argument that archivists could use the work of organizational theory to gain a “more sophisticated view of organizational processes,” and by doing so, gain a greater understanding of the “role of records and recordkeeping systems in organizations” (454). Yakel, in fact, provided such an analysis in 2001, when she examined the changing nature of records and records management roles and responsibilities as radiological records moved from analog to digital formats (2001). Terry Cook had already revived interest in the Australian continuum theory (Cumming 2010) with his well-known article “Electronic Records and Paper Minds” (1994). He predicted that by moving from a textual to an electronic environment, archivists would need to accept a new paradigm that focuses not on the record, but on the act of recording; not on the content, but on the context. For the most
part, within these discussions the focus was on dichotomies between a “paper” (or “analog”) world and an “electronic” (or “digital”) world and between the practices required to maintain and preserve paper records and those required to maintain and preserve electronic records. The continuum theorists argued that archivists must move beyond thinking of archives (and especially, the arrangement and description of archives) using the old physical models and begin to think about their activities in logical terms, distinguishing between “physical” and “logical” similarly to the way data modelers conceive of the physical design versus the logical design of databases.

By the 2000s, authors such as Fiorella Foscarini (2012) engaged in studies that accept electronic, or often, hybrid1, environments as the norm, and attempt to understand more about the theoretical frameworks and constructs that guide electronic recordkeeping environments. With the apparent convergence of traditional records management into a mixed activity that involves general management, records and information management (RIM), and information science concepts (Yusof and Chell 1998, 1999, 2002; Ryan 2006; Benfell 2007), new questions about the role of the records manager and the nature of the record have surfaced. If it is no longer the case that the records manager should handle only current, active records, in contrast to the archivist, who handles the inactive, preservation-worthy records, is there in fact (or should there be) a convergence between the two occupational fields?

One difficulty involved in answering this question rests upon the varying contexts within which recordkeeping activities occur. Functionally, the goals of an archives or digital

---

1 The term “hybrid” is used differently among different academic and professional domains. Here the term is being used in the sense typically used in archives and records management, where a hybrid environment is one which includes both electronic records management and paper-based records management within the same work environment.
repository that is aimed at collecting and preserving already-created records for the long-term appear to be different from those of a complex organization which creates records to enable the tracking of transactions but which does not consider records to be the primary component of its mission. Likewise, different organizations structure their information technology (IT) functions quite differently, thereby requiring a variety of different types of relationship-management activities, depending upon whether IT is managed in a centralized, a decentralized, or a hybrid environment², and whether shared services occur across organizational boundaries.

However, although the ARM literature is full of speculation regarding the theoretical and practical impacts of computerized technology on ARM theory and practice, only one article between 1970 and the present can be found that explicitly discusses the potential changes that evolving enterprise IT architectures will bring (Weissman 1994).³ In this article Weissman discusses changes that could be wrought by the advent of object-oriented programming, increased use of relational databases, and greater network connectivity, suggesting that archivists will increasingly need to become “information-engineering experts” (34). Moreover, since publication of that article in 1994, technological environments have evolved even more and the days when relational databases and object-oriented programming were ARM workers’ only technical challenges may well be characterized as “the good old days.” For example, cloud computing represents an emergent

---

² Here the term “hybrid” is used as it is used in organizational management, where a hybrid environment is one in which the IT work is managed neither in a centralized way nor a decentralized way but rather, with some elements of work managed centrally and some managed by local business units themselves.

³ One could appropriately argue, however, that the work of the Monash Clever Recordkeeping Metadata (CRKM) project, which is geared toward automating metadata creation and sharing metadata between business systems, current recordkeeping systems and archival systems also addresses the issue of new enterprise IT architectures, although it does not explicitly present this as a goal. Rather, it focuses on its ties to records continuum theory (Evans, McKemmish, and Bhoday 2004; Evans, Reed, and McKemmish 2008).
technology that may require radical changes in ARM activities beyond the question of whether recordkeepers handle paper or electronic documents, or think with “electronic minds” (Upward 1996).

In addition, with the exception of continuum theorists (Hurley 1995, 2011; Iacovino and Reed 2008; Reed 2005; Upward 1996, 1997; Upward and McKemmish 1994) and some proponents of digital curation ((DCC) 2010; Beagrie 2006; Lee and Tibbo 2011; Tibbo and Lee 2012), previous studies have often characterized ARM functions in a manner suggesting that they are performed by a specific and readily-identifiable set of practitioners (“archivists” or “records managers”), whose functional responsibilities are clearly defined within relatively well-bounded organizations or collaboration agreements. Within much archival literature, the focus of activities naturally tends to be on the archival functions, and authors assume that records management is merely a necessary component of good archival practice with goals that are subsumed by the needs of the archival practice (Yusof and Chell 2002). Other researchers have noted that within the archival literature, there appears to be a separation between archives and records management (Cox 2000b), with many records management articles expressing a RIM, or corporate orientation (Kim and Lee 2008). Yusof and Chell (2002, 55-56) argue that this is because records management as a field has derived from several different “streams of thought”: archival theory, information science, and management information systems (MIS).

This study will show that in complex organizations which have a primary goal that is not long-term preservation, not only has the distinction between the work of the records managers and the archivists already begun to disintegrate, but the activities that have traditionally been associated with archives and records management are dispersed among a
variety of occupational groups. This has created potential issues with data and process ownership that places the goal of ensuring that records-related accountability continues to occur within these organizations at risk. It also points to the need to examine the actual changes in the roles and responsibilities of recordkeeping stakeholders within organizations using new technologies in order to clarify the ways in which changing practices and expectations may need to be reflected within ARM theory and practice.

This study addresses gaps in previous literature in two ways. First, the study examines the actual environments of workplaces that are not archives— the recordkeeping stewards interviewed either work within state governments or collaborate with others who work in state government. The study highlights a range of electronic recordkeeping practices within those organizations. Second, the study examines an emerging technology (i.e., cloud computing) that has rarely been examined in the ARM literature with respect to its impacts on ARM theory and practice in actual organizations. By painting a picture of actual recordkeeping environments using cloud computing, the study examines how this emerging technology places stress upon academic depictions of the work of archives and records management, and points toward areas of research that need to be examined in order to ensure that the theoretical work in archives and records management remains consistent with the continually evolving technological environments that organizations face.

---

4 Here I use Barbara Reed’s definition of an archives as “an organisation or part of an organisation not immediately connected to the creation and management processes of the records which is deemed responsible for the continued management and preservation of a selection of records” (Reed 2005).

5 Two exceptions exist in academic literature: Askhoj, Sugimoto, and Nagamori (2011) attempt to show the inadequacy of the Reference Model for an Open Archival Information System (OAIS) for cloud computing environments and to construct a layered model that is relatively consistent with the OAIS model, except for relying on services rather than entities. Also, Stuart and Bromage (2010) investigate some of the risks of cloud computing for records management, although they tend to blur the distinction between activities on the World Wide Web and cloud computing and also mistakenly conflate cloud computing with third-party storage or application hosting.
1.1.1. Why cloud computing?

Cloud computing is a form of information technology (IT) provision that treats computing as a set of services that can be purchased on-demand through networks. Individuals and businesses can thereby use the amount of service they want to use without having to make large infrastructure investments that may lead to idle resources when computing needs are lower than maximum capacity. This arrangement allows computing infrastructure, hardware, and software to be treated as largely modular services that can be scaled up and down easily and with minimal ongoing interaction and negotiation with one’s computing resource provider. This arrangement also allows organizations to minimize the number of information technology professionals necessary to meet the IT needs of the organization.

Cloud computing has become associated with the notion of economic sustainability in a number of IT circles. Because it allows organizations to make use of large amounts of computing power without having to make correspondingly large capital investments in IT infrastructure (Creeger 2009), numerous organizations have found shifting to “the Cloud” to be both cost effective and operationally straightforward. Some have even argued that the movement to the Cloud is an inexorable, overarching shift from one mode of industrialization to an entirely new mode (McAfee 2009), similar in nature to the shift from steam power to electrical power during the late nineteenth and early twentieth centuries (Carr 2008).

The Obama administration has made cyber-infrastructure in general and cloud computing in particular key priorities for the federal sector (InformationWeek 2009). In September 2009, the federal government announced its “Cloud Computing Initiative,” outlining both the rationale for the initiative and some of its key components, such as its major characteristics, its delivery methods, and its deployment models (Lewin 2009).
2010, then first Chief Information Officer (CIO) of the United States of America Vivek Kundra published a report on the state of public sector cloud computing, describing thirty high-profile implementations that were in place or in process (Kundra 2010). Kundra has since left his position as the U.S. CIO, but new U.S. CIO Steven VanRoekel announced very early in his tenure that he intended to continue Kundra’s “Cloud First” policy (proofpoint 2011), which requires federal agencies to evaluate “safe, secure cloud computing options” before making any new investments (Kundra 2011, 2). Since that time, VanRoekel has reported,

With our Cloud First initiative, agencies identified 79 services to move to the Cloud in order to reap savings and service improvements. This year, agencies successfully migrated 40 services to the Cloud and were able to eliminate more than 50 legacy systems in order to save taxpayer dollars while expanding capabilities. As part of this effort, agencies created six services in the Cloud that they weren’t previously able to provide (2011).

In late 2011 the U.S. General Services Administration (GSA) launched the FedRAMP (GSA 2012), “a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services which every agency will be required to use” (VanRoekel). In addition, in 2012 the CIO Council, in conjunction with the Chief Acquisition Officers’ Council, published a set of best practices for agencies acquiring cloud computing services (2012).

An examination of the websites of state governments in the United States (US) shows that at least three-quarters of all states have adopted cloud computing or are in the process of adopting cloud computing within one or more of their state-level agencies (excluding basic website hosting services). For example, Wyoming has shifted the email of all of its approximately 10,000 state government employees to Google Apps for Government, “putting them on a single email platform for the first time” (Office of Governor Matt Mead 2011).
Minnesota implemented Microsoft Office 365 to deliver the state’s Enterprise Unified Communications and Collaborations services (MN OET). New Jersey’s Transit Authority implemented an on-demand, cloud-based customer relationship management (CRM) service from Salesforce.com, allowing it to handle more than five times the number of customer inquiries than it had before implementing the service. The transit authority’s response time reportedly dropped by more than 35 percent, and its productivity increased by 31 percent (Kundra 2010). The State of Utah implemented a hybrid solution that uses an internal private cloud (Towns 2009) in conjunction with Force.com for CRM, Google Earth Professional for shared Geographic Information System (GIS) planning, and Wikispaces for self-provisioned wikis. These contracts are managed centrally through the state’s Department of Technology Services (DTS). Initial estimates suggested that Utah, which has an IT budget of $150 million per year, will save four million dollars per year as a result of moving to the Cloud (Kundra 2010).

However, if cloud computing sometimes seems straightforward from the point-of-view of IT costs and service provision, its benefits for records management are murkier. The U.S. National Archives and Records Administration (NARA) has suggested that agencies should be aware of the potentially adverse effects of cloud computing on records management work (2011). NARA noted that some cloud architectures “lack formal technical standards governing how data is stored and manipulated in cloud environments,” thereby threatening the long-term trustworthiness and sustainability of the data. In addition, a lack of portability standards might make it difficult to dispose of or transfer records in accordance

---

6 This is yet another disciplinary-specific use of the term “hybrid.” Here, “hybrid” refers to a cloud computing implementation which includes both a public cloud and a private cloud jointly. The organization thus provides and manages some cloud computing resources in-house and some externally.
with recordkeeping requirements. ARMA International pointed out that a wide range of potential records management risks are associated with cloud computing, such as the potential failure to meet recordkeeping regulatory requirements, jurisdictional issues regarding data storage, vendor continuity concerns, lack of clarity surrounding data ownership, and interoperability challenges (2010). Although some vendors are beginning to produce records management applications that can be integrated with cloud computing services (Miller 2011; Alfresco 2011; SpringCM 2011), state government professionals must be diligent to ensure that their agencies adopt appropriate records management policies and procedures when cloud computing services are contracted in order to satisfy state and federal recordkeeping requirements.

As threats to the long-term trustworthiness and sustainability of data, the risks that NARA and ARMA International associate with cloud computing strike at the heart of ARM goals. Richard Cox argued that records management exists to “support accountability, the protection of crucial evidence, and the nurturing of corporate memory” (2001, 13). He pointed to records’ role in ensuring organizational accountability and providing evidence of organizational transactions and decisions. Likewise, Anne Gilliland-Swetland (2005), examining the development of a research agenda for electronic records management, noted that since the late 1980s ARM professionals have largely accepted that recordkeeping as an activity is geared toward ensuring the evidentiary quality of records. The international recordkeeping standard ISO 15489-1, which “provides a framework for any organization, public or private, to adopt and use to manage its records, irrespective of the medium on which they are created, captured and maintained” (McLeod 2003), adds that in addition to providing evidence and ensuring its accountability, an organization has the responsibility to
manage its records so that it can support the continuing conduct of its business and remain compliant with its regulatory requirements. To do this, ISO 15489-1 says an organization must “create and maintain authentic, reliable and useable records, and protect the integrity of those records for as long as required” (ISO 2001, 6). Both the Society of American Archivists (SAA) code of ethics (2012) and the ARMA International code of ethics (2011) agree, stating that protecting the authenticity, integrity, and accessibility of records is a key ethical and professional responsibility of ARM workers. Thus, to the extent that cloud computing presents a risk to an organization’s ability to effectively manage and maintain its records, it threatens the goals of the organization’s ARM professionals.

In addition, as will be seen in the next chapter, cloud computing represents a new and unfamiliar technology for most recordkeeping professionals, increasing the risk that lack of knowledge about the nature of the technology will lead workers to overlook some of the procedures in which they should engage when attempting to ensure the authenticity of records over time or fail to address all of the potential risks that inhere in this technological approach.

1.1.2. Why state government?

In order to govern adequately, a government must manage and maintain its records effectively. Authentic, reliable, and accessible records provide the organizational memory necessary for the government to fulfill its mission legally and effectively. Moreover, they provide the accountability and evidence that assure the public that the government is doing its job legally and effectively. The Council of State Archivists (CoSA) describe a number of reasons why state governments must ensure that they maintain authentic, reliable, and accessible records (2007):
• Government records are essential to protect life in the case of disaster. They provide the means for rescue workers to find and save lives and for rebuilding necessary infrastructure when destruction has occurred.

• Government land records protect property by proving “ownership, boundaries, and other essential information for home and business owners” (11) and provide proof of inheritance, property distributions, and educational attainment.

• Government records document and verify the rights of individual citizens such as “the right to vote (voter registrations), the right to government services (military service records, birth records, employment records, education records), and the right to justice (court records, criminal justice records)” (11).

• Government records document the rights of communities and groups, including “civil rights (employment regulations, laws, court records), community welfare (land records, transportation records, public health records), and civil protection (military records, criminal justice records)” (11).

• Government records document the “transactions, decisions, and precedents” (12) of government activities, allowing the government to maintain its operations on an ongoing basis and providing accountability to citizens.

• Government records provide citizens with a “sense of community, a sense of belonging, a sense of place,” (12) by maintaining their accessibility to the temporally structured and geographically bounded artifacts that link them together as a community.

Although theoretically a government could do a good job of managing and maintaining its records and still fail to perform its mission or sustain legitimacy with the
populace, it cannot provide assurance that it is performing its mission or acting legitimately without authentic, reliable and accessible records. Thus, to the extent that cloud computing poses risks to government’s ability to effectively manage and maintain its records, it threatens the legitimacy and effectiveness of government itself.

However, no research has previously examined in detail the ways in which a cloud computing adoption will impact the nature of recordkeeping work within state agencies. In fact, since Ernst Posner’s (1964) survey of state agency ARM programs in the early 1960’s, no detailed examination of state ARM work has occurred at all, with the exception of the Council of State Archivists’ (CoSA) report “The State of State Records” (2007) and their recent State Electronic Records Initiative (SERI) Report (CoSA SERI Committee 2012). Posner’s study examines only the structure of the state ARM programs, providing information about the structure of the various state programs, their budgets during the period studied, and state laws that govern the disposition of records. It does not offer a detailed examination of recordkeeping functions within any state, nor does it describe technological systems used for ARM purposes (as would be expected, given its publication well before electronic recordkeeping became common). The CoSA State of State Records report provides updated state archival program information, but again, this report focuses on ARM programs themselves, offering more general knowledge about funding arrangements, program structure, and holdings. Because the report aims to provide a comprehensive understanding of state government archival programs throughout the U.S., it does not delve into a detailed examination of the relationship between recordkeeping functions and the sociotechnical systems within which those functions reside. The SERI report focuses upon the status of state electronic records programs within the fifty states, but does not examine detailed roles and
responsibilities of recordkeeping workers. It concludes that the total holdings of electronic records held in state or territory archives “… represents only a fraction of the total volume of electronic records with long-term value held in state and territorial governments’ agencies and offices that will or should come to the archives” (13). This underscores the need for deriving a greater understanding of what types of electronic recordkeeping activities are actually occurring in those agencies and offices. In addition, the primary interviewees of the SERI report are state archivists, and the report takes as a given that the preservation-worthy electronic records should all be transferred to the state archives (2012, 13). This assumption has been rejected by records continuum theorists, and thus is open to examination by this study, which seeks to understand better the relationship between the actualities of state electronic recordkeeping and the underlying assumptions of ARM theory. Indeed, this finding presents a question regarding whether it could even be possible to achieve such a transfer of all preservation-worthy electronic records to state archives or statewide records management programs, especially given increasing use of IT practices such as shared services and cloud computing. Furthermore, the lack of systematic, empirically grounded research on state recordkeeping work presents a serious gap for those interested in understanding how a new technological arrangement such as cloud computing affects existing electronic recordkeeping activities within state government.

1.2. Research Questions and Methods

As mentioned earlier, the specific goals of this exploratory study are to examine how recordkeeping stewards who work in state government or alongside other state government recordkeeping stewards in cloud computing environments perceive and act upon electronic recordkeeping requirements in the Cloud, to understand which of the functions of ARM work
Several research questions have informed the investigation reported here:

- Within the environments examined, what occupational groups are reported to act as key stewards of the information and how do members of these groups perceive and act upon recordkeeping requirements in the Cloud?

- How do the various stakeholders interact with each other with respect to recordkeeping activities within their cloud computing environments, and what do these relationships suggest about how ARM occupational roles and responsibilities are being handled in cloud computing environments?

- How do the various stakeholders perceive the roles and responsibilities of archives and records management personnel?

- What cloud computing risks does the professional and academic ARM literature report, and do recordkeeping stewards in state government cloud environments express concerns about these same risks?

- Of the main recordkeeping functions that the ARM literature attributes to ARM workers, are these functions evident in the recordkeeping environments examined and if so, are they performed by ARM workers?

The project described here uses an interpretive strategy and a multiple-case study research design. Data collection methods include semi-structured interviews conducted with a wide set of state and federal government personnel, private consultants, and representatives...
of professional organizations, the collection of internal and published state government documents, and an examination of 682 published articles from six ARM journals over a 42 year period. Data analysis uses the constant comparative method from Grounded Theory and documentary content analysis. Final reporting is presented within the framework of Structuration Theory. The results of the individual components of the overall study are triangulated to develop the study’s conclusions.

Chapter 2 reviews in more detail the literature on how ARM professionals have characterized the effects of new technologies – specifically, computerized technologies – on ARM theory and practice, focusing on ARM roles and responsibilities and key constructs from ARM theory, theories about the effects of new technologies on occupational roles & responsibilities, and cloud computing as an emerging technology with direct impacts on organizational recordkeeping. Chapter 3 describes the methods used in this study. Chapter 4 discusses the results of the documentary analysis and multiple case study, considering each case separately, and then comparing and contrasting findings across all cases and interviews. Chapter 4 also analyzes these results in light of the five research questions given above. Chapter 5 discusses the implications and importance of this research and offers suggestions for future research.

---
7 The professional organizations were the National Association of State CIOs (NASCIO) and RTI International.
2. LITERATURE REVIEW

Over the past fifty years archivists and records managers have been facing a continual onslaught of new and previously unfamiliar technologies for handling their information. These new and emerging technologies can influence their work in a variety of ways. To understand the impacts of a given technological adoption on a specific set of workers, however, one must understand who these workers are, why they engage in the activities in which they engage, and how these activities and reasons for engagement impact and are impacted by the new technology. This literature review provides the core concepts and empirical understandings that situate this study. It provides a framework deriving from technology- and work-related research in multiple disciplines, offers a historical and conceptual examination of the occupational commitments of ARM workers, and discusses cloud computing as a new and emerging technology.

2.1. The Work Environment

2.1.1. Work in Complex Sociotechnical Systems

Work environments are social systems which exhibit varying degrees of complexity. When an organization brings technology into the work environment, workers enact that technology and social relations simultaneously, creating non-random, yet not entirely predictable, outcomes (Orlikowski 2000). Within such a system people and technology combine to create products or services jointly. To understand the outcomes that result when a new or emerging technology is introduced, one must consider both the human element and the technical elements jointly, considering the requirements of both for engaging in goal-
directed work and understanding how these two elements interact (Fox 1995). One must also understand the wider historical and environmental structures that shape human action during the performance of work (Orlikowski and Robey 1991). In fact, one can consider a computerized work environment to be a “sociotechnical system,” if one uses that term in a broad sense to indicate that human action, work context, and technology interact together to influence outcomes (Dillon 2000).

Sociotechnical systems tend to be complex. Vicente (1999, 13-16) highlights the characteristics of a complex sociotechnical system. All complex sociotechnical systems will exhibit some or all of these elements. However, the degree to which any given feature is exhibited will vary across types of systems. As he points out, the characteristics defining a nuclear power plant are significantly different from those defining a small office environment.

A complex system, according to Vicente, tends to reflect a large problem space in which numerous elements interact so that it is difficult (or even impossible) for people within the system or external to the system to enumerate all the key variables that influence the operation and the outcomes of behavior within the system. Likewise, such systems are social in nature, requiring significant human-to-human and human-to-technology interaction. Because of the social nature of these systems, they are characterized by heterogeneous perspectives, whereby the people within the system are strongly motivated by different value

---

8 Since every worker relies upon both discursive and tacit knowledge, the distinction between a sociotechnical system and a non-social system is an analytical distinction rather than an ontological distinction. Even the lone writer uses the language and tacit skills learned while growing up in a social system. Similarly, virtually any enabling artifact could be considered a technology, as far as it requires technē. One can understand Vicente’s cognitive engineering view, however, as one of degree rather than one of kind.
systems. In spite of the difference of values, some degree of coherence occurs through cooperation and competition, governed by tacit and explicit authority mechanisms.

The various elements of complex systems also tend to be distributed, both geographically and functionally. This can make it difficult to monitor and control the system as a whole without specific mechanisms designed to do so. In addition, the goal-relevant properties of the system tend to be dynamic and often the overall system takes considerable time to reflect change in response to changes in these properties. The whole system will evolve as its properties evolve, but often with a lag in response time.

Complex systems can often be hazardous as well, in that the outcomes of operational errors in the elements of the system or its inability to adjust to changing internal and external factors can lead to harmful outcomes, whether those outcomes be in terms of health and safety, economics, or some other variable. For example, in the realm of nuclear power, Vicente shows how inaccurate mental models of the technical processes led to a nuclear power plant meltdown. Likewise, in the realm of digital curation, human action, machine failure, or human-machine breakdowns can lead to catastrophic data loss. For example, on March 30, 2012, personally identifiable healthcare information of more than 750,000 Utah citizens was stolen when the computer system on which it resided was allowed to retain its out-of-the-box (i.e., “default”) password by a systems administrator (Towns 2012; Henetz 2012; Gibson 2012; Utah Department of Health 2012).

Complex sociotechnical systems are generally composed of yet other systems which are coupled in a variety of configurations themselves. In other words, they are often systems of systems (Ackoff 1971). Because of the complexity of the interactions between all of the components and participants, it is generally difficult or impossible to predict fully the
outcomes of changes in these interactions or to predict the effects on the whole system when
a single significant change occurs in one of the component systems. This is because of the
large number of possible interaction effects that can occur and because it is not always
possible for any given individual to disentangle all of the different ways in which the
component systems are coupled. Moreover, because such systems comprise human agency,
where agency is defined as “the capability of engaging in action” (Giddens 1984), choice
regarding how people will use the technology cannot be fully predicted; thus, the overall
impacts of a group of interrelated agents may engender unanticipated consequences. “People
adapt systems to their particular work needs, or they resist them or fail to use them at all; and
there are wide variances in the patterns of computer use and, consequently, their effects on
decision making and other outcomes” (DeSanctis and Poole 1994, 122).

In addition, complex systems are often highly automated, with machinery or
computerized technology performing many of the detailed operations, and use of the
technology involving mediated interaction with it. For example, computer interfaces may be
used by people in order to present to the people the relevant variables to which they must
direct system inputs in order to perform their goal-directed activities.

Finally, complex sociotechnical systems are often characterized by a tendency toward
disturbances, where a disturbance is an unpredicted factor that influences the operation of the
whole system. For example, financial institutions are prone to economic shocks in the form
of unanticipated resource depletion (such as an earthquake influencing supply lines), or
external funding resources suddenly drying up due to political shocks (Vicente). Because of
this tendency toward disturbances, it is desirable that the system be prepared to respond to
unanticipated factors that directly affect some or all of its component properties. In an
exemplary system, people and machines will react flexibility and appropriately to unpredictable shocks.

2.1.2. The Structuration of a Social System

Until now, the term “system” has been used as it is used in everyday speech. As a common language habit, there is nothing particularly problematic with this; we often talk about living “in” social systems, as if the system itself is some sort of structure or thing that exists independently of our membership in it. However, this approach to defining systems creates conceptual problems when used theoretically. By treating a social system as if it has some sort of independent existence apart from the people within it, one reifies the concept in a manner which makes it seem as if social reality and physical reality are two entirely separate, concrete entities. Such treatment has caused numerous arguments over whether social reality is really “real,” whether it is “socially constructed,” or whether there is such a thing as a so-called “objective” reality, and has led to a theoretical inability to explain how the two different aspects of our existence (i.e., subjective and objective, or ideal and material) can act upon each other at all.

In order to reconcile this dualism, sociologist Anthony Giddens treats systems as “reproduced relations between actors or collectives, organized as regular social practices” (25). Structures are “rules and resources, or sets of transformation relations, organized as properties of social systems” (25). Giddens examines the ways by which social systems are reproduced via a recursive process in which the structures that mediate human action are simultaneously reproduced by that action. Thus human action and structures are not separate phenomena but rather, are co-constituted when agents engage in interaction via structures of

---

9 I use the term “reify” in this paper in its philosophical sense of treating an abstract concept as if it has concrete existence. This is in contrast to other senses that have been adopted, for example, in computer science or natural language theory.
signification (i.e., interpretive schemes), structures of legitimation (i.e., norms), and structures of domination (i.e., facilities). In other words, when humans interact with each other and the world, the interaction “implies the interlacing of meaning, normative elements and power” (28-29) and it is within these routine\textsuperscript{10} interactions that structure comes into being. Structure is thus in a constant state of becoming. Social systems, of which structures are properties, are reproduced as a result of human action yet the acting individuals view the systems as constraining that action. Since human agency allows choice with respect to taking particular actions at any point in time, and since humans may engage in actions that have unintended impacts on the world, the social system not only reproduces but also evolves with changing purposes and as a result of unintended interaction effects. This leads to modifications in the structure of society over time.

Giddens’ original Theory of Structuration is generally considered a “meta-theory” in that Giddens does not attempt to explain in detail all the micro-relations in the behavior of individual people or in organizational behavior. Rather, he elaborates how society and its members co-evolve through human agency. In addition, he does not discuss technology per se, but several organizational theorists use his theory to discuss the micro impacts of technology on organizational structure (Barley 1986; DeSanctis and Poole 1994; Orlikowski 1992, 1996, 2000; Orlikowski and Barley 2001; Orlikowski and Robey 1991; Orlikowski and Yates 1994). For example, Barley examines the ways in which CT scanners affect the organizational and occupational structures of radiological work, treating technology as a social object. Orlikowski criticizes this conception, noting that while it may be appropriate

\textsuperscript{10} Giddens used the notion of “routine” to indicate that when people act and interact they rely upon both tacit and discursive knowledge. He made a point of noting that the tacit element of our knowledge, that which he called “practical knowledge,” provides a frequently underestimated impact on our day-to-day existence.
for CT scanner technology, which exhibits “relatively fixed and standardized functions and features” (1992, 402), it is generally not appropriate to assume technology is a fixed entity, especially for information technologies, which vary frequently as a result of updates, reconfiguration, learning, and innovative usage. In addition, by treating technology as an object, a duality of agency and structure is reintroduced into the theory, a situation inconsistent with Giddens’ theory. DeSanctis and Poole develop a modified form of structuration theory called Adaptive Structuration Theory (AST). Using this approach, they describe structures as “templates for planning and accomplishing tasks,” where structures are “found in institutions such as reporting hierarchies, organizational knowledge, and standard operating procedures;” they elaborate, “Designers incorporate some of these structures into the technology” (125). However, by treating structure as embodied within technology, they also depart from structuration theory, which argues that structure has “no existence independent of the knowledge that agents have about what they do in their day-to-day activity” (Giddens 1984, 26). Giddens explains:

To say that structure is a ‘virtual order’ of transformative relations means that social systems, as reproduced social practices, do not have ‘structures’ but rather exhibit ‘structural properties’ and that structure exists, as time-space presence, only in its instantiations in such practices and as memory traces orienting the conduct of knowledgeable human agents (Giddens 1984, 17).

By treating structure as embodied within technology, DeSanctis and Poole also re-introduce the dualism between subject and object that Giddens explicitly tries to avoid (Orlikowski 2000). Even Orlikowski originally treated structures as if they were embedded within technology, thereby assuming technology was a material artifact and that structure could have an existence within it (Orlikowski 1992, 2000). However, after some modifications of her theory (2000), she begins to treat technology as a facility (in Giddens’ sense of the term, which means that it is a mode of typification incorporated within actors’ power to employ
resources to engage in action). She explains that when agents use technology regularly, engaging with “some or all of its prescribed properties,” (407) they enact emergent structures. She refers to these enacted structures as “technologies-in-practice.”

The Theory of Structuration is particularly useful for understanding how history, employee knowledge, norms, and power relations influence the use of technology within organizations. It will be helpful in explaining not only the perception of cloud computing by state government recordkeepers, but also in explaining the nature of the record within ARM theory and practice and the reasons why ARM researchers in North America have had difficulties in the past century elucidating professional identity and explaining the boundaries between archival practice and records management practice.

2.1.3. Changing Patterns of Work

2.1.3.1. Increasing Collaboration in State Government

Workplace collaboration has increased across all fields since the 1960s (Moody 2004), largely as a result of the changes in communication structures that have come about with the changing information and communications technologies (i.e., ICTs). Archivist Helen Samuels (1986), remarking upon the increasingly distributed and networked nature of work within organizations, offered suggestions for appraising such networked information for archival purposes. Also around this time, ARM researchers began commenting in more depth about how the changes required new ways of thinking about records. For example, in 1995 Anne Gilliland-Swetland remarked upon then-prevailing perceptions of how networking could change social and organizational environments. She noted that researchers were then hypothesizing “that networking encourages intra- and inter-institutional as well as transnational collaboration; that it is a democratizing force that is breaking down
hierarchical, organizational and scholarly structures; that it can function almost as an
instantaneous, synchronous medium and as a result may be more reflexive, spontaneous,
candid, and informal” (2004, 595).

Increasingly distributed work patterns have also developed in the public sector.
Sharon Dawes and Ophelia Eglene share information about the growing collaborative
 provision of government services, including collaboration “across two or more distinct public
 sector agencies, or between public and private or nonprofit entities, to deliver government
 services” (2008, 2). They point not only to changing models of service provision, but also to
 changing functional and organizational requirements for the performance of government
 activities, revealing that a key motivating factor for increased engagement in collaboration
 has been the introduction of new technology and the perception within the government that
 the “technological expertise necessary to implement public service delivery systems” (2) lies
 in the private, not the public, sector. The strengthened desire to engage in collaborative
 service provision and the increased sophistication of computer technologies, however, has
 also led to a greater distribution of records management and archival functions within the
 public sector.

Kwon and colleagues (2008) report that digital preservation service provision in state
 government is widely dispersed, with the IT unit standing out “across all three branches of
government as holding a significant role in the standards-setting process and in providing
 services related to digital preservation” (188). In addition, parallel standards-setting and
 service-provision efforts occur across the executive, legislative, and judicial agencies in state
governments, often with little or no cross-branch cooperation and communication. The
authors reveal that
units other than the state libraries, archives, and records management (LARM) units have the authority to set standards for digital information created and maintained by government agencies. Even within the areas generally considered to be within the realm of state LARM units – retention and disposal – legislative and judicial agencies are operating independently to a great degree (188).

In the 1990s, public service provision became more distributed and collaboration increased when the National Performance Review recommended that government agencies reengineer government activities, making “full use of computer systems and telecommunications” to revolutionize services delivery (Dawes and Eglene, 1). Dawes and Préfontaine (2003) examine some of the themes that emerged from these collaborative projects. For example, they note that each project had a very distinct set of norms and expectations regarding how key roles and functions should be implemented. In addition, the relationships were quite dynamic and the roles and responsibilities evolved throughout the project, although early performance strongly influenced later “actions, performance and results” (41). In order to effectively work together, such collaborations need a clear institutional framework reinforced by some mechanism of authority, whether that be regulation, law, or formal, contractual agreements between partners. The authors relate, however, that experimentation and creativity are strengthened by the development of informal relationships. Performance, communication, and an ability to adapt to changing conditions are also strongly affected by the nature and flexibility of the technical tools. However, data presents a key issue: data ownership and rights among partners is a continuing conundrum and projects must tackle the question of the “stewardship11 responsibilities of multiple public partners” (42).

11 Information stewardship is defined here as the “active involvement in the management, including the preservation, of digital data for future use.” This definition corresponds directly to Lee and Tibbo’s (2011) definition of digital curation. Stewards are those individuals or groups that engage in stewardship activities.
One might wonder why ownership of public data would be a particular issue among public sector entities; there are two reasons. The first is the question of who “owns” the responsibility for engaging in the stewardship role. When multiple agencies crossing multiple jurisdictions engage in shared services, who should maintain responsibility for the jointly created data? This is an organizational and sometimes a legal question. Kristin Martin and Jan Reagan (2003) point out that with more state government information being presented on the World Wide Web, the boundaries between publications and records are blurring, leading to some confusion as to whether the stewardship of particular documents should reside under the purview of a State Library or a State Archives and Records Section. Their research is specific to North Carolina, but this perception holds true for any state in which the two entities maintain separate spheres of control. In absence of clear-cut statutes that regulate what constitutes a “publication” as opposed to a “record,” state librarians and archivists need to collaborate in order to ensure appropriate stewardship coverage.

Martin and Reagan are not alone in this sentiment. During 2005, the Library of Congress convened three workshops that included representatives from all 50 states. They engaged in mutual discussion and learning about how state governments were handling the digital preservation challenges that existed at that time. During the workshops one of the key ideas upon which participants agreed was that “working across different institutional and professional communities was one of the most important issues facing digital preservation" (Library of Congress, 20).

In fact, working across different communities has also been found to be an issue for any shared services or collaborative arrangement. Sue Richards (2001) and others (Walters 2010) share important prerequisites for successful joint working arrangements. Richards
notes that clearly defined goals, the monitoring of and response to performance metrics, adequate resources, the presence of a clearly identified leader, and mutual support and trust are all crucial elements for collaborative success. Tyler Walters adds that a key ingredient for leadership is a well-defined governance structure that outlines roles and responsibilities for both the leader or leaders and the joint participants. Although Walters favors having the leader be a separate entity that is not a member of the collaboration, H. Brinton Milward and Keith Provan (2006) point out that three different governance models can hold: the lead organization model, the network administrative model, and the self-governance model. The network administrative model is Walters desired form, where a separate leader (i.e., organization or administrative entity) is hired to perform governance activities; the lead organization model is one in which a generally powerful organization within the shared arrangement also acts as the lead; the self-governance model is one in which all members of the collaboration engage in governance (Milward and Provan, 22). Christine Ryan and Peter Walsh suggest that in public sector shared services, clear expectations are essential, and that this requires an explicit agreement between partners in early stages regarding what constitutes acceptable performance and what constitutes “credible” reporting of that information (2004, 623). Depending upon the nature of the service provision and the type of sharing (e.g., cross-agency collaboration vs. private outsourcing of public services) these could be agreements such as memos of understanding (MOUs) between agencies (Reed) or contracts like service level agreements (SLAs) with external vendors (Richards).

The second ownership issue related to distributed service provision is a property rights question related primarily to the rights of access and sharing when one party holds specific property rights over data. For example, the attendees of the Library of Congress’
2005 state government workshops were individuals who engaged in a wide variety of digital preservation activities as librarians (43%), archivists (34%), records managers (8%), IT professionals (13%), and other workers such as clerks and agency information custodians (2%) (Library of Congress, 42). A number of these participants reported that issues with the copyright status of information can be a key factor in a state’s preservation decisions. For example, one attendee “described situations where state digital information was subject to the copyright of commercial contractors that either generated the information on behalf of the state, or that packaged existing state information for sale to the public” (12). Although the state was allowed copies of the newly copyrighted content, it was not allowed to provide public access. In situations such as this, public information stewards need to be aware that they must negotiate with vendors for the rights to the data that the information stewards have created or collected. Another workshop participant reported that California state law does not explicitly declare that public records reside in the public domain. As a result, states are subject to federal copyright ownership requirements. Although Section 105 declares federally created records to be in the public domain, this is not necessarily the case for state created information. Cobb and Palmer report, “Generally, when a government remains neutral, or when there is the absence of a claim of ownership, the provisions of the Federal Copyright Act apply to a publication. Thus, copyright ownership would appear to be in place for agencies per federal law absent any statement for public domain in California law” (2004, 20). Although the California Legislature has tended to lean toward protecting and providing free public access to state created documents, the actual legal status of these records is murky, and the ownership of individual agency records by the agencies has created confusion around appropriate and inappropriate uses of records and documentary materials from a
copyright-standpoint. Such confusion serves as a disincentive to preserve materials. Representatives from yet another state commented that public information that was perceived to be under copyright law in their state was typically not preserved because of “legal barriers to reproduction and access” (Library of Congress 2005).

However, another possible disincentive for responsible stewardship of records in state agencies may reside in accountability concerns. In an environment in which ownership of processes and data is murky, employees may fear reprisal in the event of a disaster or accident if they take responsibility for activities for which accountability has not been clearly designated.

2.1.3.2. Accountability Concerns in Society and at Work

Issues related to accountability have become increasingly problematic in the United States during the past century. Conceptions of “accountability” have shifted, as a result of two primary factors. First, increasingly sophisticated information technologies have changed the structures of work in most professions, and certainly within archives and records management. A substantial body of ARM research since 1970 has focused on the impacts of new information technology on ARM theory and practice (Bearman 1989a; Brothman 1991; Burke 1981; Cook 1983; Cook 1994; Cox 2000a, 2005; Gilliland-Swatland 1992; Ham 1975, 1981, 1984; Hedstrom 1991; Jimerson 1989; Lee 2005; Lytle 1980; Lytle and Dürr 1980; Mason 1981; Plavchan 1980; Stout and Baird 1984; Sundin and Winchester 1982; Weldon 1983; Duranti 2007, 2000). With newer networked technologies, organizations in all sectors (i.e., public, private, non-profit, non-governmental organizations) have engaged in shared services, leading them to question the nature of accountability in these arrangements (Reed 2004). Concepts such as “shared or distributed” (Fitzpatrick 2000, 9) accountability have
been developed to deal with inter-agency collaboration, inter-government partnerships, outsourcing to the private sector, and integrated service delivery (Reed 2004, 143-147; Apro 2006; Anderson, Dovey, and New Zealand 2003). With records, or parts of records, shared between partners from different jurisdictions, or between public and private organizations, determining who is accountable for all or parts of the creation, management, and preservation of those records becomes more complicated, as does determining who is responsible for the different components of service delivery.

Second, according to Giddens (1999) and Ulrich Beck (1992), a wider, overarching change in the structures of society have come about in the past century and have led to increasingly problematic relations between risk and responsibility, both of which are closely connected to accountability. These researchers argue that the structure of modern society has undergone a fundamental change as a result of the increasing general knowledge of science and increasing saturation of complex technologies. Giddens points to the existence of large scale technologies which are so complex that no individual or single group can understand fully either their internal functioning or the possible long-term outcomes of their use. He claims that the inability to assess the risk of many modern outcomes has derived from two features of society which he calls “the end of nature” and “the end of tradition” (1999, 3).

The concept “end of nature” represents the fact that there is virtually no physical environment that humans have not touched. This recent situation has led to a fundamental shift in how people view nature. Previously in history, people worried primarily about how they potentially could be harmed by natural phenomena such as weather, illness, natural disasters or injuries sustained when one ventured into nature. Now, however, social attention has shifted to what impact humanity is having on nature. This has led to a newfound sense of
responsibility toward areas of life that previously appeared to be controlled by “fate.” In addition, as a result of the globalization allowed by scientific advances and nearly ubiquitous technology, we no longer feel constrained by tradition, a situation Giddens calls the “end of tradition.” Increasing knowledge of alternative ways of doing things has widened the realm of choice in our lives. This reinforces our sense of individual responsibility.

However, an unintended consequence of the combination of greater knowledge, highly complex technologies, and new ways of perceiving mankind’s place in the world is that new risks have surfaced as a result of the widespread, goal-directed use of nature. This is occurring at the same time people feel greater responsibility for both themselves and the world. Mad cow disease resulted from the crowded and unhealthy overpopulation in transporting the animals, something no one predicted but was nonetheless a result of human activities. Birth defects resulted from the widespread medically prescribed use of thalidomide by pregnant women. Environmental issues result from increased use of a variety of technologies, particularly those using depletable natural resources. For example, global warming may have occurred as a result of the build-up of greenhouse gases due to the use of gas in automobiles and factories, the production of electricity, the use of nitrous oxide in fertilizer, the use of gases for refrigeration, and deforestation (which would store some of the excess CO² created by the other activities) (National Geographic na).

In fact, global warming presents a relevant example for Giddens’ assessment of the new nature of risk. He calls the risk that results from humanity’s actions upon nature “manufactured risk,” in opposition to an earlier view in which risk was seen as the likelihood of an outcome of nature (or fate) acting upon humankind.
The earlier conception of risk supported the development of insurance as an institution. Insurance is able to provide some security against adverse events such as illness, fires, hurricanes, etcetera, because they occur with enough regularity and because people have a long enough history of their reoccurrences that individuals can calculate their probability. In the “new” world people are faced with risks for which they do not have a history from which to compute the level of risk of “manufactured” disasters. Often, because the adverse events are a result of human interaction with brand new technologies people do not even anticipate their occurrence at all. In fact, in the case of many of these risks, such as global warming, the situation is so complex that some experts do not even believe the risk exists! As a result, the previous safety net of insurance is no longer able to guarantee that one can be insured against these risks for which all people are at least partially responsible.

However, in a world in which multiple, shared, and ambiguous responsibilities exist (but one nonetheless tends to believe some identifiable human agent must be responsible), who is to be held accountable when disaster does strike? The shared sense of unease about this has, according to Giddens, contributed to the increasing litigiousness in society and to an ever-increasing concern with identifying responsibility and allotting accountability.

In fact, one can now see that the direct link between accountability and the actual cause of many disasters has been broken. For example, in the Utah data breech mentioned earlier, the State CIO Stephen Fletcher was forced to resign as a result of the error that caused this breach (Vijayan 2012). The “cause,” of the breach, however, was a complicated set of correlated events. An individual system administrator failed to change the default password, but he or she was several managerial layers removed from Fletcher. In addition, the server that was provided by an independent contractor was supposed to have been
supplied with encryption safeguards, but was not. Finally, according to Fletcher, adequate funding for security was consistently denied for budgetary reasons (Williams 2012). Fletcher said that he was unable to obtain funding to adequately secure all systems prior to the breach because it was not considered as important as other state priorities. In addition, even if he had received permission for additional funding when the server was purchased, the purchase occurred too late in the 18-month budgetary cycle for funding to have been received in time to avert this disaster. On top of these contributing factors, “cyberattacks targeting Utah…spiked by [an unprecedented] 600 percent” during the four months prior to the incident, Fletcher said in April, 2012. In fact, the governor’s office reported that there are “nearly a million attempts every day to infiltrate the state’s IT network” (Williams 2012).

When Governor Herbert was asked about Fletcher’s forced resignation, however, he replied, “There needs to be some accountability for the lack of oversight and leadership” (Deseret News 2012). This is a situation in which an individual who “was not directly responsible for the data breach” (Goldman 2012) was nonetheless held accountable.

The expected link between responsibility and accountability often does not hold in the modern risk society. Robert Behn (2001) reports that this is true for modern government service in general. He comments that at one time if a president asked someone to work in his administration, the general feeling was “you don’t say no to the president” (2001, 16). He adds, however, that the first President Bush had difficulty filling administrative positions because candidates were concerned about “the sheer complexity of the federal ethics laws” and “the fear that a simple, honest mistake could lead to a public nightmare” (16). Confusion about accountability has led to disincentives to engage in public service, and may be leading
to hesitation within public organizations to take responsibility for activities that do not have clearly designated ownership of responsibility or clearly understood processes.

In fact, determining what the term “accountability” even means has become a difficult endeavor, since the term is situational (Yakel 2001; Apro 2006; Mulgan 2002; Reed 2004) and often vague. According to Mulgan (2002, 3), “it needs to be specified in context: who is accountable to whom and for what?” For many authors the notion of accountability includes four components: a person or group who will be held accountable for their activities, a (different) person or group that has the right to receive information regarding the past actions of the accountable party, the right of the information receiver to hold the accountable party up to a set of specific standards, and the existence of someone who has the power to reward or punish the accountable party if the accountable party does not meet those standards (Anderson, Dovey, and New Zealand 2003; Behn 2001; Heeks; Lerner and Tetlock 1999; Mulgan 2000, 2002; Schedler 1999).

Behn remarks that a crucial component of accountability is the ability to impose sanctions upon the individual or group that is held accountable (4). He criticizes definitions that do not explicitly include the notion of sanctions, such as those of Kearns (2011) or Barata and Cain (2001). Kearns deliberately offers a broad definition of accountability, suggesting that the narrower definition focuses specifically on compliance needs but fails to include “responsiveness to the needs and interests of stakeholders inside and outside of the organization in an effort to maintain the credibility of the organization and sustain public trust in its activities” (198). He says most stakeholders would prefer that those who are held accountable will behave in a manner that is not just oriented toward meeting the letter of the law, but also towards satisfying and responding to the stakeholders’ expectations, values and
perceptions. Barata and Cain define accountability as “the capacity of the people, through an
elected legislative body, to measure and verify the performance of government, particularly
with regard to its obligations under this ‘contract’” (2001, 248). Other authors point out that
the expectations to which someone will be held accountable need to be shared and explicitly
agreed upon (Fitzpatrick 2000) or that there needs to be a “general recognition” of the
standards of accountability (Grant and Keohane 2005, 29).

2.1.3.3. Accountability in the ARM Literature

In archives and records management, accountability is often mentioned but rarely
explicitly defined. However, many ARM theorists have indicated that a primary goal of
archives and records management as an occupation is to ensure organizational accountability
(Bearman 1995; Eastwood 1993; Cook 1994; Cox 2000a; Duranti 1994; Meijer 2001a; Reed
2004), or what McKemmish refers to as the “recordkeeping-accountability nexus, which
flows from an emphasis on records as evidence of social and organisational activity” (2001,
338). Joseph, Debowski, and Goldschmidt (2012) argue that recent shifts in recordkeeping
responsibilities suggest that increasing concern with accountability will characterize the
recordkeeping professions in the future.

Albert Meijer (2001b) comments on the lack of a clear definition of the term in
academic literature. He chooses a definition put forward by Barbara Romzek (2000) that
defines accountability as “a relationship in which an individual or agency is held to answer
for performance that involves some delegation of authority to act” (Meijer, 259). The theme
of assignment, or delegation, is a common one within the accountability literature, and a

12 By “contract,” they appear to be referring to the notion of a social contract, an ethical theory that suggests that
people consent (either explicitly or tacitly) to surrender some of their freedoms by submitting to the authority of
the state in exchange for protection of their remaining rights.
number of authors suggest that it is a key component of accountability (Fitzpatrick 2000; Grant and Keohane 2005; Meijer 2001b; Romzek 2000). Noting that information sharing is a component of accountability, Meijer points out that the digitization of records most strongly affects the information sharing phase of accountability processes. During the information phase, a “forum reconstructs the actions of an individual or agency in order to form an opinion” (260). Accurate records and a systematic records management program are important because they are required to support accountability goals. Meijer’s discussion of the relationship between records and accountability suggests that he considers records to be artifacts that can be examined and interpreted in order to allow one to “reconstruct” a series of past actions – they become representatives of objective scenarios and facts. Accountability for him is thus a “thing” that is an outcome of reconstructing a set of interconnected facts as represented by records.

Barbara Reed (2004) speaks of the idea of shared or distributed accountability within public sector shared services environments. She accepts Mulgan’s definition of accountability as “a relationship in which one party, the holder of accountability, has the right to seek information about, to investigate and to scrutinize the actions of another party, the giver of accountability” (140). Although Mulgan also includes the notion of sanctions in his definition, Reed does not mention this aspect of accountability. Her treatment follows a more legalistic notion of accountability, whereby it is defined as a legal right that can be held or assigned. She does not specifically discuss the nature of the resulting responsibility on the part of the accountable entity, although she does focus on the ways in which accountability is shared among those individuals in real-world public sector settings.
Elizabeth Yakel (2001) uses a very different approach in her analysis of accountability, criticizing some authors’ tendency to treat accountability as a thing. She adopts ideas from social constructionism\textsuperscript{13} and from Weick’s ideas of organizational sensemaking (1995) to develop a process-oriented theory of accountability that she uses to analyze radiological reports from a large tertiary care medical center. She introduces her analysis by discussing the tendency in ARM literature to use the notion of warrant, which she defines to be “law, customs, standards, and professional best practices accepted by society and codified in the literature of different professions concerned with records and recordkeeping” (2001, 233). She notes that if one uses the notion of warrant it needs to be extended to include uncodified practices and tacit knowledge because it is often the case that these uncodified practices and knowledge conflict with, compete with, or are melded into local practice in ways that do not directly relate to the written warrants. She treats accountability as both a process of accounting and as a form of narrative, which allows individuals to justify their actions within their specific organizational contexts. She also allows ethical beliefs, morals, and responsiveness to be added to the narrower compliance-oriented definition of accountability, arguing that records are intermediaries in the accounting process. As intermediaries, records both shape the justifications of past behavior and, through sensemaking processes, lead to expectations of similar future behavior, thereby shaping the nature of accountability over time. She asserts that because every individual context has its own unique accountability process, multiple accountabilities exist.

\textsuperscript{13} Social constructionism attempts to explain how social phenomena are created, institutionalized and reproduced by human interaction. It does not take so-called “reality” as given and objective but rather, treats objects as subjectively perceived and their apparently objective characteristics as jointly created via social interaction and intersubjective agreement.
The recognition that accountability is both situational and evolutionary is important, but when it is placed within the context of social constructionist theories, one treats it as a social entity that is inherently subjective and separate from so-called objective reality. Nonetheless, recognizing the processual nature of accountability environments provides an important step in developing a theory of accountability that can help found ARM professionals’ identity as supporters of accountability. If one treats accountability as a social structure in Giddens’ sense, one can capture this sense of the term. As a social structure, accountability exists only when enacted on a routine basis by human agents, and their situated actions constitute it and are simultaneously shaped by it. It is a particular enactment of rules and resources that are regularly treated as a property of social systems. It is founded on the normative and meaning-laden structures of each given social environment and reflects the dominance structures found within that environment. There are not “multiple accountabilities.” Rather, there are multiple instantiations of the social relations that help make up the repetitive structure we interpret as accountability.

The fact that accountability has received so little explication from a group of professionals who claim it is a primary basis for their profession is somewhat troubling, however. The International Organization for Standardization’s ISO 15489-1 does define the term as the “principle that individuals, organizations, and the community are responsible for their actions and may be required to explain them to others” (2001a, 2).\(^\text{14}\) However, ARM professionals have alluded to their allegiance to accountability for much longer than this.

\(^\text{14}\) The ISO 15489-1 standard specifies what it means to perform records management. According to McLeod, it was designed in order to standardize international best practices.
standard has been in existence, so the authors of the standard apparently interpreted what they believed to be the already-common use of the term by recordkeeping professionals. Also, as Yakel pointed out, accountability varies from organization to organization, depending upon a combination of written warrants, tacit knowledge, and ethical and moral orientation. By treating accountability as self-evident, ARM professionals may fail to recognize the conception of accountability held by other occupational groups within (or outside) the organization in which they work. To the extent that their conceptions are fully in sync with those of other occupational groups with whom they share records responsibilities, this may not be problematic. Without a clear assessment of various individuals’ and groups’ conceptions within concrete settings, however, one cannot know whether the “ARM view” of accountability maps well to the view of other occupational groups with whom they correspond.

Accountability itself, as a structure in which dominance relations are enacted both tacitly and explicitly, is likely to be a source of political struggle within organizations because it involves an asymmetry of power due to the delegated responsibilities and ability to provide rewards or threaten sanctions (DeSanctis and Poole; Orlikowski 1992). It is required in situations in which someone is granted the freedom to perform actions on behalf of another individual or group; in other words, power over resources is delegated to someone. In return, the accountable party takes on the responsibility of ensuring that he, she or they can provide an explanation, or accounting, of their activities to show that they have acted in good faith and according to mutually understood expectations (Parkinson 1993). Understanding more about the perceptions of dominance and delegation within an organization could

---

15 ISO 15489-1 was first published in 2001.
therefore help to clarify the extent to which ARM professionals’ understanding of this term maps to the understanding of other occupational workers who are co-stewards\textsuperscript{16} of the information that is presumably evidence of accountability or lack thereof.

Barbara Craig describes archivists’ accountability:

A legitimate question for archivists to ask would be how their professional accountability is expressed, especially as it relates to the assertion of competence to do appraisal and their largely understood power to make keep-and-destroy decisions responsibly, with a view to serving the needs of society as a whole. Clearly, the broader notion of “accountability” should include a dimension of archival accountability, that is, both a recognition of the principle and the provision of a means for rendering an account for the responsibilities to the profession and to society for the decisions we make on records (2007, 28).

She also asserts that the results of a survey she conducted indicate that “there is ambiguity in the archivist’s practices and beliefs concerning when and how the society they serve, as represented by the general public and users, have access to appraisal assessments and decisions” (28).

Terry Eastwood argues that “Accountability is a property of the institutional structure of a democracy” (69) [my italics]. Unlike Yakel, he distinguishes accountability from responsiveness by stating, “responsiveness is a measure of how much accountability an institutional structure permits,” being “a consequence of interaction within such structures” (69).

Wendy Duff (2001) agrees that archivists are responsible for ensuring accountability and remarks that “accountability depends upon access to trustworthy records” (230). This raises the question of what roles and responsibilities archivists must accept in order to fulfill their professional functions. The issue of ARM roles and responsibilities and ARM identity will be covered in the next section.

\textsuperscript{16} For example, IT workers could be considered one co-steward.
There are a number of archivists who dispute the notion that the primary function of an archives is, or should be, accountability (Dirks 2004). These writers argue that focusing primarily upon the accountability and evidence of organizations creates a risk of forgetting that archives also exist to support social memory and historical research. According to this view, archivists must therefore always consider the historical value of records as well as their potential evidentiary value. However, James O’Toole points out that thinking of archives as supporters of accountability is not inconsistent with thinking of them as keepers of social memory insofar as they do in fact support historical accountability (O'Toole 2004). He says that the notion of historical accountability leads ARM professionals to examine records for possible ethical and moral issues that could allow the records to hold past individuals, groups, or nations accountable for actions performed in the past, such as war atrocities.

ARM researchers could benefit from more awareness of how they interpret accountability in general and how they are accountable to supporting it within their organizations. To learn this requires venturing into the organizational settings of practitioners to understand better how common understandings and practices lead various recordkeeping professionals (including, but not exclusive to, records managers and archivists) to identify with their roles and responsibilities.

2.2. The Workers

2.2.1. ARM Occupational Identity in the Pre-Computerization Era

Yakel points out that accountability comprises identity work, whereby an individual develops his or her sense of identity as accountability is given and requested. The delegation of responsibility and consequent accountability expectations in particular functional areas involve selecting both for personal characteristics and for occupational skills and knowledge.
The interaction between competence, work achievement, choice and social relations within
the organization form occupational identity (Phelan and Kinsella 2009).

Historically, the occupational identity of ARM workers has been contested ground
among North American archival and records management workers and researchers, who
began questioning their occupational identity long before computerization established a firm
foothold in the field. Within this pre-computerization time frame, one can find a tension
between the roles and responsibilities associated with the fields of records management and
archival administration. According to some, this tension is related to the historical
development of archives and records management within the American context. Richard
Cox (2000b) states that modern records management principles developed between 1943 and
1985, a period that straddled both pre-computerized and post-computerized ARM
environments (although Cox does not remark on this fact). He argues that the development of
archives as a field of occupational practice is linked to the development in the late nineteenth
Gilliland-Swetland (1991) agrees, suggesting that Progressive ideology led to an overarching
belief during the early part of the twentieth century in the efficacy and superiority of
“scientific” modes of management and organization and that these rational modes of
organization strongly impacted both the development of and the tools for organizational
management, including the methods of recordkeeping and the development of the archival
profession. His argument is consistent with that of Joanne Yates, who notes that the desire
for internal control developed from 1850 through 1920, where ‘control’ is comprised of “the
mechanisms through which the operations of an organization are coordinated to achieve
desired results” (1989, xvi). During this timeframe, claims Yates, formal methods of internal
communication in organizations grew in quantity and complexity, eventually leading to “control through communication” (xvii). She paints a picture of increasing reliance on wider scale communication media and changing genres as a means for achieving corporate goals, in place of previous oral communication and less structured written communication. Organizational communication became an objectified managerial tool, which allowed greater control over operations and increased scale. Systems supplanted individuals and many leaders and social scientists saw the ability to engage in system thinking as a primary requirement for organizational success.

Terry Cook suggests that it was this “rationalization and bureaucratization of office work” (1994, 408) that led to the structures of modern records management and archival duties. By using lower-paid, female secretaries with typewriters, the senior administrators who made the important business decisions no longer had to involve themselves with the creation of records. Thus, records and their creation were separated from the act of decision making that engendered them. Archivists, whose domain (as “secretaries”) had previously involved them in both decision making or advisement responsibilities and the recording of the resulting decisions, thereby fell from relatively high-powered positions to caring for huge volumes of records on behalf of the real decision makers. The need for managing these huge volumes, and the recognition that the records must be protected for those in power, led archivists to objectify the record, according it a status akin to a holy relic that must be preserved because of its connection with past decisions and actions of the real power wielders.

In fact, after the rationalization of communication was successfully achieved, from the 1940s well into the 1960s, archivists debated the nature of their profession, focusing
heavily on the concepts of professionalization and efficiency, and on managerial principles. In the late 1960s, Frank Evans (1967) argued that archivists working within the profession during the early days of the National Archives in the 1930s did not view their records management duties as a separate undertaking from their archival duties. He commented that for the previous two decades the professional literature was filled with pleas for a closer relationship between archivists and records managers. This suggests that the literature from the mid- to late-1940s recognized that archivists in general perceived there to be two separate professions – archiving and records management and also perceived that these two professions did not work together as much as would be desirable. In fact, the articles Evans himself cited reflect a uniform discomfort with the question of how distinct archives as a profession is from records management as a profession.

Luke Gilliland-Swetland, Richard Cox, and others (Berner 1983; O'Toole 1990; O'Toole and Cox 2006) note that this discomfort reflects an ongoing debate about the nature of archivists in society. Luke Gilliland-Swetland argues that the dispute, framed in terms of a dichotomy between a “historical manuscripts tradition” and a “public archives tradition,” was the consequence of a group of strong-minded archivists who followed the ideas of Margaret Cross Norton during the 1930s. According to Gilliland-Swetland, the public archives tradition advocated the concept of provenance.17 This was in contrast to the historical manuscripts tradition that favored practices such as “item-level descriptive control, the imposition of predetermined classification schemes for cataloging purposes, and the reliance on several types of nonintegrated access tools” (161). In addition, the public archives tradition focused upon the efficient management of public records whereas the historical

---

17 At this time, archivists generally defined provenance to be “‘keeping records ‘as nearly as possible in the same order or classification as obtained in the offices of origin’” (Gilliland-Swetland 1991, 161).
manuscripts tradition focused upon maintaining the primarily private records of individuals for historical purposes.

Rebecca Hirsch (2010) suggests that the dichotomy between the historical manuscripts tradition and the public archives tradition is ideological in origin, and founded on Norton’s belief in “archives as legal records” and her belief that “the historical profession should have no influence whatsoever on their treatment or retention” (67). She, like Gilliland-Sweetland, points out that the rules of “provenance, original order, public access, and government support” (66), primary principles of the public archives tradition, came directly from the scientific historians of the early twentieth century, who adopted the precepts of nineteenth century European archivists. Gilliland-Sweetland and Hirsch both argue that the ideological disagreement between the public archives tradition and the historical manuscripts tradition still affects the archives and records management professions, separating them. The former claims that modern debates around professionalization and certification ultimately collapse into “two traditionally competing ideals of the archivist as professional: humanist historian-scholar or expert documentary manager” (171). Hirsch maintains that decisions such as Mark Greene’s choice “to make professional identity the topic of his presidential address [during] the 2008 SAA annual meeting shows that the question is not settled” (69).

An examination of articles from the professionalization debates (Boltenko 1985; Hull 1980; McCrank 1979; Spadoni 1983-84; Taylor 1977) supports that assertion for the period of time during which the debates occurred - the late 1970s and 1980s. Whether it is still the case, however, does not seem so clear.

Evans traced the development of records management as a professional activity of the government from 1941, when SAA renamed its Committee on Reduction of Archival
Material to “Committee on Records Administration” and when the National Archives instituted a “records administration program” (45). He provided a variety of cases showing disagreement between those who believe that archivists and records managers provide essentially the same function and those who argue that they either serve two different ends, or require two different skillsets in order to perform their jobs. He closed his article with the argument that records managers determine both the quality of archives and the nature of society’s involvement with archives, urging both archivists and records managers to accept the commonalities that bind them together.

Thus, one can see that disagreements about the appropriate distribution of roles and responsibilities between archivists and records managers occurred even before computer technology affected any aspects of the occupational identities of these two groups of workers. In addition, this literature supports Yusof and Chell’s previously mentioned contention that the archival literature provides one source of lineage for records management activities and the RIM and MIS literatures provide a different lineage.

2.2.2. Occupational Identity of ARM Workers in the Computerized Era

Many of the early articles discussing computer technology focused on pragmatic issues related to the practical application of computer technology to already-existent archival practice (Clarke, Edelglass, and Williams 1975; Cunha, Poole, and Walton 1977; Edelglass, Strom, and Turner 1977; Kula 1977; Poole 1977), with very little explicit recognition that archival practice may change dramatically in the wake of the new technology and with very little discussion of either records management activities outside of archives or of the professional identity of the two occupations. Rather, the primary concerns with computerization in the 1970’s revolved around techniques for engaging in more pragmatic
tasks like automated indexing and the development of systems that would allow automated
indexing (Hickerson, Winters, and Beale 1976; Torchia 1976), “automated data processing”
(Fishbein 1975; Rieger 1976), the appraisal of machine-readable records (Dollar 1978;
Robbin 1979), and automated finding aids (Bearman 1979; Calmes 1979; Dewhitt 1979;
Dollar 1978). Dewhitt noted that in a survey of archivists in the United States and Canada the
crucial issue related to machine-readable records was that of ensuring control. Dewhitt was
also one of the earliest to point out what has since become an ongoing theme in ARM
literature - difficulties in communications between archivists and the technical personnel
upon whom they must rely when constructing and maintaining automated systems.

A few authors, however, did express concerns with the potential impacts of
computerization on professional identity, such as Regehr’s “Counterpoint” letter to
Archivaria, in which he suggested that automated indexing could lead archivists to become
mere “technical officers” (1976). Likewise, Gerald Ham (1975) wrote his now-classic “The
Archival Edge,” an influential piece that discussed the impacts that technological
advancement, in addition to other factors, was likely to have on the role of the archivist. He
suggested that archivists would need to develop more subject-specific archives, would need
to build state archival networks to counteract the disorganization of records, and would need
to develop more urban archives, using historical techniques to keep a clear view of
continuing changes in society.

In the 1980s, authors were still concerned with pragmatic issues such as the appraisal
of automated records (Cox and Samuels 1988; Naugler 1984) and with indexing and retrieval
(Michelson 1987), but the discussions were becoming more nuanced and often reflected
concerns about provenance- versus subject-indexing (Lytle 1980; Lytle and Dürr 1980;

In addition, a number of researchers began to contemplate how computerization would affect occupational identity, and questioned how technological change might impact archival theoretical foundations (Bearman 1989a; Burke 1981; Cook 1983; Ham 1981, 1984; Jimerson 1989; Lytle and Dürr 1980; Peterson 1984; Weldon 1983). A number of researchers expressed anxiety about the status of the archival profession and archivists as technology advanced, such as Sundin and Winchester (1982), who discussed the characteristics needed by an “intelligent database” and the possibility that such a database would be capable of replacing archivists. At the 1981 Annual Meeting of the Society of American Archivists, Frank Cook suggested that advances in technology had shaken the confidence of archivists to such an extent that they may not “even know what it is that we are supposed to be preserving.” He asked, “What is an archivist?” (1983, 11) and urged that archivists begin to recast archival theory to ensure that it still mapped to the changing world of information.

Gerald Ham continued his influential conceptual argument that the environmental changes brought on largely by technological change and a greater volume of information was leading to a new era for archivists, now referred to as the “post-custodial era” (1981, 208), in which “every individual…will become his own records manager; and scheduling, as we now know it, will be difficult if not impossible” (209). He called for a review of the foundations of the profession, pointing out difficulties with the traditional concept of provenance in
electronic environments. He asked, “...how does the traditional concept of provenance apply to a data base management system where information is stored without regard to administrative or functional context? Is not the notion of original order irrelevant to records stored in a random access file?” (1981, 209) He also argued that archivists would need to become activists as an increasingly greater need for early selection of materials became standard. Both of these themes would be picked up by postmodernist archival thinkers (Brothman 1991; Brown 1991-92; Cook 1994) and the Australian continuum theorists (Hurley 1995; Iacovino and Reed 2008; Iacovino and Todd 2007; McKemmish 1994, 2001; McKemmish and Upward 1994; Reed 2008; Upward 1996, 1997, 2000) in the 1990s and 2000s. The work of Ham and others pointed to the changing social environment that coincided with the technological change. For example, Peterson (1980) and Kirby (1986) both noted the increasing consideration that archivists would need to give to the tensions between privacy and freedom of information as the legal framework struggled to catch up with the technological changes.

Among the concerns about the changing role of the archivist in society came more calls for archivists to be open to and conversant with contemporary information technology (Plavchan 1980) and more able to collaborate with related occupational personnel, such as historians, records managers and information technologists – that is, if they wished to remain relevant to society (Ham 1984; Mason 1981; Stout and Baird 1984). For example, Margaret Hedstrom (1998) made one such call and simultaneously provided help for archivists to begin to develop conversance with IT. She wrote a manual that explained the steps one needed to follow to “locate, appraise, accession, process, and preserve machine-readable records.” A component of the openness to change, argued several authors (Dearstyne 1987;
Jimerson 1989; Lytle and Dürr 1980; Mason 1981), was the need to pay more attention to the nature and needs of archives’ users.

In 1985, when Sidney Levy and Albert Robles conducted a survey of resource allocators’ perceptions of archivists, they found that others perceived that “archivists’ professional identity is unclear,” according to Randall Jimerson (2000, 6). Jimerson noted that several aspects of this confusion of identity among American archivists can be seen in the thematic content of debates within the field. While early debates centered on the relationship between archives and records management and the historical manuscripts and public archives aspects of the profession, later debates have centered around professionalization itself and around the impacts of technological development on the field and its practitioners’ roles and responsibilities.

Cox (1994) noted that the disagreement about professionalization and in particular, certification, standards, and accreditation, is partially due to the archival community’s origins in both academic history and librarianship. He commented that records management “has had a separate identity for more than thirty years” and that some records managers have been “arguing for a merger and integration of their field with information technology and resources management” (74). Cox (2000a) remarked that the “predominance of electronic records has increased the span between humanists and technocrats in the field, fueling the continuing controversy between those who see themselves as historians and those who see themselves as information scientists” (178). Indeed, in the early 1990s Terry Cook had already noted the presence of a new breed of archivists who often performed both archival and “high-tech” computer processing tasks (1991-92).
Cox laments the apparent disjunction between the professional identity of archivists and that of records managers, referring to it as the result of an “unfortunate schism” (34). He believes this schism leads both sets of professionals to fall short of their responsibility to manage records properly. He says, “…archival appraisal and records scheduling are closely related (in fact, one cannot really succeed without the other) and…such functions have some common links in theory and knowledge” (21). For Cox, inadequate education supports the continuing professional division between archivists and records managers. He asserts that the education of archivists and records managers must stem from a similar theoretical base, saying, “It is unwise to develop a records management program lacking an archives component because it skews basic principles such as the life cycle or records continuum concept and harms an organization's ability to administer its most important records - current and historical, active and inactive” (2001, 20). He discusses archival education in the United States and remarks upon “immense gaps in knowledge about virtually every aspect of archival work...affecting the ability of each professional records manager to do much more than guess about what decisions they are making within their institutions” (2000a, 233).

Cox’s study at the beginning of the twenty-first century followed more than a decade of efforts on the part of archivists to cement their place as records professionals. During the 1980s, driven by a desire for greater professional status and clearer standards to aid in delineating the boundaries of the profession, many archivists began to take part in the previously mentioned “professionalization movement” (McCrank 2001, 589). Besides the certification offered since 1989 by the Academy of Certified Archivists, in 2002 the Society of American Archivists’ (SAA) revised their recommendations for the components of a Masters of Archival Studies degree and published them as Guidelines for a Graduate
Program in Archival Studies (2002). These recommendations broke coursework into categories of “core” and “complementary” knowledge, where core knowledge “provides the theoretical and practical basis necessary to work as a professional archivist” and complementary knowledge “introduces students to other disciplines, knowledge of which will deepen their understanding of archival work and support its accomplishment” and “allows students to specialize in specific aspects of archival work or to function in truly cross-disciplinary settings” (2002a). Core knowledge includes knowledge of archival materials and functions, knowledge of the profession, and contextual knowledge. Complementary knowledge includes knowledge of allied professions and disciplines such as information technology and organizational theory and information about specific aspects of archiving such as records and information management, reference, and appraisal and acquisition.

Six years after Cox’s examination of American archival education, Jeannette Bastian and Elizabeth Yakel (2006) reviewed the state of archival education in North America. Like Cox, they reported that a large number of programs offer only one archives core course, and the average number of core courses is 3.5, “close to the standard three-course sequence that was considered state of the art in the 1970s and early 1980s” (141). This three course requirement was initially set forth in SAA’s 1977 guidelines, and included “a course in basic archival theory, a practicum, and an independent study” (Riggs 2005, 63). Bastian and Yakel related that even among the courses offering a similar topical focus, across the different schools “there was little indication that professional education has reached agreement on its core literature apart from the use of the SAA Fundamental Series” (149). They further noted

---

18 They were updated again in 2005 and in 2011 (SAA 2012).
that conceptions of professionalization generally include the formalization of the cognitive base of a discipline, as evidenced through a core body of literature and standardized educational offerings. They suggested that their study, which revealed a lack of both a standardized core body of literature and consistent standards of education, shows that the field of archival science has only a tenuous hold on any common standards of professionalization. However, because they did not offer a comparison to other disciplines and fields that are readily and clearly recognized as “professions,” it is difficult to compare just how tenuous this hold is in comparison to that of other professions.

ARMA International, like SAA, offers educational opportunities to practitioners of RIM. They offer a certificate training course, online classes, web seminars (ARMA International 2012a), and a certification for an individual to become a Certified Records Manager (ARMA International 2012b). They also offer a certificate in “Generally Accepted Recordkeeping Principles,” or “GARP” (ARMA International 2012c). These programs focus on a wide range of principles, from general management principles, records creation and use, records systems, storage, and retrieval, records appraisal, retention, protection, disposition, and technology, to (for the GARP certificate) understanding the ARMA maturity model for information governance.

This historical look at the development of archives and records management work indicates some contradiction between, or at least confusion with, the structural logics shaping the outlook and professional duties of records managers and those shaping the outlook and professional duties of archivists, where the logics are the “historical patterns of material practices, assumptions, values, beliefs, and rules by which individuals produce and reproduce their material subsistence, organize time and space, and provide meaning to their social
reality” (Thornton and Ocasio 2008, 101). These patterns include the underlying dominance patterns that structure resource use within the organizational environments in which occupational workers reside.

In the archives and records management fields, the historical manuscripts tradition, steeped in the maintenance of private records, focused upon the goal of furthering historical research and developed tools and techniques of management that came from librarianship and from the practices of academic historians. Alternatively, the public archives tradition focused on a distinctly different set of goals, usually expressed in terms of providing an efficient management of records in order to establish and maintain control of the business of government. In order to provide efficiency, this tradition expected that records managers would be part of the management teams of government organizations and would focus on records disposal as a means for dealing with an otherwise uncontrollable growth in records. The efficiency focus treated records managers not as custodians of culture so much as service-oriented professionals geared toward helping their agencies to understand how to handle records, and when to dispose of them or send them for permanent storage to the archives. As computerized techniques became more common in organizations, an expectation developed that people practicing recordkeeping activities would need to have a better understanding of these technologies and how they would influence their own work roles and responsibilities. However, the rapid shifts in technology in the past decade toward increasingly networked and interactive technologies raise the question of how acting recordkeepers gain that knowledge within their own organizational contexts.
2.2.3. Current Thoughts on ARM Identity

As mentioned earlier, Cox says that the goals of records management and archival practice are essentially the same, with focus on particular activities being the main difference between the two occupations. The continuum theorists, on the other hand, deny an essential distinction between the activities of archivists and records managers (Hurley 1995; McKemmish 1994; Upward 1996, 1997; McKemmish and Upward 1994). The continuum model questions the life cycle theories that prevailed in the U.S. for more than forty years which suggest that a record is created (“birth”), goes through a sequential series of changes and eventually is either disposed of (“death”) or sent to an archives. For the continuum theorists, the conceptual separation between the “active” and “inactive” record is inappropriate, as is the separation in responsibilities between the records manager and the archivist. Both categories of workers are “recordkeepers.” Records do not move in a sequential process from birth to death (or preservation). Rather, there is a continuum of possible actions that can be taken upon a record, and preservation should be considered and planned at the time of creation. Both archivists and records managers share similar goals. Within organizations, there is no guarantee, in fact, that units will send all their records of continuing value to the archives, and the need to provide access to citizens may well occur long before a record reaches the archives. However, the Australian recordkeeping context and history differs from that within the United States. A number of archivists and records managers have weighed in on the question of whether the two occupations serve the same purposes, and general agreement has not been found.

---

19 Gilliland-Sweetland (2005) does comment that in Australia, however, the community of “archivists and records managers” does not entirely coincide with the community of “recordkeepers.”
Another relatively new concept that attempts to bridge boundaries between records managers, archivists, and information managers is that of digital curation. The term “digital curation” has been used to reflect the lifecycle activities related to research or e-science data (Yakel 2007), although there are no inherent conceptual reasons why this term should exclude non-research data management and preservation. According to Neal Beagrie (2006), the term was coined to avoid some of the confusion related to the ways that other terms such as “digital preservation,” “digital archiving” and “records management” vary across different professional circles. Beagrie posits that it first arose in 2001 during the "Digital curation: digital archives, libraries and e-science seminar" sponsored by the Digital Preservation Coalition and the British National Space Centre to establish dialogue between archivists, library and information management specialists, and data managers in e-Science.

JISC defines digital curation to be “maintaining and adding value to a trusted body of digital information for future and current use; specifically, the active management and appraisal of data over the entire life cycle” (JISC 2006, 1). The Digital Curation Centre (DCC 2010) describes the components of digital curation as “maintaining, preserving and adding value to digital research data throughout its lifecycle,” and asserts that it comprises a number of actions that promote the maintenance, accessibility, and preservation of data. Such actions include conceptualizing and planning the creation of digital objects, creating or ingesting the objects and assigning metadata to them, ensuring access, appraising and selecting which objects are to be curated and preserved for the long-term, disposing, transferring to a long-term archive or repository, preserving, reappraising, storing, and transforming (e.g., through migration). Thus, “digital curation” refers to the range of activities required to manage and maintain digital objects from the conceptualization of their
creation until their eventual disposal or preservation, and includes activities that ensure continued maintenance and potential access after moving to an archives.

Lee and Tibbo (2011) accept Yakel’s definition (2007) of digital curation, that is, “the active involvement of information professionals in the management, including the preservation, of digital data for future use” (Lee and Tibbo, 124). They take an explicitly postcustodial position, recognizing that although archivists can and sometimes do work entirely within the realm of archives and digital repositories, a great deal of recordkeeping activities and information management requires continuing curation across the lifecycle of the information and occurs in organizational environments that are not exclusively oriented towards preservation. As part of their continuing mission to develop an appropriate educational program for information stewards working in such environments, they have developed a “Digital Curation Matrix” that highlights the skills and competencies necessary for professionals who engage in the process of curating information objects, including records (Lee and Tibbo; Tibbo and Lee 2012; DigCCurr Project na).

2.2.4. Recordkeeping Roles and Responsibilities

The literature on records management shows some divergence of opinion regarding the more specific goals of records management, or records and information management (RIM) as some call it, as a form of activity. For example, David Stephens, using the latter designation, claims that the main purpose of RIM is “to provide better management of organizational records systems and the information they contain” (2007, 1). He says that RIM is a specialized discipline “primarily concerned with the systematic analysis and control of recorded information, which includes any and all information created, received, maintained, or used by an organization in accordance with its mission, operations, and
activities” (1). Cox, like many other archivists, argues that records management exists to “support accountability, the protection of crucial evidence, and the nurturing of corporate memory” (2001, 13). He focuses on the record’s role in ensuring accountability and providing evidence, while keeping in mind that records are created for the operational purpose of documenting a transaction or decision. Similarly, Anne Gilliland-Swetland (2005) notes that by the late 1980s ARM workers largely accepted that recordkeeping is geared toward ensuring the evidentiary quality of records.

The *ISO International Standards 15489-1: Information and Documentation-Records Management* formalizes this focus on records as evidence by defining a record to be “recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context, and structure sufficient to provide evidence of that activity” (ISO 2001a, sec. 3:3). Philip Bantin (2008) asserts that treating records as a consequence or product of an event highlights the activity of defining “more precisely and conceptually when the record is created by the business event or personal activity,” thereby placing “greater emphasis on understanding functions and processes and on precisely linking the records to the events that created them” (27). David Bearman (1994) suggests that ARM professionals identify generic forms of documentation that are associated with the various functions of the organization, using “the relationship between these functions and forms to ‘schedule’ records,” that is, to “determine how long the information in each needs to be kept” (15). He stresses the priority of establishing functional requirements for electronic records management. Thus, the ARM professional must know what data comprises a record and be able to identify the records of his or her organization. He or she does this on the basis of organizational policy, “legal requirements, known needs
for the records, and calculated risks associated with their destruction” (17). He or she must also “articulate criteria for retention” of these records (17), or in other words, engage in appraisal. In addition to assessing what constitutes a record, determining the relationship between organizational functions and documentary forms, and appraising and selecting records for retention and disposal, Bearman insists that ARM professionals must take part in the assessment of recordkeeping systems and collaborate with other information professionals to develop appropriate technological methods for ensuring that the evidentiary quality of the organization’s records is maintained. To do this, he or she must be able to take part in evaluations of “hardware, software, storage media, and documentation techniques” (21).

Bearman argues that the ARM professional must engage in four separate tactics to ensure that accountable management of records occurs: policy, design, implementation, and standards. These tactics must be grounded in a firm understanding of the risks to the organization’s records, since “the risk factors change with each system migration” (29). Policy provides the guidelines for use of information systems. Design provides the basis for specifying requirements and implementation ensures that these requirements are appropriately instantiated. Standards related to archival functional requirements ensure that systems consistently support those requirements.²⁰

According to this view, ARM workers must be prepared to act as service-oriented intermediaries that can communicate the nature of records to members of offices and agencies outside their own, and they must be able to offer advice regarding the management of records to ensure that the evidentiary quality of the records is maintained, so that

²⁰ However, according to CoSA’s SERI report mentioned earlier, more than half of state archives are currently not engaging in functional requirements assessment.
organizational accountability occurs. To do this, they must be able to identify for their organization what constitutes a record, relying on policy, legal requirements, organizational information needs, and an understanding of the risks associated with maintaining the records in their own organization. They must appraise and select records for retention and disposal. They must also understand technical recordkeeping systems well enough to assess the risks to records that will be stored within those systems, to offer recommendations for design and implementation of recordkeeping systems, and to identify the functional requirements for recordkeeping. Finally, they must be able to perform these duties in a manner consistent with the efficiency requirements of their organization.

In 1994 Bearman also argued that archivists rely upon standards for data interchange, for information about structure and for information about data context. He thereby linked standards requirements to the three components of a record as he (and later, ISO 15489-1) defined it: content, context, and structure. He also remarked that as of the time of that article, such interchange standards were quite undeveloped. Bearman’s recommendations provided an agenda that has since been taken up by a number of individuals.

2.2.4.1. ISO 15489-1

ISO 15489-1 is composed of a variety of sections that highlight (1) the benefits of records management and its regulatory environment, (2) topics concerning policy and responsibilities, (3) requirements, (4) design and implementation, (5) process and controls, (6) monitoring and auditing, and (7) training (ISO 2001a). It also defines the necessary characteristics of trustworthy records. Specifically, they must be authentic, reliable, exhibit integrity, and be useable (ISO). An authentic record is one that “can be proven to be what it purports to be, to have been created or sent by the person purported to have created or sent it,
and to have been created or sent at the time purported” (7). A reliable record is “one whose contents can be trusted as a full and accurate representation of the transactions, activities or facts to which they attest and can be depended upon in the course of subsequent transactions or activities” (7). A useable record is “one that can be located, retrieved, presented and interpreted” (7). Finally, a record’s integrity refers “to its being complete and unaltered” (7).

The standard also offers a description of a best practice process. Such a process involves conducting a preliminary investigation of the organizational context and capabilities, assessing existing systems, analyzing business activity, identifying requirements for records, identifying strategies to satisfy these requirements, designing a records system, implementing it, and conducting a post-implementation review. By sharing widely accepted best practices, it provides a normative set of guidelines for records managers and organizational management. Although the ISO 15489-1 standard has been adopted by NARA, it is not clear to what extent state governments currently use it.

ISO 15489-1 also specifies a number of specific activities that an organization should include in the conduct of its records management program, such as determining what records should be created in each business process and what information to include in those records, determining the form and structure of the records and the technologies used to create and maintain them, assessing and implementing metadata requirements, assessing and implementing access and retrieval requirements, assessing the risk to records, preserving records for future access, complying with legal and regulatory requirements and organizational policy, implementing disposition schedules, and engaging in quality and efficiency control of its recordkeeping processes. The ISO 15489-2 Technical Report notes that records management professionals should “have primary responsibility for the
implementation of ISO 15489-1,” establishing “the overall records management policies, procedures, and standards for the organization” and implementing the processes outlined in the standard (ISO 2001b, 2). The technical report also highlights the specific tasks that should occur in order to implement the requirements outlined in the standard.

2.2.4.2. DoD 5015.02

Besides the ISO 15489-1, the U.S. Department of Defense (DoD) created a standard that is directly applicable to the RIM operations of the U.S. government. This is DoD 5015.2, originally issued in 2002, and re-issued in 2007 as 5015.02. This standard sets forth mandatory functional requirements for Records Management Application (RMA) software used in the federal government sector. It specifically applies to “the Office of the Secretary of Defense, the Military Departments, the Chairman of the Joint Chiefs of Staff, the Combatant Commands, the Inspector General of the Department of Defense, the Defense Agencies, the DoD Field Activities, and all other organizational entities with the Department of Defense” (Department of Defense 2007, 1). It has been endorsed by the National Archives and Records Administration (NARA) as a de jure federal standard. According to Stephens, “An RMA’s primary functions are defined by the standard as categorizing and locating records and identifying those due for disposition as provided by an organization’s retention schedules,” as well as storing, retrieving, and disposing of electronic records (208). In contrast, document management software is “a software application used for managing documents that allows users to store, retrieve, and share them with security and version control” (208). DoD 5015.02 prescribes design criteria for both types of software. Stephens and the DoD have said that many state agencies require their electronic records systems to comply with this standard, which defines a number of mandatory functions, including managing records,
maintaining backward compatibility, filing email messages, searching and retrieving, to name a few (Stephens 2007; DoD 2007). However, only six state archives reported to CoSA that they use these standards (CoSA SERI Committee, 10). It is unknown how many records management applications (RMAs) are used in state agencies currently.

DoD 5015.02 refers to functions that must be available in records management and document management software rather than focusing upon the roles and responsibilities of records management professionals. However, it does explicitly allocate records management responsibilities to records managers, who are defined to be “the individuals who are responsible for records management administration” (21). The standard defines records management to be “the planning controlling, directing, organizing, training, promoting, and other managerial activities involving the life cycle of information, including creation, maintenance (use, storage, retrieval), and disposal, regardless of media” (20-21), showing that within U.S. government, the life cycle model of records still holds sway and records managers are formally designated as the occupational group responsible for records.

2.3. The Technology

2.3.1. Cloud Computing as a New and Emerging Technology

Businesses, market forecasters, and academic researchers have identified cloud computing as an emerging technology (Buyya et al. 2009; Xu 2010; Johnson et al. 2010; Kroski 2009; Gartner 2009; Voas and Zhang 2009; Hutchinson, Ward, and Castilon; Vizard 2011; Wakunuma, Stahl, and Ikonen 2011; Wu et al. 2010). The term “emerging technology” is empirical in focus, referring to the status of technologies with respect to actual markets and in comparison to substitutable technologies. Cozzens and colleagues (2005) examined nearly 2,000 articles and conclude that there was no clear consensus on the meaning of the term,
however. The major concepts that they found in the literature related to “emerging technologies” suggest that such technologies show fast recent growth, are in the process of transition and/or change to something new, have market or economic potential that is not exploited fully yet, and are increasingly science-based (18). Day and Schoemaker (2000) define emerging technologies to be those technologies for which “(1) the knowledge base is expanding, (2) the application to existing markets is undergoing transformation, and (3) new markets are being tapped or created” (2). They add that these technologies have the potential to create or transform an industry.

The notion of emerging technologies relies upon theories related to technology evolution. According to Everett Rogers (2003), technological evolution comes about through a process by which innovative technologies diffuse throughout a social system. “Diffusion is the process by which (1) an innovation (2) is communicated through certain channels (3) over time (4) among the members of a social system” (Rogers, 11). Innovation is “an idea, practice, or object that is perceived as new by an individual or other unit of adoption” (12). The market accepts new ideas or products insofar as people perceive them as being better than the technology previously available to them; being compatible with their values, experience, and needs; being relatively easy to understand and use; being capable of being experimented with or tried out “on a limited basis” (16); and having results that are publicly visible. Rogers focuses on the perception of novelty as opposed to attempting to provide an objective criterion of novelty.

Although some new technologies dramatically arrive on the scene with no commercial (or technical) precursors, such a situation is rare (Adner and Levinthal 2002). Often, emerging technologies appear to be dramatically new but in fact have existed for some
time in a small market niche. When market characteristics in a different niche from that in which they have developed are appropriate, they can then be transferred to that new niche and taken up by the public. If the new niche has adequate resources to support the technology and if the technology serves market needs, the dramatic appeal of the technology makes it appear as if it is brand new. In other words, one can distinguish between a technology’s technical development and its market application (52). A technology can represent a “convergence of existing technologies,” (Voas and Zhang 2009, 16) and still be characterized as an emerging technology if it has the potential for dramatic disruption of markets.

Technological novelty can come about by three separate mechanisms: platform innovation, component innovation, or design innovation (or a combination of these) (Sood and Tellis 2005). Platform innovation occurs when a new technology is developed using scientific principles that are obviously different from those of currently existing technologies. Component innovation occurs when new parts or materials are introduced to an existing platform. Design evolution occurs when the linkages and layout of components are reconfigured within the same technological platform (153). These three types of innovation can potentially lead either to incremental or radical innovation and the subsequent development of an emerging technology. Thus, emerging technologies sometimes engender a large change in technology and sometimes engender only a relatively minor change or merely an introduction to a new market niche followed by large-scale public adoption.

When examining cloud computing for its impacts on recordkeeping professionals, it suffices to treat it as merely a “new” technology. Much of the literature on emerging technologies focuses either on how the technology emerges (Hamilton and Singh 1992; Adner and Levinthal 2000, 2002; Avila-Robinson and Miyazaki 2011; Barnes, Buckland, and
Brancheau 1992; Sood and Tellis 2005), upon the macroeconomic impacts of emerging
technologies in general on nations (Hung and Chu 2006; Cozzens 2009; OECD 2007), or on
management techniques for discovering and managing nascent technology in order to take
advantage of the potential growth for organizational purposes (Kostoff, Boylan, and Simons
2004; Hutchinson, Ward, and Castion 2009; Day and Schoemaker 2000; Marmor, Lawson,
and Terapane 1979; Wheatley and Wilemon 1999). However, the literature that examines the
impacts of new technology adoption on workers often examines change management issues
such as role change and stress caused by technology adoption (Axtell et al. 2002; Fisher and
Wesolkowski; Lau et al. 2001; Nelson 1990; Parker, Wall, and Myers 1995; Salanova, Cifre,
and Martin 2004; Tarafdar et al. 2007; Venkatesh, Morris, and Ackerman 2000; Wahlstedt
and Edling 1997; Weil and Rosen 1997; Zorn, Hector, and Gibson 2008; Zorn 2002),
productivity impacts (Mahmood and Mann 1993; Guthrie 2001), and structural impacts and
linkages (Barley 1986, 1990; Burkhardt and Brass 1990; Guy and Skottz 2005; Hall 2002;
Hall 2005; Hector 2003; Huber 1990; Lau et al. 2001; Levy and Murnane 2004; Myers and
Young 1997; Nelson 1990; Wall and Clegg 1981; Weil and Rosen 1997). Although the
various writers often treat the term “technology” in idiosyncratic ways that reflect their
methodological stance, the conceptual descriptor “new” is relatively straightforward. By
“new,” these authors simply mean a form of technology that the workers being studied have
not previously encountered within their organizations.

Because of its relative novelty, however, identifying what comprises cloud computing
can be difficult, since cloud computing appears to have come about through incremental
changes that have largely been caused by component innovation and design innovation. That
is to say, a number of already familiar technologies exist within the Cloud paradigm, but
have been modified slightly to scale upward and downward more easily, have been combined in ways that were not previously done, and use a new business model. Two key questions, therefore, are what is cloud computing and what is new about it?

2.3.2. Defining Cloud Computing

2.3.2.1. Cloud as Hardware, Architecture, or Market

The National Institute of Standards and Technology (NIST) defines cloud computing to be “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” (2009c, 2). According to NIST, cloud computing is composed of five essential characteristics:

- **On-demand self-service.** This implies that automatic provisioning of services can occur without any need to interact personally and individually with the various providers of services. It also implies that new contracts and agreements do not need to be renegotiated each time resource levels need to be changed.

- **Broad network access.** A number of platforms and devices can access the services over a network. This is equivalent to what Weiss (2007) calls distributed computing.

- **Resource pooling.** This refers to a “multi-tenant” model, whereby consumers of services share resources without being aware of the other users of the services and without knowledge of the exact location from which services are being provided.

- **Rapid elasticity.** Consumers can scale up (or down) their resource usage rapidly and dramatically, with an associated increase (or reduction) in costs to match the scaling. This is equivalent to Vaquero and colleagues’ (2009) “scalability.”
• **Measured service.** Resource usage and service provision are monitored and measured, and the consumer pays on the basis of the service provider’s per unit pricing scheme. This is typically a utility, or “pay per use,” pricing system (Mell and Grance). It is equivalent to the utility models mentioned by Weiss and Vaquero et al.

The NIST approach focuses heavily on the benefits to consumers in a market-exchange system; when one examines the NIST definition eight key aspects are highlighted:

- Ubiquitous nature,
- Convenience,
- On-demand service,
- Shared pooling of resources,
- Configurable resources,
- Rapid provisioning,
- Minimal management effort required, and
- Minimal service provider interaction.

Although some possible costs or risks inhere in some of these aspects (e.g., shared pooling of resources increases the risk of unauthorized access to an organization’s data from outside the organization), most of these qualities represent what it is about cloud computing that makes it desirable. Ubiquitous: one can access it anywhere in a convenient fashion. On demand: one can access it when he or she wants it. Shared pooling: although other users can access the same resources, they are logically separate and any one user is unaware of the others while accessing the resources. Configurable: one can configure the resources to his or her specifications. Rapid provisioning: one can access the resources rapidly and can scale up or down rapidly. Minimal management effort required: when one wants to access the
resources or scale up or down, he or she can do so with little effort. Minimal service provider interaction: when one want to access the resources or scale up or down, he or she does not have to engage in continual communication or negotiation with the service provider.

Many authors define cloud by the hardware or the architectural characteristics of the technology (i.e., “technology-based” definitions) or by the particular ways in which cloud service providers deploy it and offer it to cloud consumers (i.e., “market-based” definitions), an approach researchers have often used to distinguish technological innovations (Chandy and Tellis 1998). For example, Delic and Walker (2008) defined computing clouds technically, as “huge aggregates of various grids (academic, commercial), computing clusters and supercomputers” (Delic and Walker 2008), begging the question of how cloud computing can be distinguished from these other architectures. Alternatively, Kevin Jones of Dell Services said, “It's not a new architecture, new technology or a new methodology – it is, however, a radically new way of doing business” (Jones 2012). A list of a variety of definitions of cloud computing found in addition to the NIST definition examined here is available in Appendix A.

One of the benefits of NIST’s depiction of cloud computing is its comprehensive nature. It points to both technical and business-related aspects of the Cloud: besides defining five distinguishing characteristics of cloud computing, the authors also offer detailed descriptions of the typical delivery and deployment models.

2.3.2.1.1. Delivery Models

The Cloud delivery models are Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as-a-Service (IaaS) (Armbrust et al. 2009; Buyya, Shin, and Venugopal 2008; Buyya et al. 2009; CCUCDGroup 2010; Creeger 2009; Foster et al. 2008;
SaaS is as a form of service provision in which the consumer is able to use “the provider’s applications running on a cloud infrastructure and accessible from various client devices through a thin-client interface such as a web browser (e.g., web-based e-mail)” (Wyld, 12). Under this service type, “the consumer does not manage or control the underlying cloud infrastructure, network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings” (12). In other words, instead of installing and maintaining one’s own software, one can simply use the software applications that are hosted on a network by someone else (CCUCDG; Creeger; Hinchcliffe 2009; Klems 2009; Kroski 2009; Lockheed Martin, Alliance, and Market Connections 2010; Mell and Grance 2009a; Vaquero et al.). Creeger pointed out, however, that SaaS is not unique to cloud computing but rather, is a precursor technology of which cloud computing takes advantage. Cloud services are infrastructure and services that the organization can rent (Patterson 2010). SaaS services are “applications running on top of a cloud computing environment” which include applications such as email services, photo and video services, and services such as file storage and computational processing (Jaeger et al. 2009). Salesforce.com CRM, for example, is a popular SaaS offering (Schuller 2008).

A PaaS service allows the consumer to build and deploy consumer-created applications using programming languages and tools supported by the provider (e.g., java, python, .Net) onto the Cloud infrastructure (Wyld; CCUCDG; Creeger; Jaeger et al.; Knorr and Gruman; Youseff, Butrico, and Da Silva). In other words, one does not need to have any
tools on his or her server to build the applications. Wyld remarked, “The consumer does not manage or control the underlying cloud infrastructure, network, servers, operating systems, or storage, but has control over the deployed applications and, possibly, application hosting environment configurations” (12). For example, the Google Application Engine allows users to build and host web applications on the Google infrastructure (Google Developers 2012). Similarly, Force.com is a PaaS offering (Schuller).

IaaS provides the consumer with processing, storage, networks, and other fundamental computing resources with which the consumer can deploy and run software, including operating systems and applications (CCUCDG 2010; Lockheed Martin, Alliance, and Market Connections 2010; Mell and Grance 2009a; Myerson 2009; Vaquero et al. 2009; Wyld 2009; Youseff, Butrico, and Da Silva 2008). “The consumer does not manage or control the underlying cloud infrastructure, but has control over operating systems, storage, deployed applications, and possibly select networking components (e.g., firewalls, load balancers)” (Wyld, 12). Examples include Amazon's Elastic Compute Cloud (EC2) (Youseff, Butrico, and Da Silva 2008). Amazon’s Simple Storage Service (S3) is also an IaaS service, offering data storage infrastructure (Amazon Web Services 2012).

Consumers often adopt layered combinations of these three delivery models.

2.3.2.1.2. Deployment Models
Cloud computing is offered via four deployment models: private cloud, public cloud, community cloud, and hybrid cloud (Mell and Grance 2009c; Kundra 2010; Marinos and Briscoe 2009; McCafferty and McAlpine 2010; Sun Microsystems 2009; Wlodarczyk, Rong, and Thorsen 2009; Wyld 2009; Zhang, Cheng, and Boutaba 2010). A private cloud is “operated solely for an organization,” although “it may be managed by the organization or a
third party and may exist on premise or off premise” (Mell and Grance 2009c, 2). A public cloud, on the other hand, is available to the general public or a large industry group and is owned by an organization that sells cloud services. A community cloud is “shared by several organizations and supports a specific community that has shared concerns” (2). It may be provided by a third party organization or by the organizations themselves. Likewise, it can reside onsite or offsite. A hybrid cloud is “a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability” (2).

2.3.2.2. Cloud as Structure

Although examining delivery and deployment models can help one classify various actual services offered on the market, this view of technology can also create conceptual confusion. Some have argued that cloud computing is mere “hype,” with nothing to distinguish it from other pre-existent computing architectures than the business model it uses (Ellison 2009; Schneier 2009). This view ignores two aspects of computing evolution. First, it implies that there are no differences at all between the configuration or architecture of cloud technology and that of previous computing models. Second, it implies that a new business model is a trivial change. Neither assumption is entirely correct. In addition, by focusing purely upon underlying hardware or configuration or upon financial arrangements, this conception of technology can lead one to ignore the ways in which a new technological adoption requires modifying individual processes and procedures, causes changes to intra-organizational power dynamics, and is affected by the employees’ perceptions of the technology, thereby impacting their use of it. If one considers technology as structure, however – that is, as co-constituted with the human actions that allow the technology to
operate – one realizes that a new technological arrangement in the organization will engender a number of potentially evolutionary impacts that go well beyond hardware, architecture, or business model. In fact, the notion of cloud computing rather directly reflects Giddens’ idea of structuration in that even with relatively incremental changes in technology, the consequent changes in other social structures, such as business models, can begin to take on an added importance, leading people to arrange themselves socially and organizationally in entirely novel ways. Furthermore, through their explicit recognition of that novelty, they can then begin to engage in creative innovation (or destruction) (Schumpeter 1950) that restructures many individual characteristics of the social system, thereby leading to an evolution of society.

One difficulty with thinking of a specific technology as a structure is that the term “technology” itself is somewhat slippery. For example, in the paragraph immediately preceding this one, the term is sometimes used to reflect the artifactual set of properties that comprise what we think of as that technology (e.g., “cloud computing”). Alternatively, it is also sometimes used to reflect the emergent rules and resources that come about when human beings routinely act upon these artifactual properties in regularized, ongoing enactment of the technology itself. Orlikowski (2000) refers to the latter view of technology as technologies-in-practice. From this viewpoint, cloud computing can be seen not merely as the set of material and social properties that comprise what we call “cloud computing.” It is, in fact, existent only insofar as individuals engage in recurrent action with those properties and thus constitute the technology-in-practice that is called cloud computing. Nonetheless, it is useful to understand how computing technologies materially evolved into cloud computing and the ways in which it is different from other computing technologies with respect to architecture.
or hardware, since some of these differences will influence the possible ways in which people can use the technology and the limits upon their usage of it (that is, its affordances).

2.3.2.3. Comparing Cloud to Earlier Forms of Computing

2.3.2.3.1. The Evolution of Cloud Computing

Cloud computing has evolved from technology that was conceived in the 1950s when early time sharing environments were developed to allow multiple users to access a single computer at once, and also from later developments associated with high-throughput computing. Discussing the history of timesharing, Garfinkel (1999) relates that an underlying goal of the original Advanced Research Projects Agency (ARPA)-funded21 project to develop time sharing capabilities was to create a computer operating system that would allow computing to operate as a utility, that is, as something that is always available and is highly reliable.22 He explained that although the formal project to develop a time sharing machine began in 1963,23 one of its key developers, John McCarthy, first began to develop time sharing in 1957 in order to overcome the limitations of batch processing. Unlike batch processing, which allows only one job to be processed at a time, time sharing allows multiple users to access the same computer processor at once, with the computer switching between

---

21 ARPA, now referred to as DARPA, or the Defense Advanced Research Projects Agency, a U.S. defense department agency responsible for the development of new technology for use by the U.S. military.

22 Paul Edwards points out, however, that a driving force for the funding of this vision lay in military concerns about ensuring that the speed of interaction between humans and computers was great enough that a tightly coupled decision-making man-machine “team” could be created and effectively implemented for military purposes (1996, 265).

23 The term “time-sharing” did exist prior to McCarthy’s use of it, but Edwards notes that prior to McCarthy’s use of the term, “time-sharing” actually referred to “multiprogramming,” a technique in which “several programs could be run at once under the direction of an ‘executive program’” (Edwards, 256). The earlier use of the term “time-sharing” is not consistent with the later notion that implies multiple users. Rather, “multiprogramming” still relies upon batch processing and a single-user environment, albeit providing greater speed than earlier approaches.
different jobs so quickly that it gives “the illusion of real-time, interactive use of the machine” (Garfinkel, 4). The MIT research laboratory that resulted from this project made use of the services of a public utility engineer to design its original backup system in order to ensure the kind of reliability that a utility would provide. The research from the ARPA grant was largely successful, and now, after 50 years of time sharing evolution, we are able to conceive of computing as a pervasive and highly distributed phenomenon.

Speaking at the MIT Centennial in 1961, McCarthy said, “If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility....The computer utility could become the basis of a new and important industry” (Garfinkel, 1). However, the original vision of utility computing expressed by John McCarthy could not be achieved entirely by the successful development of time-sharing capabilities. Specifically, McCarthy’s vision of computer as utility could not be developed merely by designing into it the technological capacity of time sharing. The technology needed organizational and funding capabilities that could sustain such a distributed system. In short, utility computing is not just an infrastructure, it must also engage a business model “in which computing resources, such as computation and storage, are packaged as metered services similar to a physical public utility, such as electricity and public switched telephone network[s]” (Foster et al. 2008, 2). The utility business model rests upon necessary technological capabilities, but the technological capabilities are not sufficient to create the utility.

In the early 1960s, Paul Baran of Rand Corporation, developed packet switching in order to enable more robust communications networks for military purposes (Abbate 1999). By the 1980s, concerns about America’s continuing ability to engage in the global society
dominated cultural discourse (Pomerantz, Choemprayong, and Eakin 2008). In 1981 the Department of Defense issued a report stating that “the military power of the United States is inextricably tied to the programmable digital computer” (1988, 99-100). By the late 1980s Americans expressed concerns about both the United States’ military and economic place in the world. For example, in 1989, then-senator Al Gore stated that “the nation which most completely assimilates high-performance computing into its economy will very likely emerge as the dominant intellectual, economic, and technological force in the next century” (Gore 1990). By the late 1980s and early 1990s, the business sector recognized that a strong computer infrastructure was necessary in order to allow America to compete internationally (Yudken and Simons 1988; Winograd 1991; Savage 1994). Likewise, key funding stakeholders emphasized the need for computing power and speed sufficient to allow large-scale sharing among researchers in the sciences (FCCSET 1994).

Researchers have used the term “metacomputing” to describe their relatively early explorations into geographically distributed computing systems. Baraglia (1997) noted that they used this term specifically to describe the notion of distributing multiple resources over a network so they could be used as if they were a single computer. One could distinguish a metacomputer from a “typical” parallel processing machine because the latter generally consisted of “tightly coupled processing nodes of the same type, size and power” whereas the former had “loosely coupled and heterogeneous” nodes (224). Grimshaw (1994) used the term “metasystem” in place of “metacomputer” to describe “a distributed computing system composed of a heterogeneous group of autonomous computers linked together by a hierarchical network” (257).
By the mid-1990s, partly as a result of the passage of the High Performance Computing and Communications Act (i.e., HPCC) (FCCSET), the computer science community focused much of its efforts on developing a strong infrastructure that would allow researchers to share applications, infrastructure, and data within the scientific community. The result of these efforts culminated in strengthening the capacity for high performance and data throughput via parallel processing within network architecture, hardware systems, and software environments (Savage 1994; Buyya, Shin, and Venugopal 2008). What has come to be known as “grid computing” arose from these efforts and still provides a widely used and beneficial environment for many communities looking for a way to share resources widely but efficiently.

2.3.2.3.2. Grid, Supercomputers, Cluster, Web 2.0 and P2P
During the late 1990s the Partnerships for Advanced Computational Infrastructure (PACI) program, the National Partnerships for Advanced Computational Infrastructure (NPACI), and several hardware partners acted as centers of the grid, and together with other Partners for Advanced Computational Services established the National Technology Grid, which linked together a number of regional grids (Foster and Kesselman 1999b). These regional grids, also referred to as GigaPOPs (“gigabit network points of presence”) connected to a research backbone called the vBNS (i.e., “very high-speed backbone network service”). The first working GigaPOP, Ameritech’s Metropolitan Research and Education
Network (MREN), allowed a number of research institutions to connect (Foster and Kesselman 1999b).

During this evolution of computer systems and networks, technology developers and the scientific community sought to develop computing capabilities that were transparent to the user, dependable, consistent, pervasive, and inexpensive (Foster and Kesselman 1999b; Bote-Lorezo, Dimitriadis, and Gómez-Sánchez 2003). At the same time, they focused on developing the ability to share computing resources, whether mainframes or clusters of computers, architecture, data sets, or even people (Bote-Lorezo, Dimitriadis, and Gómez-Sánchez; Chetty and Buyya 2002; Čibej, Sulistio, and Buyya 2009). Within the grid community, the individuals or institutions sharing such resources make up what is called a virtual organization. A virtual organization is central to the idea of a grid in that “the real and specific problem that underlies the Grid concept is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” (Foster, Kesselman, and Tuecke 2001).

Virtual organizations (VOs) have been defined variously from the early 2000’s onwards. Within early conceptualizations, a virtual organization was typically considered to be a group of individuals or institutions, as reflected by the now classic definition of Foster, Kesselman, and Tuecke, who focused upon the sharing of resources via strictly defined rules:

The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem solving and resource-brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is

---

24 Participating institutions were University of Illinois at Urbana-Champaign, University of Illinois at Chicago, University of Chicago, Northwestern University, University of Minnesota, University of Wisconsin, University of Michigan, Michigan State University, Purdue University, Indiana University, Argonne National Laboratory, and Fermi National Accelerator Laboratory (Foster and Kesselman 1999b).
allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization (200-201).

A looser conceptualization is that of Mietzner and colleagues (2009), who defined VOs to be “dynamic collections of individuals, institutions and resources” that allow “flexible, secure, coordinated resource sharing” (138).

Foster (2002) provides a list of three criteria by which an entity can be recognized to be a grid (or not). He says a grid is a system that “(1) coordinates resources that are not subject to centralized control…(2) using standard, open, general-purpose protocols and interfaces…(3) to deliver nontrivial qualities of services” (3). A list of a variety of definitions of grid computing found in the literature examined here is given in Appendix B, according to which grids appear to comprise the following set of characteristics: dependable, persistent, consistent, and inexpensive access; high-end computational capabilities; geographically distributed computing; shared, heterogeneous resources; scalability and large scale problem-solving; decentralized control; the use of standard, open, general-purpose protocols; provision of non-trivial qualities of service (i.e., “QoS”); transparent computing services; and virtualization.

Because resource sharing in grids was meant to cross traditional organizational boundaries, grid architecture evolved to avoid the centralized ownership and administration of cluster computing (Čibej, Sulistio, and Buyya). This allows individuals from disparate organizations in distributed geographical locations to collaborate and share on projects that require high-performance or high-throughput processing (Buyya et al. 2009; Chetty and Buyya; Vaquero et al.; Mietzner, Karastoyanova, and Leymann; Klems, Nimis, and Tai 2009) or widespread data sharing (Moore, Rajasekar, and Wan 2005; Moore et al. 1999). Within a grid, the heterogeneous resources reside on a variety of nodes, sometimes
numbering in the thousands, and each node can have a different owner and different policies (Čibej, Sulistio, and Buyya), although it is not unusual to find grids composed of various kinds of clusters (Delic and Walker 2008).

Figure 1 shows the overlap of five common types of computational systems: supercomputers, clusters, grids, clouds, and Web 2.0, according to Foster et al. (2008).

The term “supercomputer” is well known, if not generally defined clearly. Wikipedia suggests that a supercomputer is “a computer that is at the frontline of current processing capacity, particularly speed of calculation” (http://en.wikipedia.org/wiki/Supercomputer). Willard (2008) lists other definitions that have been proffered, including “a computer used to address the most demanding problems,” “any computer that turns a compute bound problem
into an I/O bound problem,” and more whimsical offerings such as “any computer built by Seymour Cray” and “a computer that is only one generation behind what the users want.” With the advent of grids and clouds, supercomputers can be made into super supercomputers, via massively parallel and distributed processing techniques that allow incredibly rapid calculations to occur by sharing the workload among a pool of very powerful machines.

Supercomputers, however, can also be interconnected to create clusters of computational resources. Buyya et al. define a cluster as “a type of parallel and distributed system, which consists of a collection of inter-connected stand-alone computers working together as a single integrated computing resource” (Buyya et al. 2009, 602). However, clusters are often composed of commodity computers – low-end, inexpensive units that are likewise linked together (Buyya et al.). Unlike grids, clusters are generally located in the same geographic area and managed by a single entity. In addition, cluster schedulers tend to focus on “enhancing the overall system performance and utility as they are responsible for the whole system” (Buyya et al., 603). Foster and Kesselman (Foster and Kesselman 1999a) define a cluster to be “a collection of computers connected by a high-speed local area network and designed to be used as an integrated computing or data processing resource.” Grid computing, on the other hand is “more loosely coupled, heterogeneous, and geographically dispersed” (Marinos and Briscoe, 476).

On the services-oriented side of Foster and colleagues’ diagram in Figure 1 reside cloud computing, Web 2.0, and some grid computing implementations. Popular literature often uses “Cloud computing” and “Web 2.0” as interchangeable terms, although Rhoton (2009, 11) notes that there is no intrinsic connection between the two. According to Tim O’Reilly (2006), “Web 2.0 is the business revolution in the computer industry caused by the
move to the Internet as a platform, and an attempt to understand the rules for success on that new platform.” According to Mell and Grance (2009a), Web 2.0 is “the trend of using the full potential of the web,” by “viewing the Internet as a computing platform,” “running interactive applications through a web browser,” leveraging interconnectivity and mobility of devices,” exploiting the “long tail,” that is, gaining profits in “selling specialized small market goods,” and deriving “enhanced effectiveness with greater human participation” (62). Because cloud computing is a highly service-oriented architecture, it generally takes advantage of Web 2.0 functionality, but strictly speaking even server farms provided in a private cloud computing environment without Web 2.0 functionality could be considered a cloud. Because Web 2.0 functionality can be provided without cloud computing architectures (and vice versa) the two concepts should be distinguished from each other.

Peer-to-Peer (P2P) is a distributed architecture in which “participants share a part of their own hardware resources (processing power, storage capacity, network link capacity, printers)” where “these shared resources are necessary to provide the Service and content offered by the network” (Schollmeier 2001, 101). P2P computing involves peer nodes (computers) sharing content directly with one another in a decentralized manner (Buyya et al. 2009). Within P2P architectures, resources are available equally and on demand to every user (Kurdi, Li, and Al-Raweshidy 2008). In a P2P environment, the provider of services cannot readily be distinguished from the consumer, since any individual computer can play both roles (Taylor and Harrison 2008). Such a system “focuses on resource sharing in environments characterized by potentially millions of users, most with homogenous desktop systems and low-bandwidth, intermittent connections to the Internet” (Crowcroft et al. 2004).
2.3.2.3.3. Comparing Cloud and Its Relatives

A variety of physical and architectural differences are found between cloud computing and the other computing systems just discussed. As with most IT system concepts, however, these differences generally tend to be a difference of degree rather than type. Those who argue that cloud computing is just “the modern version of the timesharing model from the 1960s” (Schneier) miss the fact that one does not need a mainframe to enter the cloud and in fact, the Cloud came about because of the increased speed allowed by newer technology and the existence of the Internet. At the same time, the idea of connecting multiple users to what appears to them to be a single source has been around for a long time. So, how can cloud computing be distinguished from other types of computing?

Scalability is primarily achieved in grids by increasing the number of working nodes, whereas in the Cloud it is achieved by automatic resizing of virtualized hardware resources (Vaquero et al.). Clouds can perform this more easily than grids because they typically tend to be centrally controlled and managed, whereas grids tend to have decentralized control and self-management. In addition, service level agreements and quality-of-service guarantees are an inherent feature of cloud computing, whereas they are generally layered on top of architectures such as the grid, and entirely absent from cluster computing. Because cloud computing offers isolation of resources (as opposed to the Grid’s resource sharing), pricing becomes simpler for the consumer and more problematic for the service provider. The consumer sees a pay-per-use model; the service provider must assess the best approach for ascertaining appropriate pricing.

Although the kind of dynamic, runtime provision of resources seen in the Cloud could potentially be available in some grids (Čibej, Sulistio, and Buyya 2009; Kurdi, Li, and Al-Raweshidy 2008), batch processing has tended to be the norm for grid computing (Jones
2008; Foster et al. 2008). Also, according to several authors, within grids high-performance and throughput are achieved with both parallel and distributed processing arrangements that often use high-end computers, as opposed to cloud and cluster computing which tend to rely on commodity servers to reduce cost, albeit with a potential sacrifice of performance (Armbrust et al. 2009; Buyya et al. 2009; Čibej, Sulistio, and Buyya 2009; Agrawal et al. 2010). However, Chetty and Buyya (2002) suggested that grids composed primarily of commodity machines do exist.

Vaquero and colleagues also noted that grids have historically been developed for scientific purposes. Scientific projects have typically received public funding for particular projects, leading to more centralized approaches to resource allocation. They are usually billed using a fixed rate per service or require different organizations to share idle resources, creating virtual organizations which are allocated resources on a fair use basis. They do this by using resource brokers to determine fair use on the basis of automated policies (Buyya, Abramson, and Venugopal 2005; Buyya, Shin, and Venugopal 2008; Murphy et al. 2010; Kotrotsos et al. 2010). Clouds, on the other hand, typically provide commercial services and “are usually billed using a pay-per-use model” (Vaquero et al. 2009, 54).

Although grids attempt to ensure a fair share of resources across organizations, clouds use virtualization to provide an illusion of a single dedicated resource. They do not rely on explicit sharing but rather, provide resource isolation through virtualization. Both grid and cloud support the aggregation of heterogeneous hardware and software resources, but grids typically virtualize data and compute resources whereas clouds typically also include hardware virtualization (Vaquero et al. 2009).
Grids offer such services as metadata search and data transfer, whereas that type of service is still underdeveloped in the Cloud (Vaquero et al. 2009). Nonetheless, when it comes to quality of service (QoS), grids tend to lag behind clouds, with service level agreements (SLAs) needing to be created via applications that reside on top of the grid. SLAs are an inherent feature of clouds (Vaquero et al. 2009). In addition, the overall service orientation of grids in comparison to clouds is somewhat different: “Grid computing specifically refers to leveraging several computers in parallel to solve a particular, individual problem, or to run a specific application. Cloud computing, on the other hand, refers to leveraging multiple resources, including computing resources, to deliver a ‘service’ to the end user” (IBM 2009, 6).

Foster and colleagues (2008) note that the ability to manage and track provenance has typically been built into grids via workflow systems, and has also been built for grids as a standalone service, PreServ. They define provenance to be “the derivation history of a data product, including all the data sources, intermediate data products, and the procedures that were applied to produce the data product” (6). They add, however, “Provenance is still an unexplored area in Cloud environments, in which we need to deal with even more challenging issues such as tracking data production across different service providers (with different platform visibility and access policies) and across different software and hardware abstraction layers within one provider” (2008, 7). At least one attempt to deal with provenance in the Cloud has occurred, however. Muniswamy-Reddy and colleagues (2009, 2010; Muniswamy-Reddy and Seltzer 2010) describe what they call a Provenance-Aware Storage System (PASS) to augment the Amazon Web Service (AWS) for backend storage in order to provide the capability of tracking and managing provenance in the Cloud.
Table 1 provides a detailed comparison of cloud computing and three other architectures discussed in the literature, (i.e., cluster, grid, and peer-to-peer), highlighting a variety of features.

2.3.2.1. Handling Recordkeeping Risks in the Cloud

As mentioned in Chapter 1, NARA and ARMA International have remarked on the potential recordkeeping risks associated with Cloud Computing. A variety of researchers and governmental agencies has also looked into possible risks, and has provided more detailed information regarding both risks and best practices. For example, the U.S. government’s CIO council, in conjunction with the Federal Compliance Committee published *Creating Effective Cloud Computing Contracts for the Federal Government: Best Practices for Acquiring IT as a Service* (CIO Council 2012). This document highlights ten areas in which federal agencies “require improved collaboration and alignment during the contract formation process...when acquiring cloud computing services” (3): selecting a cloud service; CSP (i.e., “Cloud Service Provider”) and End-User Agreements; Service Level Agreements (SLAs); CSP, agency, and integrator roles and responsibilities; standards; security; privacy; e-Discovery; Freedom of Information Act (FOIA); and e-Records. Although “e-Records” obviously relates to recordkeeping, several other of these areas can potentially impact recordkeeping activities in the organization. The European Union’s Article 29 Data Protection Working Party (Article 29 Data Protection Working Party 2012) attributes many of these risks to “a lack of control over personal data as well as insufficient information with regard to how, where and by whom the data is being processed/sub-processed” (2).

Recordkeeping risks associated with cloud computing and the steps needed to mitigate these risks according to the CIO Council and the Article 29 Data Protection
### Table 1 - Comparing Cloud to Related Architectures, by Feature and Source

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cluster</th>
<th>Grid</th>
<th>Cloud</th>
<th>P2P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Server Types</td>
<td>Commodity</td>
<td>High-End</td>
<td>Commodity computers and high-end servers and network attached storage</td>
<td>Edge of Network (i.e., desktop)</td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td></td>
</tr>
<tr>
<td>Ownership</td>
<td>Single</td>
<td>Multiple</td>
<td>Single</td>
<td>Multiple</td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td></td>
</tr>
<tr>
<td>Discovery</td>
<td>Membership Services</td>
<td>Centralized Index &amp; Decentralized Information</td>
<td>Membership Services</td>
<td>Decentralized</td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Buyya et al. 2009)</td>
</tr>
<tr>
<td>Ease of Use</td>
<td>Di fficult</td>
<td>Easy</td>
<td>Easy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Jones 2008)</td>
<td>(Jones 2008)</td>
<td>(Jones 2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Klems 2008)</td>
<td>(Klems 2008)</td>
<td>(Klems 2008)</td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>Cluster</td>
<td>Grid</td>
<td>Cloud</td>
<td>P2P</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------------------------</td>
<td>---------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Security/Privacy</td>
<td>Traditional</td>
<td>Public/private key pair based</td>
<td>Each user/application provided a virtual machine. High security/privacy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>login/password-based. Medium level</td>
<td>authentication and mapping a user to</td>
<td>guaranteed. Support for setting per-file access control list (ACL).</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of privacy – depends on user</td>
<td>an account. Limited support for privacy.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>privileges.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Credential delegations and</td>
<td>Security through isolation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>user authorization</td>
<td>(Vaquero et al. 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Simple Use of Webforms (over SSL)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Foster et al. 2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Public Key Infrastructure (PKI) and X.509 SSL certificates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Youseff, Butrico, and Da Silva 2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resource Sharing</td>
<td>Collaboration (VOs, fair share),</td>
<td>Assigned resources are not shared</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>policies &amp; procedures</td>
<td>(Vaquero et al. 2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Vaquero et al. 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>Cluster</td>
<td>Grid</td>
<td>Cloud</td>
<td>P2P</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Resource Management</td>
<td>Centralized</td>
<td>Distributed</td>
<td>Centralized</td>
<td>Distributed</td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Taylor and Harrison 2008)</td>
<td></td>
</tr>
<tr>
<td>Capacity</td>
<td>Stable and guaranteed</td>
<td>Varies, but typically high</td>
<td>Provisioned on demand</td>
<td>Varies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Buyya et al. 2009)</td>
<td>(Agrawal et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td>(Buyya et al. 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Wang et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Zhang, Cheng, and Boutaba 2010)</td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>Low latency, high bandwidth</td>
<td>High latency, low bandwidth</td>
<td>Low latency, high bandwidth</td>
<td>High latency, low bandwidth</td>
</tr>
<tr>
<td>(Latency/Bandwidth)</td>
<td></td>
<td>(Buyya et al. 2009)</td>
<td>(Buyya et al. 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td></td>
</tr>
<tr>
<td>Application Development</td>
<td>Local</td>
<td>In the Cloud</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>Cluster</td>
<td>Grid</td>
<td>Cloud</td>
<td>P2P</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Business and/or Funding Model</td>
<td>Limited, not open market</td>
<td>Limited, not open market</td>
<td>Pay as you go; utility pricing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public good or privately assigned; project-oriented resource sharing; policy</td>
<td><em>(Buyya et al. 2009)</em> <em>(Foster et al. 2008)</em> <em>(Weinhardt et al. 2009)</em></td>
<td>Tiered, per-unit, and subscription-based pricing</td>
<td></td>
</tr>
<tr>
<td>Standardization &amp; Interoperability</td>
<td>Virtual Interface Architecture (VIA)-based</td>
<td>Some open grid forum standards</td>
<td>Web Services (SOAP and REST)</td>
<td>No Standards <em>(Stockinger 2007)</em></td>
</tr>
<tr>
<td>Provenance Management</td>
<td>Done via workflow systems</td>
<td></td>
<td>Relatively unexplored</td>
<td><em>(Foster et al. 2008)</em></td>
</tr>
<tr>
<td>Feature</td>
<td>Cluster</td>
<td>Grid</td>
<td>Cloud</td>
<td>P2P</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Computational Model</td>
<td>Batch</td>
<td>Interactive</td>
<td>(Foster et al. 2008)</td>
<td>(Foster et al. 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Various (e.g., batch, interactive, distributed, parallel)</td>
<td>Long-lived services based on hardware virtualization</td>
<td>(Klems 2008)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Short-lived batch-style processing (job execution)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Klems 2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalability</td>
<td>100s</td>
<td>1000s</td>
<td>100s to 1000s</td>
<td>Millions</td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Jones 2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nodes and sites scalability</td>
<td>Elastic</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Vaquero et al. 2009)</td>
<td>(Agrawal et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Wang et al. 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Zhang, Cheng, and Boutaba 2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Nodes, sites, and hardware scalability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Vaquero et al. 2009)</td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>Cluster</td>
<td>Grid</td>
<td>Cloud</td>
<td>P2P</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Virtualization</td>
<td></td>
<td>Virtualization of data and computing resources</td>
<td>Virtualization of hardware and software platform</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Vaquero et al. 2009)</td>
<td>(Wang et al. 2010)</td>
</tr>
<tr>
<td>Sef-Management, Failure Management</td>
<td>Limited (often failed tasks/applications are restarted)</td>
<td>Limited (often failed tasks/applications are restarted)</td>
<td>Strong support for failover and content replication. VMs can be easily migrated from one node to another</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Buyya et al. 2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reconfigurability</td>
<td>Reconfigurability, self-healing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
</tr>
<tr>
<td>Software Dependencies</td>
<td></td>
<td>Application domain-dependent software (grid middleware required)</td>
<td>Application domain-independent software</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Taylor and Harrison 2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocation/Scheduling</td>
<td>Centralized</td>
<td>Decentralized</td>
<td>Both centralized and decentralized</td>
<td>Decentralized</td>
</tr>
<tr>
<td></td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td>(Čibej, Sulistio, and Buyya 2009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Stockinger 2007)</td>
<td>(Stockinger 2007)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Working Party are:

- **Lack of availability of one’s data.** Because many cloud providers rely on proprietary technology and because there are still no widely accepted interoperability standards for cloud computing, one may be subject to vendor lock-in. Under such a situation, if one chooses to terminate the contract with one’s CSP, or if the CSP goes out of business or merges with another company, one may be at risk of losing access to one’s data. The CIO Council recommends that any agency ensure that it explicitly specifies data ownership in its service contract and that it validates that the CSP has the capability (and intent) to transfer data back to the agency in the event of contract termination or to a different CSP if and when the agency desires.

- **Lack of data integrity.** Although cloud providers engage in resource isolation techniques to ensure that information cannot be accessed by other cloud consumers than those who have authority to access it, the underlying computing resources (e.g., hardware) are nonetheless shared among a variety of consumers. An agency needs to ensure that it understands how resource isolation occurs and what security safeguards the CSP provides. The CSP should provide auditing capabilities and should be given clear metrics that the agency needs to evaluate on a regular basis. In addition, processing logs should be available to the agency on a regular basis.

- **Lack of Confidentiality.** A cloud provider may receive a request or demand for information from law enforcement agencies. According to the CIO Council and Article 29 Data Protection Working Party, the agency needs to understand how such a request could impact its own information. For example, if the agency receives an e-Discovery request, it needs to be confident that the cloud provider can ensure that
they can meet that request in a timely manner. In addition, if the cloud provider stores information outside the agency’s jurisdiction, it may be subject to legal requirements that conflict with the agency’s privacy requirements, especially if the information is held outside the country. If the data needs to be stored in a particular jurisdiction, this needs to be specified in the cloud contract as well. In addition, the CSP should be required to sign a confidentiality agreement and to follow the same confidentiality rules that the agency’s personnel have to follow.

- **Lack of Intervenability.** Recordkeepers sometimes need to access records to make corrections or to erase erroneous information. The agency needs to be sure that the cloud provider can offer this ability. Contracts need to specify clearly what roles and responsibilities all parties to the contract play to minimize the risk that the agency will not be able to intervene with its records according to its policies and mandates, according to the CIO Council and Article 29 Data Protection Working Party.

- **Lack of Isolation.** Because underlying resources are shared within a cloud computing arrangement, it is physically possible for cloud providers to link information from several different customers. If an administrator has sufficient access rights, he or she could link these pieces of information in ways that are not acceptable to one or more of the customers. Therefore, agencies must make sure that the CSP ensures that no one has access to more information than they actually need to have. In addition, the CSP needs to let the agency know what technical measures it takes to isolate the information so that the risk of unauthorized access is minimized.

- **Lack of Transparency.** The CSP needs to be transparent about its policies and procedures regarding security, privacy, and handling of the data. In many cases, cloud
providers layer their services. That is, some portions of their services are actually provided by yet other vendors. In such cases, the agency not only needs to know who all the parties that touch the data are, they need to ensure contractually that all of these vendors are willing and able to meet their recordkeeping requirements. In addition, the agency needs to understand how the cloud provider can comply with its retention periods. The agency needs to ensure that the data is erased or anonymized acceptably. To do this requires either destroying or demagnetizing storage media, or overwriting the data sufficiently. Special software tools exist that will overwrite data multiple times to ensure it is unrecognizable (Article 29 Data Protection Working Party 2012). However, one needs to recognize that 100% secure infrastructure is not possible in the Cloud. This is why full transparency of the CSP is essential, according to the CIO Council and Article 29 Data Protection Working Party.

Of course, any agency needs to ensure that the CSP is aware of all recordkeeping requirements, including laws and regulations by which the agency must abide when handling records. The CIO Council notes that risks can be lowered by including recordkeeping personnel in the requirements definition process and including them in communications channels (32). In addition, the CSP must be willing and able to transfer records of long-term value to an agency-specified archive according to retention period requirements.
3. METHODS

The specific goals of this exploratory study were to examine how recordkeeping stewards who work in state government or alongside other state government recordkeeping stewards in cloud computing environments perceive and act upon electronic recordkeeping requirements in the Cloud, to understand which of the functions of ARM work described by ARM academic literature occur in the recordkeeping environments examined, and to determine whether these functions are performed by ARM workers or by other recordkeeping stewards when they do occur. I addressed the following research questions:

- Within the environments examined, what occupational groups are reported to act as key stewards of the information and how do members of these groups perceive and act upon recordkeeping requirements in the Cloud?
- How do the various stakeholders interact with each other with respect to recordkeeping activities within their cloud computing environments, and what do these relationships suggest about how ARM occupational roles and responsibilities are being handled in the Cloud?
- How do the various stakeholders perceive the roles and responsibilities of archives and records management personnel?
- What cloud computing risks does the professional and academic ARM literature report, and do recordkeeping stewards in state government cloud environments express concerns about these same risks?
• Of the main recordkeeping functions that the ARM literature attributes to ARM workers, are these functions evident in the recordkeeping environments examined and if so, are they performed by ARM workers?

Question one addresses how key recordkeeping stewards in cloud environments in which state agencies operate define and address the recordkeeping requirements of their organization or program. Question two addresses the explicitly recognized relationships between recordkeepers, both with their own occupational group members and with members of different occupational groups. It also compares and contrasts the occupational identification of recordkeepers within the organization to the traditional ARM depictions of ARM workers found in ARM literature. In addition, the question seeks to discover the affinities and disparities between individual occupational members, especially as it may relate to disparities in interests and power across different occupational groups. Question three attempts to understand how the various recordkeeping personnel view the roles and responsibilities of those individuals with the job titles “archivist” and “records manager” within their organizations or programs. This helps to clarify whether they perceive their own jobs as similar to the traditional ARM depictions of those jobs. It also helps to clarify the ways that archivists and records managers distinguish their roles and responsibilities from those of other occupational workers and the ways they may be pressured to distinguish themselves. In addition, it compares the roles and responsibilities within the organization or program to traditional ARM research depictions of the roles and responsibilities of archivists and records managers. Question four attempts to understand the extent to which the recordkeepers understand what risks cloud computing can present to the performance of their duties and to determine whether members of different occupational groups share the same, or
different, concerns about cloud risks. Finally, question five addresses the extent to which ARM functions as found in the literature are performed in organizations or programs and ascertains who (occupationally) performs these functions in the case study environments. Question five further seeks to understand whether the ARM literature’s depictions of occupational roles and responsibilities map to actual roles and responsibilities. The overall result of answering these five questions is a better understanding of the structure of recordkeeping in environments in which a recent cloud adoption has occurred.

3.1. Methodological Strategy

The research strategy used a multiple case study design embedded within an archival analysis of primary and secondary literature. This is based on Yin’s (2008) recommendations for selection of a research strategy. Yin says one should consider the type of research question posed, the extent of control the investigator has over actual behavioral events, and the degree of focus on contemporary as opposed to historical events, coming up with a matrix that one can use to determine appropriate research strategy. This matrix is reproduced in Table 2.

<table>
<thead>
<tr>
<th>Method</th>
<th>Form of the research question?</th>
<th>Control of behavioral events required?</th>
<th>Question focused on current or historical events?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>How, why?</td>
<td>Yes</td>
<td>Current</td>
</tr>
<tr>
<td>Case Study</td>
<td>How, why?</td>
<td>No</td>
<td>Current</td>
</tr>
<tr>
<td>History</td>
<td>How, why?</td>
<td>No</td>
<td>Historical</td>
</tr>
<tr>
<td>Survey</td>
<td>Who, what, where, how many/how much?</td>
<td>No</td>
<td>Current</td>
</tr>
<tr>
<td>Archival Analysis</td>
<td>Who, what where, how many/how much?</td>
<td>No</td>
<td>Historical or current</td>
</tr>
</tbody>
</table>

Table 2 - Yin's Criteria for Selecting Research Design (8)
The research questions addressed in this study are “how” and “what” questions, and they address a subject matter which is both current and over which the researcher neither needs nor has control of the behavioral events. This suggests a possible range of mixed methods, including case study, history, survey, or archival analysis. Because the subject of cloud computing is still very new and very little research has been conducted on the organizational impacts of cloud computing, a case study was chosen over historical analysis, for which very little primary source materials would have been available. Because the individuals chosen to take part in the analysis are reflecting upon their perceptions, a semi-structured interview process was chosen in order to gain as much contextual understanding of those perceptions as possible. Charmaz (2006, 28-29) notes, “Qualitative interviewing provides an open-ended, in-depth exploration of an aspect of life about which the interviewee has substantial experience, often combined with considerable insight. The interview can elicit views of this person’s subjective world.” Archival analysis of primary and secondary source documents and published materials also added to the contextual foundations of the study. Thus, the study itself engaged three types of analysis - semi-structured interview analysis, documentary analysis of source materials, and content analysis of published literature within the archives and records management field - the results of which were triangulated to address not only the similarities and conflicts between the three approaches, but to take advantage of multiple modes of knowing about the subject matter at hand (Glaser and Strauss 2009).

Cresswell (2009) asserts that triangulation involves the comparison of results from two or more data collection activities to determine whether their results support or contradict each other. Small (2011) points out that triangulation is used both to reconcile findings from multiple data sources and from the use of multiple methods of analysis. Triangulation can be
used either to confirm (or disconfirm) findings across different methods (Small). It can also be used to compensate for the potential weaknesses of one analytical or data collection approach by including a different type of approach. Researchers often compensate for the relative shortcomings of purely quantitative or purely qualitative data sources and analyses (Small) by engaging in mixed methods approaches, triangulating results to strengthen their analyses. Jick (1979) argues, in fact, that mixing data sources can provide one with a more holistic and contextual understanding of the units being studied. It is thus particularly well suited for research studies that attempt to gain a comprehensive understanding of how and why individuals choose to engage in particular courses of action (Jick). Glaser and Strauss clarify by pointing out that if one uses multiple sources of data, discovering the emerging categories found in that data during the process of research, one can gain a more comprehensive and potentially creative understanding of the subject; by allowing the researcher to take advantage of the different modes of understanding that different types of data provide (68), multiple data sources lead to a greater likelihood of truly novel discoveries.

3.2. Rationale for Case Selection

This research study used a multiple case study design which included examination of state government recordkeepers in three separate cloud computing implementations. Following both Stake and Glaser and Strauss, I attempted to maximize diversity across cases while minimizing diversity within case environments in order to “learn about complexity and contexts” (Glaser and Strauss 2009; Stake 2005). I did this by focusing internally on the recordkeeping stakeholders and their perceptions within each cloud implementation, while selecting for as great a degree of variety among the structural and governance elements of the
cloud implementations. Thus, each cloud computing implementation represented a different type of cloud arrangement: (1) a statewide cloud implementation within a recently consolidated IT collaboration and communication structure; (2) a cloud implementation undertaken by a single state agency that used a statewide private cloud communication structure existing within a longer-established consolidated IT structure; and (3) a cloud implementation that involved information sharing among several different state agencies, a federal agency, and numerous local, public and private health care organizations.

The first level of analysis in this study involves examining cloud computing implementations in which state government agencies take part. Following that, it examines more closely the relationships between records managers within these environments and the other recordkeeping stewards. In the cases in which records managers are notably absent, the discussion focuses upon the current recordkeeping stewards, with the conclusion (in Chapter 4) dwelling more thoroughly upon the potential implications of and reasons for the absence of records managers.

3.3. Selection of Cases

The first necessary step in a case study design is to determine the boundaries of the case (Yin 2008). The study described here examined the perceptions and relationships of recordkeeping stewards, particularly as they pertain to the traditionally defined functions and activities of records managers. Specifically, it looked at occupational group dynamics under conditions of recent technological change. Several boundary conditions define the cases explored. First, the particular technological change environment examined was the adoption of a cloud computing service in which one or more than one state-level agency participated. Second, the occupational groups were examined after the adoption had “gone live.”
Intergroup dynamics are in a constant state of flux in organizations (even virtual organizations), but the degree of fluctuation is particularly high during the early phases of technological implementations when requirements are not fully defined, technical bugs are still being reported relatively frequently and open bug reports typically present a higher degree of adverse impact to the organization, its employees, or society than those that occur later in the implementation process or after go-live. Likewise, explicit changes in roles and responsibilities also occur during the pre-go-live state and a period of testing and sometimes, trial-and-error process changes occur during the prototype and test phases. In order to reduce the impacts of these temporary changes, only organizations for which the cloud computing adoption had already gone live were included for selection. Still, it was desirable to capture adoptions that had not reached a firm sense of closure or completion for the interviewees in order to ensure that the perceptions of how the cloud implementation had changed (or were still changing) group dynamics would still be fresh in participants’ minds. As a result, only adoptions that had gone live no earlier than 2010 were included.

In order to identify the population of state-level cloud computing implementations available for study, I first examined the state website of all fifty states to determine which states had published cloud computing implementation information on their website. In addition, I performed a wide internet search of online journals and blogs using keywords “state,” “statewide,” “cloud computing” and “The Cloud.” Finally, I performed searches on several well-known computing journals, websites, and blogs that often discuss cloud computing technologies. In addition, I examined the implementations that were highlighted in former U.S. CIO Kundra’s State of Public Sector Cloud Computing (2010). Although
these latter cases were too old to meet the implementation date criterion for inclusion in the case studies, they sometimes led to additional sources for review.

In January, 2012, after identifying a wide variety of states in which statewide or local cloud computing implementations had occurred or appeared to be occurring, I re-examined the websites of all the states to update cloud implementation statistics. All local-only implementations were removed from the list of potential cases, in order to reduce the number of potential confounding factors coming from the variety of fiscal, cultural, and political climate effects that separate state from Table 1 decision making. For the same reason, higher-education (i.e., university) implementations were removed, although implementations from states’ Departments of Education were included. From the articles and the state website searches, I was usually able to determine which of these implementations were still in process and which had gone live (and when). If it was not clear what stage a cloud implementation was in, I contacted agency personnel (as listed on the states’ websites) either by email or by telephone. As a result of this research, I determined that by February, 2012 thirty-nine states were either considering the implementation of a cloud service, in the Request for Proposal (RFP) stage, in the implementation stage, or had gone live. By early April, 2012 I had compiled a list of states involved in cloud computing implementations, along with the nature of the implementation, the specific agencies sponsoring the implementation, and the stage of the implementation. Of the implementations from these 39 states, 30 implementations fit the criteria of involving state agencies and of having gone live since 2010. They are shown in Table 3.

At this point, to validate my understanding gleaned from web-based accounts of these cases, I attempted to contact a wide variety of individuals from all of the states that
### Table 3 - States That Meet the Selection Criteria

<table>
<thead>
<tr>
<th>State</th>
<th>Agency</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>Florida House of Representatives</td>
<td>CES (California Email Services) Cloud-based email consolidation via Microsoft Business Productivity Online Suite (BPOS) of services offered to state agencies; involves email with legal eDiscovery capabilities. State agencies were given the option to use the cloud-based email or the state's in-house email. Although still ongoing, a number of agencies have already implemented the cloud-based service.</td>
</tr>
<tr>
<td>CA</td>
<td>Department of General Services (DGS) Department of Fair Employment and Housing (DFEH)</td>
<td>LogicBit's HoudiniESQ Legal practice management product and HoudiniESQ Right-to-Sue system. This system allows legal practice management, including workflow, document management, email, and calendaring capabilities.</td>
</tr>
<tr>
<td>CO</td>
<td>Statewide</td>
<td>Consolidating its 40 data centers into only three (6 have currently been consolidated), forming a private state cloud.</td>
</tr>
<tr>
<td>CO</td>
<td>Statewide</td>
<td>Consolidated its email and collaboration into Google Apps for Government. It will also consolidate the remaining 15 data centers into one single private cloud.</td>
</tr>
<tr>
<td>DE</td>
<td>Department of Education; Department of Technology and Information (DTI) led a statewide implementation</td>
<td>Entire state adopted Microsoft Live@edu for email, calendaring, and web apps. It provides Microsoft Exchange Server, Microsoft Outlook, and a variety of Microsoft apps.</td>
</tr>
<tr>
<td>DE</td>
<td>Department of State/Division of Professional Regulation</td>
<td>Cloud-based CRM product (i.e., &quot;Constituent&quot; Relationship Management).</td>
</tr>
<tr>
<td>FL</td>
<td>Southwood Shared Resource Center (SSRC)</td>
<td>SSRC, the state's first primary data center, is moving to a private cloud, utility-based computing model. Many of the components have already gone live.</td>
</tr>
<tr>
<td>FL</td>
<td>Florida House of Representatives</td>
<td>Myfloridacensus.org is a statewide census data repository, with GIS data stored and accessibility via relational database on Microsoft Azure.</td>
</tr>
<tr>
<td>State</td>
<td>Agency</td>
<td>Implementation</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>FL</td>
<td>All executive branch agencies and law enforcement</td>
<td>Private cloud email archiving service via Xerox's ACS for all state email. In addition, the FBI's Criminal Justice Information System (CJIS) law enforcement private cloud email went live in March. Thus far, the CIO's office has gone live, some questions remain about how many other agencies will go live, if any.</td>
</tr>
<tr>
<td>GA</td>
<td>Statewide, managed by the Georgia Technology Authority</td>
<td>Moved state agencies to a private government cloud with IBM and AT&amp;T.</td>
</tr>
<tr>
<td>HI</td>
<td>Hawaii State Public Library System</td>
<td>Offers all public libraries (which are part of the State Library System) access through Lenovo's Secure Cloud Access.</td>
</tr>
<tr>
<td>ID</td>
<td>Department of Labor</td>
<td>Department email pilot went live using Microsoft Azure. They are currently updating the infrastructure to move Office 365 to all email users in the state.</td>
</tr>
<tr>
<td>IL</td>
<td>Partially funded by Department of Education; offered statewide; run &amp; operated by IT staffers around the state</td>
<td>IlliniCloud: community infrastructure and cloud services offered to over 150 school districts by a non-profit consortium; data warehouse and reporting capabilities; IaaS services for schools to run their information systems and curriculum applications. This also offers disaster recovery and storage services.</td>
</tr>
<tr>
<td>KS</td>
<td>Statewide</td>
<td>Statewide IT consolidation, offering a statewide private cloud email solution for all agencies; two agencies already host their email in the Cloud.</td>
</tr>
<tr>
<td>KS</td>
<td>Department of Wildlife and Parks</td>
<td>Google Apps for Government.</td>
</tr>
<tr>
<td>KY</td>
<td>Department of Education</td>
<td>The Department began with a Microsoft Live@edu email, communications and collaboration for all 174 school districts. In 2012, they upgraded to Microsoft 365.</td>
</tr>
<tr>
<td>KY</td>
<td>Department of Education</td>
<td>Financial Software Package for all 174 school districts moving to the Cloud.</td>
</tr>
<tr>
<td>KY</td>
<td>Department of Education</td>
<td>Kentucky Student Information System (KSIS).</td>
</tr>
<tr>
<td>MD</td>
<td>Statewide (Department of Information Technology)</td>
<td>Cloud Messaging and Collaboration Services with Google Apps for Government.</td>
</tr>
<tr>
<td>MD</td>
<td>Maryland Education Enterprise Consortium (MEEC)</td>
<td>Google Apps for Education.</td>
</tr>
<tr>
<td>State</td>
<td>Agency</td>
<td>Implementation</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>---------------</td>
</tr>
<tr>
<td>MI</td>
<td>MI Department of Technology, Management and Budget</td>
<td>MiCloud provides storage and hosting services for agencies statewide, using a statewide private cloud infrastructure.</td>
</tr>
<tr>
<td>MN</td>
<td>Statewide; let by Office of Enterprise Technology</td>
<td>State's Enterprise Unified Communications and Collaborations services delivered through Microsoft 365.</td>
</tr>
<tr>
<td>NB</td>
<td>Statewide</td>
<td>Human capital management (through Workday &amp; Cornerstone).</td>
</tr>
<tr>
<td>ND</td>
<td>ND Information Technology Departments and K-12 School Districts</td>
<td>PowerSchool student information system.</td>
</tr>
<tr>
<td>SC</td>
<td>Department of Health and Environmental Control</td>
<td>Google Apps for Government messaging and collaboration services.</td>
</tr>
<tr>
<td>SD</td>
<td>Statewide</td>
<td>Intermap's NEXTMap GIS service.</td>
</tr>
<tr>
<td>UT</td>
<td>Department of Community and Culture</td>
<td>CRM through Salesforce.com.</td>
</tr>
<tr>
<td>VA</td>
<td>Virginia Information Technologies Agency</td>
<td>Enterprise Application Development Platforms, offering a virtualized software development platform, using Amazon EC2 service.</td>
</tr>
<tr>
<td>WA</td>
<td>Washington Sound Transit</td>
<td>Microsoft 365 hosted email and collaboration.</td>
</tr>
<tr>
<td>WY</td>
<td>Statewide</td>
<td>Google Apps for Government Email and Collaboration.</td>
</tr>
</tbody>
</table>
appeared to be engaging in or considering cloud computing, after receiving approval from UNC Chapel Hill’s Institutional Review Board (IRB) in the form of an exemption under the category labeled “4.Existing data, public or deidentified, 2.Survey, interview, public observation” cited under 45 CFR 46.101(b) of the Office for Protection from Research Risks (OPRR) at the U.S. Department of Health and Human Services (http://www.hhs.gov/ohrp/policy/exmpt-pb.html).

I contacted individuals both to assess the most appropriate implementations to include as case studies and to validate my understanding of the cloud computing implementations. In addition, I wanted to speak with individuals whose states did not necessarily meet the requirements of inclusion as a case study but whose reports could provide comparative information to be used when analyzing the chosen cases. In other words, when questions regarding relations in cloud computing came up, I wanted a basis for comparing the perceptions of those individuals who were part of the case study with a more general population of individuals who were not part of the case study. Thus, I sent requests for participation to individuals at all of the original 39 states identified. I sent requests to the state archivists, to CIOs when this contact information could be discovered, to members of the IT groups involved in the cloud implementations when their contact information could be discovered, and to any specific individual that was listed on the state website as affiliated with the cloud implementation when their contact information could be discovered. In addition, I spoke with an individual at the National Association of State CIO’s to determine more about what is occurring within the various states and to gain information and recommendations on further approaches to use to initiate contact with state employees. Although contact information could not be found for all state cloud implementations, I was
able to identify 69 individuals from: (i) a state professional organization involved in one of my cases, (ii) two independent contracting firms working directly on an implementation in one of my cases, and (iii) the U.S. Centers for Disease Control (CDC), which is an integral partner in the case involving a shared services cloud implementation. The list of states or districts from which I was able to garner interviews is given below. Those states that were chosen as case study subjects are emphasized via underlining and bold font.

- Florida
- Georgia
- Kansas
- **Kentucky**
- **Minnesota**
- Nebraska
- New Jersey
- **North Carolina/Washington, D.C./Independent Consultants**
- South Carolina
- Wyoming

To request interviews I sent an email letter to all potential interviewees requesting participation in my study, as well as a “Consent to Participate” form. The template of the letter of request is provided in Appendix D and the Consent to Participate form is provided in Appendix E. After obtaining consent, I scheduled telephone interviews with the participants. Typically I called the interviewees at their offices, although on a few occasions they called me. The interviews were recorded and the audio transcripts were saved to a password protected folder.
Between April and July, 2012 I conducted 25 interviews. I conducted a follow-up interview with a new participant (on the basis of a later recommendation made by a previous interviewee) in early February, 2013 to gain some crucial missing information on a project, two conversations with a new participant in April, 2013 and another new one in June, 2013. I also conducted a follow-up conversation requested by one of my earlier interviewees when new information became available at his implementation site. In total, I conducted 38 telephone interviews throughout the project. Of these, 29 were with participants that represented the implementations included in the case studies. During the course of each interview, I used snowball techniques to elicit other potential interviewees from the current interviewee, based upon the current interviewee’s belief that the referred individual could help me understand a variety of aspects and points-of-view associated with the cloud implementation. In addition, I engaged in follow-up email queries to the original interviewees, resulting in twenty-one email “conversations” (i.e., twenty-one threads, not counting the back and forth responses on both sides) in July and August, 2012. Finally, as a result of interest in the project, some of the original interviewees spontaneously sent me further documentation on the project or about their roles and responsibilities within the project and organization. This occurred on thirteen separate occasions between July and October, 2012.

Those interviewed represented a wide variety of occupational groups, including: State Archivist, Archivist, Records Manager, Data Practices Liaison, Compliance Official, Chief Technology Officer, CIO, Senior Director, Director, Manager of IT, Data Architect, Epidemiologist/MD, Program Director, Independent Consultant, Implementation Director, Collections Assistant, Senior Policy Analyst, and Program Director. Backgrounds and
training included records management and/or archives, law, computer science and/or IT, and medicine.

3.4. Post-Interview Process

After each interview, I took brief notes of the key themes that I had perceived, and then I used transcription software to transcribe all conversations to Word documents. These transcripts were then compared against the audio files once again when I exported the audio files to NVIVO 9.0 and copied the text documents line-by-line to NVIVO in “Transcribe” mode. This functionality allows one to select any particular portion of conversation and “follow along” with the text by listening to it, or to find particular parts of the interview to review both textually and aurally. It also then allows one to code not only at the line level but also to capture all portions of audio files that have been coded at the same nodes.

Nodes are what each code category is called in NVIVO. They can represent different topics, themes, relationships, causes, people or other entities, as well as time periods and geographical locations (although one can also use classifications in NVIVO to classify entire nodes or collections of nodes). The nature of one’s qualitative coding, however, depends a great deal upon one’s theoretical outlook, as Charmaz shows when she compares Glaser and Strauss’s ideas of appropriate coding techniques. I initially engaged in line-by-line coding, as recommended by Charmaz, who points out that this type of coding not only helps one remain “open to the data and see nuances in it,” but also can help one “identify implicit concerns as well as explicit statements,” thereby allowing one to refocus interviews (50). Using the insights from this detailed coding effort, I was able to discover a number of relationships between codes, participant responses, and case study sites, which finally emerged as key categories, or themes, in the project. To do this, I followed Glaser and
Strauss’s advice to engage in simultaneous data collection, coding, and analysis (101). The themes discovered will be explored in more depth in chapter 4.

I engaged in both in vivo coding and thematic coding. In vivo coding involves relying specifically upon the participants’ own terms and phrases and using those terms and phrases as a code. For example, one phrase that was used by both IT and non-IT personnel in interviews across several different states undergoing cloud computing at the same time as their state’s IT consolidation, was the phrase “sucked into.” Specifically, in three different cloud implementations (two of which are case studies and one of which involved interviews with individuals from a state that was not a case study site but did have a simultaneous cloud adoption and IT consolidation), interviewees referred to local IT employees as being “sucked into” the central IT department’s control when consolidation occurred. Although one could classify this as “consolidation” or “being consolidated,” the phrase “sucked into” conveys much more of the emotional ambivalence portrayed by the interviewees. When I began creating code hierarchies, however, this term was itself “sucked into” the higher-level category “IT consolidation” for purposes of keeping track of the number of consolidations that occurred simultaneously with cloud implementations. I refer to this phrase again in Chapter 4.

Coding was an iterative process which involved returning to past interviews to compare the codes with each fresh interview’s data in order to see if new or continuing themes had emerged. I coded each interview as a single entity and compared it to the pre-existing coding and themes from the pre-existing interview transcripts. Thus, I moved from the initial line-by-line coding during the first interview at each site to focused coding during the analysis of the line-by-line coding and during the comparison of further interviews to the
pre-existent line-by-line coding. Focused coding involves sifting through the detailed codes and determining the most significant or frequent codes (Charmaz). It is more directed and conceptual coding that comes from synthesis of individual detailed codes into higher order themes or categories that map to the data. The process is not linear, however, and when new codes are discovered it can lead one back to more detailed line-by-line coding. Thus, the entire coding process was highly iterative. In several cases, new material arose in subsequent interviews that led me to do line-by-line coding of portions of the later interviews as well as re-assessment of the earliest interviews.

3.5. Unit of Analysis

The definition of the unit of analysis, which is the primary topic of the study, and is “the same as the definition of the ‘case’” (Yin, 24), is to be distinguished from the data sources (Yin), which can provide either direct or contextual information regarding the unit of analysis. This study examined the perceptions of recordkeeping stewards regarding their roles and responsibilities and relationships with stewards from other occupational groups within a particular cloud computing environment. It compared these perceptions to previously existent literature representing the perceptions of ARM academic professionals about ARM recordkeeping roles and responsibilities and the occupational relationships under changing technological environments. Hence, there are two primary units of analysis embedded in the overall study of recordkeepers’ perceptions of cloud computing: recordkeeping stewards who have engaged in cloud computing implementations involving state government and ARM academic professionals who have discussed ARM roles and responsibilities under conditions of changing technologies. “The Cloud,” however, is treated as appropriately representative of one particular type of technological change and using a
cloud computing implementation that has already gone live within the past three years
allowed me to temporally bound the study and to avoid muddying the analysis with a variety
of different technologies. In order to analyze these units of analysis, a variety of data sources
were used, including interviews with recordkeeping personnel who work outside of the
particular cases that were studied.

3.6. Data Sources
As mentioned earlier a variety of data sources were used to analyze recordkeeping in
state cloud environments:

- Interview data from occupational group members that act as recordkeeping stewards
  within the cases identified, such as IT employees, records managers, data compliance
  officials, archivists, lawyers, data practices personnel, independent consultants,
  agency directors and senior directors, members of state professional associations, and
  MD/epidemiologists.
- Interview data from occupational group members that act as recordkeeping stewards
  external to the individual cases identified, including IT employees, records managers,
  archivists, agency directors and senior directors, and members of professional
  associations.
- State statutes and state records laws relevant to the particular cloud computing
  investigations.
- Federal statutes and laws relevant to the particular cloud computing investigations.
- Published literature from the cloud implementations included as cases, including
  requirements documents, internal emails and memos, states’ Central IT directives,
  informational materials, published internal audit results and press releases.
• Published articles from a variety of ARM journals (discussed below).
• Information Technology policy documentation.
• Emails from state employees (to me).
• Information regarding the implementation and organization from state websites.
• Internal job descriptions.

3.7. Documentary Analysis of Site-Specific Materials
   When I examined the documentary material gleaned from the three case study sites, I engaged in a similar process of coding as occurred during the interview process. However, I did not engage in line-by-line coding of the documentary material. Rather, I engaged in focused coding and constant comparison of the documents with interview responses. (The line-by-line coding was used early in a site’s interview process to help me determine themes, categories, and relationships about which I could further investigate with interviewees.) The documentary material was analyzed to gain background information on the particular implementations, to understand the legal and cultural constraints placed upon the implementation participants, and to determine the outward-facing projection that the organizations’ upper management wanted to present to employees and to the public at large. Requirements documents provided a technique to validate interview responses and also provided further information regarding the particular values selected to define the necessary activities and constrain the scope of the individual implementations.

3.8. Content Analysis of Published ARM Literature
   I conducted a content analysis of published literature from six archives and records management journals over 43 years. This literature was examined using qualitative and
descriptive techniques to gain an understanding of how ARM research has historically viewed the roles and responsibilities of ARM workers, how it has delimited these workers’ spheres of control, and how it has depicted the impacts of technological change on ARM roles and responsibilities.

3.9. Literature Selection

I began searching articles from 1970 onward, or from the date of first publication of the journal if that date fell after 1970. The six journals selected were: *The American Archivist*, *Archivaria*, *Archival Science*, *Records Management Journal*, *Archives and Museum Informatics*, and *Journal of the Society of Archivists*. Because of the variation in abstract availability and full text searching capabilities, several techniques were used to discover the relevant literature.

*The American Archivist* is available for full-text search through JSTOR’s *Data for Research* functionality (http://dfr.jstor.org/). Because early versions of this journal do not contain abstracts, I performed searches for literature that contained one (or more) of the following text strings somewhere within the body of the text (including titles):

- Machine-readable
- Automat*25
- Continuum
- Cloud comput*
- Comput*

25 The asterisk (*) reflects a “wild card” character, which enables one to search for any potential conclusion to the term after the asterisk. For example, searching for comput* would bring back search outcomes that include computer, computers, computing, and so forth.
The articles were provided in comma-separated values (CSV) format and I exported them to Excel. I cleansed the data to remove unnecessary columns, merged the separate files that were downloaded, removed duplicates, and sorted the information ascending using Volume, Issue, and First Page as sort criteria. A number of items that occur regularly in this journal were also removed. These included: books reviews, news notes and front- and back matter, technical notes, bibliographies, abstracts of foreign periodicals, SAA committee and meeting notes, and errata. What remained were research articles and letters to the editor. I allowed the latter to remain on the grounds that in this journal these letters frequently represent researcher viewpoints and thus can provide a variety of opinions about materials related to technology. After the download, cleansing, and removal of unnecessary components, 1,029 articles remained for further review. I was aware that a number of these articles may contain one or more of the above mentioned words and yet still be irrelevant for the purposes of this study. (For example, if the word “digital” were to occur prior to the beginning of the article to indicate that a digital copyright is held, the article may contain that word and yet represent a topic entirely unrelated to technology or technological change in the ARM profession.) Because of this, I manually skimmed these articles individually to determine which were relevant and which were not. 156 articles were found to be relevant.

After this initial review, I examined the remaining 156 articles to classify them based on the following factors:
• Whether they discuss archivists or records managers (or both);
• Whether records management is the main topic of the article;
• Whether ARM functions are discussed;
• Whether roles and responsibilities are discussed; and if so, whether they discuss roles
  and responsibilities or functions as being archival, RIM, or records management-
  oriented, or conjoined;
• Whether computer technology or its effects on ARM occupation(s) are discussed;
• Whether the article discusses its primary topic in relation to the operation or mission
  of an archival- or of a non-archival organization.

I conducted a similar search on the content of Records Management Journal, using
the same keywords and criteria: 54 articles were determined to be relevant. Likewise, I
conducted a similar search on the content of the Journal of the Society of Archivists, resulting
in 42 relevant articles from that journal.

Archivaria and Archival Science both contain online abstracts. These were manually
reviewed to determine the relevancy of the articles to the subject at hand. The first
assessment made was to verify that the content type was, in fact, a research article. Some
content types were excluded by their nature: book reviews, editorials, communications
regarding meetings and society formations, submission instructions for authors, issues
introductions, lists of contributors, and author indices, as well as volume tables of contents.
The next assessment was to examine the abstract to determine relevancy, using the criterion
given below. If the article proved immediately to be irrelevant on the basis of that criterion, it
was categorized simply as “no.” If it appeared immediately relevant or potentially relevant
on this basis, it was tracked in an Excel spreadsheet that contained fields showing: online
identifier (if available), the content type (i.e., book review, research article, letter to editor, etc.), the subject-matter category (e.g., technological, professional identity, recordkeeping, etc.), author, journal title, article title, and keywords, where available. Categories allotted to the articles were “technology,” “professional identity,” “nature of the archives,” “recordkeeping,” “business archives,” “archival education,” “digital convergence,” and “organizational culture.” During the process of categorization I examined more deeply the articles that were categorized as “technology,” “professional identity,” “nature of the archives,” “recordkeeping,” and “archival education.” 270 relevant articles were found in Archivaria and 114 were found in Archival Science.

I also assessed the individual articles in the journal Archives and Museum Informatics for information relevant to the impacts of technological change on archives and records management, however, this proved to be somewhat of a conundrum. Virtually all of the articles in this journal were related to technology in some way. However, the vast majority of articles focused on technical reviews and how-tos, in contrast to research or theoretical exposition on recordkeeping or professional identity, education or roles in the face of changing technology. Ultimately, after a detailed review, 46 were found to be relevant.

Thus, in total, I compiled 682 articles from these six journals for further assessment and review. The documentary analysis conducted used techniques to identify themes as reported by Ryan and Bernard (2003). I used five eight techniques of analysis described by these authors: repetition, or “topics that occur and reoccur” (89); “indigenous typologies,” which are the equivalent of Glaser and Strauss’s in vivo coding; deduction of themes from metaphors and analogies presented by authors; examination of similarities and differences between different texts that discuss the same topic; and searches for missing data by asking
of the articles, “What is missing?” (92). By engaging in detailed examination of the texts I was able to develop a representation, or story, of the roles and responsibilities of archivists and records managers from the framework of the academic literature. This representation was used as a proxy for perceptions of various strands\textsuperscript{26} of archival thought and when compared to the results of the interviews and the various states’ documentation, allowed me to compare the academic portrayal of ARM roles and responsibilities to the “real-world” roles of recordkeeping stewards as perceived by the stewards themselves.

\textsuperscript{26} For example, sample “strands” are “post-custodial,” “continuum theory,” or “life cycle theory.”
4. RESULTS

This section discusses the results of three cases examined for this project. In order to help ensure anonymity of interviewees, when referring to specific individual respondents, their gender is reported in a random fashion. That is, in one section Person 14 may be treated as male, while in another section he or she may be treated as female. However, within the course of any one discussion point, I maintain the currently assigned gender. The random gender assignment seeks to make it difficult to deduce the identity of individuals. Gender for an interviewee in any given discussion point was chosen by random number generator (http://www.random.org/integers/), where the integers one to five were treated as female and six to ten were treated as male.

The discussion is structured by case, and is then followed by the key findings that arise when attempting to answer these questions. Table 4 shows the original five research questions and the chapter’s sections that map to each question.

An interpretive method such as the one used here portrays the respondents’ subjective understanding of their environments and offers hypotheses about why they believe and behave as they do in that environment. However, the method does not guarantee generalizability of the findings. As a result, the findings derived in this study will be presented along with hypotheses (in Chapter 5) that can inform future research into recordkeeping in cloud computing environments specifically, or recordkeeping in new technological environments in general. When relevant, related theoretical approaches that
<table>
<thead>
<tr>
<th>Question</th>
<th>Section(s)</th>
</tr>
</thead>
</table>
| Question #1: Within the environments examined, what occupational groups are reported to act as key stewards of the information and how do members of these groups perceive and act upon recordkeeping requirements in the Cloud? | 4.1.2 – Recordkeeping Stewards (Minnesota)  
4.1.4 – Requirements and Actions (Minnesota)  
4.2.2 – Recordkeeping Stewards (BioSense 2.0)  
4.2.4 – Requirements and Actions (BioSense 2.0)  
4.3.2 – Recordkeeping Stewards (Kentucky)  
4.3.4 – Requirements and Actions (Kentucky)  
4.5.2 – Reported Recordkeeping Stewards (All cases jointly)  
4.5.3 – Perceived Requirements in the Cloud (All cases jointly) |
| Question #2: How do the various stakeholders interact with each other with respect to recordkeeping activities within their cloud computing environments, and what do these relationships suggest about how ARM occupational roles and responsibilities are being handled in the Cloud? | 4.1.7 – Interactions between Recordkeeping Stewards (Minnesota)  
4.2.7 – Interactions between Recordkeeping Stewards (BioSense 2.)  
4.3.7 – Interactions between Recordkeeping Stewards (Kentucky)  
4.5.4 – Interactions Between Recordkeeping Stewards (All cases jointly) |
| Question #3: How do the various stakeholders perceive the roles and responsibilities of ARM personnel? | 4.1.5 – Stewards’ Perceptions (Minnesota)  
4.2.5 – Stewards’ Perceptions (BioSense 2.0)  
4.3.5 – Stewards’ Perceptions (Kentucky)  
4.5.5.2 – Stakeholders’ Perceptions of ARM Roles and Responsibilities (All cases jointly) |
| Question #4: What cloud computing risks does the professional and academic ARM literature report, and do recordkeeping stewards in state government cloud environments express concerns about these same risks? | 4.1.4.2 – Acting Upon Requirements (Minnesota)  
4.1.7.2 – Perceptions of Changes Brought by the Cloud (Minnesota)  
4.1.8 – Synopsis of Case 1 Findings (Minnesota)  
4.2.1.1 – The Decision to Move to the Cloud (BioSense 2.0) |
<table>
<thead>
<tr>
<th>Question</th>
<th>Section(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.3 – Legal Environment Affecting Stewards (BioSense 2.0)</td>
<td></td>
</tr>
<tr>
<td>4.2.5.2 – Perceptions of “Cloud Computing” (BioSense 2.0)</td>
<td></td>
</tr>
<tr>
<td>4.2.8 – Synopsis of Case 2 Findings (BioSense 2.0)</td>
<td></td>
</tr>
<tr>
<td>4.3.1.1 – The Decision to Move to the Cloud (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.3.5.2 – Perceptions of “Cloud Computing” (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.3.7.1 – Perceptions of Working Together in the Cloud (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.3.7.2 – Perceptions of Changes Brought by the Cloud (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.3.8 – Synopsis of Case 3 Findings (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.4.2 – The ARM Literature on Cloud Computing</td>
<td></td>
</tr>
<tr>
<td>4.4.6 – Cloud Risks</td>
<td></td>
</tr>
<tr>
<td>4.1.5.1 – Perceptions of “Records” (Minnesota)</td>
<td></td>
</tr>
<tr>
<td>4.2.5.1 – Perceptions of “Records” (BioSense 2.0)</td>
<td></td>
</tr>
<tr>
<td>4.3.5.1 – Perceptions of “Records” (Kentucky)</td>
<td></td>
</tr>
<tr>
<td>4.4.1 – The ARM Literature on Recordkeeping Roles and Responsibilities (Discourse Analysis)</td>
<td></td>
</tr>
<tr>
<td>4.4.5.1 – ARM Functions and Their Occupational Allocation</td>
<td></td>
</tr>
<tr>
<td>4.5.6.2 – Stewards’ Perceptions of ARM Roles and Responsibilities</td>
<td></td>
</tr>
<tr>
<td>4.6.1 – “A variety of individuals from different occupational groups act as recordkeepers within their organization”</td>
<td></td>
</tr>
</tbody>
</table>
support (or contradict) the findings from this study are also provided in Section 4.4, the chapter’s discussion section. These theoretical explanations can provide a foundation for future research and offer potential frameworks for interpreting the results.

4.1. Case 1: Statewide Email and Communications Cloud

4.1.1. Introduction

In 2012, Minnesota completed its implementation of Microsoft 365, a cloud computing communications and collaboration service. The email services of the entire executive branch of the state migrated to the cloud service and the state has now also begun implementing Microsoft Share Point in the cloud. Recordkeeping stewards and executive management discussed their perceptions of the rationale for and effects of this implementation, as well as sharing detailed information about their roles and responsibilities as recordkeepers. Taking part in these conversations were individuals who self-affiliated with occupations within the Executive Branch’s organizational hierarchy and agencies, including the CIO’s staff, the Information Policy and Analysis Division (IPAD), the Minnesota Historical Society and State Archives, MN.IT Services (the state’s central IT service provider), and records management personnel from five agencies. These individuals all identified themselves and the members of the other groups as comprising the state’s key recordkeeping stewards. Interview responses were supplemented and validated by internal documentation, internal change management materials, legislative statutes, and literature published on the web in order to come to an understanding of the recordkeeping environment within which this cloud computing service arrangement operates.
4.1.1.1. The Decision to Move to the Cloud

Prior to 2005, the state operated within a federated information governance structure, with IT service provision spread among the agencies and operating largely independently. One interviewee referred to this arrangement as highly “organic” (P-19), with pockets of growth springing up spontaneously in a number of separate locales. In 2005, however, under the authority of Executive Order 05-04 and Minnesota Statute 16E, Minnesota’s Office of Enterprise Technology (OET) was created. At this time, the State CIO’s office was given agency-level standing on then-Governor Tim Pawlenty’s subcabinet, helping to further his 2004 “Drive to Excellence Initiative” (Minnesota Statutes 2005, Sections 16E.01 - 16E.20; MN Exec. Order No. 05-04 2005; State of Minnesota 2004). The goal of this initiative was to enact a long-term plan that would consolidate the many highly autonomous elements of Minnesota’s distributed agency structure into a more cohesive, yet still federated, enterprise structure. The stated motivation behind this initiative was to achieve greater collaboration, clearer communication, and less redundancy of work between the diverse elements of state government, a goal still driving the state’s strategic planning. The Drive to Excellence Transformation Roadmap (State of Minnesota 2005) that presented the business case elements and necessary steps to achieve the initiative’s goals asserted that the creation of an Enterprise IT Governance Structure was a “fundamental element” required to see the plan through to its successful completion (18). The Drive for Excellence team members also determined that the state needed to mitigate the risk of insufficient adherence on the part of the agencies to a standardized and consolidated IT Governance Structure. To achieve this, they recommended “compliance incentives” such as shared funding (229).

Shared funding was not forthcoming at that time, however. Although the 2005 legislation provided additional central oversight of IT projects, agencies still retained a great
deal of autonomy in their purchasing power, since they were able to make purchases of less than $1,000,000 without obtaining any explicit permission from the CIO’s office (Hrdinová, Helbig, and Raup-Kounovsky 2009). This provided minimal centralized control over project funding and hence, over agencies’ abilities to undertake independent IT projects (P-12). Furthermore, then State CIO Gopal Khanna determined that the immense scope of change management would best be served by moving toward consolidation incrementally, relying on a collaborative planning process that included “Minnesota’s agency heads and their chief information officers or directors of Information Technology in decisions affecting their agencies and agency customers” (MN OET 2007b, 1). At that point in time, numerous, separate email, voice and voicemail systems remained. As of 2007, the Executive Branch had between 30-40 different email systems, as well as 30-40 different voicemail systems (P-19). Cohesive, shared communication within and between agencies was not yet possible.

As a result, in 2007, OET requested and received legislative approval to undertake a comprehensive consolidation of the executive branch’s communications infrastructure (MN OET 2007a), and commenced the project in 2008 after selecting Microsoft Exchange Server 2007 as the platform (MN OET 2009; Microsoft 2012). The project took 18 months to implement and when it was complete, senior executives in OET realized that in spite of having achieved project goals successfully (P-19), the state was already falling behind technologically again. An executive level interviewee noted,

So, the RFP was fulfilled, we built it, and at the tail of that project … we've got everybody on the same platform and here come some of the reasons why we considered the cloud solution. We're looking at the tail end of a project that tried to consolidate and was, at the origin, the most modern platform for communication … And when we were at the tail end of that, we were already behind, right? So we've completed a project and we already have an obsolete system. We look at any

---

27 This rule will be called the “million dollar rule.”
upgrades to that system and we look at another million bucks. We see that this thing doesn't scale very well. We've got a lot of difficulties in understanding exactly what will this take. We want to take that same communication platform to other non-state agencies, such as other branches of government - the judicial branch, and the legislative branch, perhaps - and perhaps other levels of government, and we kind of scratch our head and go, “Wow. We're going to be doing this e-mail system, this e-mail stuff, for quite a bit because, wow, it takes a lot of effort to do so.” So we … see the tail end of the project as pretty much a successful project. But then it evolves, because back in ’07 … nobody was really seeing SharePoint as a collaboration infrastructure, a platform … The whole idea or activity around social networking and group netware internally didn't seem needed, but three years later in 2010, when we were finishing up the project, everybody's saying, “Yeah, not just e-mail, but how about SharePoint?” And same problem, right? It's growing organically all over in the agencies, and now … if we look at both capacity, core competency, and simply limitations of being in this never-ending support role of basic communication infrastructure, … right around that time is when the majority of the major vendors were saying, “Cloud, cloud, cloud” (P-19).

The decision to implement a cloud service, according to the above interviewee (who reports directly to the CIO), was a strategic decision that involved considerations of core competencies and opportunity costs. The executive managers at OET realized that in order to provide a cohesive enterprise-wide communications structure they needed a system that could scale relatively easily and cost effectively. In addition, they judged that the core mission of the government is to provide services for citizens, and that the OET needed to find the most effective way to allow this. P-19 noted,

Is the vision of the Governor that we become the best e-mail service provider? Or is the mission of the government that we use e-mail in order to communicate more effectively on the services that we provide? So … there's a core competency item, and there's also an opportunity cost item. There's an opportunity cost in the context of … do I really want to spend time with the best and brightest, including my own, on making sure this thing runs, or do I just want to forget about it and have somebody else do it on our behalf, that it is their core competency and it is their time and investment of time to polish the vehicle, if you would?

P-19 indicated that the decision to implement the cloud computing service was not made totally on the basis of cost and that the total costs of ownership (TCO) between an on-premise service and the cloud service did not differ significantly, although the cloud service
may have been slightly less expensive.\textsuperscript{28} Rather, the decision was a long-term strategic decision that would allow a new form of IT activity in which OET could begin to act not only as cloud customers, but also as cloud brokers for other state entities, and as a cloud provider, since it had already consolidated some of its data centers and was providing private cloud services through those data centers, its large mainframe service, and its large-scale telephony services (P-13, P-19). Moving to the Cloud represented a decision to begin moving into an integrator role, wherein the OET would make decisions as to whether to purchase, build, or broker IT products and services for the entire state’s needs in a more cohesive, consolidated way to achieve long-term strategic business requirements. For example, in June, 2012 the city of St. Paul, MN contracted with OET to place its email onto the Office 365 service (Heaton 2012, P-13, and P-19). Chargeback occurs on a cost-of-service basis,\textsuperscript{29} by providing the services for predetermined, publicly accessible rates (MN.IT Services 2012).

The state’s cloud implementation was ultimately geared toward enabling cross-organization collaboration; increasing the efficiency of state workers by allowing real-time communication and collaboration; allowing the creation of more standardized processes and policies across agencies, thereby making it easier for workers who transfer from one agency to another to avoid excessive re-training; and keeping pace with continuous upgrades in software and hardware procurement (e.g., mobile technology) in a cost-effective manner (MN OET).

\textsuperscript{28} The project implementation team did not create a business case, but did compute a Total Cost of Ownership (TCO) at a per person level. The team also compared the computed Cloud-TCO to the then-known TCO of non-cloud services. Although the initial TCO figures were available for current review, the original non-cloud TCO calculations are no longer available. Because there is no comparator for the TCO figures obtained, they are not included here.

\textsuperscript{29} MN.IT Services does not earn a profit on what they charge entities. They charge only the calculated cost of service provision.
The movement toward consolidation of IT services within the state received a boost during the legislative session effective July, 2012, when the legislature passed a bill that consolidated the entire state’s IT personnel into OET, which has now rebranded itself as MN.IT (pronounced “Min-it”) Services. As of now, all the various agency CIOs report directly to the MN.IT Services CIO and all IT personnel have been folded into OET, causing that organization to grow from one comprising about 350 people to an organization of about 1,800 people (P-13). In addition, all funding to the various agencies now flows directly from MN.IT Services, giving the newly consolidated organization considerably more control over purchasing decisions on the part of agencies (P-12). Although still a federated system, the State of Minnesota has moved towards a more centralized IT governance structure, in terms of IT resource control, communication structures, and budget. For example, the state has a number of approved vendors on state contract. If an agency wishes to engage in a project with one of these vendors it can do so. If it wishes to engage in a project with a vendor that is not on the approved list, however, it must go through an RFP process, which will require formal approval and ultimately, CIO sign-off (P-13). The million dollar rule mentioned earlier still holds, but when an agency seeks to implement IT projects, the funding passes through the state CIO’s office, so the link between authority structures and financial structures is explicitly cemented in decision makers’ assessments and practices (Minnesota Statutes 2012, Section 16E.14 - 16E.145).

This provides a new locus of domination (or facility in Giddens’ terminology) for employees. By strengthening the authoritative characteristics of the IT governance structure, IT has received greater control over the internal actions of other employees, both within and external to the IT group itself. Because many of the functional responsibilities of non-IT
occupational group members rely upon IT infrastructure, in essence IT has gained power over the functional capabilities of the non-IT state employees.

4.1.2. Recordkeeping Stewards

4.1.2.1. Minnesota’s Information, Data, and Recordkeeping Environment

In the state of Minnesota, the main occupational groups serving as recordkeeping stewards are records managers, data practices officers and liaisons, IT professionals, and archivists. In an ideal recordkeeping world, one would include all the records creators as well, but in fact there are sound reasons for considering the records creators to play a different stakeholder role than those who officially act as stewards. Records creators in the state, as in most organizations, rely upon the other occupational groups to educate them about issues such as retention, long-term value, and access policies or to provide them with the technical resources and training to create the records used in day-to-day transactions. Few creators have a clear understanding of why and how retention and disposition occur (P-8, P-12, P-21, P-22, P-23, P-24, and P-25). If questions about rights of access or potential violations of state data practice laws arise, a Data Practices professional is called upon to give advice or write an opinion. Records management personnel draft the retention policies. The State Archives, in conjunction with the state’s Records Disposition Panel, approves the retention policies and determines which records will be transferred to the state archives when they are no longer needed for transactional purposes. IT personnel assign and provide access to the infrastructure and applications, and ensure that security and privacy of information, among other legally mandated data and recordkeeping requirements, are met. Hence, individuals from these stewardship groups provide the advice and actions necessary to ensure that records are maintained according to organizational, legal, and societal expectations.
4.1.2.2. Minnesota Historical Society and State Archives\textsuperscript{30}

The Minnesota Historical Society (MHS) “collects, preserves and tells the story of Minnesota’s past through museum exhibits, libraries and collections, historic sites, educational programs and book publishing” (MHS 2012a). The Society is a private, not-for-profit organization that was chartered in 1849 by the Minnesota Territorial Legislature (MHS 2012b). This structural arrangement is important because the Minnesota State Archives is a department of MHS and therefore not a state agency. When asked about the relationship between the state archives and the other recordkeeping stewards, who operate from directly within the state’s executive branch structure, MHS interviewees reported that MHS acts somewhat outside the regular channels of the state because it is more of a “quasi-agency” (P-8).

MHS is physically and administratively separate from the state’s executive branch agencies. Although the state has maintained archives since 1913, when MHS was designated the state’s “Department of Archives and History” (MHS 2012b), the current formal assignment of the state archives as a department of MHS occurred in 1971, by statute (Laws of Minnesota 1971, Chapter 529, H.F.No. 2670). This statute also created the Records Disposition Panel, composed of the attorney general, the state archivist, and either the legislative auditor (in the case of state records) or the state auditor (in the case of local records) to authorize the destruction or sales of records and the movement of records of permanent value to MHS.

As a non-profit organization that was independently set up and later assigned responsibility for maintaining the state archives, the state archives has no records

\textsuperscript{30} The Historical Society’s website can be found at http://www.mnhs.org/.
management authority in the state. It does, however, have the partial authority to determine which records are deemed to have long-term preservation value and therefore should be sent to the archives when their transactional use is complete. The archives can do this because it receives and officially approves the retention schedules from agencies via the State Archivist’s place on the State Records Disposition Panel. However, the MHS interviewees reported that they play a non-custodial role over electronic records. Sometimes (as in the case of the I-35W bridge collapse31) they become aware of records that clearly will have long-term value and will contact an agency and collaborate to make sure that the records are preserved for the long-term. Usually, however, they merely offer “guidance to the agencies so that they can keep these records within their agency structures, because they have enduring value to the agencies” (P-8). Interviewee P-8 indicated that the agencies generally have more resources required to maintain the records than the State Archives has. In other words, the State Archives does not take on a greater role in current long-term recordkeeping because it does not have the resources to do so.

One key theme that arose in interviews with MHS staff was that of being external or outside the state agency structure and therefore neither directly involved in state agency IT or recordkeeping decisions nor directly aware of agencies’ uptake of information technologies. Another theme was that of non-custodialism. Many records of continuing value to the agencies are kept on a permanent basis by the agencies themselves, rather than being transferred to the state archives. Instead, the archives provides guidance and advice on the preservation of these records via informal relationships with agency staff. The third key theme was that of archives personnel acting as consultants, by providing advice and influence

31 On August 1, 2007, the Mississippi River crossing in downtown Minneapolis, called the I-35W bridge, collapsed, killing 13 people and injuring 145 others (FEMA 2007; Louwagie 2012).
through relationship-building, via their work on the Information/Data Domain Team, and by offering themselves as experts on records management. The Archives reviews and approves retention schedules, via the state archivist’s placement on the State Records Disposition Panel, but otherwise plays no role in retention decisions made by individual agencies. Neither does it have any authority over the records management operations of any state agency. Thus, the State Archives does not manage the pre-archival records management activities of the state.

4.1.2.3. Records Management (within the Department of Administration)

Although the records management functions were transferred to the Department of Administration in the 1971 law referenced above, there is no centralized records management program in Minnesota (P-8). The 1971 law states that the Commissioner of Administration oversees the state records center and has the power to establish “standards, procedures and techniques for effective management of government records,” and that the head of each state agency and the governing body of “each county, municipality, and other subdivision of government” is expected to cooperate “in conducting surveys and to establish and maintain an active, continuing program for the economical and efficient management of the records” (Laws of Minnesota 1982, Chapter 573, H.F.No. 534). In practice, however, P-8 reported that during budget cuts several years ago, the Department of Administration “took themselves out of that business.” Now, “there is no one central office coordinating records management or anything like that in the state” (P-8) and each agency conducts its own records management.

32 The Information/Data Domain Team is one of four teams that operate under the purview of the Enterprise Architecture group, which reports to the State CIO. The state archivist is a member of this domain team, which reviews and makes recommendations on requests for policies, standards and guidelines. A data architect in the Enterprise Architecture group of MN.IT Services is another member. As a result, the state archives has a direct, albeit somewhat informal, line of communication with MN.IT Services.
activities, without centralized direction (P-8, P-9, P-21, P-22, P-23, P-24, P-25). A number of agency records managers play a dual role, with a proportion of their time assigned to records management responsibilities and the rest of their time assigned to data practices or other responsibilities. Some agencies have no records management personnel at all, or have clerks performing records management duties.

4.1.2.4. Information Policy Analysis Division (IPAD)

The IPAD personnel are responsible for assisting and consulting with individuals, government entities, businesses, and associations regarding the Minnesota Data Practices Law (Minnesota Statutes 2012) and the Minnesota Open Meetings Law (Minnesota Statutes 2012, Section 13D). Almost but not all IPAD analysts have a law degree (P-16). They research and draft Commissioner’s opinions, evaluate applications for temporary classification and requests for approvals of new uses of data, process appeals of challenges to the accuracy and/or completeness of data, and also “consult with the information technology community to ensure that information systems are developed that comply with data practices laws” (IPAD 2012).

4.1.2.5. Information Technology

MN.IT Services is the executive branch agency headed by the State CIO. The statute which created this agency (Minnesota Statutes 2012, Section 16E) declares that the OET

---

33 The nature of classification here refers to accessibility of the information according to Minnesota’s Data Practices Act. According to this act, information is either (i) data about individuals or (ii) data not about individuals. For each of these two categories, there are three possible levels of accessibility: (a) accessible to anyone, (b) accessible to data subjects and to government officials whose duties reasonably require access, and (c) accessible only to government officials whose duties reasonably require access. This creates six possible classifications. For data about individuals the classifications are (a) public, (b) private, and (c) confidential. For data not about individuals, the classifications are (a) public, (b) nonpublic, and (c) protected, non-public.
“provides oversight, leadership, and direction for information and telecommunications technology policy and the management, delivery, accessibility, and security of information and telecommunications technology systems and services in Minnesota” and manages “strategic investments in information and telecommunications technology systems and services.” Since the consolidation of information technology services, all IT personnel in all executive branch agencies now report to the centralized MN.IT Services and are under the ultimate authority of the State CIO.

4.1.3. Legal Environment Affecting Stewards

Several statutes directly affect information management, records management and data practices in the state. A key difference between Minnesota’s information management regulations and those of many other states is that Minnesota has an explicit “Data Practices Act” (Minnesota Statute 13) which regulates access to government information and attempts to balance the right to privacy with the responsibility for transparency in government action. The law states that all government information in the state is considered open to public inspection unless it has been explicitly classified as non-public by law or by temporary classification. The classification “non-public” must be defined by legislative statute or by temporary classification. Although an agency can temporarily classify a piece of information as confidential (in the case of data about persons) or protected, non-public (in the case of data about non-person entities), this temporary classification must be approved by the Commissioner of Administration and by the legislature in order to remain non-public.

34 In particular, these are Minnesota Statute 13, the “Data Practices Act”; Minnesota Statute 138.17, The “Records Management Act”; Minnesota Laws 1971, Chapter 529; Minnesota Statute 15.17, the “Official Records Act;” and Minnesota Statute 325L, the “Uniform Electronic Transactions Act.” Descriptions of these laws can be found in Appendix F.
(Gemberling and Weissman 1982). If this permission is not forthcoming, it automatically becomes public after two years. Thus, all government information is considered public and therefore accessible by the public unless there is a specific law that classifies the information as non-public (or an agency temporarily classifies it as non-public).

Information which is a government record falls under the rules of the Records Management Act. Government records are defined to be state and local records, including all cards, correspondence, discs, maps, memoranda, microfilms, papers, photographs, recordings, reports, tapes, writings, optical disks, and other data, information, or documentary material, regardless of physical form or characteristics, storage media or conditions of use, made or received by an officer or agency of the state and an officer or agency of a county, city, town, school district, municipal subdivision or corporation or other public authority or political entity within the state pursuant to state law or in connection with the transaction of public business by an officer or agency” (Minnesota Statutes 2012, Section 138.17).

The Records Management Act gives the state’s Records Disposition Panel the power to direct the sale or destruction of records deemed not to be of permanent value and to direct the disposition (“by gift”) to the Minnesota Historical Society of records that are deemed to be of permanent value. It also specifies when agencies must submit records to the State Archives, allows the State Archives to inspect records that are listed on a state records disposition schedule, and designates that all records in the State Archives are open to the public unless they are specifically classified as non-public by the archives.

The implicit assumption that information is public unless explicitly categorized as private carries recordkeeping consequences. For example, suppose a government employee produces a report. During the creation of this report, a number of versions may be generated. Suppose further that the final version is complete and has been classified as a public record, but the government employee who has created this report keeps the earlier versions on his or her computer. If a member of the public makes a request to see all versions of this report, the
employee is legally bound to provide all existing versions to that citizen (according to Statute 13). If the final version of the report itself (which is a record) is not classified as non-public (according to Statute 138.17) and is either still within its retention period or has passed its retention period but not been destroyed or transferred to the archives, the employee must share that record with the citizen. In addition, the draft reports are considered government data and thus fall under the ruling of the Data Practices Act. That is, if they are maintained and a member of the public requests to see them, the government employee is legally bound to share them with the requesting citizen. If, however, the employee has destroyed the drafts, then there is no sanction associated with not being able to provide the non-record versions to the citizen since those versions were under no retention requirement.

For email this can create some confusing and potentially risky outcomes. Not all government created emails are government records, but all government created emails are government data. Thus, if they have not been classified as non-public, they are by law open to disclosure to a requesting individual (or to a litigant’s lawyers through eDiscovery) if they have not been destroyed (IPAD 2000). Likewise, if they are records, they must be maintained according to their retention periods and they must be available for public access unless classified non-public.

There are two risks associated with this scenario. If individuals keep everything for an indeterminate period, citizens are free to access both the public records and any other email residing within the government’s ownership. If individuals therefore delete emails en masse, and thereby inadvertently delete records, they have also violated state law and are subject to legal sanction in the event these records are requested or subpoenaed. If either records or data
are requested that are available and public, an employee is bound by law to avoid destroying them until the agency has responded to the request.

Such a scenario makes the question of retention periods, classification, and automatic deletion of emails a significant factor when moving to a cloud-based system. The cloud system must be able to enact the appropriate retention schedules in order to avoid placing the state at risk legally and to reduce potential e-Discovery search costs. This project’s interviews revealed that the records managers who were interviewed largely considered automatic retention and deletion to be a problematic retention-based aspect of the movement to the cloud service. Understanding the process by which the cloud recordkeeping requirements were determined will thus be helpful for understanding the various stewards’ responses to the final product from a recordkeeping point of view.

4.1.4. Requirements and Actions

4.1.4.1. Defining the Recordkeeping Requirements

4.1.4.1.1. Recordkeeping Requirements as Seen by IT

The requirements definition process as a whole involved several layers of activity. As mentioned previously, Minnesota’s office of the CISO conducted an initial security and legal assessment of the ability for a cloud service to meet the requirements of an enterprise EUCC system. Then the Project Implementation Team held a series of meetings with IT employees to investigate what requirements they believed they needed, what things Microsoft indicated they could and could not provide, and what were the “show stoppers,” the things that “if they don’t agree to it, we’re going to walk away from the table” (P-13).
In addition, the requirements analysis team brought together a group of individuals composed of eight to ten subject matter experts (SMEs) from different agencies to ascertain their particular email requirements. Information technology manager P-13 explained,

We worked together over a period of a couple of months to define what the requirements were - standard language stuff, definitions - those kinds of things. We then did an RFP-type of request out to our prospective vendors to say, “Hey, we’ve got this need, how do you guys address these needs? Can you fulfill them?” and similar types of activity. We put a “desired” and “required” on the internal document. We didn’t share that with the … prospective vendors; we just said, “Hey, some things will be required, some things will be desired, and we’ll just go from there.”

When asked how the particular agencies on this team were selected, P-13 replied, “Based on their need,” and then clarified that the selection decision was based on elements such as stringency of retention requirements and the agencies’ risk of a legal action that would require legal holds. For example, the Department of Transportation (DoT) was included on the team due to the I-35W bridge collapse; as a result of that disaster, the Attorney General’s office required the DoT to keep all of its email (P-13). In addition, various agencies have different approaches to retention, perhaps having one retention period for the emails of individuals’ who are at high risk for legal action or for people who are associated with cases under legal hold, with a different retention period for everyone else.

In other cases, messages pertaining to particular topics (such as those associated with greater litigation-risk), are retained for a different period of time. Such requirements exist in addition to (or as a component of) standard agency retention requirements. Therefore, there are very few state-level general retention requirements\(^{35}\) and a great variety of agency-specific requirements. According to P-13, the agencies from which SMEs were selected for participation on the requirements team were those that would have the greatest need for

\(^{35}\)Hyperlinks to the state’s general retention schedules are available online from MHS and can be found at [http://www.mnhs.org/preserve/records/retentionsched.html](http://www.mnhs.org/preserve/records/retentionsched.html).
unusual or particularly risk-mitigating retention requirements. However, other than the DoT, P-13 did not identify the agencies that participated in the requirements assessment or the occupational groups or status of the participants within the organizational structure who acted as SME’s, although one records manager interviewed indicated that they thought that the CIO of their agency was included in the requirements process, but did not know in what capacity (P-23).

Another individual from MN.IT Services (who was not a member of the requirements team) indicated that she believed that in order to ascertain storage needs, IT personnel went to a variety of different agencies to collect retention schedules, both to predict storage needs over time and to accommodate the various retention periods (P-12). When asked whether she thought that any records management requirements need to be considered during a cloud implementation, this individual remarked that she believes that several records management-oriented issues should definitely be considered when managing email in the Cloud. For example, when managing records in-house

… it might be easier to kind of figure out what the original records are because the whole issue around retention – not “issue,” but the whole approach – around retention is you want to keep your original. Determining what those originals are may be a little bit different when it goes to the Cloud. So that’s the other reason why they were … really taking a focus on the storage capacity issue because trying to figure out how to do your de-dup,36 how to do synchronization, and how to, in some cases, pull some of that data back, especially when you’re looking at attachments to emails … So, yeah, I think it’s something definitely to be considered … (P-12).

This individual reported that the group of SMEs from across the state agencies represented the records management viewpoint during requirements definition.

P-13 indicated that many of the state’s requirements were informed by federal mandates and by particular business sector needs.

---

36 Deduplication is a technique for eliminating duplicate copies of repeating data.
… when you start to look at the Department of Revenue and some of those types of things, you look at Health and Human Services, you look at HIPAA and those kinds of pieces, you look at Public Safety and the FBI relationship they have and you’ve got to look at CJIS [Criminal Justice Information System] and some of the other things that are there, so I think it sort of depends upon which sector of business that you’re in, as far as what will drive your retention requirements … look at the Department of Corrections as another good example … they don’t have the best people in their facilities as far as their detainees, if you will. And they get a lot of lawsuits as far as, “Hey, so-and-so did this and that” and therefore they typically have a much shorter retention policy just to help limit themselves in their exposure to some of those kinds of things.

P-13 also re-stressed that various layers of retention requirements exist across agencies. Some agencies insist on keeping everything, some want to keep things with certain keywords, “some will do all people within an agency, some will just do certain people that are under investigation or legal hold or whatever; it’s a hybrid approach” (P-13). P-13 also noted that during requirements gathering, access requirements played a big role for some agencies that needed quick and ready access and whose legal counsel needed access to the data. MN.IT Services had to be able to provide that access but simultaneously needed to segment the data in such a way that the agency personnel could only see their own agency’s information.

4.1.4.1.2. Recordkeeping Requirements as Seen by Records Managers

Of the five Minnesota records managers interviewed all of them share a perception that no records manager or records expert was included in any requirements assessment activities prior to implementation (P-21, P-22, P-23, P-24, P-25). One did note, however, that although the state did not include him during the requirements definition process, after the implementation had begun within their agency he was “brought on board right away” to aid in the implementation (P-23).
The records managers employed during the implementation period were not particularly satisfied with how the requirements definition process went, as illustrated by P-24, who said, “during that process, I got the impression that there wasn't a lot of consideration of that [i.e., records management impacts of moving email into the Cloud] . . . I'm not sure records management was really thought about.” Another remarked, “When we deal with our information technology office … they exclude us in those conversations and then it becomes a real challenge for us to run behind them and catch up and force the issue about protection of information, the requirements of information – and it's just frustrating” (P-21). Yet another commented,

It’s like we talk amongst ourselves about it but, it was hard to – even when people were brought in or showed us Office 365 – it’s like the decision was already made. We weren’t consulted … and it’s basically just that we have to deal with it, work around it … IT does its own thing and records management is never involved. They are not part of that process. So there's a huge gap there between us and … to also identify the key contacts or the key people that we need to be in alliance with … so that we know what the criteria are, what their requirements are, so that when the systems are developed … the recordkeeping requirements are embedded so that things can be managed a lot easier, rather than labor-intensive, manual processes … We don't work together, that's the problem (P-22).

4.1.4.1.3. Recordkeeping Requirements as Seen by Executive Management

One of the MN.IT Services executive interviewees offered advice to other states considering cloud computing implementations and this advice may offer some clues about the reasons for the approach taken during requirements analysis. P-19 advised that potential implementers need to

… go to the source of what needs to be complied with, not necessarily take folklore understanding of how things need to be done, but rather, challenge the exact limitations and/or requirements that need to be complied with. We found that a lot. Especially in state government there is that, "Well, that can't be done via a third-party provider," and then you ask, “Why?” “Because we need to maintain a record.” “Okay, why?” “Well, because it says here.” “Well, if we maintain them, if they’re our
records, the fact that they're housed somewhere else … doesn't mean that they're not ours.” “Well, it's less ours then.” “Why?” – Tying it back to real requirements I think is really, really important … So that would be a huge recommendation to anyone embarking on it. Don't take anybody's interpretation of what should be. Go back and show me the regulation because otherwise you're going to get caught up into interpretation of what they think should and should not be the overall delivery stand (P-19).

This executive thus distinguishes between the recordkeeping requirements and the “delivery stand,” that is, the way in which the requirements are implemented and supported. The delivery stand represents how requirements are met as opposed to specifying which requirements need to be met. What specific tasks are enabled and how key (abstract) requirements such as “maintaining ownership” are provided through concrete system capabilities comprise the delivery stand. Thus, arguments that cloud computing will threaten the capacity to retain ownership over records was not accepted by this executive because according to P-19 because “the state still owns the information.”

4.1.4.1.4. Recordkeeping Requirements as Seen by Archivists

The archivists interviewed had little or no knowledge of the Microsoft implementation, noting that because they are not a state agency, they are not required to participate in MN.IT Services project implementations. They emphasized their separate status and indicated that they had very little (if any) knowledge of the specifics of either the implementation itself or of state employee responses to it. They stated, “We’re a quasi-state agency, so our IT functions have not been sucked up into the centralized unit” (P-8). They stressed that they consider themselves to be a deeply concerned stakeholder in agency recordkeeping activities but that their influence over such activities is necessarily non-authoritative and consultative.
Nonetheless, they did seem aware of potential strains in relationships between agency personnel and MN.IT Services personnel and spoke of widespread perceptions that MN.IT Services “often likes to set initiatives for the rest of the organization instead of the other way around,” offering the hypothesis that this occurs because the possession of the data on MN.IT Services servers leads to a sense of responsibility or ownership for the stewardship of that information:

You know, this was never an issue before electronic records . . . but now our information resides on their [MN.IT Services] servers, in their systems, and so they see themselves as the stewards of this, and responsible for its management, whereas, you know, we should be also having a say in that because we are the content owners as it were. We're responsible for that content as well. There's a constant tension there between whose stuff is it and who gets to manage it (P-18).

The perception that implementations of new technology have reduced the power of non-IT agencies is not limited to non-IT personnel. Speaking about the cloud implementation itself, MN.IT Services employee P-13 said, “Initially, I think that there was some heartburn around it. Um, you know, I can certainly see it’s a little bit of freedom that’s been given up by them.”

4.1.4.1.5. Recordkeeping Requirements as Seen by IPAD

An individual from the Information Policy Analysis Division (IPAD) indicated that he believed that one of his colleagues in the division had been consulted “on issues more relating to security, making sure that … the contract for the implementation would be in compliance with the Data Practices Act and probably, peripherally, as that relates to records retention obligations” (P-16). The colleague P-16 referenced, however, has since left the employ of the state and could not be reached for comment.
4.1.4.2. Acting Upon Requirements
There was some disagreement among the various interviewees as to whether the apparent lack of records management input has created records management problems. Most of the records managers’ concerns regarding the cloud implementation appear to revolve around the potential risk to records management practices rather than to actual process or procedural impacts that have occurred. One records manager, in fact, said that he has felt no impact (P-22). Another spoke of her agencies’ implementation of a 90-day automatic purge of emails during the cloud implementation, saying it was “fine for my records management purposes, because I just tell people that if you need to save an email longer, you just need to save it” (P-25). However, some did express concerns related to the automatic purge. P-22 asked, “How do we know that that’s [the purged records] been disposed of properly?” This uneasiness echoes a common concern discussed in published literature regarding records management in the Cloud. The Article 29 Data Protection Working Party in the European Union stresses the need for organizations to know exactly how cloud providers will destroy data, whether via demagnetizing storage media or multiple overwriting of the data. If the latter strategy is used, an organization should know what tool is being used to destroy the information (Article 29 Data Protection Working Party). The CIO Council also stresses this point, and describes two measures to help ensure that appropriate deletion occurs: (i) including a records manager in the requirements definition process to ensure that a disposition date for categories of records within the system is set and that the system has the ability “to automatically execute itself or send a file owner a notice when it is time to delete certain records” (32) and (ii) “including an entire records management component as a part of a cloud computing contract” or allowing electronic records management bidders to include an integrated email management solution that supports the maintenance of records’
“functionality and integrity throughout the records’ full lifecycle” (32). Although it is not clear that (i) was followed during requirements definition, Minnesota’s written archiving requirements suggest that the latter approach was, in fact, attempted. Among the requirements specified within the state’s formal requirements documents the following archival requirements are relevant to this need:

2. Mail that is deleted from the user's mailbox could still [be] accessible from the archive based on the retention rules…
3. The solution is capable of managing retention rules that span from 1 day to indefinitely.
5. Within the solution, archived mail and journaled mail includes all metadata associated with the content, including user specified categories.
6. The solution provides the ability to create rules/policies that import content from journaled mail to archived mail. Triggers for this process can be based on content age, content size and/or percent of mailbox quota used. Rules/policies should allow for content to be stubbed or removed from user's view (but is still available in the archive for the duration specified by the retention rule).
7. Within the solution, archived mail and journaled mail cannot be altered or allow for spoliation.
10. The solution provides the ability to apply retention rules to specific groups of people (e.g. an agency).
11. The solution allows for tiered administration of retention rules for specific groups of people (e.g. an agency), while preventing access from other groups (e.g. a different agency).
14. The ability for administrators to modify which retention rule applies to a specific piece of content, if it is assigned an incorrect retention rule.
15. The solution provides an automated method to remove all archived mail from a mailbox and the archive based on agency defined retention rules, unless an administrative override is defined (e.g. legal hold).
18. The solution allows users with the proper permissions to access archived mail for resource and shared mailboxes. Additionally, these users can manually select content to be imported into the archive…
(MN OET 2009, 3-4)

The complete Archiving Requirements Document can be found in Appendix C.

Within this document a variety of requirements associated with authenticity, integrity, and usability are highlighted. MN.IT Services has set up access requirements based on passwords, supporting the attempt to satisfy authenticity requirements. Reliability cannot really be satisfied automatically by any software alone, and must rely upon appropriate policies, procedures, and business processes, except insofar as the system automatically assigns the email date and password-based access, thereby ensuring that workers cannot falsify that information or easily purport to be someone other than themselves. However, the requirements do not specifically address the means used to prove that destruction of obsolete records occur, suggesting that some of the requirements recommended by experts such as the CIO Council have not been met in this implementation.

Records managers’ primarily expressed their concerns about requirements, however, as general apprehension about not clearly understanding new processes and fears of not knowing exactly what is happening on the back-end of the system, suggesting that the concerns are change management, training, and interoperability issues, alongside a lack of full trust of IT, rather than cloud-specific issues.

When asked about the archiving capabilities of the Microsoft 365 emailing service, OET employee P-13 indicated that although Microsoft 365 offers email preservation and

37 Here I do not use the term interoperability in the narrow sense of ability of the technical hardware and software to allow linking of machines. Rather, I use Landesbergen and Wolken’s broader idea of interoperability as information sharing, a concept that includes the narrower technical notion, but also includes the idea of people talking and sharing information (Landesbergen and Wolken 1998).
search capabilities, the service does not meet the state’s business requirements so the state still engages in long-term email preservation on-premises. They do this by using Microsoft Exchange’s journaling function, which is a feature within Exchange itself. They use the journaling function to capture records that need to be preserved, and then transfer the records to the state’s long-term preservation servers. The journaling function allows one to specify individuals and groups (to which individuals are assigned via business rules) for which all email will be captured when it is routed through the Microsoft Hub Transport. When an email message comes in or goes out through the Hub Transport, the specified rules are checked and if the individual or group is set up for journaling, a copy of the message and its associated metadata (e.g., to:, cc:, and bcc:) and attachments are captured and sent to the journaling server (Technet.microsoft.com 2010). Rules can be set up to specify that different groups can be sent to different journal archives, making later search and retrieval easier. When the copy is made at the Hub Transport, the email is then sent to its recipient. The journaling function allows retention policies and legal holds to be specified (Microsoft 2011). Although it is possible to apply the policies to folders, conversations or individual messages, P-13 did not specify the details of how the retention policies were set up, other than to note that every agency had its own policies and retention periods, which were applied to individuals, groups, whole agencies, and also to certain keywords. The various groups’ email copies are sent to different mailboxes that are each associated with the particular groups that have been set up, and for which different retention periods hold. At the end of retention periods, messages and associated metadata are automatically destroyed.38

---

38 Again, P-13 did not specify the method of “destruction,” so it is not clear whether the information is truly “destroyed” or just deleted.
4.1.4.3. Vendor Selection

After the requirements were defined and the RFP had been publicized and vendors examined, MN.IT Services invited Microsoft onsite for “two to three weeks” (P-13) to go through the service capabilities. In preparation for the visit, requirements participants drafted a list of five to ten “show stoppers,” things that would lead the state to refuse to contract with Microsoft. The show stopper that proved to be relevant to the service adoption was that at that time Microsoft did not require its employees to go through FBI security checks. This did not comply with the stringent federal (i.e., FBI) and state requirements associated with CJIS usage. Upon the insistence of MN.IT Services, Microsoft contractually agreed to abide by this requirement.

4.1.5. Stewards’ Perceptions

4.1.5.1. Perceptions of “Records”

The archivists’ and records managers’ perceptions of “record” were very similar. P-21 and P-24 both referred to a record as a piece of information that provides evidence of business transactions, while P-23 added to this definition the idea that records also contain information about the reasons behind those activities. P-25 discussed records as being a by-product of an activity, that is, any information that is gathered during the course of government business as part of a transaction. The key concepts pertaining to the various similar definitions and descriptions given by the records managers are:

- Record as evidence of a transaction;
- Information that is gathered during the course of business;
- Information that is a by-product of activity;
• Information that shows the beginning, middle, and end of a process of decisions such that full evidence of those decisions is provided; and,

• Information that reveals the reasons behind business activities.

Both P-8 and P-16 specifically remarked that the definition of record used by employees of the state is derived directly from the notion of “public record” found in Minnesota law, which defines public records to be:

state and local records, including all cards, correspondence, discs, maps, memoranda, microfilms, papers, photographs, recordings, reports, tapes, writings, optical disks, and other data, information, or documentary material, regardless of physical form or characteristics, storage media or conditions of use, made or received by an officer or agency of the state and an officer or agency of a county, city, town, school district, municipal subdivision or corporation or other public authority or political entity within the state pursuant to state law or in connection with the transaction of public business by an officer or agency” (Minnesota Statutes 2012, Section 138.17).

The IPAD respondent (P-16) also defined a record to be any information or data, regardless of form, that is created, used, or received by a state entity in connection with the transaction of public business which provides a full and accurate knowledge of official activities. P-16 also added, however, that to be considered an “official” record this information must be subject to retention rules. That is, if the information is classified by MHS and required by state law to be retained and scheduled, then that information is a record or records, regardless of its format. Otherwise it is information or data, but not a record. This definition is consistent with the formal government definition which excludes data and information that does not become part of an official transaction, library and museum material made or acquired and kept solely for reference or exhibit purposes, extra copies of documents kept only for convenience of reference and stock of publications and processed documents, and bonds, coupons, or other obligations or

39 The term “public” as used here should not be confused with the previously mentioned concept of classifying information as public according to the Data Practices Act. Here “public” simply refers to the fact that the information is created and held by a public entity.
evidences of indebtedness, the destruction or other disposition of which is governed by other laws” (Minnesota Statutes 2012, Section 138.17).

These definitions show similarities to definitions and descriptions of records found in the ARM literature. For example, the International Council on Archives (ICA) defines a record to be “recorded information produced or received in the initiation, conduct or completion of an institutional or individual activity and that comprises content, context and structure sufficient to provide evidence of the activity regardless of the form or medium” (ICA, 7).

Bantin (2008) argued that treating records as a consequence or product of an event highlights the activity of defining “more precisely and conceptually when the record is created by the business event or personal activity,” thereby placing “greater emphasis on understanding functions and processes and on precisely linking the records to the events that created them” (27). Interviewee P-21 described the contextual nature of such links by describing a record as “a story that shows the beginning, middle, and end of a process of decisions, providing evidence of those decisions.” Furthermore, the idea that a record is a “story” that provides evidence of an entire process brings to mind the ICA’s and David Bearman’s (1994) insistence that the evidence provided within information must comprise content, context, and structure in order to be a record. Bearman stressed this in 1994 when he argued that ARM professionals identify generic forms of documentation that are associated with the various functions of the organization using “the relationship between these functions and forms to ‘schedule’ records,” that is, to “determine how long the information in each needs to be kept” (1994, 15).
All records should be subject to retention schedules that indicate the disposition of the records after their primary use value has passed, according to ARM theory (Fischer 2006). This is one area in which the ARM theoretical definition diverges somewhat from the state public records definition, however. In ARM literature, if information provides evidence of a transaction and comprises content, context, and structure it is considered a record, regardless of whether or not it is subject to a legally mandated retention schedule (although in theory, *all records should* be subject to retention schedules according to that theory). That is, from an archival point of view it is not the act of placing a retention rule on a piece of information that makes it a record. Rather, being a record places the responsibility of assigning it to a retention schedule.

Part of this divergence in definition comes about because, for state government, records retention concerns are focused on a subset of all archivally-defined records (i.e., public records) and these are defined primarily for legal compliance requirements. For example, according to the archivally-conceived notion of a record, even those records which, by law, are subject to obligations or evidence of indebtedness would be considered to be records if they provide evidence of a transaction. By statute, however, such pieces of evidence are not considered “public records.” Thus, the only records required to meet retention rules are those records that meet the criteria provided in the definition of public record given by Minnesota Statutes 2012, Section 138.17. The ARM theoretical definition of a record, on the other hand, includes *all* information created in the transaction of business for the purpose of providing evidence of those transactions, and includes content, context, and structure. In other words, the ARM theoretical definition of record is broader than the

---

40 Of course, in practice no organization maintains perfect retention and can be considered successful if it attains 80 to 90 percent retention, according to Stephens (2007).
definition of public record used within state government\textsuperscript{41}. This may seem to be a small
distinction, and in most situations it is, but it may help to highlight one reason why the public
archives tradition and the historical manuscript tradition remain at odds to this day. The
historical manuscript tradition, focused on history, would seek to retain all records that
provide evidence of government transactions. The public archives tradition, focused on
efficiency, would seek to retain all records that meet the governmentally defined notion of a
“public record.” The former tradition would preserve for posterity a much greater volume of
information because it more broadly defines the “continuing value” of records and is more
broadly inclusive of what types of information are record- versus non-record material.

The distinction between these two different approaches to treating information as
record may represent a conceptual gap that is full of political tension. Typically, for example,
a governor’s private records are considered as part of the public record, but many other
public executive managers’ correspondences are not. In fact, because the decision of which
record is subject to retention is defined jointly by the state archives and by legal compliance
requirements, political processes and tensions within any given state can influence the
selection of information as record or non-record material (P-8). In fact, in Minnesota the
State Records Disposition Panel is composed of two other members besides the state
archivist, increasing the possibility of varying interpretations of what information sources are
records and what retention requirements should hold for them.

To be sure, many pieces of information not subject to state archival permanent
retention are nonetheless retained permanently within the state agencies themselves, and are
thus “archived” by records managers and IT personnel maintaining electronic records within

\textsuperscript{41} And Minnesota’s definition of a public record is not significantly different from that of most other states.
their systems or the Cloud (P-8; P-9; P-21). In other words, whereas the traditional notion of records manager implies that they will manage short-term and long-term retention, but not permanent retention, the reality in Minnesota is quite different. Many records are maintained “permanently” not within the archives, but within the agencies themselves (P-8; P-9; P-21).

IT employees did not define the term “record” in their interviews but rather, pointed to the existence of a records law that defines “public record.” Furthermore, the IT employees, including the IT executive manager, used the term “record” differently than the records managers and the IPAD interviewee did, both semantically and in terms of frequency of use in discourse. For example, Table 5 shows the frequency of use of the term itself in the overall responses given by the different personnel types throughout the course of their interviews.

### Table 5 - In-Interview Frequency of Conversational Use of the Stub “Record,” by Occupation

<table>
<thead>
<tr>
<th>Interviewee's Occupational Affiliation</th>
<th>Interviewee Frequency of Use of Terms Containing the Stub “Record”</th>
<th>Interviewer Frequency of Use of Terms Containing the Stub “Record”</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>IT</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>IT</td>
<td>11</td>
<td>20</td>
</tr>
<tr>
<td>Records Management</td>
<td>24</td>
<td>30</td>
</tr>
<tr>
<td>IPAD</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>Records Management</td>
<td>41</td>
<td>34</td>
</tr>
<tr>
<td>Archives</td>
<td>51</td>
<td>27</td>
</tr>
<tr>
<td>Records Management</td>
<td>59</td>
<td>36</td>
</tr>
<tr>
<td>Records Management</td>
<td>85</td>
<td>35</td>
</tr>
<tr>
<td>Records Management</td>
<td>105</td>
<td>37</td>
</tr>
</tbody>
</table>

42 The term “archives” is placed within quotation marks here to highlight the loose use of the terms. If a record is “archived” by a state agency, it often does not meet many of the strict preservation technical requirements as those records maintained permanently by the state archives itself.
What this table shows is that in spite of being asked very similar questions on the same topical areas, the MN.IT Services respondents used terms containing the stub “record-” about three times per interview, on average, whereas the records managers and archivists used terms containing “record” about 62 times per interview on average. Of course, the use of terms containing “record” on the part of the interviewee’s would be subject to the interviewer using this term. However, the basic set of questions asked of the different occupational groups were the same, although follow-up clarification depended upon the responses given by the interviewees. In other words, the interviewer asked the same base questions to each occupational type. If one occupational type responded using the term “record,” a follow-up question by the interviewer would be more likely to include that term as well. Thus, if the interviewee responds with the use of a word containing the term “record,” the interviewer would be more likely to follow-up that response with a sub-question that also used the term “record.” Table 5 illustrates that overall, the interviewer used the term “record” more frequently in the interviews with the IPAD representative, the archivists, and the records managers than with the MN.IT Services personnel. However, in all cases but one, these three occupational groups used the term “record” more frequently than the interviewer and in two out of three interviews the interviewer interjected the term “record” but still had fewer responses that used this term. In the third MN.IT Services interview (that is, the one in which the MN.IT Services person used the term “record” more frequently than the interviewer), two things were revealed through detailed examination of the contextual uses of the term “record”:

- This individual, although working with MN.IT Services, corresponds regularly with the State Archives by co-participating in a data-oriented workgroup and by taking
part in regular conversations about records-related impacts of cloud computing with members of the State Archives. The other two MN.IT Services personnel do not appear to have as frequent a contact with the State Archives.

- Of the eleven times this individual used terms containing “record,” eight of these uses occurred in a single reply to the (follow-up) question, “do you think that the issue of security of the information is an IT issue or a records management issue? Or do you think it’s both (or neither)?”

- The three uses which occurred spontaneously (that is, not in response to the use of the term by the interviewer) on the part of the interviewee do not appear to refer to “record” in the sense of either public record or the archival definition of record (as evidence of a transaction that has context, content, and structure) but rather, map more closely to a technical definition of a record as “a database entry that may contain one or more values” (http://www.techterms.com/definition/record).

An interviewee from the State Archives discussed the evolving nature of the meaning of the term “record,” noting that this term has been problematic “over the course of maybe the last ten years as the technology has evolved” (P-8).

...each time there’s new technology or greater adoption of technology there are more and more questions that come up. For instance, with voicemail, instant messages, um, I had somebody call me the other day asking about internet browser histories. That sort of thing, so I think that invariably as people start to incorporate these technologies more into their routine workflows they start thinking to themselves, “Wow, wait a minute. Do I need to be managing this information somehow? And then we have to look and see how that fits into the retention scheduling process (P-8).

---

43 Specifically: “…you can just go in there and you can add these records but we don’t talk about changes and deletes and some of those changes and what is the SLA around that, and are there going to be integrity reports to . . . identify, how many users or how many people or records don’t have contact information . . . and what are you going to tie that to?” (P-12) Also, “…with this on-premise thing you can go through the back door or you can do some update statements and you can fix those records or update the data” (P-12).
4.1.5.2. Perceptions of “Cloud Computing”

Stakeholder definitions of the phrase “cloud computing” were more consistent across occupational boundaries than were descriptions of the term “record.” The general theme that held for all interviewees except one was “outside my realm of support” or “outsourcing.” In their interviews, the IT personnel also displayed the common theme of “not having to worry about that (e.g., server or infrastructure management) anymore” (P-12; P-13; P-19), as did one of the archivists (P-9). The IPAD interviewee did not provide a definition of cloud computing.

Two of the five records managers did not provide a definition, but of the three who did, two described “cloud computing” as associated with being “outside” one’s local area and with access that is not geographically constrained: “it’s a spot you can go to that’s outside your local area to compute” (P-25); “. . . it doesn’t matter if the server is in the building or if it’s in Chicago or wherever. You can access it, but your staff doesn’t have to support it” (P-24). The third records manager appeared to conceive of cloud computing more as a system which gives the appearance of infinite resources – “an infinite hole where information can live” (P-21).

All responses to “how would you define or describe cloud computing” point to the “outside” nature of the information and its management, to the idea that internal employees no longer support the hardware and software or the information itself while at the same time, internal people can access it anywhere, anytime. These descriptions all imply the insertion of a third (external) party into what was previously considered an internal process or an internally managed relationship (between non-IT and IT employees). Given that their cloud computing vendor is a large scale provider of cloud services and also provides large scale consulting services, the similarity in definitions across all occupational types and
organizational hierarchies suggests either that the external vendor played a strong role in educating personnel about the nature of cloud computing or that overall change management provided a strong message regarding the “outside” nature of cloud computing to employees, or both. It certainly would be worth a cloud vendor’s time and effort to ensure that when an organization thinks “cloud” they simultaneously think it must be externally offered.

4.1.5.3. Perceptions of Roles and Responsibilities

Records managers described their roles and responsibilities in rather similar ways. Table 6 shows the duties described by these personnel, indicating with a check mark whether the records manager listed in that column mentioned that duty within his or her interview.

All of the records managers pointed to retention schedules as their primary responsibility, although one did stress that he was disturbed that everyone saw records managers as primarily responsible for retention schedules only. (When asked his primary responsibility, this person also indicated that retention schedules are primary, but are only one among several activities that are associated with ensuring compliance and that records managers should be seen as supporting compliance, not just as creating retention schedules.) Four of the five also highlighted taxonomy creation or classification as a key responsibility, focusing on the need for such activities in order to enable the appropriate disposition of records over time. Three out of five felt that a crucial component of their jobs was that of education. They felt that their responsibility is to educate business users and IT regarding the definition of records, the appropriate management of records, and the need for final disposition actions to occur accurately and at the correct time. Three of the five also cited
Table 6 - Self-Described Responsibilities of MN Records Managers

<table>
<thead>
<tr>
<th>Duties</th>
<th>Records Managers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Develop and/or assist business units develop and update retention</td>
<td>✅</td>
</tr>
<tr>
<td>schedules</td>
<td></td>
</tr>
<tr>
<td>Engage in taxonomy development and/or classification</td>
<td>✅</td>
</tr>
<tr>
<td>Provide education on records management and information management,</td>
<td>✅</td>
</tr>
<tr>
<td>(including training people what a record is and how to use it)</td>
<td></td>
</tr>
<tr>
<td>Monitor, collaborate with MHS, and/or train users about preservation</td>
<td>✅</td>
</tr>
<tr>
<td>Manage off-site and/or hard copy storage</td>
<td>✅</td>
</tr>
<tr>
<td>Monitor retention holds on departing employees and/or clean up their</td>
<td>✅</td>
</tr>
<tr>
<td>mailboxes when necessary</td>
<td></td>
</tr>
<tr>
<td>Develop a plan for upgrading the records management program</td>
<td></td>
</tr>
<tr>
<td>Assist with compliance education</td>
<td>✅</td>
</tr>
<tr>
<td>Inventory records</td>
<td></td>
</tr>
<tr>
<td>Monitor compliance with retention schedules (including destruction)</td>
<td>✅</td>
</tr>
<tr>
<td>Teach people how to use an EDMS system</td>
<td></td>
</tr>
<tr>
<td>Manage other records managers</td>
<td></td>
</tr>
<tr>
<td>Assess software tools</td>
<td>✅</td>
</tr>
<tr>
<td>Coordinating records clean-up</td>
<td></td>
</tr>
<tr>
<td>Act as agency SharePoint site coordinator</td>
<td></td>
</tr>
<tr>
<td>Educate people on disclosure</td>
<td></td>
</tr>
<tr>
<td>Sit on committees such as the Data Domain Workgroup</td>
<td></td>
</tr>
</tbody>
</table>

some type of preservation activity as key to their position, although none clarified exactly which aspect of preservation they felt was within their occupational purview. Several cited other activities such as updating their current records management programs, teaching business units about software or making software assessments, helping with records clean-up, and dealing with compliance issues such as disposition or disclosure rules. All of the activities fit within one of six types of activities, however:

- Retention scheduling;
- Educating business units and/or IT;
- Developing taxonomies and/or classification schemes;
• Assessment of IT tools;
• Helping with or training people about some preservation activity or activities; and
• Management of (some records’) storage.

The IPAD respondent P-16 described his role as advisory in nature. Specifically, P-16 indicated that the IPAD personnel issue advisory opinions on all issues related to data practices and the Open Meeting Law and advise people regarding data access rights, data privacy concerns and the Open Meeting Law. P-16 stressed the purely advisory nature of his position and limited his activities to those that are associated directly with data (in contrast to records) practices and open meeting law issues.

Because the IT respondents reside in a variety of hierarchical levels and positions within MN.IT Services, the range of responsibilities cited by them varies quite a bit as well. In addition, based on their descriptions of roles and responsibilities, it appears that employees within MN.IT Services vary their roles and responsibilities over time, switching from certain types of activities to others as the need for services changes within the organization. Consistent with that assessment, the IT personnel expressed their job positions in ways that suggest more mobility across roles and responsibilities than was suggested by the non-IT respondents. Since some of the IT personnel explained that the nature of their IT role changed during or immediately after the cloud implementation, I describe here all the responsibilities reported as having occurred both during and after the cloud computing implementation occurred. Such responsibilities include:

• manages a particular portfolio of IT services;  

---

44 That portfolio is not listed here in order to protect anonymity.
• engages in issues related to product management, financial management, technology management, and preparation of budgets;
• works on product features and functionality to ensure IT meets its goals;
• engages in requirements definition for new systems;
• works on change requests and service template changes;
• mediates between business technology users and Microsoft Corporation for Microsoft 365-related issues or requests;
• supports IT services for all state executive branch employees;
• manages IT application upgrades, where necessary and relevant to product line;
• engages in implementation activities with business users and vendors;
• monitors bugs and bug fixes;
• handles data-related services in the Enterprise Architecture group, for example, data delivery, standardization, design, reporting, business intelligence;
• provides policy recommendations related to data use and management;
• ensures compliance with data practice laws;
• examines retention-related issues in the Cloud;
• sits on Data Domain Workgroup;
• helps to resolve cloud-related questions and issues post-implementation;
• engages in strategic decision making for MN.IT Services;
• communicates with the media regarding MN.IT Services activities; and
• participates in high-level implementation decision making;

The respondents from the Minnesota State Historical Society/State Archives described their activities as consultative in nature. They consult with the agencies as records experts,
providing the agencies information on records management best practices. In addition, they work to increase awareness and knowledge of the State Archives and of the need for diligent records management and retention compliance. In addition, specific duties of individual respondents are:

- manages the archival duties within the State Archives;
- participates in review of retention schedules for the state agencies;
- manages the access and preservation decisions for the State Archives;
- participates in the Data Domain Workgroup;
- helps to make budgetary decisions (with others) for the State Archives;
- identifies materials of historic significance in the collections, making selection and acquisition decisions;
- enhances catalog records;
- helps build and maintain project web site;
- processes collections; and
- works with the State Historical Records Board to encourage and evaluate proposals originating in the state and to recommend action to NHPRC.

4.1.6. Concerns and Perceptions of Risk

When they initially considered cloud computing, MN.IT Services executive leadership examined whether or not service providers would be able to meet the state’s security and legal requirements. They specifically considered their recordkeeping and data practices mandates and requirements and the legislation surrounding these mandates, according to P-19. She referred to the potential “project killers” as “security and legal.” After
conducting a three month Chief Information Security Officer (CISO) study, they determined it could meet their requirements, given that they chose a private cloud environment (i.e., the Microsoft service that was then referred to as BPOS-D, where “D” stood for “dedicated”) and that Microsoft would be willing to engage in specified security practices. In particular, Minnesota required that Microsoft employees undergo a U.S. Federal Bureau of Investigations (FBI) security check of the same nature that any state employee would have to have before gaining access to the data.

After the security and legal evaluations were complete, OET determined that moving to the Microsoft service would, in fact, meet the legal and security requirements and furthermore, would improve their risk posture. The Enterprise Unified Communications and Collaboration (EUCC) strategy enacted via the cloud contract complies with the Statement of Auditing Standard (SAS) 70 Type II requirements, ISO 27001, Sarbanes-Oxley (SOX).

---

45 It is now referred to as Microsoft 365-G, where “G” stands for the government version of the service, but still implies that one’s data resides on servers dedicated to your contracted environment. Although the infrastructure and services are controlled by Microsoft, they remain entirely separate from other customers. The only entities allowed on Minnesota’s service are governmental entities in the state of Minnesota.

46 A Statement on Auditing Standards (SAS) No. 70, Service Organizations, or SAS 70 audit, is an auditing standard developed by the American Institute of Certified Public Accountants (AICPA) (AICPA 2013a). It provides a uniform way by which service organizations can report their internal control activities and processes to their customers. It also allows an independent auditor to produce an opinion on the internal controls the audited organization exhibits (AICPA). Of the two types of SAS 70 report (i.e., Type I and Type II), the Type II report is the more stringent. It requires both the service organization’s description of controls and also detailed testing of those controls over a period of at least six months. It includes (at a minimum) the independent auditor’s opinion, the service organization’s description of controls, the auditor’s description of the tests of operating effectiveness and the tests’ results, as well as other information the service organization may choose to report (AICPA 2013b).


48 The Sarbanes-Oxley Act of 2002 created the Public Company Accounting Oversight Board (PCAOB) to oversee the activities performed by auditing companies. It also mandated a number of reforms to enhance financial disclosures and to combat financial fraud within companies.
and Health Insurance Portability and Accountability Act (HIPAA)\textsuperscript{49} (MN OET 2011). In addition, the data centers have badge and smart card restricted access, biometric scanners, on-premise security officers, and continuous video surveillance (MN OET).

Employee concerns about the implementation of cloud computing revealed both concerns about perceived risks of cloud computing and general concerns associated with the implementation but not necessarily specific to cloud computing itself. Table 7 shows the concerns and risks that have been expressed, according to whether the individual sharing the risk is an ARM worker or a non-ARM worker. Within that table, “professional affiliation” represents the particular occupation which an interviewee has expressed as their primary affiliation in the organization. The table also shows which characteristic essential to good recordkeeping is at risk, according to the recordkeeping standards defined by ISO 15489-1.

4.1.7. Interactions between Recordkeeping Stewards

4.1.7.1. Perceptions of Working (Together) in the Cloud

The State Archives is an entity separate from the rest of the state physically, organizationally and fiscally. (See Figure 2 for a depiction of recordkeeping stewards’ organizational location within Minnesota State Government.) The Archives can advise, and can have some influence through the State Records Disposition Panel. However, it has no authority over records creators, records managers, or IT personnel within the state agency structure (P-8; P-9). Records managers and IPAD interviewees agreed that they have little direct contact with archives personnel, noting that their primary contacts are “IT, legal, and just business staff” (P-24), and that contact distribution is relatively equally spread between those three groups (P-24). P-21 noted that much of the paper-based work the records

\textsuperscript{49} See page 194 of this document for more details on this act.
Table 7 - Employee Concerns Regarding the New Cloud Computing Service

<table>
<thead>
<tr>
<th>Stated Concerns, ISO 15489-1 Risk Area and Operational Risk Area</th>
<th>Professional Affiliation of Interviewee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concern</td>
<td>Operational Risk Area</td>
</tr>
<tr>
<td>More layers of activity or intermediaries increases the risk of data corruption</td>
<td>Authenticity, integrity</td>
</tr>
<tr>
<td>With multiple cloud vendors and data sharing, how interoperable will the data be across vendors, in order to ensure easy and quick access to citizens or lawyers?</td>
<td>Access</td>
</tr>
<tr>
<td>Getting forensically sound copies of email accounts is a more cumbersome process and riskier with third (and fourth) party intermediaries</td>
<td>Authenticity, integrity, access</td>
</tr>
<tr>
<td>SLAs that ensure up-time commitments ensure only that systems are up and available, but don’t ensure that the data is available, which requires a separate OLA-level agreement. Has such an agreement been enacted?</td>
<td>Access</td>
</tr>
<tr>
<td>Access rights and authority over processes are no longer clear</td>
<td>Incentive to curate, information ownership</td>
</tr>
<tr>
<td>The risk of data leaks increases with third-party handling of data</td>
<td>Security and privacy</td>
</tr>
<tr>
<td>It is difficult or impossible to ascertain whether recordkeeping requirements are being met: “There is no clarity around what the backend handling is: How is data accessed and stored? How can it be audited? How does one measure performance on recordkeeping requirements?</td>
<td>Provenance, access, performance monitoring</td>
</tr>
<tr>
<td>Concern</td>
<td>Operational Risk Area</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
</tr>
<tr>
<td>If a security breach occurs, who will be held liable? There is no legal precedent on this yet, so no apparent repercussions of breaches are present, increasing the risk of breach in absence of incentives for cloud providers to shore up their security practices.</td>
<td>Security, privacy, accountability, control over legal outcomes</td>
</tr>
<tr>
<td>Lack of clarity around what the processes are for eDiscovery now</td>
<td>Accountability, process ownership</td>
</tr>
<tr>
<td>Data synchronization is more complicated because it requires going through a third party</td>
<td>Process controls, provenance</td>
</tr>
<tr>
<td>Excessive storage availability creates a “save everything” mentality, with two negative results: ~ increased legal risk for data that is now saved that previously would have been disposed of in a timelier manner. ~ more difficulty finding relevant information in litigation cases, due to increasingly huge quantities of unclassified data</td>
<td>Search and retrieval, accountability, information volume concerns, litigation risk</td>
</tr>
<tr>
<td>Increased storage could lead to increased information available for discovery, and increased chance of litigation</td>
<td>Accountability, litigation risk, information volume concerns</td>
</tr>
<tr>
<td>Concern</td>
<td>Operational Risk Area</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Increases in work load due to automated deletion when employees leave the organization: if they didn’t properly classify it before departure, someone needs to go and classify the information within the 28-day grace period</td>
<td>Accountability, records disposition, litigation risk, employee performance due to work overload</td>
</tr>
<tr>
<td>Concerns that data may not be properly disposed of according to retention schedules and lack of knowledge about how this disposal occurs in the Cloud</td>
<td>Authenticity; reliability</td>
</tr>
<tr>
<td>Lack of knowledge about how de-duplication of data occurs</td>
<td>Incentive to curate, data and process ownership</td>
</tr>
<tr>
<td>Data ownership has become less clear</td>
<td>Authenticity, reliability, litigation risk, data and process ownership</td>
</tr>
<tr>
<td>How do you find or create a trustworthy relationship with a cloud provider so that the special requirements that government records carry are met?</td>
<td>Vendor relations, privacy and security</td>
</tr>
<tr>
<td>Are any state entities actually using cloud services for long-term preservation and if so, how are they dealing with security and data practices issues?</td>
<td>Security, access, process controls</td>
</tr>
</tbody>
</table>
Figure 2 – Organizational Location of Stewards in Minnesota’s Microsoft Cloud
managers perform is relatively solitary but that handling electronic records requires much more collaborative work, which can include collaboration between the records manager, a systems developer, a systems administrator, and the legal department. They also work with IPAD on issues specifically related to data practices. P-21 pointed out that the Archives “give us some guidance and give us some direction and give us some training if we need it, and oversee some parts of it, but they don't want to own it. And they can’t; they don't have the resources…even though the Historical Society is a state department, it's autonomous. It's a quasi-state [agency].” Other occupational personnel referred to the State Archives with an apparently distanced, but politely respectful, tone. The IPAD interviewee (P-16) reported that the general public and internal state personnel often ask IPAD retention-oriented questions because they know that IPAD covers data practices and because that group’s name is still attached to a number of the pre-existing retention schedules. When retention questions are presented to them, however, they pass the questioner over to the State Archives personnel, whom P-16 referred to as the “government records specialists.”

Although Minnesota interviewees reported that they do perceive archivists and records managers to be recordkeeping personnel, when speaking about the context of recordkeeping in the Cloud they excluded archivists. Instead, respondents included archivists as experts at records management and records stewards, but did not include them in their descriptions of recordkeepers in the Cloud. Three reasons are likely for this:

1. Personnel (including archivists themselves) perceive archivists to be “external” to the state structure, since they work under the auspices of MHS. As a result, although they provide expert consulting services and are among those who provide final approval
for retention schedules, they are not really seen as part of the state organization itself and they maintain a non-custodial role over recordkeeping functions.

2. Permanent retention of archival email is not maintained in the Cloud. Rather, when email is kept for longer periods of time, these records are migrated to an internal preservation server.

3. As employees of MHS, the archivists do not reside on the same email system as the rest of the state. Although their email is also cloud-based, they correspond via their own Gmail cloud. Thus, they are not viewed as integrally tied to the state’s email cloud.

That said, neither do state employees exclude the archivists from all things cloud. Members of the State Archives sit on committees and workgroups such as the Data Domain Workgroup, which does discuss cloud computing and its use within the boundaries of the State’s information architecture. Nonetheless, the archivists do not participate in ongoing (or even one-off) issues related to the new Microsoft cloud. Thus, it appears that although the archivists do not engage in day-to-day recordkeeping tasks, they do play more strategic roles alongside MN.IT Services.

Comments regarding inter-occupational relations primarily centered on the relatively recent consolidation of IT governance, or to agency or occupational relations with MN.IT Services (P-8). In particular, records management perceptions of IT were described in the most problematic manner during the interviews. The respondents expressed frustration that they feel they cannot communicate with IT or that IT will not act upon such communication (P-21). The records managers also expressed frustration with what they perceive to be MN.IT Services personnel’s lack of understanding or concern about records management issues (P-
For example, P-21 commented of IT, “everybody's awesome at their job but they are so technology-focused that they don't see the greater picture of the content and so it's very frustrating.” This interviewee also commented on a meeting that MN-GRIN had with Minnesota’s Chief Information Officer Carolyn Parnell:

> We did ask her to attend a meeting and this was probably six, eight months ago and she did grant us an audience [my italics]… we approached her about the lack of contact that records managers have with the IT division, the OET. We would really like to participate because we have concerns, and she was like, ‘Oh, okay!’ And, you know, nothing happened. Well, we don’t know who else you can talk to if you can't talk to the big wig.

From the reported perspective of most of the records managers, tension exists between records management and IT. P-25, a relatively new employee of state government, did not want to discuss this tension directly but did say, “There are issues with IT,” adding that the records managers who have worked within P-25’s agency for several years have been talking about the issues with IT and records management “for years now.” P-24 remarked that there “is some tension” between him and IT, specifically referring to questions about who is responsible for ensuring that there is adequate storage:

> Our IT department, in the past, has called me and said, “Such and such drive is getting full, which is a records management problem, so you need to go and talk to the people who have things in this drive.” And I recognize that it might be a records management problem, so I go and I talk to the people and by and large I've found out that actually these people are keeping things that they do need to keep. So there's a little tension between me and IT as far as then I come back to IT and I say, “Well, actually, this is an IT problem, you're not supporting what they… it's a storage problem.” And so they understand that I have some sort of role in managing what's kept and what's not kept but that sometimes things need to be kept and we need to figure out smart ways to keep them. And so I've tried at various points to kind of push them to think about smarter backup and smarter storage. And I know that they're doing that right now, and that's part of the whole cloud thing - that we can store what we need to store without a huge burden on our internal IT resources” (P-24).

In addition, P-24 perceives that IT staff [members] seem to think that more storage and forever storage is what people want. They just want to store everything forever and then they'll be happy with IT.
And there are certainly those people who just want to store everything forever, but my struggle is to emphasize, no, we don't want to store everything forever and we need tools that will help us to get rid of things when we should (P-24).

P-22 echoed many of these concerns about IT, noting that not only does she operate in “a separate environment,” but also “there's no communication whatsoever. We don't get involved in designing of new systems to ensure that records management requirements are embedded, so there are already systems that have been decommissioned but do require to be kept because there had [originally] been no sort of consideration to incorporate the retention and dispositions of records.” P-22 also remarked that although the role of records management is much more about multiple forms of compliance, IT tends to see records managers “as more paper-based” and as “the people that help develop the retention policies” (P-22). Pointing out that, “we don't work together, that's the problem,” P-22 added, “IT does its own thing and records management is never involved.” She then pointed out that this non-communication makes it virtually impossible, when new systems are implemented, for records management and records retention requirements to be included in the new system requirements, so that “things can be managed a lot easier, rather than [through] labor-intensive, manual processes.” P-22 believes the records managers are forced to engage in such labor-intensive processes right now. She also believes that due to the lack of communication, those individuals working as records managers are not provided with the necessary background and opportunities to develop the technical knowledge that will allow them to communicate clearly the full requirements of records management. According to P-22, this leads IT staff to misunderstand records and the goals of records management and to restrict records managers from being able to perform all of their necessary functions on both paper and electronic records. It also reinforces an already-present and implicit belief that records management is “about paper” while IT is “about electronic” (P-22).
However, not all records managers share this discomfort with IT, perhaps partly because some of them are IT. P-23 is a records manager who reports to her agency’s IT department, which has recently been pulled into MN.IT Services as part of the IT consolidation. P-23 reported that this reporting structure has worked very well for her and that it has allowed her to develop a better understanding of technology, which allows her to manage electronic records better. This comment echoes the belief of P-22 that some technical knowledge is necessary in order to ensure more effective records management.

Another concern raised by records managers, especially with respect to performing records management in the Cloud, was the concern that the specific roles and responsibilities of records managers are not always clarified. When discussing how people in the organization tend to think of records managers as “hard copy” workers while IT deals with the “electronic” information, P-22 said, “I think that's what needs to be defined: the role of IT and the role of records management. People don't really understand the two and they distinguish it by [assuming] that records managers manage paper, and then it's like, who manages the electronic and then IT is like, well no we don't manage the electronic; we just provide the space and the systems.” Records manager P-24 concurred: “Because people say, oh, well, that's data, that's a technology issue. IT manages storage and, and if we need to get rid of things, we get rid of things because IT says we don't have the space rather than because we should be keeping them or we should not be keeping them.” P-24 has also had issues in her agency in getting people to recognize that database data is also a record and thinks this difficulty is related to the fact that IT is considered the “owner” of the information within

50 However, P-22 did exhibit some equivocation. She said that she thought IT was the best fit for records management given that “we don’t have a specific legal department in our agency,” (perhaps unintentionally) implying that if the agency had had a legal department, it may have been the best fit for records management.
databases. People think that because it is “data” it must be a technology issue. P-24 also believes that the changes in roles and responsibilities have led to a declining sense of ownership on the part of records creators and records managers, asking “Is it really ours anymore?”

According to P-22, “OET's operating exactly the way that IT has been operating within different organizations as well. Basically, they'll bring in the tool – ‘Here’s the tool. Use it.’ – and then basically don't provide anything, [any] roles around that. Because maybe it's not even their responsibility, either” (P-22). When it relates to the retirement of old systems, this can exacerbate records management issues. “They [IT] bring a system and off they go to use it during its active stage, but then what happens is, years later . . . we need to do something about it and IT says well, we're not going to support the system anymore. Because it's old and it's outdated . . .” (P-22). P-22 also asserted that in such a situation the divisions [agencies] are expected to come up with a solution for the retention of the information within those systems, even though they don’t have the technical knowledge to know how the information can be migrated to a new system.

We're not like IT people. How are we supposed to know? So, it's something back and forth and it's not thought out, who's responsible for what are we capturing, where we [are] capturing it, what happens to it from the beginning through to the end, where is the retention that says they can be destroyed or, if the organization is supposed to close up for whatever reasons what will happen to that record?” (P-22).

Some of this confusion is exacerbated by agency decisions about who the liaisons to records managers (or in some agencies, records coordinators, who spend about 5% of their time helping records managers) will be.

It is not a formal thing where it's in their job description. It's just something that they've been nominated [for], and they go through training, but most of the time what I see is that people that have been nominated it's again because of the misconception of what records management is. The people that get elected, they're usually the people
that are on the bottom and they each have minimum skills, and if you train them, then
when we go to developing a retention or any sort of analysis work, it is someone else
who gets sent, so then the problem [arises] of needing to retrain. It's [due to a]
misunderstanding of [what] records management is really (P-22).

MN.IT Services manager P-13 (IT) did not mention relationships between IT and
other occupational personnel at all, except to note that a number of the users were unhappy
that they could no longer create and edit their own .pst files now that the email is in the
Cloud. “We have SLAs and stuff like that in place that say, “You put this request in and, and
you’ll get results in three to five business days . . . So, .pst files - they could create them
themselves; they can’t do that themselves now, that’s a request to us. They can still get the
data. Is it as timely? No” (P-13).

Although the State Archives personnel reported no issues with MN.IT Services
themselves, they did recognize that some of the business personnel and records managers did
feel some tension:

IT often like[s] to set initiatives for the rest of the organization instead of the other
way around, which sometimes leads to difficulties…You know, IT’s the driver
instead of business being the driver . . . this was never an issue before electronic
records but now, our information resides on their servers, in their systems, and so
they see themselves as the stewards of this, and responsible for its management,
whereas, you know, we should be also having a say in that because we are the content
owners as it were. We’re responsible for that content as well, so there’s a constant
tension there between whose stuff is it, and who gets to manage it . . . (P-8)

P-21 finds one source of role stress51 to be related to scenarios in which P-21 is
requested to manage people in the organization that reside in positions that are higher in
status:

51 Role stress is the stress experienced by people in organizations specifically because of their role in the
organization (Tarafdar et al. 2007). One way in which role stress has been associated with technology adoption
is through role conflict, which occurs when an individual “is exposed to contradictory, incompatible, or
incongruent role requirements” (306). This happens when an individual is asked to fulfill the requirements of
more than one role, where the expectations from a given role may contradict the expectations from the other
roles. Another way in which roles stress has been associated with technology adoption is via “role overload,”
In the hierarchy of this organization, I'm kind of low on the pole. That's how I view my place in the hierarchy. I am expected to tell - not ask - but to tell all the people above me what the status is, what the direction is … and I sometimes feel really uncomfortable doing that because I'm low man on the pole and I'm telling the uppers, this is what we're going to do and this is how much it's going to cost, and if you don't do this what's going to happen. And I think that's some of what I should be doing but I really feel uncomfortable doing that.

Part of this discomfort appears to be a result of the response that is sometimes provided when P-21 does offer direction. “There are times when I'm called in to render an opinion and the people that have asked for my opinion don't like it, and then they pull me and send me on my way” (P-21).

In fact, the apparent lack of concern exhibited by IT personnel with respect to their own status suggests that the IT personnel perhaps do see themselves at the top of the food chain. Perhaps the clearest indication of this came from the IT executive manager. At one point, this individual said, “We are IT, so we're representing the hosting of e-mail, collaboration for all branches of government.” This statement is certainly true, but also speaks to the perception that for all activities associated with technology (which is virtually all activities), IT is the internal (and external) face of the government and the representative of the State.

4.1.7.2. Perceptions of Changes Brought by the Cloud

Very few cloud-specific changes were cited by respondents. The most prevalent change noted, however, was the fact that roles and responsibilities had become confused and blurred as a result of the cloud computing adoption. Although this frequently occurs with new technology adoptions, in this case it appears that some of the confusion is related to the
distributed nature of the information processing itself in a cloud environment. In addition to having a third party (and for some people, the suspicion or actuality of more than one additional party) added to the work flows, some activities for which records management personnel are still held accountable are no longer within their control. For example, they no longer have the capacity to go directly into terminated employees’ emails to “clean up” the folders and ensure that retention requirements are met as they did prior to the implementation (P-21, P-24). This increases the risk of the terminated individuals’ records being treated as if the individuals had not been terminated. However, if one of those individuals has been named in a discovery case, this can create problems for the records managers. How do they get access to these folders? Will Microsoft provide the information in a timely enough manner to satisfy the legal requirements, especially given a perception on the part of records managers that service provision has slowed down?

Additional concerns and confusion about data ownership have resulted from the cloud adoption. MN.IT Services still wants the records managed by the records managers, suggesting they (i.e., MN.IT Services) are the “highway” only, while the records managers are the drivers (P-25). However, the records managers sometimes push back, effectively saying, “We don’t have access to the servers or the technological ability to manage these records now; we don’t have the authority to handle these records now. You own the machines and you control the movement of service requests to Microsoft. You own this responsibility.” (P-25)

Other changes brought by this implementation appear to be change management issues more than cloud adoption issues. Records managers lack knowledge about how decisions were made during requirements analysis and implementation. They also lack
knowledge about the genuine risks of cloud computing to records and recordkeeping activities. Some of them have a somewhat mistaken understanding of “where” the data resides when it is “in” the Cloud. For example, P-21 said, “Now with the migration to Microsoft, our information, according to our statutes information needs to stay within our boundaries, that's our interpretation of it, so it's in our boundaries. So Chicago and India, and Canada – wherever the data centers are – are not within our boundaries.”

4.1.8. Synopsis of Case 1 Findings

Overall, several themes predominate in this case. Ownership is a primary theme running through virtually all conversations: ownership of data, ownership of systems, and ownership of job responsibilities. Another theme that arose was that of having had “agency” IT personnel “sucked up” by MN.IT Services. In addition, several interviewees reported that it was difficult to disassociate the effects of the IT consolidation with that of the cloud adoption; they felt that the changes that had occurred as a result of each of the two measures were tightly coupled. In fact, it is difficult to know whether this is true or not, especially given that virtually all respondents from this case, from the other two cases, and from the other interviewed states that have adopted cloud computing services, have all had cloud computing adoptions occur at about the same time as they have had their IT consolidations occur. Within the interviews with personnel in the other two cases presented here, and from other interviewed states that have adopted cloud computing services, every interviewee reported that his or her cloud computing adoptions and consolidations occurred very close together temporally. The overall effect has been a feeling on the part of individual agency personnel that they have lost power within the organization. Finally, whether directly causing it or simply amplifying an already present attitude, the Minnesota cloud computing adoption
has brought to the surface expressions of resentment on the part of agency personnel towards MN.IT Services. They have complained that IT “won’t listen” to them or that they are locked out of decision making, and that mutual communication is lacking with respect to information systems decisions.

In addition, archivists appear to identify themselves as “outsiders” in the state communication- and power structures. They do not find this problematical, however. In fact, from the point of view of archival theory, one might hypothesize that this is a desirable place for the archives to be, since it allows them greater objectivity and great ability to maintain the “archival bond” (Duranti 1997). In the case of Minnesota, however, the Archives takes a consciously chosen non-custodial approach, primarily due to resource constraints.

The archivists appear to recognize the power structures within the agencies implicitly, however, in that they realize that they can exert influence through the Data Domain team, a team affiliated with MN.IT Services. By tying themselves to the real power brokers, they derive a status and are able to market themselves as the “records experts” for the organization. By engaging in the cross-department team, they are also able to influence members of MN.IT Services and garner an audience for their own issues and concerns regarding records management and preservation.

Records managers did not speak as “outsiders.” Rather, they expressed their relationship with IT in a rather competitive manner. They argued that people do not recognize their roles within the organization but instead assume that IT has control over roles and responsibilities that are more properly records management roles and responsibilities. They included IT in their assessment of those who think that Records Managers are primarily associated with “paper.” Unfortunately, some individuals did not have a clear understanding
of the technology driving cloud computing and thus would be unlikely to be given full audience by IT. Assuming, for example that the data centers might be in “India or China,” shows a lack of clarity around the actual architecture and thus the actual risks toward records. Given that one generally needs to speak the language of one’s intended audience, it is unlikely that IT personnel would take much stock in communications about the entire level of risk if some of the records managers’ stated concerns were clearly not relevant or were based on faulty assumptions.

Records managers did, however, voice concerns about the incentives and disincentives of managing records in this new, distributed environment, noting that their process ownership had been taken away from them.

4.2. Case 2: Multi-Jurisdictional Cloud Implementation

4.2.1. Introduction

The CDC BioSense Program was initiated in 2003 as a result of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (CDC na), itself a response to the terrorist attacks on 9/11/2001 and the subsequent anthrax attacks ((CDC 2008); P-26). Originally, BioSense was set up to provide a national public health syndromic surveillance system that would afford early detection and assessment of potential bioterrorism-related health incidents (CDC na). In 2011, it was identified as a desired component in the updated Public Health Emergency Preparedness (PHEP) cooperative (CDC 2011). The original PHEP cooperative agreement, for which Congress authorized more than $5 billion between 2002 and 2007, was designed to help “public health departments at the state, local, tribal, and/or territorial levels work together to improve preparedness” for “human illness from chemical, biological, radiological agents, and naturally occurring health
threats” (CDC 2008c). The revised PHEP cooperative agreement defined a set of necessary
preparedness capabilities and recommended interventions that would help to achieve these
capabilities. The BioSense system was identified as a component that would aid in achieving
the PHEP cooperative agreements capabilities #6: “Information Sharing” and #13: “Public
Health Surveillance and Epidemiological Investigation” (CDC 2011, 4, 120).

“Syndromic surveillance” refers to “methods relying on detection of clinical case
features that are discernible before confirmed diagnoses are made” (Mandl et al. 2004, 141)
and “without regard to the specific diagnoses, if any, that are assigned to them by clinicians”
(Reingold 2003). The International Society for Disease Surveillance (ISDS) defines it as “the
real-time (or near real-time) collection, analysis, interpretation and dissemination of health-
related data to enable the early identification of the impact (or absence of impact) of potential
human or veterinary public-health threats which require effective public health action”
(Kass-Hout, Barr, and Alletto 2012). Buehler et al. (2009) provide detailed characteristics of
such surveillance. They say that it is “surveillance for human health-related events or
outcomes, including pre-diagnostic events or diagnoses … for the purposes of early event
detection or situational awareness (i.e., monitoring disease trends or other markers of
community health in situations where there is a need for prompt information…),” implying
the need for real-time data (166). Such discernible features could include monitoring in the
aggregate individual activities such as purchasing more facial tissues, orange juice, or cold
medicines, increased calls to nurses and clinics, increased health-related queries on the
Internet (Gesteland et al. 2003), rising emergency department visits, or a wave of influenza-
like illness, all of which could, taken together, indicate the beginnings of an epidemic
resulting from bioterrorism (Buehler et al.; Mandl et al.) or natural pandemic (P-26). Such
features are generally “pre-diagnostic” in the sense that they represent the earliest discernible phases of widespread health events that often occur before any diagnoses of specific conditions are made. BioSense relies upon both pre-diagnostic and diagnostic data sources (Loonsk 2004). Even in early discussions of the BioSense 2.0 redesign, however, a key question was whether a new syndromic surveillance system was, in fact, even necessary (P-26). Would its benefits exceed its costs?

BioSense 2.0 is “the first Department of Health and Human Services system to move completely to a distributed cloud computing environment” (CDC 2013b). Although the initial 2003 version of the CDC’s BioSense system did not reside in the Cloud, in November 2011, BioSense 2.0 moved from its testing phase to an initial go-live in Amazon’s AWS “Government Cloud through Amazon” environment (http://aws.amazon.com/govcloud-us/), which offers “security controls and certifications such as FISMA, SAS 70, ISO 27001, FIPS 140-2 compliant end points52, and PCI DSS Level 153,” as well as an additional layer of permissions “that restricts access to those on an approved list of US Persons” (Amazon Web Services na). It also meets HIPAA requirements (Dublin 2012).54

This case was selected as a means for examining a cloud computing implementation in which state governments took part but which was not entirely delimited by a state

52 The Federal Information Processing Standard (FIPS) Publication 140-2, or FIPS 140-2, specifies the security requirements for a cryptographic module within a security system that protects sensitive but unclassified information. It includes four increasing levels of security and supports the implementation and design of cryptographic modules in a wide variety of environments and for a wide variety of applications (NIST 2001).

53 The Payment Card Industry Data Security Standards Level 1, or PCI DSS Level 1, created by the private PCI Security Standards Council, is a framework for organizations to use to ensure that their payment card data security process provides “prevention, detection and appropriate reaction to security incidents” (PCI SSC nd). “Level 1” refers to the size of the organization, in terms of number of payment card transactions processed per year. Those organizations that process more than six million Visa transactions per year are Level 1, the largest level defined by the PCI DSS.

54 These regulations and other legislation affecting the BioSense 2.0 environment are discussed in more detail in Appendix G.
government’s legal, cultural or administrative boundaries. This allowed some examination of themes that may otherwise appear to represent state government environments alone, but may in reality go beyond the limits of state government workplaces or may be separate from but dependent on state government activities. Five individuals representing the Centers for Disease Control in Washington, D.C., the North Carolina Department of Public Health, a consulting company hired to manage the implementation, and a statewide surveillance program, North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) were interviewed. All individuals worked with BioSense and BioSense data both before and after the move to a cloud-based system (i.e., they worked with both BioSense 1.0 – as it is now being called by participants – and BioSense 2.0).

4.2.1.1. The Decision to Move to the Cloud

One CDC respondent reported that “multiple factors played into” moving BioSense into the Cloud (P-56). This person also suggested, “we could always look at it as version two because one didn't work, or we can look at it as a 2.0 movement where there's more context … there's more collaboration, there's more concerted effort to create data just like what happened with Wikipedia, and many other examples for the 2.0 movement” (P-56). In fact, both reasons appear to hold true. One problem associated with the first version of BioSense was the result of a combination of factors, including architecture, governance, and general lack of use (Buehler et al. 2009; P-56). For example, an interviewee shared that during the first iteration of BioSense, the CDC “totally bypassed state and local health departments and that didn’t win them a lot of friends . . . so they’re trying to do a more user-centered approach this time around” (P-14). In addition, BioSense 1.0 initially focused on event detection, which was not a successful strategy for encouraging use by state and local jurisdictions. Most
states already have their own syndromic surveillance systems and their own models and requirements for event detection and reporting (P-14). By engaging in its own event detection and reporting, BioSense 1.0 often would report events to State Health Departments that did not meet the criteria that the local systems followed. Many state healthcare personnel told interview respondents that duplication of efforts and “false warnings” occurred and that BioSense was not really meeting their particular needs (P-14). Instead of event detection, these individuals wanted a “broader view of the data,” whereby they could see other jurisdictions’ aggregate data in addition to their own (P-14).

Yet another challenge in garnering support for BioSense is related to skepticism around the benefits of automated syndromic surveillance systems, which have generally been perceived to provide value only when used in conjunction with more collaborative communications mechanisms, such as direct note-taking and person-to-person communication by physicians and/or epidemiologists in emergency departments, during potential high-risk scenarios (such as the Olympics), or rapid laboratory results reporting (Osaka, Takahashi, and Ohyama 2002). Others have raised questions about the accuracy of single-institution syndromic surveillance systems, suggesting that such systems need to cross jurisdictions and regions, existing at least at a city-wide or regional level in order to provide accurate early detection (Weber and Pitrak 2003). Reingold questioned the ability to confirm that syndromic surveillance systems support any of a wide variety of purported benefits that have led to their increasing popularity in the United States. In addition, since most states and many of the jurisdictions\textsuperscript{55} that are collecting syndromic surveillance data already have their own systems and did not anticipate benefits from a duplicative effort on the part of CDC,

\textsuperscript{55} Jurisdictions can be defined as states, countries, territories or, in some cases, geographical regions.
they instead focused on the primary perceived risk of sending local data to the CDC – the potential loss of control and ownership of their own local data (P-27). CDC executive manager P-26 said,

The original system didn't quite work from a quality perspective, because the only player or the only master was CDC. Even though CDC was bringing other partners in, making decisions on the data was primarily a data-run operation at CDC. All the data pretty much ended with one entity, which was CDC. And for the many partners, even if they chose to participate, [this] was really a high barrier for entry, as well as [a] high barrier for collaboration and sharing.

Many jurisdictions had no pre-existing data-sharing agreements or standard language to develop usage agreements, so the incentive to engage in sharing was quite low and the risks were perceived to be high. Thus, when BioSense was redesigned as BioSense 2.0, the designers primarily focused on allowing and encouraging data sharing by providing ownership and control of data to the jurisdictions themselves, and on taking governance out of the hands of the CDC (P-14; P-26; P-27). Currently, the Association of State and Territorial Health Officials (ASTHO) acts as the negotiating body of BioSense 2.0, contracting on behalf of BioSense 2.0 with Amazon Web Services (P-26). It also negotiates the data use agreements (DUAs) with jurisdictions being “onboarded.” The CDC provides funding for the initiative on behalf of those jurisdictions that are awarded grants for their participation, paying for the cloud services but otherwise having no managerial or control rights (P-26; P-27). Jurisdictions are entirely free to share or not to share their data with other jurisdictions when they participate in the BioSense 2.0 program. Table 8 shows the number of state-, county-, and city- level jurisdictions at various stages of onboarding as of mid-July, 2013.

56 Onboarding” is the term BioSense 2.0 participants use to refer to a jurisdiction beginning to prepare technically to send and store its data in BioSense 2.0 after signing the DUA, thereby becoming a full BioSense 2.0 “partner.”
Table 8 - Status of BioSense 2.0 Jurisdiction Onboarding, July 2013

<table>
<thead>
<tr>
<th>Status</th>
<th>State/D.C.</th>
<th>County</th>
<th>City</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully onboarded</td>
<td>13</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Onboard in process</td>
<td>24*</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>DUA signed, but onboarding not started</td>
<td>--</td>
<td>2</td>
<td>--</td>
</tr>
</tbody>
</table>

*Two of the counties in the process of onboarding reside in states that are not included in the “state” column. Thus, jurisdictions from 26 states were in the process of onboarding as of July 2013.

Although a stated rationale for participating in BioSense 2.0 is data sharing, in practice the individual jurisdictions tend to hold tightly to the access of their own information (P-27). Several interviewees reported that little data sharing occurs across jurisdictional lines (P-26, P-27). In fact, although the DUAs originally provided for optional data sharing with the CDC, such a low rate of sharing with the CDC occurred that the BioSense 2.0 Governance Committee had to change the DUA verbiage (P-27). The verbiage now requires that any jurisdiction receiving grant funding by the CDC for participation in BioSense 2.0 must share their data with the CDC. Besides concerns about maintaining control of their data, some states – like Florida – require that all state health data remain within the physical boundaries of the state (P-27). For others, the time and expense of labor creates a resource issue that the state cannot justify undertaking for what they perceive to be a small incremental increase (or no change) in value (P-27). BioSense 2.0 is an example of a network good – one in which widespread sharing would create huge value both overall and to individual states, but for which reaching that level of sharing is difficult because the individual benefit received is much smaller before such sharing becomes widespread.

Based on the DUAs between ASTHO and the jurisdictions, the data may remain viewable only by the data owner/jurisdiction or it may be shared with other jurisdictions, with the CDC, and/or the general public. Jurisdictional personnel may view other jurisdictions’ aggregated data according to two limitations: (1) A jurisdiction may only view
other jurisdictions’ aggregated data if it provides access to its own aggregated data; and, (2) a jurisdiction may only view other jurisdictions’ aggregated data down to the same level of aggregation at which it provides access to its own data. For example, if the state provides other jurisdictions county-level access to its data then it can view the data of other sharers at the county-level. If it only provides access to its data at the state-level, then it can only receive access of others’ data at the state-level. This requirement is referred to as the “principle of reciprocity” in the BioSense environment (Spears 2013, 2): no one can receive an access level more granular than the level at which it shares its own data.

4.2.2. Recordkeeping Stewards

Stewardship for BioSense 2.0 is complicated, and relies upon a highly distributed set of roles and responsibilities. Ownership of the BioSense 2.0 information is also widely distributed. Individual medical facilities (including pharmacies) create medical records. In addition, two federal entities continue to participate, as they did in the BioSense 1.0 system – the Department of Defense and the Veterans Administration. Information from the latter two organizations’ records are captured and uploaded at least twice daily to these jurisdictions’ “storage lockers” on the Amazon Web Service (AWS), which serves as a platform for individual jurisdictions to engage in analysis, transformation, and visualization through the BioSense 2.0 web interface and allows safe storage of data. These storage lockers remain privately accessible to the owners (i.e., jurisdictions) alone (P-26; P-27). In addition, within the system is a shared space in which jurisdictions can place the data that they are willing to share with other jurisdictions.

A narrow set of personnel within a relatively wide set of organizations are both primary users and potentially, primary stewards of information that is sent to BioSense 2.0.
Because jurisdictions plan to share some data with the general public, the public could be considered a stakeholder, but it is not responsible for actual information stewardship. However, when speaking of stakeholders, the interviewees did not identify other stakeholders and stewards either by individual name or by occupational name or role and responsibilities. Rather, they consistently and continually referred to the name of the organizations participating in the BioSense 2.0 program when speaking of data stewardship. For example, the North Carolina Department of Health is identified in all official communication regarding BioSense 2.0, even though individual physicians and employees within the Department of Health work within the BioSense collaboration and the data itself originates from individual facilities within the state.

4.2.2.1. United States Centers for Disease Control and Prevention (CDC)

The CDC is the primary funding agent for the BioSense 2.0 system (and was the owner and funding agent of the original BioSense system). A variety of participants within the CDC engage in BioSense-related information activities. For example, the CDC’s Public Health Surveillance and Informatics Program Office (PHSIP)\textsuperscript{57} evaluates data from the BioSense project. The CDC also funds ASTHO to host BioSense 2.0. In addition, it coordinates with the Council of State and Territorial Epidemiologists (CSTE), the National Association of County and City Health Officials (NACCHO), and the ISDS to facilitate their joint participation in the BioSense Governance Committee (CDC 2013c). Personnel within

\textsuperscript{57} Specifically the Division of Notifiable Diseases and Healthcare Information (DNDHI) within the PHSIP is responsible for leading “the integration of CDC’s statistical, epidemiologic, and informatics methods for public health surveillance and evaluation” (CDC 2013d).
the CDC also monitor and manage the data within the BioSense 2.0 system on an ongoing basis, ensuring that access requirements and constraints are met.

4.2.2.2. State Departments of Health

Each state follows a potentially different configuration of data submission roles and responsibilities. For example, in North Carolina, hospitals and other local facilities submit the data to the State Division of Public Health (DPH), which submits all the data to BioSense 2.0. The DPH, however, subcontracts out their syndromic surveillance system to an entity called NC DETECT, which is housed within the University of North Carolina, Chapel Hill (P-14). Some states house their syndromic surveillance system within their Department of Public Health itself, however. The systems themselves may be built in-house, such as NC DETECT, or may be purchased off-the-shelf, like Essence, which is the one of the most popular syndromic surveillance systems in the country.58 In North Carolina, NC General Statute § 130A-480, enacted in 2005, mandates that jurisdictions send their public health threat data to the State Division of Health. In 2007, the legislature added verbiage that made it clear that it is legally permissible for them to allow the DPH to share this data with the CDC, if the jurisdiction agrees to share it. That is, there is no requirement for the individual hospitals or jurisdictions to share their data with the CDC, although there is a requirement to share public health threat data with the state itself. The jurisdictions are allowed, however, to share their data with the CDC, with the DPH acting as intermediary.

Not all states require jurisdictions to send their data to the state public health department, and not all states provide a legal provision for the health department to share that

58 The CDC is currently developing a registry of all the currently used syndromic surveillance systems used nationwide. It is not complete as of the publish date of this document, but currently includes thousands of different systems being used in the United States (P-14).
data with the CDC. Each state has its own laws and norms regarding the sharing of syndromic surveillance data. As a result, when ASTHO negotiates DUAs with jurisdictions, the appropriate level of negotiation must be determined and the specific language of the DUA contract must be approved by all of the individual stewards.

The data itself is generated by individual emergency departments, urgent care centers, pharmacies, labs, etc. The nature of the sharing arrangement is constrained by a combination of state law and the wishes of the jurisdictions and facilities. Thus, some individual facilities send their data directly to BioSense 2.0. Some send their data to their regional facility or State Department of Health, which then chooses whether to share with BioSense 2.0. Some share with their State Department of Health and simultaneously choose not to share with BioSense; others allow the Department of Health to share. Yet others are required by state law to allow their Department of Health to share their data with the CDC. The wide variation means that new DUAs must be negotiated whenever a new facility is brought on board, whether that jurisdiction is an individual facility within a legal jurisdiction or an aggregate of facilities within a region or state.

No interviewee discussed levels of stewardship more granular than the level of facility. Neither was this information available in any documentation. When specifically asked about personnel, some stewards agreed that individual records managers or records creators may be monitoring or creating the primary data itself, but the interviewees also stressed that the jurisdiction data that BioSense 2.0 receives is secondary data, where the term “secondary data” refers to a subset of data that is created after the data is first collected for its primary medical uses (P-14). Although the BioSense data is de-identified, at the storage level keys exist that allow the people or systems that manage the information to trace
it to its source and validate it for correctness or, if a syndromic event occurs, to trace back to the individuals affected. Although the recipients of the secondary data “have no control over how the data are entered,” and cannot vouch for the correctness of the data itself, they can perform comprehensive data quality checks to discover data problems and use standardization processes to improve interpretability. Nonetheless, the fundamental approach they must take with this data is a “garbage in, garbage out” stance: the data in BioSense 2.0 is only as good as its source.

4.2.2.3. Association of State and Territorial Health Officials (ASTHO)

ASTHO is a national nonprofit organization that represents public health agencies throughout the United States, the U.S. Territories and the District of Columbia. Its primary role is “to track, evaluate, and advise members on the impact and formation of public or private health policy which may affect them and to provide them with guidance and technical assistance on improving the nation’s health” (ASTHO 2013). ASTHO acts as the “owner” of the BioSense 2.0 system in that it provides hosting for the system through Amazon Web Services. It engaged in the early vendor selection and acts as the primary governing body for the collaborative effort associated with onboarding organizations and maintaining the relationship with Amazon. ASTHO signed the DUA agreement with Amazon and works with individual jurisdictions that are negotiating their DUAs with BioSense 2.0. For example, in North Carolina the “DUA is between ASTHO and the Division of Public Health” (P-14). Beyond these governance duties, however, ASTHO does not act as a data steward itself.
4.2.2.4. National Association of City and County Health Officials (NACCHO)

NACCHO is an association of 2,700 local health departments across the United States (NACCHO 2013a). It works with ASTHO and several other health-oriented associations and societies to help govern the BioSense 2.0 collaboration. Like these other associations, NACCHO is also helping to recruit jurisdictions to take part in the BioSense 2.0 collaboration through informative workshops and advertising of the benefits of BioSense 2.0 (NACCHO 2013b). Also like ASTHO, NACCHO is not a data steward.

4.2.2.5. The International Society for Disease Surveillance (ISDS)

ISDS is a nonprofit organization “dedicated to the improvement of population health by advancing the science and practice of disease surveillance” (ISDS 2013b). Its membership is comprised of “professional and academic subject matter experts in the fields of public health surveillance, clinical practice, health informatics, health policy, and other areas related to national and global health surveillance” (ISDS). They facilitate collaboration among health professionals and researchers. As mentioned earlier, ISDS wrote Final Recommendation: The Core Processes and EHR Requirements of Public Health Syndromic Surveillance. It also worked with the CDC to “develop a draft PHIN Messaging Guide for Syndromic Surveillance” (ISDS 2011). It works with ASTHO and the CDC to onboard and educate potential and current BioSense 2.0 members but otherwise plays no data stewardship role.

4.2.2.6. Local Pharmacies

A pharmacy claims system which was providing information to BioSense 1.0 continues to provide pharmacy claims information to BioSense 2.0. These data “are currently only used to provide additional data for influenza-related syndromes and sub-syndromes;
however, the use of these data will be expanded to other conditions as part of BioSense 2.0” (Gallagher 2012, 8). In fact, “VA, DoD, pharmacy, and laboratory data received by CDC will be used internally to conduct anomaly analysis, and will only be shared in aggregate format containing no individually identifiable information with the CDC EOC59 and applicable public health jurisdictions when the BioSense Program is engaged in conducting surveillance on high profile events or public health emergencies” (9). Thus, there are two types of stewards who handle this data: The initial data creators, records managers, and IT personnel who handle the data at the point of source, prior to sending it to the CDC and the personnel who manage the data on a continuing basis once it has been received by the CDC and automatically transferred from the source system into BioSense 2.0.

4.2.2.7. States and Local Jurisdictions

As of June 2013, 13 state and local jurisdictions provide live data to the BioSense 2.0 Data Warehouse nationally. In addition, another 24 are in the process of setting up their live data feeds (as well as two county jurisdictions from yet another two states). The jurisdictions are responsible for meeting the data transport requirements and sending the minimum data set. As mentioned before, the providers of data may be individual facilities, data aggregators providing data on behalf of the state, state Health Departments, or a combination of these parties. They are responsible for the authenticity, integrity, and reliability of the data at the point of creation and transfer limited responsibility to BioSense 2.0, in that BioSense 2.0 will engage in some clean-up of the information it receives, but cannot verify the accuracy of the original primary records that were created.

59 The CDC’s Emergency Operations Center (EOC) is the CDC’s command center for “monitoring and managing emergency response to public health threats in the United States and around the world” (CDC 2012).
4.2.2.8. Department of Defense (DoD); Veterans’ Administration (VA)
Both the Department of Defense and the Veterans Administration play the same role as the other data providers in the BioSense 2.0 system. Individuals within each agency prepare the primary data according to BioSense 2.0’s core processes and EHR requirements mentioned earlier. BioSense 2.0 respondents did not refer specifically to any particular individual or occupational type when referring to the stewardship role of these organizations; they referred to the overall source organization as the primary stakeholder and pre-submission information steward.

4.2.2.9. National Laboratories
Two national laboratories provide data to BioSense 2.0 (Gallagher). As with the other data providers, interviewees considered these laboratories to be data stewards of the primary data (as opposed to considering the particular employees within the laboratories to be the stewards).

4.2.2.10. Research and Data Management Organizations
Some of the states that provide data to BioSense 2.0 rely upon data aggregator organizations to manage, cleanse and/or analyze the data for them before it is sent to BioSense 2.0. For example, in North Carolina the state’s Division of Public Health contracts with NC DETECT, the North Carolina Disease Event Tracking and Epidemiologic Collection Tool, to manage their data. NC DETECT collects, cleanses, and provides reports on the data to the NC DPH before feeding the data to the BioSense 2.0 warehouse. NC DETECT was originally set up in 2005 as a result of a state mandate that Emergency Departments report chief complaint data to the state (Rath 2012). Rath reports that NC
DETECT now reports “near-real-time statewide surveillance capacity to local, regional and state-level users across the state, with twice daily data feeds from 117 emergency departments, hourly updates from the statewide poison center, and daily feeds from statewide EMS runs and select urgent care centers.” Overall, NC DETECT receives information from more than 400 sources around the state (P-14). Because most of the data received is in free text fields, NC DETECT conducts natural language processing. They also standardize the data since many facilities use their own specific terms for a variety of the fields they send. In addition, NC DETECT checks to insure the completeness of data received and to verify that hospitals send their data according to the agreed-upon schedules. They work with the hospitals when issues are discovered (P-17). NC DETECT has been sending data to BioSense since 2007.

4.2.3. Legal Environment Affecting Stewards
The BioSense 2.0 program was developed to allow information sharing among participants from all levels of government, private organizations, researchers, and potentially even the public at large. Thus, the number of potential legal restrictions and requirements that may impact the storage, management, and use of the data residing in the repository as a whole is too large to list within this document. The most impactful of the related regulations or legislation mentioned here are described in more detail in Appendix G.

As mentioned earlier, the BioSense program itself arose out of the Public Health Security and Bioterrorism (BT) Preparedness and Response Act of 2002, which requires the CDC to work together with other federal agencies to monitor records and information related to potential public health risks. As the product of a Federal government agency, BioSense 2.0 is also required to meet FISMA and other privacy and security legislation, such as HIPAA. It
is also subject to the punitive actions of legislation such as the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, which assesses monetary penalties in the event of health information breaches and offers monetary incentives for medical facilities to use meaningful electronic health records (HER).

FISMA is particularly important, since it affects all electronic government information produced or managed by federal government entities. Enacted in 2002 FISMA (“FISMA,” 44 U.S.C. § 3541 et seq.) is Title III of the E-Government Act (Public Law 107-347), passed by the 107th Congress and signed by President George W. Bush in December 2002 (section 3541 title 44). This act requires each federal agency to “develop, document, and implement an agency-wide program to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source” (NIST 2002). The act defines information security to be “protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction” in order to provide integrity, confidentiality and availability (44 U.S.C. § 3541 et seq., 2002). FISMA affects many aspects of storing and managing information in networked environments, and BioSense 2.0 had to ensure that it was FISMA-compliant, which it did.

In addition, each state participating in BioSense 2.0 has its own legislation and requirements associated with electronic health records. Appendix G only discusses one particular state (i.e., North Carolina) for illustrative purposes, but similar laws related to privacy, security, and sharing of health information can be found in every state.
4.2.4. Requirements and Actions

Although the data feeding into BioSense 2.0 is considered to be the property of the individual jurisdictions, it is subject to requirements of the Freedom of Information Act:

All data/information that external organizations provide to CDC or other federal agencies, are subject to the Freedom of Information Act (FOIA) requirements. Data shared with CDC for analysis using BioSense 2.0 are no exception. Jurisdictions will ultimately have control of what data are shared with partners including CDC, and should always consider FOIA and open records laws when they determine what and how much data they share (BioSense Redesign Team 2013).

The original BioSense 2.0 partners, the CDC, and the Governance Committee negotiated a minimum set of data elements (shown in Table 9) to be shared during a highly user-centered requirements gathering phase. None of the data is personally identifiable information (PII), although some of the fields are masked at the source to hide personally identifiable data that is transferred to BioSense 2.0. For example, the facility will mask the medical record number associated with each field sent, but will retain the information that allows backwards re-identification. Then, if a surveillance event occurs, the BioSense information can be remerged with local records data to determine who the affected individuals were. The data is raw secondary data\(^{60}\) and is uploaded to the jurisdiction’s storage locker, which resides on the BioSense 2.0 data warehouse in Amazon S3. Thus, the CDC was provided legal mandate and funding to create both BioSense 1.0 and BioSense 2.0. In the case of many of the states, there is a legal requirement that jurisdictions share data with their state health department. However, legislation is also used, as in the case of NC General Statue § 130A-480, to legitimate the sharing of data to a degree that exceeds the legally required interstate and federal (i.e., CDC) sharing. The legitimating verbiage includes the

\(^{60}\) “Secondary data” is data which is derived from another source that has created the data to satisfy their business purposes. In this case, hospitals create data for use in treating patients and the data is stored in the patients’ medical records. However, a subset of this primary data is also sent to the State Health Department and/or BioSense 2.0 (in the case of those jurisdictions that are participating in BioSense 2.0).
**Table 9 - BioSense 2.0 Minimum Set of Shared Data Elements (ISDS 2011, 56-65)**

<table>
<thead>
<tr>
<th>#</th>
<th>Data Element Name</th>
<th>Description of Field</th>
<th>Usage</th>
<th>Cardinality</th>
<th>Value Set</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Facility Identifier (Treating)</td>
<td>Unique facility identifier of facility where the patient originally presented (original provider of the data)</td>
<td>R</td>
<td>[1..1]</td>
<td>National Provider Identifier</td>
<td>Use facility identifier for state or local reporting only. This is due to agreements with many health data providers that explicitly state that states or localities will not expose them to a third party like the federal government when reporting above state level. This number should be specific for each facility location (not a number representing an umbrella business) It is recommended that National Provider Identifier (NPI) be used for the Facility Identifier.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>----------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2</td>
<td>Facility Name (Treating)</td>
<td>Name of the treating facility where the patient originally presented</td>
<td>O</td>
<td>[0..1]</td>
<td></td>
<td>Use facility identifier for state or local reporting only. This is due to agreements with many health data providers that explicitly state that states or localities will not expose them to a third party like the federal government when reporting above state level. This number should be specific for each facility location (not a number representing an umbrella business). It is recommended that National Provider Identifier (NPI) be used for the Facility Identifier.</td>
</tr>
<tr>
<td>3</td>
<td>Facility Location 0 [0..1]</td>
<td>Street address of treating facility location</td>
<td>O</td>
<td>[0..1]</td>
<td></td>
<td>Use facility identifier for state or local reporting only. This is due to agreements with many health data providers that explicitly state that states or localities will not expose them to a third party like the federal government when reporting above state level.</td>
</tr>
<tr>
<td>4</td>
<td>Free text 6</td>
<td>City of treating facility location</td>
<td>O</td>
<td>[0..1]</td>
<td>Free text</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Free text 6</td>
<td>County of treating facility location</td>
<td>O</td>
<td>[0..1]</td>
<td>Free text</td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>----------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>6</td>
<td>State of treating facility location</td>
<td>O</td>
<td>[0..1]</td>
<td>FIPS 5-2 Use numeric codes</td>
<td>This number should be specific for each facility location (not a number representing an umbrella business) It is recommended that National Provider Identifier (NPI) be used for the Facility Identifier.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Facility / Visit Type</td>
<td>Type of facility or the visit where the patient presented for treatment</td>
<td>RE</td>
<td>[0..1]</td>
<td>TBD</td>
<td>Use facility identifier for state or local reporting only. This is due to agreements with many health data providers that explicitly state that states or localities will not expose them to a third party like the federal government when reporting above state level. This number should be specific for each facility location (not a number representing an umbrella business) It is recommended that National Provider Identifier (NPI) be used for the Facility Identifier.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>Report Date/Time</td>
<td>Date and time of report transmission from original source (from treating facility)</td>
<td>R</td>
<td>[0..1]</td>
<td></td>
<td>Use facility identifier for state or local reporting only. This is due to agreements with many health data providers that explicitly state that states or localities will not expose them to a third party like the federal government when reporting above state level. This number should be specific for each facility location (not a number representing an umbrella business). It is recommended that National Provider Identifier (NPI) be used for the Facility Identifier.</td>
</tr>
</tbody>
</table>

Patient Demographics
<table>
<thead>
<tr>
<th>#</th>
<th>Data Element Name</th>
<th>Description of Field</th>
<th>Usage</th>
<th>Cardinality</th>
<th>Value Set</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Unique Patient Identifier</td>
<td>Unique identifier for the patient</td>
<td>R</td>
<td>[1..*]</td>
<td>HL7 Table 0203</td>
<td>Examples of Unique Patient Identifiers are Patient Account number or a Master Patient Index (MPI) number. This data element may be used as the unique identifier used between the data sender and receiver to identify the record. The cardinality allows multiple identifiers to accommodate situations where a data provider sends multiple identifiers, such as patient MPI number in addition to patient account number. In addition, if the message goes through a data intermediary, such as an HIE, then multiple patient identifiers may exist. In such cases, it is important that all intermediaries retain and provide all associated patient identifiers for the patient.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>---------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>10</td>
<td>Medical Record #</td>
<td>Patient medical record number</td>
<td>O</td>
<td>[0..1]</td>
<td>HL7 Table 0203</td>
<td>It is recommended that data providers submit the patient medical record number to facilitate identification of the patient, in the event of a required follow-up investigation. Without the medical record number, the work required to follow-up on the records of interest greatly increases for the data provider and may cause unacceptable delays in public health response. In addition, the medical record number may aid in record de-duplication efforts and may often aid in the resolution of apparent transcription errors.</td>
</tr>
<tr>
<td>11</td>
<td>Age</td>
<td>Numeric value of patient age at time of visit</td>
<td>R</td>
<td>[1..1]</td>
<td>LOINC Code 21612-7</td>
<td>Note: Sending DOB is may not be an acceptable alternative to sending age due to possible restrictions in data privacy. Data providers and receivers should determine specific data restrictions on age for their jurisdiction. The data requested is the patient’s age at time of visit. The age should not update over time as the patient ages.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>12</td>
<td>Age units</td>
<td>Unit corresponding to numeric value of patient age (e.g. Days, Month or Years)</td>
<td>R</td>
<td>[1..1]</td>
<td>UCUM Age Units</td>
<td>Relevant Age Unit values: Days Weeks Months Years Use the unit that is applicable to and describes the numerical age value.</td>
</tr>
<tr>
<td>13</td>
<td>Gender</td>
<td>Gender of patient</td>
<td>RE</td>
<td>[0..1]</td>
<td>HL7 v2.5.1 Adminis-trative Sex (Table 0001)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>City/Town</td>
<td>City/Town of patient residence</td>
<td>O</td>
<td>[0..1]</td>
<td>Free text</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>State</td>
<td>State of patient residence</td>
<td>RE</td>
<td>[0..1]</td>
<td>FIPS 5-2</td>
<td>Use numeric code</td>
</tr>
<tr>
<td>17</td>
<td>Country</td>
<td>Country of patient residence</td>
<td>RE</td>
<td>[0..1]</td>
<td>ISO 3166-1 Country Value Set</td>
<td>Use 3 character codes</td>
</tr>
<tr>
<td>18</td>
<td>Race</td>
<td>Race of patient</td>
<td>RE</td>
<td>[0..*]</td>
<td>CDC Race Category Value Set</td>
<td>The patient may have more than one race defined.</td>
</tr>
<tr>
<td>19</td>
<td>Ethnicity</td>
<td>Ethnicity of patient</td>
<td>RE</td>
<td>[0..*]</td>
<td>CDC Ethnicity Group Value Set</td>
<td></td>
</tr>
</tbody>
</table>

Patient Health Indicators
<table>
<thead>
<tr>
<th>#</th>
<th>Data Element Name</th>
<th>Description of Field</th>
<th>Usage</th>
<th>Cardinality</th>
<th>Value Set</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Unique Visiting ID</td>
<td>Unique identifier for a patient visit</td>
<td>R</td>
<td>[1..1]</td>
<td>HL7 Table 0203</td>
<td>A visit is defined as a discrete or unique face-to-face clinical encounter within a service department or location. This data element may be used as the unique identifier used between the data sender and receiver to identify the record.</td>
</tr>
<tr>
<td>21</td>
<td>Visit Date/Time</td>
<td>Date/Time of patient presentation</td>
<td>R</td>
<td>[1..1]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Date of onset</td>
<td>Date that patient began having symptoms of condition being reported</td>
<td>RE</td>
<td>[0..1]</td>
<td>LOINC Code 11368-8 (Illness / Injury Onset Date / time)</td>
<td></td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>---------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>23</td>
<td>Patient Class</td>
<td>Patient classification within facility</td>
<td>RE</td>
<td>[0..1]</td>
<td>HL7 v.2.5.1 Patient Class (Table 0004)</td>
<td>It is recommended that PHA constrain the transmitted data using the patient class code set (example: only transmit records where patient class = E, I, O). There is a potential for a large amount of data if not constrained. If the PHA does not choose to constrain these data with separators, this field will be critical to process, constrain, and/or filter the data as needed by the PHA. Relevant Patient Class values: E: Emergency  I: Inpatient  O: Outpatient  P: Pre-admit  R: Recurring patient  B: Obstetrics</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>-----------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>24</td>
<td>Chief Complaint / Reason for visit</td>
<td>Short description of the chief complaint or reason of patient’s visit, recorded when seeking care</td>
<td>RE (see notes)</td>
<td>[0..*]</td>
<td>LOINC Code 21612-7: Free text (Preferred) Or ICD-9 Clinical Modification diagnosis code (including E-codes and V-codes) Or ICD-10 Clinical Modification diagnosis Code Or SNOMED Disorder/Disease domain</td>
<td>This value is critical for PHSS and is considered REQUIRED. However, there are settings or scenarios where this field may be blank (e.g. trauma patient). Therefore, the Usage value is ‘RE’. This field needs to be the richest and most complete free text description of the patient's chief complaint. If both the free text chief complaint text and drop down selection chief complaint text is available, send only the free text chief complaint. If the chief complaint is only from drop down list fields, then concatenate all drop down list chief complaints selected for that record/visit and submit. For updates: Some hospital systems automatically overwrite chief complaint with final diagnosis when the final diagnosis code is assigned. The chief complaint text should NOT be replaced with other information either manually or by the data provider’s system. It is imperative that chief complaint text remains how it was captured in the ED.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>----------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td>25</td>
<td>Triage Notes</td>
<td>Triage notes for the patient visit</td>
<td>O</td>
<td>[0..1]</td>
<td>LOINC Code 54094-8 (Triage Note): Free text</td>
<td>Triage notes should be sent as free text. This field should NOT include patient identifiable information. This may require practitioner education and training for the proper / intended use of this field. Triage notes may benefit from additional processing (e.g. negation processing, natural language processing, etc.) in order to maximize the utility of the data.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>26</td>
<td>Diagnosis/External Cause of Injury Code</td>
<td>Diagnosis code or external cause of injury code (for injury-related visits) of patient condition</td>
<td>RE</td>
<td>[0..*]</td>
<td>ICD-9 Clinical Modification diagnosis code (including E-codes and V-codes) Or ICD-10 Clinical Modification diagnosis code Or SNOMED Disorder/Disease domain</td>
<td>Do not delay sending of patient data for diagnosis or verification procedures. Patient data should be sent even if the diagnosis/injury code is not available. Any new data can be sent as an update to correct errors or to transmit data that was previously unavailable. Include V-codes and E-codes. This field is a repeatable field so multiple codes may be sent. The first diagnosis code should be the principal diagnosis. When the first-listed diagnosis code (principal diagnosis) is an injury, also provide one or more supplemental external-cause-of-injury codes or E-codes. E-codes provide useful information on the mechanism and intent of injury, place of occurrence, and activity at the time of injury.</td>
</tr>
<tr>
<td>27</td>
<td>Clinical Impression</td>
<td>Clinical impression (free text) of the diagnosis</td>
<td>O</td>
<td>[0..1]</td>
<td>LOINC Code 44833-2</td>
<td>This field is typically a free text field and is distinct from the diagnosis code.</td>
</tr>
<tr>
<td>#</td>
<td>Data Element Name</td>
<td>Description of Field</td>
<td>Usage</td>
<td>Cardinality</td>
<td>Value Set</td>
<td>Notes</td>
</tr>
<tr>
<td>----</td>
<td>----------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------</td>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>28</td>
<td>Diagnosis Type</td>
<td>Qualifier for Diagnosis / Injury Code specifying type of diagnosis</td>
<td>RE</td>
<td>[0..*]</td>
<td>[0..*] HL7 v2.5.1 Diagnosis Type (Table 0052)</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Discharge Disposition</td>
<td>Patient's anticipated location or status following ED/UC visit</td>
<td>RE</td>
<td>[0..1]</td>
<td>National Uniform Billing Committee (NUBC) – Patient Status UB04 codes</td>
<td>This field will update with multiple submissions. Include both the code and text description of the code. Discharge disposition should not be updated once the patient becomes an inpatient.</td>
</tr>
<tr>
<td>30</td>
<td>Disposition Date/Time</td>
<td>Date and time of disposition</td>
<td>RE</td>
<td>[0..1]</td>
<td></td>
<td>Transmit this field as empty if the patient has not been discharged. Do not wait to transmit data elements until patient is discharged.</td>
</tr>
<tr>
<td>31</td>
<td>Initial Temperature</td>
<td>Patient’s first recorded body temperature, including units</td>
<td>RE</td>
<td>[0..1]</td>
<td>LOINC Code 11289-6 (BODY TEMPERATURE) UCUM for Coded Numeric Units</td>
<td>Temperature may provide value in classifying certain conditions, such as pandemic flu. Units of the temperature should also be included.</td>
</tr>
<tr>
<td>32</td>
<td>Initial Pulse Oximetry</td>
<td>Oximetry value</td>
<td>RE</td>
<td>[0..1]</td>
<td>For Generic Pulse Oximetry: Use LOINC Code 59408-5 UCUM for Coded Numeric Units Units = % (percent)</td>
<td></td>
</tr>
</tbody>
</table>
**Legend: Table 9 - Biosense 2.0 Minimum Set of Shared Data Elements (ISDS 2011, 56-65)**

**Key Terms and Definitions:**

**Data Element Name** = Name of the minimum data set element.

**Description of field** = Description of the data element

**Usage:**

Refers to whether an element is a required or optional field. The Usage codes are:

**R – Required:** Indicates that the field is a required field and must be supported by the EHR system. A value must be present in the field in order for the message to be accepted.

**RE – Required, but can be empty:** Indicates that the field is a required field and must be supported by the EHR system. The reporting of data is setting-specific. If data are present, then they must be reported. However, if there are no data captured in the field due to the setting (e.g. no chief complaint data for a trauma patient) and the field is blank, the message may be sent with the field containing no data.

**O – Optional:** Indicates that this field must be supported by the EHR system, but the transmission of the values captured in these fields is optional. Specific usage of these data elements shall be determined at the state or local-level jurisdiction.

**Cardinality** = Minimum and maximum number of times the element may appear.

**Value Set** = The source from which the field value can be populated (e.g., free text or an HL-7 medical information table).

**Notes** = Additional notes describing rules pertaining to the data element, processing of the data element field, or identifying relevant values for the data element.
requirement on the part of the state to take adequate measures to ensure the data’s security and confidentiality as specified by related laws.

4.2.4.1. Defining the Recordkeeping Requirements

The user requirements were gathered through a number of different processes, and the results of the requirements gathering exercises have been continuously posted on the BioSense Redesign Collaboration website since the early stages of redesign. The general public may view all documents that do not reveal potentially risky security information about the system. RTI International assessed and analyzed the system requirements and developed the system in conjunction with the other stakeholders listed above. Activities performed (P-15) in support of requirements gathering were

- Focus groups, led by trained facilitators;
- One-on-on structure interviews;
- Semi-structured interviews;
- Webinars structured like focus groups in order to develop personas; during the webinars the facilitator publicly posted responses and elicited corrections from the webinar participants on the spot;

---

61 https://sites.google.com/site/biosenseredesign/.

62 However, continuous visits to this website since 2009 indicate that documents are removed from the site on a regular basis. At any point in time, documents older than about one year are no longer available on the website, although they can be requested individually from the BioSense Redesign team. In that event, one does need to know which document is desired, however. CSTE keeps an archives of current and past webinars they have conducted at http://www.cste.org/?page=WebinarLibrary and the BioSense Redesign website offers an “archive” of a partial subset of past and current documents at https://docs.google.com/folder/d/0BxtyovMAxNHINUE0bTmZHpd28/edit?docId=0BxtyovMAxNHlbDlxTUp5ViBqa3M.
• Requirements postings on the BioSense 2.0 Redesign website about specific topics; the postings included open-ended questions, radio buttons, and rankings options to which users could respond over a publicly specified period of time;\textsuperscript{63}
• Videos presented to a pre-selected, advanced panel of users, who then answered questions about the features that the video described; and
• Individualized usability testing at public health conferences, using scenario-based testing techniques along with think-aloud procedures (P-15).

ISDS (2011) delineated a variety of transport paths and protocols that need to be used to feed the data to the lockers which reside within the BioSense 2.0 data warehouse. Protocols must meet the respective states’ and federal legal requirements for transportation of health data over networks. The various transport paths and data providers are shown in Figure 3.

4.2.4.2. Acting on Requirements

None of the BioSense 2.0 partners reported having any particular difficulties or concerns with the actions taken to elicit or to satisfy requirements. Some reported that their jurisdiction (and other jurisdictions of which they were aware) did express initial concerns about BioSense 2.0 residing in an Amazon cloud (P-26), but the design team undertook an aggressive educational campaign to ensure that all parties were provided as much information as possible about the security framework, the cloud set-up, and the ways in which BioSense met current federal government standards. As a result, they did not worry

\textsuperscript{63} Initially, the response rate was typically approximately 15 responses over two weeks. As time went on, the more typical response rate was about 30 responses in the first two days after the posting (P-15).
about security very much because of the long-term trust relationship that has built up between their organizations and the CDC (P-17). Since P-17’s organization has already been submitting data to the CDC electronically for more than five years and has had no data breaches or losses, this individual says that he believes the CDC’s acceptance of the cloud services represents a level of security with which his own organization can feel comfortable (P-17). This individual, an M.D., expressed a lack of knowledge as to the technology or architecture of cloud computing services, making a “guess” that it “will be just a different server with all these different levels of protection” (P-17).
4.2.4.3. Vendor Selection

Early in the vendor selection process the CDC was aware that the redesign of BioSense to allow the level of data sharing they desired was going to be cost prohibitive (P-26) if the system were developed in-house. A key stakeholder had already been exposed to cloud computing outside of the CDC and felt that examining the costs and benefits of moving the entire system to the Cloud was justified. In addition to cost constraints, BioSense stakeholders were primarily concerned with information security, removing ownership from the CDC and placing it directly with the participating jurisdictions, avoiding platform dependency, supporting reporting and analytical needs, and satisfying all legal requirements for health-related data (P-26). The participating jurisdictions were mainly concerned about ownership of data and the protection of privacy within the Cloud (P-14, P-17, P-27). Jurisdictions did not want federal government to co-opt ownership of their data, and had to be assured (through the governance structure and privacy requirements) that they would not lose control of their data (P-15, P-26). Because Amazon was able to meet the financial constraints and the federal compliance and certification requirements, ASTHO chose Amazon.com. However, the decision was also supported by the fact that any cloud solution would significantly lower the costs of ongoing system maintenance, according to P-26.

4.2.5. Stewards’ Perceptions

4.2.5.1. Perceptions of “Records”

Interviewees did not discuss “records” during their interviews. Although asked about records management, interviewees asserted that BioSense 2.0 contains “information” or “data” (as opposed to records). Certainly, the information residing in BioSense 2.0 is not a public record according to any legal definition of the term. However, it does map to the
concept of “recordness” held in ARM environments. ARM personnel consider a record to be any information created or received in the performance of one’s work that provides evidence of a work transaction and can be characterized as having content, context, and structure. This definition holds true for the data residing in a health-information system maintaining data for public syndromic surveillance. In fact, the data is a necessary component of performing such surveillance. Furthermore, long-term preservation of this information would support both history and evidence of activities that were taken as a result of patterns found in the data.

Currently BioSense 2.0 does not preserve its surveillance data over the long-term. Thus, if one desired to piece together activities undertaken during a BioSense-monitored pandemic, one would no longer have the same information that was used at the time of the pandemic to manage it. One could not validate relationships between past activities connected (as cause or as effect) to the pandemic. Likewise, one could not produce a reliable history or evaluation of the data’s use during the pandemic.

Interviewees were medical doctors, epidemiologists, researchers, and IT personnel. No records managers or archivists were interviewed because no records managers or archivists have played a role in the creation, maintenance or governance of BioSense 2.0. (as opposed, perhaps, to the source data feeding the surveillance system). Some jurisdictions keep their aggregated data in their own systems such that the BioSense 2.0 data is either redundant or partially redundant. For example, NC DETECT collects the information from local jurisdictions, cleanses it, and performs analysis on it. Although their data in BioSense is not preserved over the long-term, NC DETECT keeps all this collected data “forever” (P-14) in-house.

---

64 The retention for the BioSense 2.0 data is two years.
4.2.5.2. Perceptions of “Cloud Computing”

As mentioned earlier, one interviewee (P-17) reported that he did not really know what cloud computing is, saying of it, “my impression [is] it will be just a different server with all these different levels of protection.” This individual added, however, that from his perspective it did not really matter what it was because their previously “trustful” relationship with the CDC provided confidence that the CDC’s decisions regarding cloud computing would support their own organization’s requirements. In fact, all interviewees exhibited a similar lack of concern for the technical characteristics of cloud computing.

When P-26 was asked how he would define or describe cloud computing, he described it in terms of his perception of its benefits to the BioSense program:

I think it really depends. I mean, for the purpose, it's cheap storage, that's for sure… it's cheaper storage, scalable infrastructure that, for both storage and processing when you have events like pandemics crop up, you have [an] increase in volume in a short period of time, depending upon the severity of the pandemic or event. So to us it's [being] able [to] scale up high and then back down.

P-26 also pointed out that cloud computing would allow activities and functionality that the CDC could not offer if it continued to host BioSense within its own infrastructure. “Having Hadoop in the environment can really allow us to do computation-heavy algorithms. Now we’re able to process them there, and be able to really benefit. So things that we couldn't do before are really advancing the practice and now it's becoming the new norm … One trend analysis used to take anywhere from half a day to two days. And now it's literally within a fraction of a second you get it” (P-26).

Although this interviewee did not provide a definition of cloud computing, he did reveal that he had worked in a cloud computing environment prior to joining the CDC and had wanted to introduce cloud computing into the CDC for some time but “here at CDC people really want to see that other people have tried it first. And tried it and didn't fail” (P-
At the time, numerous private sector organizations were experimenting with cloud computing, but no federal agency had taken up cloud computing at an entire program level. As a result, the move to the Cloud was able to occur only after it became clear that the costs of maintaining BioSense internally were so high that if the CDC did not move to a cloud environment, the entire BioSense program would have to shut down (P-26). By moving to the Cloud, BioSense evolved from being a program in which virtually all the budget went towards the maintenance of IT systems into a program in which only 5-10 per cent of the budget goes towards the maintenance of IT systems (P-14; P-26; P-27). The excess budget is now able to fund grants to states that enter the BioSense 2.0 collaborative.

4.2.5.3. Perceptions of Roles and Responsibilities

Although people shared their job titles and “mile high” descriptions of their occupations, most did not express any relationship between their other job duties and their work on the BioSense system. On the contrary, as mentioned earlier, roles and responsibilities vis-à-vis BioSense 2.0 were typically expressed as belonging to particular organizations. For example, the CDC provides the funding, ASTHO manages the cloud relationship with Amazon, local jurisdictions are data owners, university partners are data analysts, etc. However, the respondent (P-15) from the BioSense Redesign team did speak directly and personally about her individual role on the BioSense Redesign team. Of course, P-15’s tie to BioSense 2.0 is the only tie (of all the interviewees) in which a majority of an individual’s set of occupational roles and responsibilities is founded on the BioSense Redesign. For all other interviewees, work on BioSense 2.0 is one among many components of their overall professional functions as medical doctors, researchers, and epidemiologists.
4.2.6. Concerns and Perceptions of Risk

Security of data was initially a huge consideration for states, according to all interviewees. However, the project team focused early and heavily on security training and risk analysis, and the initial concern lessened. In fact, two states within the BioSense 2.0 collaborative did “full-blown” security analyses and reported back to the collaborative that their security comparisons showed the BioSense 2.0 system to be “far more secure” than their own state environments (P-26).

This interviewee downplayed the risks of moving data to the Cloud, saying that there “are definitely risks involved, just like any other environment, even if you own your infrastructure, they are there” (P-26). Another (P-27) said, “Any computer can be hacked if it’s on the Internet.” P-27 expressed doubt about the risks, however, adding,

Let’s say that there was a widely known celebrity that some newspaper would make $s if they could crack a story about that person, or car accident involving alcohol; if somebody really had some serious technical skills (best hacker in world), they could potentially hack into any system. If they had someone out there snapping pics, or had a tip and all kinds of information from other sources, they then could deduce (potentially) look at these records and figure out that there was an accident. But the likelihood of all that happening, plus employing someone all these skills, is so infinitesimally small it is almost impossible.

P-27 added that although “authenticity, reliability, integrity, and usability were not explicitly worried about, the design lends itself to a trust factor that these characteristics are, in fact, satisfied.” See did not, however, elaborate how the design does this. She also noted that “authenticity, reliability, integrity, and usability were not explicitly worried about,” but added that the design [of BioSense 2.0] supports trust that these characteristics are, in fact, satisfied (P-27). In support of her trust, this P-27 cited the federal data security certifications and standards with which BioSense complies. She did say that the individuals who were brought in to determine requirements did have “that [curation-related] type of expertise,
because of their backgrounds.” She did not delineate the specific recordkeeping backgrounds of the requirements analysts, however. P-27 did report that some people who worked on the BioSense Redesign had worked in the past in secure, top secret environments where auditing and maintaining integrity of the data were essential. “It [i.e., the skillset] was brought in with the team” (P-27). Thus, this respondent trusted that the individuals that developed the architecture and capabilities of the system considered authenticity, reliability, integrity, and usability of the data because of the nature of their past system-building experience, not because of explicitly developed recordkeeping requirements. The designers did not, however, consider long-term preservation to be a requirement.

For P-17, the trust her organization developed during the BioSense 1.0 arrangement with the CDC aided in the comfort she felt with regards to sharing data in a cloud computing environment, as long as the CDC was still their primary partner. “Since we're already exchanging data with CDC, uh, let me see, we started sending data to BioSense, I believe, in 2007. Since it was already the way we exchanged with CDC, it didn't raise any concerns for us, and we worked with hospitals associations before and we already discussed sharing data with [the] BioSense previous version” (P-17). In addition,

We trust it [CDC] with the cloud, and of course, you know, we set up all this stuff in our state, explained the structure and how the data will be submitted to the cloud … And I think it was enough for us. And then they explained the transport mechanisms and it was assuring … UNC was also involved with more technical expertise, and we didn't have any concerns from our data management team, from our partners, from this side. So all this together – and looking through data users’ agreement with assurance that the data will be safe there, with all the protection and HIPAA rules and everything – it was enough for us (P-17).

P-16, one of the system implementers, commented that in the earliest stages of the BioSense 2.0 Redesign, the main recordkeeping risk that potential partners expressed concern about was that of security. He added,
We were able to address that via … webinars talking about the kind of security processes that we went through and also, we achieved a certification and accreditation from CDC which results in an authority to operate, which is something you need to have in order to make the system live if you’re being funded through the government. That was done back in November and so once we showed – you can only show so much security details before you’re giving away your defenses. But we, I think we convinced the states and others that we were able to meet and exceed a lot of the security requirements. So we passed the certification and accreditation without any problem” (P-16).

In essence, the primary factor that informed most participants’ concepts and acceptance of cloud computing risk was trust, in particular, trust of the CDC as an institutional steward of data.

4.2.7. Interactions between Recordkeeping Stewards

4.2.7.1. Perceptions of Working (Together) in the Cloud

As with the perceptions about requirements definition, the respondents reported largely positive relationships with their peers in the BioSense 2.0 collaborative. Two major concerns arose. The first concern relates to the ownership of data. Local and state facilities were reported to exhibit concern about sharing data with the CDC. This was expressed as a concern not with the CDC as an entity, but rather, with the CDC as a representative of the federal government. In other words, at the same time respondents expressed their trust that the CDC would manage their data in a secure and reliable manner they nonetheless also expressed concerns about allowing the federal government the opportunity to control their data.

The second concern related to collaboration with a much larger number of participants than had previously taken place. For example, P-17 revealed,

Basically, [the] BioSense 2.0 cloud belongs to ASTHO. My understanding [is] that they're managing [the] cloud, but we look at ASTHO as a CDC-grantee… The importance for us - the kind of difficulty for us - was that we had different players when I was looking at [the] data users’ agreement. Because we used to deal with [the] CDC and now CSTE comes into the play, and ASTHO comes into the play. So we
wanted this assurance that we're working under CDC, and that when we worked with data users’ agreement a little bit, we changed their wording…because it was written all about ASTHO and we wanted to make sure that our relations are still the same, that we [are] giving our data to CDC and [that] ASTHO works under CDC as grantee.

The decision makers in this organization did not want to have additional levels between them and the CDC, who provides the grants to them. For them, the concerns were not related to the technology or the data sharing, although this person did report that “the important [thing] for us was, of course, secure receiving of our data” (P-17). Rather, once security of data was shown to hold, the primary concern related to status and organizational structure, as well as to the difficulty of clarifying what all the participants’ roles and responsibilities vis-à-vis governance of the collaborative would be.

Individuals from different organizational types and hierarchical locations within their organizations exhibited different concerns. Members from the state departments of health were concerned about ensuring rapid and appropriate team response (P-17). However, the individuals from local facilities were primarily concerned with the nature of the access rights (P-17, P-26, P-27). In particular, the facilities’ workers did not want other facilities (especially within their regions) to have direct access to their line-item data. For example, facility executives reportedly did not want people from other individual facilities to be able to see their detailed, quality data on things such as number of heart attacks while in treatment or hospital-acquired infections. As a result, negotiators of the DUAs had to be “really conservative [about] what level of access to give” (P-17).

The third primary concern on the part of all users was related to the functionality supported by BioSense 2.0. As reported earlier, BioSense 1.0 would often provide event monitoring, sending (sometimes unwanted) warnings to jurisdictions even though states already monitored and applied their own criteria for event warnings. The jurisdictions wanted
BioSense to actually offer new functionality that was not already available to them. Thus, they wanted to ensure that they were offered both analytical tools for their own syndromic surveillance analyses and that BioSense 2.0 offer data sharing among jurisdictions, with sharing (of aggregate views) being the first and foremost functionality desired (P-14, P-15, P26, P-27).

4.2.7.2. Perceptions of Working (Together) in the Cloud

Although all partners expressed the desire for data sharing between jurisdictions, in practice data sharing has been very limited, according to several respondents (P-26, P-27). CDC and BioSense 2.0 personnel have speculated on why this may be the case. They suggested that lack of sharing occurs due to a culture of tightly held data ownership in the healthcare environment, a concern that individual facility data may in some way allow for quality comparisons between (potentially competing) facilities, concerns that the CDC will gain ownership of their data, and lack of definable positive benefit-cost assessments. The underlying forces driving these concerns appear to boil down to trust and economic pragmatism.

Ironically, the situation that could potentially increase mutual trust would be much wider data-sharing, such that deductive discovery of jurisdictional or facility data would not be possible.\(^{65}\) However, the necessary volume of sharing to allow such a positive network effect has not yet occurred. The CDC and the Association of State and Territorial Health Officials (ASTHO), which acts as a Governing Body for BioSense 2.0, view this as a

---

\(^{65}\) With small amounts of data and only a few recognized jurisdictions, deducing the source of the data and trends in that source’s medical provision are easier. As increasing numbers of jurisdictions enter the partnership, however, determining specific jurisdictional patterns of medical practice becomes more difficult.
particular challenge to the success of the collaborative sharing goal of the project (P-27), but are not yet sure how to achieve the required level of sharing.

On the one hand, in the introductory phase of implementation the BioSense 2.0 governance board had to make it very clear to jurisdictions that they would not lose ownership and control of their data. The DUAs were explicitly crafted to ensure that each jurisdiction would be able to choose whether or not to share their data. Over the long-term, however, jurisdictions have typically chosen not to share their data, thereby circumventing the positive network effects that widespread syndromic surveillance data sharing could allow.

In some cases, the expressed rationale for lack of sharing has been based on a relatively simple cost-benefit analysis. Since most jurisdictions have access to their own syndromic surveillance systems and are primarily interested in local health trends, they have had little incentive to use other jurisdictions’ data for analysis. In addition, P-14 reported that most of the useful data for their purposes comes from jurisdictions that are relatively geographically close to their own; when such data is needed, they all they need to do is pick up the telephone and call their counterpart in the other jurisdiction to discover syndromic surveillance trends.

4.2.7.3. Perceptions of Changes Brought by the Cloud

Interviewees revealed three primary areas of interest when asked about changes they have faced as a result of having moved BioSense to a cloud computing environment: (1) changing organizational status (with respect to the CDC) as a result of the new collaboration structure, (2) ensuring that jurisdictional ownership of information remained within the hands of the local jurisdictions, and (3) reductions in cost. As mentioned, members of facilities and personnel taking part in BioSense 2.0 have had to negotiate their changing place within the
new collaborative structure. The Public Health Department respondent and others noted that a number of people were concerned about what the addition of new collaborators within the BioSense project would mean with respect to their primary funding agent, the CDC. They wanted to make sure that they still would have the same status with respect to funding decisions as they had before. The BioSense 2.0 partners also relied upon their trust in the CDC to accept the new technological arrangement. A major topic of negotiations revolved around ensuring that the various partners continued to own and control their jurisdiction’s health data and that the federal government would have no ability to “take over” that control. Although maintaining their desired level of confidentiality and privacy certainly played a large role in their willingness to engage in the collaboration, once it became clear that they were entirely in charge of how much data sharing they would do, these concerns lessened considerably. In other words, rather than their concerns with privacy and confidentiality being primarily based on technological security concerns, the local jurisdictions were more concerned about maintaining ownership and control over their information. Technological security concerns were reduced by a variety of educational techniques. The ownership and control concerns had to be handled by a combination of in-depth negotiation and contracting to ensure that ownership was maintained at the local level.

Finally, the third important factor in the BioSense 2.0 arrangement was cost. The CDC respondent indicated that it was essential to move to a cloud computing environment in order to keep the BioSense syndromic surveillance system alive. The nature of the final design of this system was strongly determined by the recognition that the real value of the BioSense system lay less in providing surveillance outcome information and warnings and more in the value of providing a high-level view of the information environment as a whole.
The information sharing capabilities had to take priority over the provision of surveillance outcome information.

4.2.8. Synopsis of Case 2 Findings

Several themes predominated in the interviews and analysis of this case. One is that professional identity appears to play a relatively small role in the overall perceptions of cloud computing on the part of the BioSense 2.0 respondents. Rather, a group of individuals whose primary job roles are largely independent of this cloud adoption view cloud computing primarily as an objectively available “tool” that can be used to provide analytical services and data they can use in their “real” jobs, rather than as a potential source of loss of professional identity and power. However, jurisdictional ownership and control of information still remains a strong theme as well. This concern relates to larger organizational and jurisdictional ownership rather than particular job-related ownership. Respondents most concerned about cloud computing were not so much worried about the technological risks it presents but rather, with the capacity for the new collaboration to allow federal government interests to take over local jurisdictional ownership and control of healthcare data. Jurisdictions’ decision makers had to engage in a relatively complicated negotiation process for the DUAs to be accepted at the local level. The participants in these negotiations needed to ensure that the CDC would not be able to co-opt ownership and control of local health data. In addition, external healthcare associations had to be brought in to ensure the local jurisdictions would really believe the CDC on this issue, even though they largely accepted the CDC’s technical assessments of cloud computing risks.
4.3. Case 3: Agency-Specific Cloud Implementation

4.3.1. Introduction

Over a single weekend in 2010 the Kentucky Department of Education (KDE), in conjunction with Microsoft, Inc., moved more than 700,000 students, faculty, and staff from 174 distributed Microsoft Exchange Server 2003 on-premise servers to Microsoft Live@edu (P-20) (Microsoft 2010). Microsoft Live@edu is an email and collaboration service that educational institutions throughout the world can implement at no charge (Microsoft Live@edu 2013a). It offers calendars, document sharing, shared workspaces, blogs, instant messaging, video chat, mobile e-mail and document access, and address book capabilities. It also allows students to use web-based Microsoft Office applications to create documents. All services are offered free to educational institutions and students. It also offers access to the parents of students and to alumni. This cloud-based service is funded through “family-safe” banner advertisements on the parent and alumni pages, although these third party advertisements are not included in the students’ applications and services (Microsoft Live@edu).

Five Kentucky recordkeeping stewards, including individuals from the Department of Libraries and Archives, the Office of Enterprise and Technology, the Department of Education, and a large records center in a local school district discussed their perceptions of the rationale for and effects of this implementation, and shared information about their roles and responsibilities. I made several attempts to initiate an interview with the Department of Education’s records management contact that is in charge of “Physical Resources,” according to the Department of Education’s latest organizational chart. However, this person was
unavailable for comment.\footnote{Near the end of the interview process, P-28 revealed that the KDE’s records officer responsibilities were changing during the implementation timeframe and a new records officer was being brought on board.} With the exception of one technical staff member and one local archivist/records management supervisor, the participating interviewees reside at executive levels within the Commonwealth of Kentucky. Interview responses were supplemented and validated by internal documentation, internal change management materials, legislative statutes, and literature published on the Web to come to an understanding of the recordkeeping environment that operates within this cloud computing service arrangement.

The Microsoft Live@edu implementation has not been the only cloud adoption undertaken by the KDE. For example, KDE has used Infinite Campus’ State Edition to manage and share data on a single, uniform system across all of the 174 districts in the state since 2006 (P-20; Infinite Campus 2013a, 2013b). Infinite Campus is used as the source of record for student data across all school districts in Kentucky.\footnote{Infinite Campus markets itself as allowing uniform reporting and roll-up of information, and reducing administrative costs while widening administrative capabilities. Many small districts are unable to afford a data system of their own. By allowing the more affluent districts and Infinite Campus to share server resources KDE now has the resources of a small statewide server farm, allowing even the small districts which could otherwise not take part in a uniform data-sharing system reportedly to receive the same level of service as the more affluent districts. (Infinite Campus). Since 2011, the Individual Learning Plans have also been available for home school, private school, and adult students in Kentucky (KY HEAA 2011).} In 2006 KDE also implemented Career Cruising’s cloud-based Individual Learning Plan (ILP) software, which is required for students from grade 6 onwards to “explore career possibilities, set goals and track their own educational progress against those goals” (KIDS 2012, iii).\footnote{The learning plan is fully integrated with Career Cruising’s career exploration, planning, and guiding tool and is available via the Web (http://public.careercruising.com/us/en).} In addition, since 2011 KDE has used the Continuous Instructional Improvement Technology System (CIITS), another web-based SaaS application. CIITS pulls “standards, instructional materials, lesson plans, assessments, data and professional development all together into an integrated...
The CIITS online resource” (https://ciits.wikispaces.com/file/view/What+is+CIITS.pdf).69 The CIITS implementation made Kentucky the home of the largest K-12, cloud-based financial management system in the United States (Tyler Technologies 2013). Lastly, the financial software package used in all 174 Kentucky school districts in the state - MUNIS (http://www.tylertech.com/solutions-products/munis-product-suite/munis-saas) – was migrated to a cloud-based SaaS service in 2012.

4.3.1.1. The Decision to Move to the Cloud

During interviews, a Department of Education executive manager (P-20) gave strategic reasons for the agency’s widespread movement to the Cloud. Specifically, this person cited Nicholas Carr’s (2008) *The Big Switch: Rewiring the World, From Edison to Google*. P-20 asserted that the age of “Big Iron,” in which state “IT shops” build, own, and manage all their own infrastructure is no longer financially feasible, not if one simultaneously wishes to provide the same levels of service that one’s clients can receive by turning to external IT service providers. P-20 also suggested that because of resource shortfalls, no IT personnel had to be let go due to the adoption. Rather, newly freed-up IT human resources were moved to different service-oriented positions within the Office of Knowledge & Information and Data Services, or KIDS, as it is called. According to P-20, those who wanted to continue “making widgets” were told that if that was the type of job they really wanted to have, they should consider moving to cloud service providers because that is where that type of skillset would be needed in the future.

---

69 The system was adopted in response to Kentucky’s 2009 Senate Bill 1, which specified that the state had to implement centralized educational standards (KDE 2013). It provides Kentucky educators access to a wide variety of core standards in the Commonwealth (https://ciits.wikispaces.com/Background).
While discussing the agency decision makers’ choice to move to the Cloud, P-20 explicitly criticized the long-term strategic direction that COT takes with regard to information architecture and cloud computing. The criticism appears to be ideological and this appearance is probably based on the fact that the political separation of KDE from the control of COT is already assured by virtue of Kentucky’s organization structure. However, it is not possible to say whether factors other than this ideological agreement affect the relationship between the KDE and COT. One report suggests the possibility of an ongoing communication disconnect; P-29 reported, “Sometimes you get the idea that communication isn’t always the best between KDE and COT.”

Other respondents, who reported that the agencies in general are much more enthusiastic about the Cloud than the central IT agency is (P-6; P-18; P-28), suggested that cost drivers for the Department of Education’s implementations played a strong role in the KDE’s adoption of Microsoft Live@edu. A Commonwealth Office of Technology (COT) IT executive (P-18) remarked that at about the time the Department of Education was considering the Cloud, virtually all of the 174 school districts had their own hardware. P-18 added that although the Board of Education itself has a Cabinet seat and is provided centralized IT services by the Office of Enterprise Technology, the KDE (the Board’s agency that handles P-12 activities) - “is considered really outside of our realm.” Furthermore, all of the 174 school districts were facing imminent hardware failure at approximately the same time and needed a large amount of resources and support. At that point, Associate Education

---

70 COT is the centralized IT service provided for state agencies.

71 The Education Cabinet reports to an Education Commission, who in turn reports to the 11-member Board of Education (KDE 2013). The KDE has its own IT department, whereas the other agencies reporting to the Board of Education use the state’s centralized IT unit, COT (P-18).
Commissioner David Couch was able to work out a deal with Microsoft that would allow Kentucky to use Live@edu at no cost, providing Microsoft its first large-scale K-12 account. P-6 commented, “COT for years has been charging outrageous prices for storage, so naturally state agencies are going to look at these cloud computing applications and say, ‘Well we can save a whole bunch of money, particularly in this cost cutting era we’re now in if we go to cloud computing.’”

4.3.2. Recordkeeping Stewards

4.3.2.1. Office of Knowledge & Information and Data Services (KIDS)

KDE’s Office of Knowledge & Information and Data Services (KIDS), manages the information technology infrastructure, hardware, and services for the KDE (P-29). Decisions about statewide infrastructure, applications, and services are handled centrally by KIDS. Because many of the district and state records are electronic, KIDS plays a strong role in ensuring that state and local personnel can manage their records successfully. It provides assistance for applications and manages much of the infrastructure. KIDS was responsible for the decision making that led the KDE to implement not only the Microsoft Live@edu cloud service, but also the other statewide cloud services in place today.

4.3.2.2. Records Officers

Kentucky law requires that each agency appoint an individual to act as a Records Officer, who must “represent his unit of government in its relations with the Division of Archives and Records” (Ky. Admin. Regs. 725:3:045, 2013). The Records Officer helps the

---

72 Since then, Microsoft has developed a policy by which it offers Live@edu for no charge, earning revenue through banner advertisement to parents, alumni, and educators (but not students).
Division of Public Records to inventory, analyze, and schedule the disposition of his or her agency, and keeps a record of the destruction of records.

4.3.2.3. Kentucky Commonwealth Office of Technology (COT)

What is now called the Kentucky Commonwealth Office of Technology (COT) began operations in 1973 under the name Bureau of Computer Services (BCS) (KY COT 2010). Since then, several IT reorganization efforts have occurred, during which COT, which went through several name changes, moved to a more centralized governance structure and then back again to a decentralized structure. In late 2012, Governor Steve Beshear put into effect Executive Order 2004-880, which placed operational control of “all executive branch information technology infrastructure services” (Beshear 2012) under COT, which resides in the Finance and Administration Cabinet of the state. In other words, all systems and services on which many KDE records are created, maintained, or preserved are currently controlled by COT. 73 The CIO is now a member of the Governor’s Executive Cabinet (Beshear). COT describes itself as “a steward of the state’s vast volumes of data” which “understands it has the responsibility to assure its citizens and business entities that their data is safe and secure with the Commonwealth” (KY COT 2013b).

4.3.2.4. Kentucky State Archives and Records Commission

The State Archives and Records Commission determines the final disposition of the state’s public records (Ky. Rev. Stat. § 171.420, (2012)), which are defined to be “all books, papers, maps, photographs, cards, tapes, discs, diskettes, recordings, software, or other

73 Although this executive order holds for all other executive branch agencies, it specifically exempts a number of state executive branch entities, including the nine postsecondary education institutions and those Department of Education’s services provided to local school districts (Beshear 2012, Section V).
documentation regardless of physical form or characteristics, which are prepared, owned, used, in the possession of or retained by a public agency” (Ky. Rev. Stat. § 61.870, (2009). The Commission, composed of fifteen individuals, advises KDLA on ARM issues and also has the authority to review and approve records retention and destruction schedules.

4.3.2.5. Kentucky Department for Libraries and Archives (KDLA)
The KDLA\textsuperscript{74} manages all activities associated with the inventoring, scheduling, and disposition of records throughout the state. It creates policies and publishes materials that provide detailed information to state public offices about their records management and archiving responsibilities, and provides funding (through the state legislature) for a number of initiatives and projects related to these responsibilities.

4.3.2.6. Public School System Archives & Records Units
The P-12 public school districts in Kentucky must abide by the rules set out in the Public School District Records Retention Schedule (KDLA 2012a). The school districts’ records management activities are structured in various ways across the state. They may or may not have a dedicated archives and records management center; they may or may not have dedicated records managers; they may or may not have records officers who are trained in ARM responsibilities and tasks. The retention schedule provides the retention and disposition requirements for all records, both student record-related and administrative-record related.

\textsuperscript{74} KRS 171.130 (1982) specifies the establishment of the KDLA, although this statute is an update and affirmation of the original 1954 law that created the KDLA (Ky. Acts. 41:1 (1954)).
4.3.3. Legal Environment Affecting Stewards

Like other states, Kentucky provides its own legal definition of “public record.”

Public records are – “all books, papers, maps, photographs, cards, tapes, disks, diskettes, recordings and other documentary materials, regardless of physical form or characteristics, which are prepared, owned, used, in the possession of or retained by a public agency” (KRS 171.410(1)). Public records include “emails, databases, and other records electronically generated and/or stored,” as well as “public agency records that are not maintained on the agency’s premises” (5). Also like other states, Kentucky has an open records law, which establishes a right of access for citizens to public records, with several exemptions that specify which public records are not classified as “open,” and therefore not available for public access. Because electronic records are also included in the definition of a public record, this law holds for both hard copy and electronic records.

Kentucky has enacted a relatively wide body of statutes and regulations that detail the roles and responsibilities of various occupational entities in the management and preservation of electronic records, as shown in Appendix H. Although there are a relatively large number of entities that have official recordkeeping roles and responsibilities, those roles and responsibilities are rather precisely delineated in Kentucky’s laws. Even KDLA’s collection policy is mandated by statute (P-28). The key recordkeeping requirement that Kentucky relies on, however, is the statutory requirement that every employee in Kentucky who handles public records is responsible for learning and following the retention rules associated with those records. The KDLA provides a large amount of explanatory material which is sent to public agencies and posted to the public web for Kentucky workers.

One might think that such an approach would lend itself to a rather distributed, perhaps even democratically-structured, recordkeeping structure. However, this does not
appear to actually occur. A number of political and jurisdictional struggles and concerns
trump the written law and individual responsibilities for recordkeeping. That is not to say that
appropriate recordkeeping activities are not occurring at all. Rather, it is meant only to point
out that there are areas in which certain activities associated with retention are not only *not*
occurring in the Cloud, but the cloud computing adoption appears to provide actual
incentives to avoid these activities.

The information governance within Kentucky, even prior to the recent re-
consolidation of COT, is relatively centralized. While, the designated roles and
responsibilities are spread out among a wide variety of occupational roles, these roles are
delineated in detail within Kentucky statute; rights and responsibilities are also spelled out in
detail through statute. In addition, the KDLA, which is an agency within the executive
branch, is given responsibility for training public employees and records officers and for
distributing information about records management and archiving requirements across the
executive branch. In addition, the State Archives and Records Commission (SARC) has the
authority to make the final decision on matters of retention and disposition.

4.3.4. Requirements and Actions
   4.3.4.1. Defining the Recordkeeping Requirements
           4.3.4.1.1. Recordkeeping Requirements as Seen by COT
               COT played no role in the requirements assessment for KDE’s adoption of Microsoft
Live@edu, as would be expected given the KDE’s “external” placement within the statewide
governance structure.
4.3.4.2. Recordkeeping Requirements as Seen by the Department of Education

One of the two key respondents from the Department of Education was, in fact, an executive level individual. (The other was a records manager or archivist.75) Thus, perceptions about the requirements analysis are colored by that perspective. More detailed explanation is provided below in Section 4.3.6.2, but in essence, key benefits offered were reduced cost, increased services, increased service availability, and reduced risk of data loss.

4.3.4.3. Recordkeeping Requirements as Seen by Local Archivists/Records Managers

The local ARM personnel played no role in the requirements analysis or the decision to adopt the Microsoft service (P-29). Although records manager P-29 remarked that it was possible that the local district Management Information Services (MIS) personnel had some knowledge that the cloud adoption would occur, when Microsoft Live@edu was adopted, P-29 only learned about it “after the fact.”

To some extent, the relatively small knowledge that records officers, records managers, and archivists have about the various KDE cloud implementations occurs because they work with electronic records in only a “very small way” (P-29), adding that “Things have not moved to technical [sic] at such a pace as everyone thinks” (P-29). For the most part, local district recordkeepers spend much of their energies in a hard copy world.

---

75 I am being specifically vague about occupational role here because (a) only a small proportion of recordkeeping personnel in the local school districts have received either records management or archival training. By being vague I am better able to maintain anonymity for this individual; and (b) for those who play a recordkeeping role in local districts, the duties between “archivist” and “records manager” tend to be blurred, with personnel often engaging in both activities simultaneously.
4.3.5. Stewards’ Perceptions

4.3.5.1. Perceptions of “Records”

The KIDS executive respondent did not specifically define the term “record” but in all remarks used the term in a manner consistent with Kentucky’s legal definition of a public record, provided previously. P-20 also appeared to understand what retention requirements are, although this person did make additional comments about retention schedules that appeared to suggest either a misunderstanding about the nature of preservation activities in electronic environments or a lack of concern for preservation at all. This will also be discussed in Section 4.3.6.2.

Both KDLA and the local ARM professional discussed records in a manner entirely consistent with ARM usage: they did not define the term specifically but described the necessary activities associated with creating, scheduling, and preserving records and showed a sophisticated understanding of both the technological and organizational difficulties with managing electronic records in a preservation environment.

The executive level COT respondent likewise spoke of records consistent with the use that state government archivists and records managers typically speak of them – that is, according to the legal definition of the term and with an apparent concern for maintaining the context, content, and structure of the record over time. Perhaps one reason such relative consensus exists in this context is that all three of these individuals – the KDLA interviewee, the local records person, and the COT respondent – sit on the same frequently-meeting committee together. In addition, KDLA has rather close ties with COT, having both their personnel frequently engaging in joint programs (P-6).

76 The local respondent was formally trained in archives and records management in a university graduate-level LIS program. In addition, prior to joining the local district this person worked for almost twenty years as an archivist.
4.3.5.2. Perceptions of “Cloud Computing”

The respondents had very little to say about the definition or description of cloud computing. One interviewee (P-6) spoke about the NIST definition of cloud computing and, like NIST, described cloud computing in terms of its potential benefits. For example, P-6 described the Cloud in terms of services that could be beneficial to an organization that would see an advantage to “pooling a lot of their resources and using large amounts of storage that’s fairly flexible to plug in and out.” This respondent remarked that such a description “goes beyond the current scenario in the state.”

A senior KIDS executive (P-18) also used NIST’s definition of cloud computing, which he described as “Software-as-a-Service, Platform-as-a-Service, and Infrastructure-as-a-Service” (P-18). P-18 did state, however, that “I don’t really have a problem with a private or a hybrid [cloud] as long as there’s that segregation [of data] available,” noting that the KDE’s internal email cloud is part of the “big ‘edu’ cloud. Another respondent remarked, however, that a number of agencies are eager to move to cloud computing, primarily for cost reasons, since COT charges “outrageous prices for storage,” and is, in fact, unable to present prices that are competitive with outside vendor charges (P-6). P-28 explained that much of the reason that this is the case is that the services agencies receive are not identical to those that they would receive through COT – they may not have as much security or privacy, for example: “Because the agency personnel frequently do not understand the nuances of levels of security, interoperability, data ownership and location, storage jurisdiction, all they knew was [is], again, the buzzword of the Cloud was going to [will] save them phenomenal amounts of money.” If the vendor were required to meet all the necessary security and privacy regulations, they would not be that price competitive, according to P-28.”
4.3.6. Concerns and Perceptions of Risk
The knowledge of risks to cloud computing was expressed unevenly. Individuals in COT and the KDLA expressed the risks directly and clearly, speaking both of potential risks to security and to a lack of risk assessment occurring in the agencies. For example, P-18 said, “in my experience it's been the business side folks that don't understand the potential risks … they hadn't thought any of it through … all they knew was, again, the buzzword of “the Cloud” was going to save them phenomenal amounts of money.” This individual pointed to a then-recently published COT white paper (Thomas 2012) that highlighted the risks of the technology, including the legal risks of delegating responsibility for data to an outside party, concerns about confidentiality breaches, data ownership and mobility concerns, sanitization of IT equipment, the need for customization of applications, and security holes. P-18 added, “One of the other problems that I have with cloud, and I think some people are starting to realize this now, is really all you’ve done is taken your siloes of data and put them to an area where you don’t even know where they are anymore.”

4.3.7. Interaction between Recordkeeping Stewards

4.3.7.1. Perceptions of Working (Together) in the Cloud
P-6 suggested that there is “a state of combat” between state IT personnel and individuals within state agencies trying to implement cloud computing. Yet another respondent (P-29) remarked on a lack of good communication between COT and the KDE, and a KDE employee asserted that the COT people do not agree with the KDE’s public stance toward the Cloud, and vice versa (P-20). For the KDE, much of the disagreement derives from the executive-level belief that “Big Iron” is (and should be) a thing of the past (P-20). P-20 asserted, “I look at the rest of state government here, and it's really archaic…”
They're building big systems, they’re writing big systems with the personnel they have on-site and I really think that is a really archaic view of it.” The COT executive manager, on the other hand, expressed a concern about Cloud risks (P-18), as mentioned previously. At a deeper level, however, other reasons may also exist for COT management’s reluctance to have agencies moving to the Cloud. For example, P-20 revealed that COT is a “zero-based” organization, meaning that the only funding COT receives is from providing services to the state agencies. It receives no legislative funding, according to P-20. COT has already centralized the state’s own email services via an in-house, private cloud through which COT provides all email to the executive branch agencies that are constitutional agencies, charging them to recover the cost of service. From the point-of-view of COT management, much of the disagreement between agency points-of-view and the COT point-of-view is a result of the granularity at which one thinks of organizational missions and requirements (P-18). “From their [i.e., agency personnel’s] perspective, the enterprise is their organization. From my perspective, the enterprise is state government and the executive branch, which is bigger than any one, individual agency. So, while something may be beneficial to just them, if it goes against the grain for the other ten organizations, I have a hard time trying to approve that” (P-18). This suggests that when evaluating costs and benefits, COT personnel, whose clients are all the executive branch offices, have the incentive to balance the various costs and benefits of individual agencies against each other. The agencies, however, do not face the same incentives, since they typically view their clients to be Kentucky citizens in

77 By “it” this interviewee was referring to conceptions about how the state should engage in infrastructure purchases and IT services management.

78 Constitutional agencies are those that have a senior official elected in constitutionally mandated elections. These agencies, as part of the executive branch, receive IT services from COT. The KDE, although appointed by the state board since reforms in 1990, nonetheless does not answer to the governor, as would most appointed offices (P-29, P-18).
general and sometimes, the Governor (through individual directives he gives to them). The desires of these “clients” are often only observable to agency personnel through second-hand or even third-hand communication channels from upper management or through interpretation of the governor’s directives by COT, KIDS, or their own managers. In other words, agencies that use external cloud computing services realize the immediate benefits of cloud computing, in terms of per unit prices for computing services, while they do not necessarily factor in the intangible costs associated with the numerous risks of cloud computing. COT factors these risk elements in to their service provision estimates, making the net financial benefits appear much smaller to these agencies. Although the COT estimate is possibly more accurate in terms of the overall “social costs” of cloud computing to the state as a whole, there is currently no way to force agencies to “privatize” those social costs by adding in the overall risk-based estimates to their own cost-benefit analyses or estimates. Thus, the tension between the agencies and COT is often misinterpreted as merely a political or ideological difference of opinion. Even when it is recognized as being largely economic in nature, the state has as yet been unable to provide any economic mechanisms that will allow the disagreement to be resolved through voluntary decision making under well-defined rules or practices that provide agency personnel the incentives to follow voluntarily. Rather, COT has had to present regulations, policies, and mandates that increase agency costs in other

79 I say “possibly” because the risk-stance of COT also must be taken into account, and because COT has an incentive to overstate the costs they associate with risk. If an organization is more risk-averse, they have a tendency to be less inclined to accept risks that more risk-neutral or risk-taking entities would believe to be worth undertaking. It is not clear in this case whether COT or individual agencies have different stances toward risk; a quantitative test or statistical analysis that includes actual historical costs and benefits and information regarding the desired outcome on the part of all parties would be necessary to determine this. However, the presence of an incentive to avoid any particular outcome that makes agencies use their own IT services instead of COT’s services can be asserted: if an agency chooses an external service over COT’s service, all other things being equal, COT will face a revenue decrease. Since it only receives funding through these services, COT will see external services as a reduction in COT’s own ability to support its overall existence.
ways when the agencies want to undertake non-standard IT implementations. In the case of
cloud computing, COT has worked through the Enterprise Architecture and Standards
Committee (EASC) to require that agencies use COT standard software and systems (P-6). In
order to adopt software that is not an enterprise standard, an agency is supposed to present a
case to the EASC, requesting an exemption (P-6; P-18). The EASC does “let a few
exceptions go through as long as we were engaged in the drafting of the RFP or the contract”
but the agencies nonetheless must go through the extra step of requesting an exemption. The
additional step not only provides the EASC with some oversight of IT services but also
increases the cost of adopting non-standard technologies by requiring an extra layer of
decision making (and time) to occur before implementing a new technology. By increasing
these implementation-oriented costs, agencies are likely to reduce their implementation of
any non-COT provided IT service or tool, including cloud-related services and tools.

The political cycle, however, makes agencies more skeptical of the potential long-
term ability of COT to provide cost-efficient services in the future. P-28 revealed that one
factor influencing all long-term IT decision-making in the state has historically been the
question of what types of information stance COT will take when a new political regime
elected:

The legislature meets January first and does two-year budgets. The budget session is
coming up this year, so it’s possible the legislature could undo any or all of this…We
change governors in three years, so we could get a new governor who could change
everything again. In the late 90s they changed the law so a governor could serve two
terms. While a lot of the administration would be the same, initiatives would change.
So in ‘97 the governor elected served two consecutive terms. Then we changed
governors and went a complete different direction, which completely changed things.
He served one term; then we elected a democratic governor.

Thus, although the current governor and legislature is very much consolidation-oriented, IT
policy and organization could well change with the next election or budget session. Planning
IT strategy and enacting efficient policies becomes remarkably difficult in such a political environment.

In fact, even with COT standards in place, political factors still continue to play a role in some cloud adoptions and in relationships between agencies and COT. For example, some agencies have “worked around the official system” (P-6), going “behind the back” (P-6) of COT and the EASC\textsuperscript{80} by going straight to the CIO for permission, bypassing the EASC entirely. In fact, one can hypothesize that the current rush to adopt agency-level cloud computing, although primarily based upon obvious and sometimes large cost reductions for agencies, provides the additional benefit of allowing the adopting agencies to circumvent the current trend of state government IT consolidation. Speaking of both the IT consolidation and their planned cloud implementation, a respondent from one agency remarked, “The consolidation has pushed us to look at more cloud-based options just to get around the red-tape atmosphere” (P-28). In fact, for those agencies that have a lot of non-standard or open source software, which COT does not support, the new consolidated environment provides them more freedom to manage their information as they please. Adopting a non-COT service buffers them from the potential upheavals caused by large shifts in IT policy when new political regimes with different views toward IT centralization are voted into office.

The foundational issue with centrally-mandated IT standards and regulations appears to be an issue of control. Who has control and ownership of the agency’s data? “Cloud computing is all about control, who has control of the records. The state IT people don’t want to release control of the records. That is their ultimate problem. We can hire an outside vendor cheaper than we can hire them” (P-28). Although a COT respondent (P-18) agreed

\textsuperscript{80} P-6 asserted that COT is the “strongest group within the Enterprise Architectural Standards [Committee], and sometimes what they say goes.”
that the fundamental issue is one of control, arguing that being able to gain backend control and interoperability is a crucial concern when it comes to consolidation decisions (P-18), she also stressed that when it comes to cloud computing, most agencies look only at costs and do not consider the many potential risks to data that cloud computing engenders. Non-COT respondent P-6 stated it baldly, saying that COT argues that agencies are losing control of their records if they go to cloud computing because “there’s less security, there’s less accountability” in the Cloud. Although P-6 agrees with COT’s assessment, P-6 also added, however, that the problem is that there are a number of benefits to agencies from cloud adoptions such as “making sure that our archives has copies that are distributed geographically.” P-6 agrees with COT that careful negotiation with the vendor is necessary to mitigate the risks of cloud computing. In fact, at the time of the interviews, KDLA was trying to establish guidelines with respect to recordkeeping, audit, and records management in the Cloud (P-6). Since that time, the KDLA has published a set of guidelines for performing records management in the Cloud (KDLA 2012b).

Another political factor that places pressure on COT is that the Governor’s office and the legislature seems to be placing pressure on COT to “push the latest technology” by stressing the need to adopt “more innovative technology, including cloud computing” (P-6). In addition, some agencies approach COT and suggest that their personal relationship with “this legislator” or “that legislator” will lead to problems for COT if they do not allow the agency to adopt their project. COT must step relatively lightly in situations such as this (P-6). The consequence is that COT is being squeezed at both ends – from above by the Governor and legislature that control its funding and authority and from below by the individual agencies that want to maintain as much independence in IT decision making as they can.
Within KDE, however, another political struggle has occurred during cloud computing adoptions. For example, P-29 discussed the implementation of the cloud version of MUNIS, the KDE’s financial management system. P-29 noted that a “public tug-of-war” took place when KIDS implemented the cloud version of the software. At that time, KIDS argued that a major benefit of that adoption was that the disparate systems then in place across the different school districts were replaced en masse with a single statewide system. P-20 noted that the centralization not only saved money, but it also allowed greater control over the information. Some of the school districts, especially the larger districts with larger funding streams, had already adopted financial systems that met their needs very well (P-29), however. A tug-of-war ten commenced because these districts’ personnel recognized that their service qualities would decline, at least in the short-run, if they had to adopt a new system that was fully standardized across the state. Many of the unique needs that these districts faced had led them to customize their own systems in a manner that provided what to them was the greatest efficiency and effectiveness. With the new cloud-based system, they would no longer be able to perform all the financial activities (P-29) they could perform on their customized systems. In the end, the state had to mandate the adoption of MUNIS (P-29).

Once again we see a jurisdictional issue arising between a stronger political entity and a weaker entity residing under the stronger entity’s authority. Disagreements about ownership of information systems and rights to control these systems take on a political bent, stressing the already-existing power dynamics between these political entities. One sees a broad series of disputes between the state as represented by the governor and legislature, as represented by the centralized IT group (COT), and as represented by the individual
agencies. Within the single agency KDE, there have been contests between individual districts and KIDS (the KDE’s centralized IT unit) as a whole, once again with the governor’s office and legislature playing a role by supporting the KDE’s efforts to consolidate further their own IT unit’s control.

The local records officers, records managers, and archivists, however, reported that they really were not affected by the state’s overall IT consolidation and P-29 remarked that the Microsoft Live@edu adoption had until recently had very little, if any impact on their practices (P-29). Furthermore, although local archivists, records managers, and the KDLA have attempted to educate local personnel about the importance of following records retention schedules and the need to recognize that email is a record, very little attention has at this point been paid to email retention at the local level. Prior to moving to Microsoft Live@edu, KDE centrally controlled the email through the use of Outlook Web Access. As a result, local personnel have habitually considered all email issues to be “KDE issues” to the extent that, at one point, a citizen request to a local district for (local school district) records was denied, not because the records were private or protected, but rather, because the local district employees believed that such a request should go elsewhere: “That’s not us” (P-29). They were later found to have violated state access requirements by the State Attorney General’s office, but the general feeling among many local district personnel still remains that email is “not theirs” (P-29).

This is another example of confusion about the ownership of information rights and responsibilities. KDE gave the authority to control the email environment to KIDS, which also has the technical capabilities to manage that environment. For local personnel, they have neither the capability nor the technical capacity to handle many email “issues” that arise. As
a result, they do not think they are responsible for the retention and disposition of emails, despite what written statute says.

In fact, the confusion about who is supposed to do what does not reside entirely in local districts. It is a common situation after a cloud adoption. According to P-28, when KDLA adopted Tessella’s cloud-based Preservica, a lot of confusion reigned about who would take care of various implementation issues. “Nobody really knows what’s going on because they are still in the process of doing assessments. We have an electronic records committee…We talked to them; even the COT people admit that at this point they really don’t know what’s going on because they are still absorbing things.” This problem is confounded by the IT consolidation, leading to some dissatisfaction among agencies (P-29). P-28 referred to the IT consolidation as a situation in which IT personnel in agencies have been “sucked up” by COT. P-6 jokingly but colorfully described it as “ripping off an agency’s IT, putting them in COT, and then charging back to the agency the services of people who were once in their cabinet!” In fact, P-6 noted that when KDLA partners with COT, agencies tend to think that, like COT, KDLA will charge them for their services.

4.3.7.2. Perceptions of Changes Brought by the Cloud

The majority of changes due to the implementation that employees reported were not actually caused by the adoption of cloud computing. That is, many of the changes respondents cited appear in fact to be due to organizational or political factors that pre-date the cloud adoption. Some of these factors, such as tensions between agencies and COT regarding the adoption of cloud technologies, would probably be exacerbated by any technology change that has similar characteristics to those causing the tension. For example, if agencies began to want to engage in outsourcing arrangements en masse, COT would
probably be concerned about the potential security problems of allowing IT services to be managed by outside parties. To some extent, these disagreements are more strongly related to ideas of process (and IT systems) ownership and to attitudes towards “internal” versus “external” management of information than they are to cloud computing per se. The implementation of cloud computing however, serves as a legitimating platform for those arguing that services should be managed centrally by COT. Because the Cloud has indeed been marketed as having greater risk than internal storage scenarios in some areas of privacy and security, COT can bring the risks to the forefront, thereby masking the extent to which other potential concerns control their support (or lack of support) for the technology.

Nonetheless, some stated concerns with cloud computing do appear to be associated with cloud computing itself. For example, P-29 stated, “Records in the Cloud are not being destroyed according to schedules.” Although some of this is due to the fact that recordkeeping personnel have found that it’s “been hard to get people to pay attention to retention management of electronic records” (P-6), cloud computing does indeed reduce incentives to take responsibility for records disposition. In the first place, because the storage of information has suddenly become much less expensive, the monetary costs of retaining records past their retention dates appears to be virtually nonexistent to district personnel (P-18; P-20). With no one breathing down employees’ backs about servers filling up with unnecessary data, they find it easier to put activities such as disposal on the back burner, while more pressing and immediate service-provision takes the front burner.

This occurs in spite of the fact that Kentucky state public offices are subject to the requirements of the DoD 5015.02 standard (P-6). One would hope that in the event of a new adoption providing disincentives for following retention schedules, the existence of a
standard like DoD 5015.02 would help reduce the incidence of noncompliance by virtue of its strict and clearly-stated recordkeeping requirements. However, with 5015.02 that is not really possible because “nobody pays attention to that and has a tough time doing it” (P-6). In other words, the standard is so strict that few, if any, agencies have the resources available to meet the standard anyway and therefore have a strong incentive to just ignore it, even though “it’s actually in a statute that they have to do that” (P-6).

For the KIDS executive respondent, moving to the Cloud had no effect on recordkeeping roles and responsibilities. When asked whether or not the KDE had an electronic recordkeeping system, P-20 replied, “I wouldn’t buy all that,” adding, I tell people this - because they all get involved in retention: “How long do you have to retain it on paper?” Just because they're electronic doesn't mean it changes all your paper rules, so if it's required to be kept for three years, if the district wants to keep a lot of that then they can put it on a thumb drive, but we do not have big, huge, big iron retention systems. I know all the lawyers try to scare you to death: “yeah keep it for 30 decades”… the lawyers want you to keep whatever they tell you to do.

P-20 apparently believes that the nature of retention in a cloud-based system is the same as it is with hard copy records. Although this individual recognized that the retention schedules themselves are not affected by moving either to an electronic environment or to a cloud computing environment, she nonetheless has taken an additional leap to the conclusion that therefore retention activities themselves also do not change in nature or in riskiness in a cloud computing environment. Given earlier discussions of the technical difficulties with maintaining a chain of custody in the Cloud, she appears not to have a clear understanding of the techniques necessary to ensure that electronic and cloud-resident records maintain their authenticity and integrity. Alternatively, P-20 may have concluded that Microsoft is simply a more trustworthy manager of records than the internal IT department is. In fact, she explicitly said, “We had a long relationship with Microsoft. We've been using their mail system since
the early 90s…” In addition, she discussed the way the risks were considered during the requirements phase:

If your district office loses Internet connectivity in the current world you couldn't do anything. In the cloud world you can move to any building that has Internet access and you can still have access to it… So, with the three, actually four, big ones that we've done, we had to really make sure the response rate was good, and also, [that] service providers had redundancy, so if something happens - a hurricane comes through New York, a hurricane comes through Maine - that there is a backup to it if it wipes it out…with all these implementations…it was far less expensive to go to cloud. We can show it's much more reliable, that they would have 24 x 7 service and they would have redundancy, and we couldn't match that. [my emphasis]. We also saw the functionality was pretty good; typically as we moved to the service there actually was new functionality that came along with it. It just happened when they were moving to a new release of it…You know, the biggest apprehension was that Live@edu was free. That took us six months just to be able to take advantage of a free service. That was probably our biggest hurdle.

4.3.8. Synopsis of Case 3 Findings

Several key themes permeate the analysis of this research. Again, non-cloud related worries trumped actual cloud-related outcomes. Ownership and control of information management and systems came to the forefront repeatedly in conversations. The concerns about ownership and control appeared to be related to authority structures that cross jurisdictional boundaries. Concerns about KIDS taking control of local MIS activities and decision making, as well as concerns about COT taking control of agency-level information activities and decision making were mentioned by all interviewees. In fact, although the technical risks of cloud computing were eloquently raised by several interviewees, the primary point around which discussions turned were not technological concerns but rather, concerns about authority. Who has the right to make IT information management decisions? Do agencies have the right to adopt cloud computing services in the face of disapproval by COT? Do local districts have the right to refuse cloud services that are being contracted by
KDE? In addition, this case brought to the forefront the importance of strategic decision making with regards to the appropriate role of IT in modern state government. The idea of the centralized IT department having “sucked up” or “ripped out” personnel from local agencies arose, and as in the Minnesota case, a confusion around the distinction between effects from the cloud adoption and the effects from an IT consolidation played some havoc with communication flows. The ability for statutes to legitimate IT decision making also played a role in how the decision making structures in this state have lessened the ability for the less powerful units (such as agencies) to question the decisions of those units that have legislative support (like COT). In addition, however, cloud computing can be seen as disruptive in the sense that it appears to provide a means by which the less powerful agencies can circumvent decisions made by the central IT group. The combination of jurisdictional ownership issues, a sense of IT plundering the agencies, and the ability to use a new technology to avoid changes in power all serve to suggest that cloud computing serves a greater role in state government than just the obvious one of reducing IT costs. These issues suggest the possibility that cloud computing is not only a buzzword for this state; it is also a contentious conceptual platform by which power struggles can be both enacted and resolved (if only temporarily until the next voting cycle).

4.4. The ARM Literature

4.4.1. The ARM Literature on Recordkeeping Roles and Responsibilities

The ARM literature pertaining to new technologies over the past four decades reflects a great deal of uncertainty about the identity of archivists and records managers in our increasingly technological world. Articles that have touched upon changes to the ARM occupations (i.e., archivists and records managers) can be found in Appendix I. In spite of a
great deal of discussion about the changing identity of archivists and records managers, the
literature offers a consistent view of the functions of archivists and records managers. Within
this literature, ARM authors actually attributed only a few major functions to the various
ARM personnel, and these functions remain relatively constant over the forty year time span
analyzed. Although the tasks comprising these functions have changed with technology, most
ARM professionals have not represented the functions of archivists and records managers in
a significantly different manner either among themselves or over time. The primary concerns
about the roles of ARM professions in the literature center not on the workers’ functions but
rather, on the nature of their professional identity – a distinction which can be roughly
described as a distinction between “What do we do?” and “What are we about?”

The concerns about what ARM professionals are about have led to a number of
diverging viewpoints, from opinions about what the distinction between archivists and
records managers is and about whether archivists should be trained in history or trained in
some other area (such as Library and Information Science), to whether archivists should
remain true to their historical role as custodians of culture or should become information
managers, responsible primarily to their organizations. Although it is beyond the scope of
this document to investigate these distinctions more deeply than what has already been done
in the literature review, these distinctions and disagreements primarily focus on the identity
of the ARM worker. Are ARM workers closer to being custodians of culture or information
managers? Should they be more focused on conservation and collections or on access and
users? Should they focus on history and historical understanding or on information science
and evidence? The published answers to these questions have revealed a wide variety of
different opinions about the ARM worker’s identity.
When one examines the question of “what do we do?” one finds very little
disagreement, however. Figure 4 shows the results of a detailed discourse analysis of the
ARM literature from 1970 to 2010, as explained in Chapter 3. In particular, all ARM
literature that discussed technology and technological change was examined to find
references to the functions and activities of ARM workers, looking for both what authors
then argued were prevailing beliefs about ARM roles and what they themselves asserted to
be ARM roles. The figure shows the functions attributed to ARM personnel within the
articles, with the y-axis representing the number of articles in which the functions occur.81
The most frequently mentioned function resides at the left-side of the x-axis and each
function from left to right along the x-axis occurs less frequently than the one to its left.

One will undoubtedly notice that the term “function” must be taken in a quite liberal
manner on this graph. In fact, this graph includes both functions and activities and sometimes
even tasks. When someone specifically wrote that they were talking about an archival
“function,” this activity was treated as a function even if it appeared to me to be a task rather
than a functional area. In addition, some task names (e.g., processing) are also the term for a
function. Rather than attempt to place my own interpretation or definition on the terms
authors used, I kept respondents’ terms, only consolidating two or more terms in situations in
which it was very clear that two different authors were in fact referring to the same activity
(e.g., “information retrieval” was treated as the same as “finding and retrieving documents”
or “retrieving information”). Likewise, one must be careful how to interpret the number of
articles that cited particular functions. One can probably assume that if an activity or function
was listed by only one article that most ARM authors do not think of it as a key ARM

81 The data labels above each bar indicate the exact number of articles in which a particular function was
mentioned.
Figure 4 - Functions Attributed to ARM Personnel More than Five Times in ARM Journal Articles, 1970-2010

Functions Attributed to ARM Personnel
More than Five Times in ARM Journal Articles
1970-2010
function. However, one cannot meaningfully compare the number of times a function was reported with other functions that are cited at the same or similar frequencies other than to say those at the very top of the list have found a degree of consensus among ARM authors.

The reason for comparative difficulty is that the articles reviewed were not written for the goal of providing a list of all ARM functions. In fact, at various points in time, particular functions such as processing, reference, or appraisal have been studied by a larger number of authors than other functions – fads of research, one might say. For example, during a particular time period, many authors may be writing about appraisal. This does not necessarily mean they think only appraisal is an ARM function, or even that it is a more important ARM function than those which are not being discussed much in the literature at that moment; it merely reflects ebbs and flows in current research interests. Likewise, particular journal issues may focus specifically on a particular topic, such as health records, electronic records management, or archival education. Thus, although one can estimate some trends in authors’ ideas about ARM functions from the articles, one cannot really make statistically accurate comparisons among the importance of the functions to authors purely on the basis of frequency referenced, since those are cases in which the journal editors have made prioritizing decisions. In other words, when looking at this list one must paint with broad strokes.

There were also some distinctions in language use across journals that Figure 4 does not show, reflecting potential cultural differences in language or archival tradition. For example, *Archivaria* is a journal published in Canada, whereas *American Archives* is a journal published in the United States. In Figure 4 the term “conservation” was used 36 times – 17 in American Archivist and 19 in Archivaria. If one examines the usage of this term
across journals, however, one can see that a somewhat different meaning appears to be applied to the term in the former journal than in the latter. In *Archivaria*, many of the times the term is used it appears to represent what currently is separated into two terms in the American setting – “preservation” and “conservation.” Likewise, in *American Archivist* this does not appear to be the case with the usage of the term, which generally refers just to “conservation” (as opposed to preservation). It is not clear if this is a purely cultural distinction in use, a temporal change in usage, or both. Thus, one must be careful of interpreting exactly what tasks and activities go into the referenced functions.

Nonetheless, among the most frequently listed functions, the top ten would be quite familiar to both archivists and records managers as key recordkeeping duties:

- Appraisal;
- Preservation;
- Description;
- Access;
- Arrangement;
- Acquisition;
- Reference;
- Retention (i.e., retention scheduling);
- Records management; and
- Selection.

There were no apparent changes in frequency across time.

Table 10 shows the list of all functions that at least five articles reported as an ARM function, along with the number and percentage of articles that reported that function.
<table>
<thead>
<tr>
<th>Function Labels</th>
<th>Number of Articles Referencing Function</th>
<th>Percentage of All Articles Reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal</td>
<td>134</td>
<td>51%</td>
</tr>
<tr>
<td>Preservation</td>
<td>126</td>
<td>48%</td>
</tr>
<tr>
<td>Description</td>
<td>123</td>
<td>47%</td>
</tr>
<tr>
<td>Access</td>
<td>108</td>
<td>41%</td>
</tr>
<tr>
<td>Arrangement</td>
<td>90</td>
<td>34%</td>
</tr>
<tr>
<td>Acquisition</td>
<td>84</td>
<td>32%</td>
</tr>
<tr>
<td>Reference</td>
<td>61</td>
<td>23%</td>
</tr>
<tr>
<td>Retention</td>
<td>58</td>
<td>22%</td>
</tr>
<tr>
<td>Records Management</td>
<td>54</td>
<td>21%</td>
</tr>
<tr>
<td>Selection</td>
<td>50</td>
<td>19%</td>
</tr>
<tr>
<td>Information Retrieval</td>
<td>44</td>
<td>17%</td>
</tr>
<tr>
<td>Conservation</td>
<td>36</td>
<td>14%</td>
</tr>
<tr>
<td>Intellectual Control</td>
<td>26</td>
<td>10%</td>
</tr>
<tr>
<td>Classification</td>
<td>25</td>
<td>10%</td>
</tr>
<tr>
<td>Processing</td>
<td>23</td>
<td>9%</td>
</tr>
<tr>
<td>Disposition</td>
<td>21</td>
<td>8%</td>
</tr>
<tr>
<td>Storage Management</td>
<td>19</td>
<td>7%</td>
</tr>
<tr>
<td>Accessioning</td>
<td>18</td>
<td>7%</td>
</tr>
<tr>
<td>Collection Management</td>
<td>17</td>
<td>6%</td>
</tr>
<tr>
<td>Public Service</td>
<td>16</td>
<td>6%</td>
</tr>
<tr>
<td>Reproduction (e.g., copying, microfilming, etc.)</td>
<td>13</td>
<td>5%</td>
</tr>
<tr>
<td>Outreach</td>
<td>13</td>
<td>5%</td>
</tr>
<tr>
<td>Standards Development</td>
<td>11</td>
<td>4%</td>
</tr>
<tr>
<td>Information Management</td>
<td>11</td>
<td>4%</td>
</tr>
<tr>
<td>Education</td>
<td>11</td>
<td>4%</td>
</tr>
<tr>
<td>Diplomastics</td>
<td>10</td>
<td>4%</td>
</tr>
<tr>
<td>Identification of Records to Accession</td>
<td>10</td>
<td>4%</td>
</tr>
<tr>
<td>Public Programming</td>
<td>8</td>
<td>3%</td>
</tr>
<tr>
<td>Monitor Compliance</td>
<td>8</td>
<td>3%</td>
</tr>
<tr>
<td>Inventorying</td>
<td>8</td>
<td>3%</td>
</tr>
<tr>
<td>Development of Rules, Policies, Guidelines</td>
<td>7</td>
<td>3%</td>
</tr>
<tr>
<td>Electronic Records Management</td>
<td>7</td>
<td>3%</td>
</tr>
<tr>
<td>Management</td>
<td>6</td>
<td>2%</td>
</tr>
</tbody>
</table>

| Total                                               | 1256                                   |                                    |
Numerous articles referred to “Records management” or “managing records” as a key function for both archivists and records managers. However, none actually attempted to distinguish between either the functions or the day-to-day tasks of archivists and records managers. The term is very vague in many respects and can potentially comprise a wide variety of specific activities, although frequently authors have used it to mean the entire process of managing records from the moment of acquisition until final disposition. In fact, the presence of this term illustrates the archival bias of the journals selected – to a records manager, saying that one of the functions they perform is “records management” is almost tautological. To an archivist, however, treating records manager as a function indicates that they perceive records management tasks to comprise a function that is one among many components of archival work.

Some authors specifically pointed out that the primary distinction between the two occupations relates to the point in the records’ life cycle at which the workers perform their activities; others have distinguished the two occupations primarily in terms of the organizational type at which the personnel usually work. However, in none of the articles does anyone clearly distinguish between the day-to-day activities and functions of the two occupations. In fact, few articles really offer a comparison or contrast between the roles and duties of both occupational areas at all. The articles in Records Management Journal were virtually all about records management as an occupation, with only two articles specifically about archiving and only four about both records management and archiving. However, of the remaining articles in the other journals that discuss the relationship between archives and records management as occupations, all except one (Scanlan 2011) presented “the archival point-of-view” by directly discussing its topic in terms of its relevance to archivists or to
archiving as a profession. Only three articles (Yusof and Chell 1998, 1999, 2002)\textsuperscript{82} contrasted records management with archiving, and these three articles discussed the varying academic and traditional fields from which current records management evolved rather than functional roles and responsibilities of the workers. In short, in spite of the existence of articles which assert that archivists feel anxiety about their professional relationship to records managers, very little actual literature exists in the top ARM journals to support the assertion that the two fields can be clearly distinguished in terms of occupational roles and responsibilities, other than the greater focus of archivists on long-term preservation as an occupational responsibility.

When the articles did deal with the relationship between archiving and records management, only a few themes appeared, with no specific technologies discussed or specific descriptions of how the two fields are, or are not, different from each other. The various arguments are that:

- archives and records management are the “obverse sides of the same coin” (Cox, 1984-85, 188) and must work together because they essentially work towards the same purpose (Atherton 1985-86; Craig 1987; McDonald 1995; Nelson 1995);
- technology is causing the roles of archivists and records managers to merge (Duff 1995; Hopkins 1983; McDonald 1995) or leads them both to have to follow similar techniques (Dollar 1993);
- technology has removed any real difference between archivists and records managers. Availability of access is more important than location. All who work with records are

\textsuperscript{82} All were published in Records Management Journal.
recordkeepers (Flynn 2001; McKemmish 2001; Tough 2004; Upward 1996, 1997, 2000));

• archivists stand (or should stand) in an authority relation over records managers (Mumford 1970; Geda 1979);
• archivists and records managers exist in a somewhat competitive relationship as occupations (Dodds 1976; Scanlan 2011) and archivists must become records managers to remain relevant (Dodds); and
• archivists must have a knowledge of processing and records management tasks in order to develop archival automation (Dürr 1984).

In fact, the only way to discern the detail of authors’ ideas about functions performed by both ARM professions was through the discourse analysis discussed here. The primary outcome was that when functions are discussed for archivists and records managers, the functions are as listed in Figure 4 for both occupations.

Thus, an in depth analysis of the literature suggests that records managers and archivists both engage in appraisal, selection, and acquisition of records, arrangement and description, determine retention schedules, and engage in the various aspects of records management. For example, they create retention schedules and determine final disposition of records. They provide a variety of reference services and ensure appropriate access occurs (within constraints of privacy and security) and ensure the materials can be easily retrieved. For materials that have long-term value, they engage in preservation (and according to some, conservation). In addition, they must engage in public service and outreach, (both as a means to garner popular support and as an occupational duty to educate.) They also need to help develop rules, policies, and guidelines for records in their organizations. Although many of
these articles do not espouse the continuum theory explicitly and thus apparently do see archiving and records management as different occupations, their descriptions of the two areas of work are surprisingly similar. Archivists are stewards or curators of an organization’s (or society’s) records, and records managers are primarily stewards or curators of their organization’s records, unless they are records managers within government agencies. They essentially perform very similar, if not the same, functions according to the literature.

4.4.2. The ARM Literature on Cloud Computing

When reviewing the six ARM journals, very few articles focus on cloud computing from an ARM perspective. *American Archivist, Archival Science, and Archives & Museum Informatics* feature NO cloud-oriented articles (although the search mechanism for *Archival Science* is not as forthright or comprehensive as for the other journals). 83 *Records Management Journal* contains five articles with substantive comments about cloud computing and recordkeeping during the publication timeframe studied (Askhoj, Sugimoto, and Nagamori 2011; Cumming 2011; McLeod and Hare 2010; NARA 2010; Stuart and Bromage 2010), with two other articles that simply mention the term “cloud computing” without further discussion. *Archivaria* contains two articles, neither of which is about cloud computing, but also merely mention the term (Trace 2011; Buchanan 2012). Finally, the *Journal of the Society for Archivists* contains one article on cloud computing (Ferguson-Boucher and Convery 2011). Several ARM professionals and organizations have issued reports and articles on cloud computing, such as Nicole Convery, who authored the report on

83 Since the last publication date for *Archives & Museum Informatics* (under that name) was 1998, one would not expect cloud computing references, and none were found.
Aberystwyth University’s “Storing Information in the Cloud” project (Convery 2010) and Luciana Duranti, who leads the current “Records in the Cloud” project at the University of British Columbia (Duranti 2012; Duranti and Rogers 2012). The academic recordkeeping literature cites a variety of cloud computing risks, as shown in Table 11.

### Table 11 - Cloud Recordkeeping Risks Reported in ARM Literature

<table>
<thead>
<tr>
<th>Risk Area</th>
<th>Citation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access – the risk the service provider will not be able to transfer your records back to you in a usable format if you cancel contract or they go out of business.</td>
<td>(Cumming 2011; Ferguson-Boucher and Convery 2011; Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Access – the risk of unauthorized access due to information residing in a shared environment.</td>
<td>(Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Access Time – can the service provider act upon requests for records in a timely manner for eDiscovery or customer requirements?</td>
<td>(Ferguson-Boucher and Convery 2011)</td>
</tr>
<tr>
<td>Auditability – can we (or a trusted third party) audit the system on a regular basis, as needed?</td>
<td>(Ferguson-Boucher and Convery 2011; Upward et al. 2013)</td>
</tr>
<tr>
<td>Business Continuity – How much system downtime can we expect?</td>
<td>(Cumming 2011)</td>
</tr>
<tr>
<td>Compliance – Does the system use any standards that would allow us to know we comply with our regulatory requirements?</td>
<td>(Cumming 2011; Ferguson-Boucher and Convery 2011)</td>
</tr>
<tr>
<td>Contracts – Risk of failing to draft as many contracts as are needed</td>
<td>(Cumming 2011)</td>
</tr>
<tr>
<td>Control (Compliance) - Can we provide an adequate description of where the information is being kept for our legal compliance needs?</td>
<td>(Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Control (Jurisdiction) – Risk that information will be kept outside of jurisdictions with which we are comfortable</td>
<td>(McLeod and Hare 2010)</td>
</tr>
<tr>
<td>Disposal – Risk the information is not completely destroyed according to retention scheduling or our express request</td>
<td>(Ferguson-Boucher and Convery 2011; Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Disposal – Risk we will not be able to demonstrate to courts or other interested parties that disposal has occurred appropriately</td>
<td>(Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Information Lifecycle Management – Risk that organizational recordkeepers will not know who is able to (or should) “touch” the records for particular necessary tasks</td>
<td>(Cumming 2011)</td>
</tr>
<tr>
<td>Information Retrieval – Risk that the provider will not offer user-</td>
<td>(Ferguson-Boucher and</td>
</tr>
</tbody>
</table>
### Risk Area

<table>
<thead>
<tr>
<th>Risk Area</th>
<th>Citation(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>friendly or accurate standard mechanisms for retrieval</td>
<td>Convery 2011)</td>
</tr>
<tr>
<td>Interoperability – Risk that records cannot be moved from one provider to another if we change providers, due to lack of interoperability</td>
<td>(Ferguson-Boucher and Convery 2011; Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Metadata – Risk that metadata sufficient and well enough organized for ensuring context can be discovered will not be able to be created in cloud environment</td>
<td>(Ferguson-Boucher and Convery 2011)</td>
</tr>
<tr>
<td>Privacy – Risk that a breach will occur and/or that appropriate policies are not in place for notifying the appropriate people when a breach does occur</td>
<td>(Ferguson-Boucher and Convery 2011; Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Back-Ups – Risk that back-ups will not be restorable or will not be restorable in a usable format</td>
<td>(Stuart and Bromage 2010)</td>
</tr>
<tr>
<td>Retention – Risk that retention schedules will not be able to be applied in a cloud environment</td>
<td>(Ferguson-Boucher and Convery 2011)</td>
</tr>
<tr>
<td>Security – Risk that appropriate security measures are not in place or that secure techniques are not being used</td>
<td>(Cumming 2011; Ferguson-Boucher and Convery 2011)</td>
</tr>
</tbody>
</table>

#### 4.5. Analysis and Discussion

**4.5.1. Introduction**

Examination of the three different cases has revealed recordkeeping practices in which state government has been involved in the adoption of a cloud computing service. In the first case, Minnesota represents a state that follows a hybrid information governance structure. Minnesota has implemented a statewide cloud computing service for communication and collaboration. In this case, the state’s centralized IT agency led the implementation, championing it and working as the internal lead on the project. The second case examines a cloud computing service adoption in which a number of states have chosen to collaborate with each other, with local government, with federal government, and with a private, nonprofit organization to engage in an information sharing program. North Carolina was chosen as an exemplar because it represents a “typical” scenario in which individual jurisdictions send their data to the State Health Department, which uses the data for its own surveillance activities and forwards it to the CDC-funded BioSense 2.0 system. Thus,
jurisdictions must send their data to the Department of Public Health for state purposes but have the option to participate or not to participate in the BioSense 2.0 program. By selecting a scenario in which jurisdictions are free to participate or not, this case allowed some investigation into incentives and disincentives for collaboration. Finally, the third case involved an individual state agency that implemented a cloud computing service for communication and collaboration among school district employees, teachers, parents, and students statewide. In this case, the Kentucky Department of Education made the decision to move to the Cloud without consulting the centralized state IT agency. The decision was made in the face of the central IT agency’s concerns that cloud computing in general should be considered only in cases in which central IT cannot provide the IT services itself and has evaluated and approved the adoption. The implementation champion was a KIDS (agency IT department) director, who garnered support for the project at state and local levels and advocated for the needed hardware and software centralization techniques.

The original goals of this study were to examine how recordkeeping stewards who work in state government or alongside other state government recordkeeping stewards in cloud computing environments perceive and act upon electronic recordkeeping requirements in the Cloud, to understand which of the functions of ARM work described by ARM academic literature occur in the recordkeeping environments examined, and to determine whether these functions are performed by ARM workers or by other recordkeeping stewards when they do occur. The research questions informing the investigation were:

- Within the environments examined, what occupational groups are reported to act as key stewards of the information and how do members of these groups perceive and act upon recordkeeping requirements in the Cloud?
- How do the various stakeholders interact with each other with respect to recordkeeping activities within their cloud computing environments, and what do these relationships suggest about how ARM occupational roles and responsibilities are being handled in the Cloud?
- How do the various stakeholders perceive the roles and responsibilities of archives and records management personnel?
- What cloud computing risks does the professional and academic ARM literature report, and do recordkeeping stewards in state government cloud environments express concerns about these same risks?
- Of the main recordkeeping functions that the ARM literature attributes to ARM workers, are these functions evident in the recordkeeping environments examined and if so, are they performed by ARM workers?

Each of the following sections offers results and discussion related to these questions.

4.5.2. Reported Recordkeeping Stewards

Across the three cases, a variety of individuals from different occupational groups act as recordkeepers within their respective cloud environments, according to both themselves and their colleagues. Table 12 shows the reported occupational and organizational assignment of stewards as reported by interviewees. Column 1 shows the occupational group reported to be a recordkeeping steward. Column 2 shows the reporting interviewee’s occupation; column 3 shows the case(s) in which each occupational group was reported to be a recordkeeping steward. Because some occupations were reported by interviewees across cases, if a particular occupational group was reported to be a steward in more than one case (as seen in the third column), the specific cases in which an interviewee made the reports are
Table 12 - Reported Occupations of Record Recordkeeping Stewards, with Interviewee Occupation and Case(s)

<table>
<thead>
<tr>
<th>Reported Steward Occupation or Entity</th>
<th>Interviewee Occupation</th>
<th>Case(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Records Managers (includes processors and file managers)</td>
<td>Archivists (MN, KY) IPAD Personnel (MN) Records Managers (MN, KY)</td>
<td>MN Office 365 KY MS Live@edu</td>
</tr>
<tr>
<td>Records Creators</td>
<td>(one) Central IT Person</td>
<td>MN Office 365</td>
</tr>
<tr>
<td>Other Agency Personnel (e.g., Records Management liaisons)</td>
<td>Records Managers</td>
<td>MN Office 365</td>
</tr>
<tr>
<td>Jurisdiction/Facility of (data) origin (As an entity)</td>
<td>Epidemiologist/Manager External Consultants</td>
<td>BioSense 2.0</td>
</tr>
<tr>
<td>External Consultants (i.e., non-Cloud-vendor consultant)</td>
<td>Jurisdiction’s Epidemiologist External Consultants</td>
<td>BioSense 2.0</td>
</tr>
<tr>
<td>Data Practices Liaisons</td>
<td>Archivists Central IT Personnel IPAD Personnel Records Managers</td>
<td>MN Office 365</td>
</tr>
<tr>
<td>Cloud Vendor (IT) Personnel</td>
<td>Central IT Personnel (MN, BioSense) Epidemiologist/Manager (BioSense) External Consultant (BioSense)</td>
<td>MN Office 365 BioSense 2.0</td>
</tr>
<tr>
<td>Central IT Personnel (Internal)</td>
<td>Records Managers (but not unanimously) Central IT Personnel</td>
<td>MN Office 365</td>
</tr>
<tr>
<td>Archivists</td>
<td>Archivists (MN, KY) IPAD Personnel (MN) Records Managers (MN, KY)</td>
<td>MN Office 365 KY MS Live@edu</td>
</tr>
<tr>
<td>Agency IT Personnel</td>
<td>Agency IT Person (KY) Archivists (BioSense, KY) External Consultant (BioSense)</td>
<td>BioSense 2.0 KY MS Live@edu</td>
</tr>
<tr>
<td>“Individual Jurisdictions” (as entities)</td>
<td>Epidemiologist/Manager External Consultants Jurisdiction’s Epidemiologist</td>
<td>BioSense 2.0</td>
</tr>
<tr>
<td>“CDC” (As an entity)</td>
<td>Epidemiologist/Manager External Consultants Jurisdiction’s Epidemiologist</td>
<td>BioSense 2.0</td>
</tr>
<tr>
<td>“ASTHO” (As an entity)</td>
<td>External Consultant</td>
<td>BioSense 2.0</td>
</tr>
</tbody>
</table>
shown in parenthesis following the interviewee’s occupation in the second column. For example, records managers (as found in column 1) were reported to be recordkeeping stewards by archivists (2nd column) in the Minnesota case and in the Kentucky case (also 2nd column). If there is no parenthesis after a reporter’s occupation in the second column, this implies that one or more interviewees reported that steward only in the particular case that is listed in the third column. For example, the CDC as an entity was reported to be a steward by an epidemiologist from an individual jurisdiction, by external consultants, and by a managing epidemiologist; all these interviewees are associated with the BioSense project.

The individual occupation types reported to be stewards were not pressured to answer whether or not particular occupations that they had not mentioned on their own also acted as stewards. Rather, only those whom they explicitly mentioned are listed here. Thus, it is possible that if someone who did not mention IT (for example) were asked directly, “Do you think IT personnel are also recordkeeping stewards?” they might in fact respond in the affirmative. By noting only those stewards they identified on their own, we can see a little more closely their predominant, tacit predilections about the functions and roles associated with records stewardship. However, although this provides us information about their thoughts on stewardship, it may not provide complete information, since in the context of any

---

84 This was done in an attempt to avoid “leading” respondents to particular recordkeepers.
given conversation the respondents may simply have forgotten one or more occupation that they do, in fact, view as a steward.

4.5.3. Perceived Requirements in the Cloud

The reports about the process of requirements gathering showed a great deal of variation across these three cases. In Minnesota, MN.IT Services respondents reported a relatively formal and highly participatory requirements analysis process. Numerous participants within the Executive Branch participated in meetings and interviews in order to provide information to MN.IT Services that would allow the IT agency to develop comprehensive requirements documents. Nonetheless, none of the interviewed respondents outside of MN.IT Services played any role in the requirements gathering and in fact, were clearly unknowledgeable about either when and how the requirements were developed or even what the nature of the requirements were.85 The records managers in particular were unhappy about their lack of participation and reported that they had concerns about whether necessary recordkeeping requirements were ever addressed; they also reported that their ignorance of the requirements definition causes difficulties because they subsequently also do not know whether their recordkeeping requirements are or are not supportable by the new system. None of them was aware that an archival requirements document had been drafted and used during vendor selection.

In Kentucky, the ARM professionals reported that they not only did not take part in requirements analysis but also did not even know that the KDE was going to be moving to the Cloud until implementation began, well after vendor selection and requirements

---

85 This is consistent with Convery’s (2010) observation that “records and information professionals are often not part of the cloud computing consultation processes or project team from the outset” (17).
definition took place. Unlike the Minnesota case, however, the individuals did not appear to resent their lack of participation. Rather, they presented the situation as a normal fact of organizational life for them – business as usual. One individual did note, however, that in at least one case, some local employees felt that the cloud system had replaced a better-working non-cloud system. With regard to the Microsoft Live@edu system, P-29 reported:

It’s pretty clear records are NOT being destroyed. But a lot of times there are good reasons not to have things around forever…We are finding some old records that someone listed, where the person [a supervisor] has written down their [employee’s] effort isn’t good, or they [the employee] got suspended and the retention is way overdue.

Keeping records such as this past their retentions creates an additional risk of lawsuit for the school district, in the event such negative information becomes public.

The BioSense 2.0 requirements gathering was highly participative and open to public inspection, as described earlier in this chapter. A diverse governance board composed of members from all major participant groups and levels of government took part. They were joined by CDC personnel, external IT consultants, and professional organizations that are well-respected in the epidemiological community. Together, they engaged in a very open series of requirements gathering procedures, led by an external consulting firm. All decision making was either posted to the Redesign Website or, in the case of confidential documents, distributed privately to the participants. In addition, all major decisions were posted to the Website, which is largely open to the public at large, with a password-protected section for “onboarded” project partners. No respondent complained about the requirements gathering, and all stressed that it was done in an open manner to ensure that all participating jurisdictions would not only know the strategic decisions, processes, and procedures of the BioSense 2.0 Redesign, but would also be less inclined to worry about jurisdictional overstepping on anyone’s part.
In the case of Minnesota, the cloud implementation focused on email, which is entirely electronic in nature and spread throughout all agencies and units of the government. Also, the interviewed records managers work in highly collaborative environments; they rely upon IT, request information from IPAD, and educate and work closely with their agency colleagues. They reside in an environment of tight internal connectivity with their agency’s employees. The records managers reported that the implementation affected them operationally, because it led them to worry about having more work to do in some areas and to report that they now have to complete some tasks without the proper access rights to do so. It also affected them politically, insofar as they all expressed the concern that they would be placed in a less desirable relationship with MN.IT Services, upon whom they rely a great deal.

In Kentucky, the KDE was able to bypass the central IT agency COT entirely when deciding to implement their cloud services. In addition, they are still a primarily paper-based agency, which means that the majority of the archivists and records managers are working with their school district colleagues to handle records after they have been created, and to preserve them in paper form. As a result, the move to the Microsoft Live@edu cloud computing environment did not have much of an impact on them, both because they primarily focus on paper-based records and because KDE’s IT department has managed and controlled records for quite some time. These respondents were more able to speak to

---

86 For example, now when an employee terminates, the records manager must immediately peruse their emails to identify record from non-record material and in the event of a hold, make requests to MN.IT Services to get their emails sent to them.

87 For example, in the case of mistaken information, records managers could previously (i.e., pre-implementation) make updates relatively quickly and simply on their own. Now they must file a ticket to MN.IT Services and wait for action from Microsoft. Likewise, if information is needed rapidly from an employee’s email for legal hold, they are no longer able to get those records themselves, but must route it through Microsoft.
impacts of the cloud-based financial system, which had significant political and operational impact and which led the state eventually to mandate their participation.

Within the BioSense 2.0 case, respondents work at significantly different hierarchical levels within their own organizations, and they collaborate primarily with “equals” (in terms of rank) from outside their organizations. All respondents except for an external consultant were at director level or higher. The external consultant worked on implementing and maintaining technical components for the system. None felt that their primary work was interrupted as a result of BioSense 2.0, and the CDC respondent felt that the changes had been highly positive. The respondents only reported concerns regarding potential political interrelations between different jurisdictions and with regards to data sharing across jurisdictions. They did remark, however, that jurisdictions have the freedom to determine how much information sharing they wish to do. Thus, the program administrators had to engage potential participants by incentivizing people and by continually educating all participants about their freedom and their ultimate ownership of the data. The only major concern once security requirements were shown to be satisfied was a concern about status relations – that is, participants were concerned about maintaining their direct relationship with their financial benefactor, the CDC. Since BioSense 2.0 does not attempt to preserve information for the long-term, it is not surprising that no one expressed concern over potential ARM requirements except security, privacy, and standardization to enable access and retrieval.

4.5.4. Interaction between Recordkeepers

Recordkeeping stewards reported tensions with other groups in all three cases. For example, in Minnesota all respondents except the MN.IT Services personnel reported that
they believe there are communication issues with MN.IT Services. Records managers specifically reported discomfort with being “left out” of decision making when system implementations occur, with the sole exception being the respondent whose reporting structure places him within MN.IT Services. The other records managers also reported dissatisfaction with the changes in their roles since the cloud implementation and IT consolidation, citing both their inability to perform all the duties they previously performed. They also reported that IT refuses to take responsibility for those duties which they believe they can no longer perform due to lack of access rights, suggesting that the records managers believe that accountability for these tasks should rightly be inherited by MN.IT Services, which they see as the final authority in the implementation.

Their responses point to concerns about occupational status, and frustration with the apparent difference between the degree of power that IT holds within the organization and their own power. One records manager, reporting on a visit from the CIO, said, “We did ask her to attend a meeting and . . . she did grant us an audience.” The phrase “granting an audience” is telling. When the desired outcome did not occur as a result of this meeting, this records manager said, “Who else can you talk to if you can't talk to the big wig?” (my italics).

Records managers were not the only respondents to report a perception that IT is able to seize many organizational decisions, however. P-8 said, “They often like to set initiatives for the rest of the organization instead of the other way around…IT’s the driver instead of business being the driver.” In addition, a MN.IT Services employee concurred that the changes to access rights that MN.IT Services had made during the cloud implementation had

---

88 This individual originally reported to his own agency IT department, but as a result of the IT consolidation now reports directly to MN.IT Services.
led some to feel that IT had removed a portion of their powers; P-13 remarked that people felt that they had lost some of their freedom when access changes took place. When then asked if the discontent was an issue of people feeling that a certain power had been taken away rather than a concern that they could no longer perform their required activities, this individual responded affirmatively, suggesting that they could still perform their jobs, but merely had a different process to follow and different dependencies than they had before.

Kentucky’s individual agency personnel and KDE’s IT department also expressed some concerns about power. When speaking about the various cloud implementations that KDE has undergone, P-29 reported that when MUNIS\textsuperscript{89} moved to a cloud platform, some districts were quite opposed because they felt their own individual financial management programs supported their own requirements better than the new cloud service. However, after a great deal of discussion between KDE executive management, KIDS and the local school districts, P-29 reported, “The state finally just mandated it.” Ultimately, what KDE set out to be a process of persuasion became a command and control process, revealing the underlying power differential.

Although the BioSense 2.0 implementation occurred in a different type of environment (i.e., a virtual organization\textsuperscript{90}), when examining the BioSense 2.0 case, the recognized recordkeepers (i.e., the various jurisdictions, the federal government, and the external consultants) also reported that concerns about the power structure played a key role

\textsuperscript{89} MUNIS is the school districts’ financial management program.

\textsuperscript{90} One will recall from Chapter 2 that a virtual organization is typically defined as a dynamic collection of individuals, institutions and resources” that allow “flexible, secure, coordinated resource sharing” (Meitzner and colleagues, 138), in which the individuals engaged in the resource sharing define “clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs” (Foster, Kesselman, and Tuecke, 200-2001). This definition very well defines the contractual nature of the BioSense 2.0 collaborative.
in the implementation. P-17 said that the concern in North Carolina was ensuring that the previously standing relationship that the Health Department had with the CDC would remain in effect: “We wanted to make sure that our relations are still the same; that we [are] giving our data to CDC and [that] ASTHO works under CDC as grantee.” P-17 verified that the movement into a cloud computing environment was a secondary consideration for them. Their primary concerns were two-fold: working out who would perform various activities within the new BioSense 2.0 collaborative and ensuring the maintenance of their pre-existing relationship with CDC.91 P-17 also reported that another major concern was ensuring that no jurisdiction be required to engage in a level of data sharing with which they were not comfortable.

P-27 corroborated P-17’s statements about concerns with data sharing across jurisdictional boundaries, saying that most jurisdictions’ primary worry was that the federal government would be able to gain control over “their” data. ASTHO was brought in partially for the purpose of quelling these concerns. Since ASTHO has a great deal of professional reputation among medical personnel and is considered to be a very powerful organization (P-17), they were asked to act as the intermediary between the CDC and the jurisdictional partners. Technically, the individual jurisdictions own the data they supply to the system; ASTHO is the lessee in the Amazon AWS contract; CDC is the funder. The governing board created this structure to insure that jurisdictions did not avoid participation on the basis of fear that the federal government could take control of their data. The original BioSense 1.0

---

91 However, the BioSense 2.0 Redesign Team and CDC spent a great deal of time and energy providing information and education about the privacy and security capabilities of the system, so it is not clear whether the maintenance of relations would have been the highest priority had the interviewees not already been convinced of the safety of the data.
system suffered from this handicap because the CDC owned and managed the hardware and software within which the surveillance data resided.

Participants in these three cases thus focused heavily on concerns about power and control prior to, during, and after implementation of the cloud system. Records managers expressed concerns about losing ownership of their data to the more powerful IT organizations within their states; local agency employees expressed concerns about losing control over their ability to select hardware and software services best suited for their individual requirements; and state and local jurisdictions expressed concerns about loss of control over information or its management to the federal government. In all three cases, the concerns centered on the likelihood of losing power to what was perceived to be an already more powerful entity within their political or organizational environment.

4.5.5. Evidence of ARM Functions in the Cloud Environments

4.5.5.1. ARM Functions and Their Occupational Allocation

The three different cases present very different pictures of ARM functions. For one thing, only a subset of the functions can even be found in the BioSense 2.0 case. No records managers or archivists are on the staff and no long-term preservation occurs. Nonetheless, recordkeeping activities occur “behind the scenes” at the partners’ (i.e., jurisdictions’) home organizations. Respondents were unaware of either the scope or nature of these activities, however. When interviewees spoke of recordkeeping at all, they referred specifically to the activities associated with maintaining the privacy and security of the epidemiological data that feeds into and is maintained within the BioSense 2.0 repository for analysis; such data is the responsibility of the jurisdictions who submit it. Most records management and archival activities are thus not considered to be within the scope of the BioSense 2.0 virtual
organization; the participants use the system mainly for access and current syndromic surveillance analysis. Although one key aim of the system is information sharing, the degree of sharing between jurisdictions is still small.

The twenty most frequently referenced activities listed in ARM literature are shown in Table 13. This table also shows the stewards who have been reported to perform these functions after the cloud adoption for each case. (Consistent with interviewee responses, the stewards are listed only as organizations for BioSense 2.0).

Within the Minnesota and Kentucky cases, records managers and archivists both exist, but the workers’ roles and responsibilities are distributed quite differently across the two cases. Kentucky, for example, appears more like a “traditional” recordkeeping environment as mentioned in the ARM literature, insofar as appraisal and retention rules are developed by the State Archives and communicated outward to the state agencies. The KDLA determines the final retention and disposition rules, deciding which records will be accepted for long-term preservation into the archives. It sets “standards, procedures, and administrative regulations for recording, managing, preserving, and reproducing government-created or maintained records” (KDLA 2008, 10). It also supplies the agencies with updated retention schedules on a regular basis. Each state agency has a designated Records Officer who liaises with the KDLA and is responsible for “inventorying, analyzing, and advising the staff on records management procedures; participates in the agency’s information technology planning process; oversees the transfer of records and publications; and supervises the scheduled destruction of records within the agency (KDLA, 5).

The Kentucky scenario is one in which the final authority for archives and records management resides with the State Archives. Retention Schedules are drafted by the KDLA,
<table>
<thead>
<tr>
<th>ID</th>
<th>Function</th>
<th>Recordkeepers Taking Part in the Function’s Tasks</th>
<th>MN Microsoft 365</th>
<th>KY Microsoft Live@Edu</th>
<th>BioSense 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>appraisal</td>
<td>Agency records managers (on occasion of previously unclassified records types); State Archivist, Attorney General, and State Auditor (as members of Records Disposition Panel).</td>
<td>KDLA, for records retained permanently.</td>
<td>Unknown.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>preservation</td>
<td>Automated functionality managed by MN.IT Services.</td>
<td>KDLA, for records retained permanently.</td>
<td>Not performed.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>description</td>
<td>Agency records managers (on occasion of previously unclassified records types); MN.IT Services, Microsoft implement in technical systems.</td>
<td>KDLA, for records retained permanently.</td>
<td>Determination of elements by ISDS; Amazon implements within the technical systems; systems administration performed by CDC and/or ASTHO personnel and locally, by individual jurisdictions’ personnel.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>access</td>
<td>MN.IT Services; Microsoft; IPAD liaisons (for access-related briefs).</td>
<td>KIDS personnel.</td>
<td>Amazon; CDC and/or ASTHO personnel.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>acquisition</td>
<td>Automated and managed by MN.IT Services, according to previous classifications; agency acquisition performed via manual request from agency personnel.</td>
<td>KDLA, for records retained permanently.</td>
<td>Individual jurisdictions’ personnel prepare and send data to BioSense 2.0 repository. Automated acceptance.</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Function</td>
<td>Recordkeepers Taking Part in the Function’s Tasks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>reference</td>
<td>MHS archivists; agency records managers answer questions from agency personnel and MN.IT Services; IPAD liaisons (answer access-related questions via briefs and opinion reports). Agency records managers and local school district employees; technical questions answered by KIDS personnel. BioSense Redesign 2.0 project team members; CDC; BioSense 2.0 Governance Board.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>retention</td>
<td>Agency records managers; approval by State Archivist, Attorney General, and State Auditor (as members of Records Disposition Panel). Automated. Records are supposed to be sent to agency records managers, but email and communication is not yet captured electronically by the records managers. Rather, records all remain in transaction (cloud) system. No formal retention performed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>records management</td>
<td><em><strong>This name encompasses virtually all activities listed here as functions, and thus is too all-encompassing to make meaningful distinctions in activities.</strong></em> <em><strong>This name encompasses virtually all activities listed here as functions, and thus is too all-encompassing to make meaningful distinctions in activities.</strong></em> No formal records management activities are performed, although transfer and storage of data performed automatically – sent by local jurisdictions’ personnel, automated functions generated by Amazon.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Function</td>
<td>MN Microsoft 365</td>
<td>KY Microsoft Live@Edu</td>
<td>BioSense 2.0</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>---------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>selection</td>
<td>Automated, according to previously determined classification schemes by: MHS archivists for preservation; agency personnel do the basic manual movement of records into their buckets according to previously determined classification schemes. (Thus, this occurs well after an organization called an archives would normally consider this activity to occur.)</td>
<td>For email and communication, this appears not to be occurring. Rather, records all remain in transaction (cloud) system.</td>
<td>Pre-defined standards determine which data will be transferred to the BioSense 2.0 warehouse; standards developed by ISDS and the BioSense 2.0 Governance Board, in conjunction with the CDC. No preservation selection occurs, since no preservation occurs.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>information retrieval</td>
<td>Classification schemes developed by agency records managers and MHS archivists; implementation into systems performed by Microsoft employees and MN.IT Services; data structures in technical systems performed by MN.IT Services and Microsoft employees.</td>
<td>Classification schemes developed by KDLA; actual retrieval: automated, KIDS personnel, and/or local school district personnel or records managers, depending upon record.</td>
<td>Automated retrieval mechanisms developed by Amazon; problem resolution by CDC and/or ASTHO power users and administrators; discovery and retrieval performed by jurisdictions’ personnel.</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>ID</th>
<th>Function</th>
<th>Recordkeepers Taking Part in the Function’s Tasks</th>
<th>BioSense 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>intellectual control</td>
<td>- Arrangement and description performed for preservation-worthy records by MHS archivists for the State Archives; agency records managers for the state agencies; physical and logical instantiation in technical systems performed by MN.IT Services.</td>
<td>Descriptive Categories developed as a “minimum set” of data elements by ISDS.</td>
</tr>
<tr>
<td>14</td>
<td>classification</td>
<td>- State law has already legislated many of these. MHS archivists for the State Archives; agency records managers for agencies; IPAD helps determine privacy levels and writes briefs giving their legal opinions.</td>
<td>Descriptive Categories developed as a “minimum set” of data elements by ISDS. Changes and/or updates must be approved by the Governance Board, composed of representatives of all jurisdictions.</td>
</tr>
<tr>
<td>15</td>
<td>processing</td>
<td>- Largely automated, with technical system managed by MN.IT Services and Microsoft; Agency employees ensure records are assigned categories; agency records managers and agency heads determine classification and retentions to use; MHS archivists for State Archives.</td>
<td>Largely automated, with technical system managed by Amazon and the CDC/ASTHO technical personnel and several independent contractors.</td>
</tr>
<tr>
<td>ID</td>
<td>Function</td>
<td>MN Microsoft 365</td>
<td>KY Microsoft Live@Edu</td>
</tr>
<tr>
<td>----</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>16</td>
<td>disposition</td>
<td>Automated, managed by MN.IT Services; implemented by MN.IT Services and Microsoft. Education of responsibilities for disposition performed by agency records managers and MHS archivists.</td>
<td>Supposed to be performed by local school district personnel, according to KDLA retention schedules. Currently all records are kept in transaction system “forever.”</td>
</tr>
<tr>
<td>17</td>
<td>storage management</td>
<td>Microsoft and MN.IT Services.</td>
<td>Microsoft and KIDS.</td>
</tr>
<tr>
<td>18</td>
<td>accessioning</td>
<td>Automated receipt of emails; agency personnel manually add individual records to appropriate categories or use .pst files created and managed by MN.IT Services and Microsoft.</td>
<td>Automated receipt of emails; local school district personnel manually add individual records to appropriate fields.</td>
</tr>
<tr>
<td>19</td>
<td>collection management</td>
<td>This does not really occur for email, except insofar as records are automatically moved to Microsoft Exchange automated journaling functions, managed by MN.IT Services.</td>
<td>This does not really occur for this type of record.</td>
</tr>
<tr>
<td>ID</td>
<td>Function</td>
<td>MN Microsoft 365</td>
<td>KY Microsoft Live@Edu</td>
</tr>
<tr>
<td>----</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>20</td>
<td>public service</td>
<td>It is not entirely clear how this would be “translated” from the setting of an archival organization to state government. However, questions are answered by state employees; informational letters are created by state employees; agency website content created by state employees and records managers and implemented and managed by MN.IT Services.</td>
<td>It is not entirely clear how this would be “translated” from the setting of an archival organization to state government. General information provided by local school district personnel; technical questions handled by KIDS. Retention questions handled by district records managers.</td>
</tr>
</tbody>
</table>
which then presents them to the individual agencies. The agencies have their own records managers and the records management activities for the KDE are performed at the local district level. Some districts maintain the records onsite and some have central archives and records management centers. The local district archivists and records managers often differ in primary occupation from actual, trained records managers and archivists – they may be district clerks, managers, and even secretarial personnel. The functions performed are expected to be similar across districts, although actual practices vary somewhat in terms of which occupational group will complete these functions and in terms of the degree to which the districts comply with required retention schedules. Most of the functions from Table 10 are performed by archivists and records managers in the KDLA, local archivists, and records managers in the local districts. However, functions such as “access,” which in an electronic environment include setting up systems and authority control, are primarily performed by information technology personnel, either from the local Management Information Systems centers or from KDE’s central IT group KIDS. Likewise, information retrieval, although dependent upon classification schemes created by the KDLA, is managed or programmed by the IT groups. Electronic records storage management falls under KIDS’ purview as well. However, Kentucky has a central policy that states that “all state and local government employees are responsible for the records they create and maintain” (KDLA 2008, 4). The state agency heads are responsible for ensuring that “adequate and proper documentation of the organizational functions, policies, decisions, procedures, and essential transactions of the agency are created and preserved” (4).

Thus, in Kentucky the activities associated with the primary ARM functions highlighted in Table 10 are distributed between the local records managers and archivists, the
KDLA records managers and archivists, KDE’s central IT group KIDS, local agency heads, and the agency personnel who create and maintain records. Although the agency head holds ultimate responsibility within each agency, the bulk of the active records functions belong to the records creators and the local ARM personnel (or Records Center ARM personnel for those districts that have a records center).

However, KIDS handles both the management and stewardship of records that are in scope for this case study, that is, records associated with communication and collaboration for K-12 educational provision. In fact, P-29 reported that KIDS provides such a comprehensive management and control of records within the local districts that some district employees have even forgotten that provision of public email records upon request is their responsibility, and not the IT Department’s:

KDE has been in control of email for a long time…Their situation changed [due to KDE being in control] because there was a more local response, now they are more likely to point to the KDE people when discussing email policy… There was an open records request from one of the other districts. What happened was someone in the community went and said to their local school district, “we want to request these emails from these people for this certain period of time.” The small district said “Well, that’s not us”…The local agency should have said, “We will have to get KDE’s assistance.”

In Minnesota, the recordkeeping responsibilities are somewhat more dispersed than in Kentucky. Legally, the State Archivist in MHS does have final authority, in conjunction with the two other members of the Records Disposition Panel, to determine records disposition. However, the responsibility for creating disposition rules resides with the state agencies, with the help of their records managers. Once the records managers finalize the agency’s retention schedules, they forward them to the State Archives, where the State Archivist, in conjunction with the Records Disposition Panel, reviews and authorizes the retention schedules. Those records which are classified as archival are sent to MHS at the end of their active period. The
much larger number of non-archival records must be arranged, classified, and described by
the records managers. A large part of this activity occurs according to classifications pre-
existent in the agencies. However, the records managers are responsible for helping the
agencies find more effective ways to bucket their records (P-25). Of course, when new
technologies are introduced, some types of information that previously did not exist or was
not considered “record” end up being brought under the category “record” and must be
classified – potentially leading to revised classification schemes. These activities are all the
primary responsibility of the agency records managers.

The records managers in agencies in Minnesota also play a role much closer to what
the ARM literature typically describes as an archival role, in the sense that, although the
records may not be classified as “archival” for preservation purposes, they may have mid-
term retentions that are so long (e.g., 20-30 years) they may as well be considered as archival
within the agency, since archival safeguards must be adopted to ensure that authenticity,
reliability, integrity, and usability are maintained for the long-term. A relatively large number
of records are kept for decades in order to ensure that they remain available to citizens. As a
result, the records managers must ensure not only the appropriate classification and retention,
they are also expected to ensure that their records are maintained for the long haul, although
they must rely upon MN.IT Services to ensure that the recordkeeping environments are set
up and maintained in a manner that will allow such long-term retention. Perhaps this is part
of the reason the records managers in this state have shown particular concern over the
capability of a cloud computing system to retain records with authenticity, integrity,
reliability, and usability (i.e., where by usability, they mean the ability to easily find and
retrieve them when needed).
Questions of access are more complicated in Minnesota as well, because of the existence of Minnesota’s Data Practices Law. Many scenarios occur in which agency personnel or records managers do not know exactly what type of access is appropriate. Thus, IPAD has final authority for assessing the law in these cases and making recommendations. The agency personnel themselves can choose to follow IPAD’s recommendations (or not), and this will determine whether the agency is legally at fault if the law is breached. (If the agency follows the recommendation of IPAD and is later found in violation of records law, the agency will not be held responsible.) Final access decisions on data do not fall under the decision making domain of the archivists or records managers, notwithstanding that several records managers voiced the opinion that all data should be considered record and have formal classifications applied. Neither does implementing those decisions. For the Microsoft 365 system, MN.IT Services works with Microsoft to ensure the appropriate access is provided by the Microsoft employees. Agencies make requests to MN.IT Services, who forward the requests to Microsoft. Thus the set of recordkeeping stewards performing the range of ARM functions in Minnesota includes agency personnel who create and maintain the records during their use, records managers in the agencies, MN.IT Services personnel, MHS archivists (for records classified as “archival,” after they reach the end of their agency’s retention date), the State Auditor (as a member of the Records Disposition Panel), the Attorney General (as a member of the Records Disposition Panel), the State Archivist (as a member of the Records Disposition Panel), IPAD personnel (to resolve questions related to appropriate access), and employees of or contractors for Microsoft who manage the technical system and data that is part of it. There may be additional providers with whom Microsoft contracts, but if so these organizations were not revealed during interviews. Those whose
occupation does not specifically require hands-on records management responsibilities reported that the implementation was transparent. Those who need to perform these functions reported changes in processes, difficulties understanding the functionality and performance of some processes, changes in access, and general confusion about who owns which parts of processes that have had access changes.

In Minnesota, if an employee leaves the organization, the records manager is expected to determine whether that employee is under legal hold and to notify MN.IT Services if so. Previously, the records manager was able to manually ensure that retention occurred according to policy, but they no longer have those access rights. Now, they must be sure to notify MN.IT Services before the automated retention expires, or the ex-employee’s email records will be automatically purged. Storage management is performed by Microsoft via the contract with the state, creating a vendor relationship which is managed by MN.IT Services.

In short, the functional and task-level activities associated with literature-reported ARM responsibilities are widely varied within and between these three case environments. By no means is the performance of “ARM tasks” restricted only to ARM personnel.

4.5.5.2. Stewards’ Perceptions of ARM Roles and Responsibilities

When non-records managers were asked about the roles and responsibilities of ARM workers in their organizations, they most frequently responded that (a) the archivists work for the MHS and not within the state agencies and (b) records managers are the people that draft the retention schedules. Although the interviewees reported working on committees with archivists, whom they referred to as “the records experts,” they reported recordkeeping in a manner suggesting that it is more related to the management of the records before they are
transferred to the State Archives. In addition, although the Minnesota and Kentucky records managers reported that they primarily develop or help manage retention schedules, they equally frequently said that their primary responsibilities included training and educating employees in the business units to understand the nature of and compliance responsibilities towards records and developing taxonomies or classification schemes. Although retention scheduling also falls at the top of the list of ARM responsibilities discussed within the ARM publications reviewed in this study, classification was mentioned in only 25 the articles (10%), and “education” or training was mentioned by only 11, or 4% of the articles reviewed. However, the latter frequency may be an understatement because the terms “training” and “education” as given by the interviewees comprise some of the activities which are performed within archival organizations under the rubric of terms such as “public service” and “public programming.” These two terms, if taken together with “education,” comprise up to a possible 48 articles from the ARM literature, which would cause the activities “education” and “training” to fall to the top of the list of ARM functions reported by personnel in the agencies.

As mentioned earlier, interviewees from the BioSense 2.0 case referred only to organizational entities as recordkeepers and when asked about the distribution of recordkeeping tasks, responses were similar across all respondents. The CDC is expected to ensure sustainability of the repository, by funding the Amazon contract and through its financial grants to select jurisdictions. It, along with ASTHO, is also expected to ensure that

92 These terms were not initially compiled together, so the number of times all three of them – i.e., “education,” “public service,” and “public programming” may exist in slightly less than 48 journal articles, since it is possible that several of the terms may appear in the same journal article.

93 The archivists from MHS also listed “training,” “education,” and “consultation” as among their primary responsibilities.
the security and access requirements are met for the program; they are required to inform the entire community in the event of a breach of these requirements. ASTHO’s participation is less that of a recordkeeper or recordkeeping steward than of a managing entity that serves as a bridge between the other participants.\textsuperscript{94} The partnering jurisdictions are expected to engage their own personnel to ensure that data privacy meets legal standards and the requirements of the BioSense 2.0 system. Beyond that, the BioSense 2.0 virtual organization expects the individual jurisdictions to determine their own collaborative structure, management and sharing structure, and determine for themselves who will perform the roles of Data Steward and Security Steward. Apparently because no preservation activities occur, interviewees had no response about long-term preservation of the BioSense data other than pointing out that the system is unable to handle the storage requirements for more than two years of data (P-27), due to space issues. Jurisdictions have custody (within their own local systems) of the primary data that feeds BioSense 2.0 and are not concerned about the lack of long-term data for the virtual organization itself.

4.5.6. Cloud Recordkeeping Risks

4.5.6.1. Risks Reported in the ARM Literature

Although the Internet news, government news, and business news have covered cloud computing extensively – citing both benefits and costs – the ARM academic journal literature has failed to discuss cloud computing much. ARM professionals have, however, shown an

\textsuperscript{94} In 2012 a “data access issue” was reported in the 12/20/2012 Governance Board meeting notes. Although no details were publicly available, it appears that an access breach occurred which led to the possibility that some CDC user(s) with super-administrator rights may have had access to non-federal data. It was not clear that anyone actually did see the data, but apparently the issue was such that they could have. As a result, one change that they discussed in a governance board meeting is providing the super-administrator right to ASTHO rather than to the CDC. No further meeting notes reveal whether this change has (yet) been implemented. If this occurs, ASTHO may be added to the group of stewards, depending on the responsibilities of this super-administrator.
increasing interest in cloud computing; numerous organizations and projects have taken on
the issue of cloud computing. For example, within the State of Kentucky, venue for one of
the case studies presented here, the KDLA has implemented the cloud-based preservation
cloud Preservica. That particular implementation is out of scope for this report, but some of
the information gleaned about it was able to shed some light on the Kentucky case study by
offering a forum for the archives personnel to speak on the basis of close observation about
their perception of cloud computing’s risks to recordkeeping, as discussed earlier in this
chapter.

4.5.6.2. Recordkeeping Risks as Reported by Recordkeeping Stewards

When asked about the risks of cloud computing and the concerns that they and their
colleagues had about moving to the Cloud, responses were varied. However, information
security topped the list, followed by concerns about vendor response times for access
requests. Figure 5 shows the concerns and risks that the respondents reported, with the
number of respondents who reported that particular risk showing immediately above the bar
associated with each cited risk/concern. Those who attributed risks to other colleagues in
their organizations reported that those colleagues felt loss of ownership or control over their
information, and concerns about information security. The item labeled “Legal Concerns,
eDiscovery” in Figure 5 includes mentions of regulatory compliance as well as legal issues
regarding security and eDiscovery. The third most frequent concern, “Ownership/Control
Issues” relates entirely to individual concerns that someone else in the organization (or
collaborative, in the case of BioSense 2.0) will usurp one’s control over or ownership of what
is currently perceived to be “their” information or information management process. In
contrast to published reports about the importance of making sure that cloud vendors
Figure 5 - Respondents' Reported Risks and Concerns about Cloud Computing

Respondents' Reported Risks and Concerns about Cloud Computing
By Number of Respondents Citing Each Area of Concern
explicitly (i.e., contractually) recognize that the organization is the data owner, no one in any of these three cases reported organizational data ownership as a primary concern. Two individuals, in fact, explicitly said that worries about organizational ownership were unnecessary because their own organization had made it entirely clear to the vendor that the vendor had no ownership rights – they believed this rights issue was included in the SLAs. The BioSense 2.0 respondents cited information ownership as a concern that had been raised during the development of their new organizational/collaborative structure itself, but the ownership concern was not related to the ownership that the collaborative had in the information, but rather, to the individual ownership of the data contributors themselves. In addition, personnel from North Carolina cited ownership as their key concern, noting that they relied upon their excellent trust relationship (in conjunction with the BioSense 2.0 redesign team’s educational outreach) with the CDC to accept that privacy and security were being handled well. They were not concerned about the BioSense 2.0’s ownership as a virtual organization but rather, their own individual ownership of the data that they contribute to the overall system.

For records managers, the primary issues cited were individual-level information ownership and uncertainty. The records managers noted that they did not feel they knew enough about the working of the Cloud computing system to assess risks within their organizations or to understand how best to manage the records. They also cited the inability to perform some of their tasks because of changed access rules. Within this study, the repeated concerns about ownership and control suggest that the notion of “information ownership” represents contested ground. However, the themes of information ownership and process ownership are not specific to cloud computing. In fact,
many of the reported concerns in Figure 5 have nothing to do with cloud computing in and of itself, although the respondents felt that the issues were related to “this particular” cloud computing implementation of which they spoke. Table 14 shows the set of risks and concerns that ARM professionals have reported in the ARM literature, and indicates whether that specific risk area was mentioned by interviewees within the three cases. Table 15 shows the set of concerns that respondents in these cases have reported that do not exist in the cloud literature. Several of these concerns represent risks related to cloud computing that have been publicized by IT journals and professional organizations (as opposed to the ARM literature), such as the risk associated with having one’s bandwidth reduced, thereby making it difficult (or impossible) to meet the requirements for even using the cloud environment. Another concern associated with cloud computing is the concern expressed as “reduced or no compliance with recordkeeping requirements.” This is general enough, however, potentially to include more than one risk that was cited in the ARM literature.

Table 14 - Comparison of Interviewee Reports with ARM Literature Reported Risks

<table>
<thead>
<tr>
<th>Risks Reported in ARM Literature</th>
<th>Reported by Interviewees?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access – Risk the service provider will not be able to transfer your records back to you in a usable format if you cancel your contract or they go out of business.</td>
<td>no</td>
</tr>
<tr>
<td>Access – Risk of unauthorized access due to information residing in a shared environment.</td>
<td>no</td>
</tr>
<tr>
<td>Access Time – Risk of service provider being unable to act upon requests for records in a timely manner for either eDiscovery or normal customer requirements?</td>
<td>✅</td>
</tr>
<tr>
<td>Auditability – Risk of inability to assess compliance, authenticity, integrity, and reliability due to inability to conduct regular system audits.</td>
<td>✅</td>
</tr>
<tr>
<td>Business Continuity – Risk of unacceptable system downtime.</td>
<td>no</td>
</tr>
<tr>
<td>Compliance – Risk of failing to comply with regulatory requirements due to lack of standards usage.</td>
<td>no</td>
</tr>
<tr>
<td>Contracts – Risk of failing to draft as many contracts as are needed.</td>
<td>no</td>
</tr>
</tbody>
</table>
### Risks Reported in ARM Literature

<table>
<thead>
<tr>
<th>Risk Type</th>
<th>Reported by Interviewees?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (Compliance) – Risk of failure to comply with legal requirements because of a lack of knowledge of where the records are stored.</td>
<td>no</td>
</tr>
<tr>
<td>Control (Jurisdiction) – Risk that information will be kept outside of jurisdictions with which we are comfortable.</td>
<td>✓</td>
</tr>
<tr>
<td>Disposal – Risk the information is not completely destroyed according to retention scheduling or our express request.</td>
<td>✓</td>
</tr>
<tr>
<td>Disposal – Risk we will not be able to demonstrate to courts or other interested parties that disposal has occurred appropriately.</td>
<td>no</td>
</tr>
<tr>
<td>Information Lifecycle Management – Risk that organizational recordkeepers will not know who is able to (or should) “touch” the records for particular necessary tasks.</td>
<td>✓</td>
</tr>
<tr>
<td>Information Retrieval – Risk that the provider will not offer user-friendly or accurate standard mechanisms for retrieval.</td>
<td>no</td>
</tr>
<tr>
<td>Interoperability – Risk that records cannot be moved from one provider to another if we change providers, due to lack of interoperability.</td>
<td>no</td>
</tr>
<tr>
<td>Metadata – Risk that metadata sufficient and well enough organized for ensuring context can be discovered will not be able to be created in cloud environment.</td>
<td>✓</td>
</tr>
<tr>
<td>Privacy – Risk that a breach will occur and/or that appropriate policies are not in place for notifying the appropriate people when a breach does occur.</td>
<td>no</td>
</tr>
<tr>
<td>Back-Ups – Risk that back-ups will not be restorable or will not be restorable in a usable format.</td>
<td>no</td>
</tr>
<tr>
<td>Retention – Risk that retention schedules will not be able to be applied in a cloud environment.</td>
<td>✓</td>
</tr>
<tr>
<td>Security – Risk that appropriate security measures are not in place or that secure techniques are not being used.</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table 15 - Interviewee Concerns Missing from ARM Literature

<table>
<thead>
<tr>
<th>Interviewee Concerns Missing from ARM Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear of losing ownership/control of one’s data and/or work processes</td>
</tr>
<tr>
<td>Fears that privacy of information may be breached</td>
</tr>
<tr>
<td>Fear of the unknown/uncertainty</td>
</tr>
<tr>
<td>Fear of increased numbers of information gatekeepers</td>
</tr>
<tr>
<td>Fear of losing influence in the organization/loss of authority</td>
</tr>
</tbody>
</table>

95 Although no respondent mentioned metadata explicitly, several respondents did talk about the potential problems with retrieval if the information was not appropriately arranged and classified prior to moving it into the cloud.

96 Oddly, the ARM literature cited here did not specifically reference concerns with privacy. It is possible that the authors implicitly included privacy along with security when citing potential breaches. Privacy is specifically mentioned in the non-ARM cloud computing literature discussed in the literature review, however.
Interviewee Concerns Missing from ARM Literature

<table>
<thead>
<tr>
<th>Concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear of duplicative/redundant information</td>
</tr>
<tr>
<td>Fear that processes will be more cumbersome</td>
</tr>
<tr>
<td>Fear that data integrity will be lost</td>
</tr>
<tr>
<td>Job security concerns</td>
</tr>
<tr>
<td>Fear that one will not be able to perform tasks due to lack of access rights</td>
</tr>
<tr>
<td>Disaster recovery</td>
</tr>
<tr>
<td>Concerns that retentions will not be able to be applied to information in the cloud</td>
</tr>
<tr>
<td>Concerns about managing cloud vendor when one does not understand their processes</td>
</tr>
<tr>
<td>Reduced accountability</td>
</tr>
<tr>
<td>Reduced or no compliance with recordkeeping requirements</td>
</tr>
<tr>
<td>Fear that bandwidth will not remain sufficient</td>
</tr>
</tbody>
</table>

The ARM literature mentions 24 separate types of recordkeeping risk in the Cloud. Respondents voiced eight of these risks as concerns for them individually. In addition, however, they reported another 16 concerns, only three of which are found in any cloud computing literature and only one of which is an actual risk of cloud computing itself, albeit not even specifically a recordkeeping risk. That is, concern about privacy breaches is an actual risk in a cloud computing environment. So is the fear that retention schedules may be difficult to manage in a cloud environment. In addition, the “disaster recovery” concern appears to map to the ARM literature risk cited as “recoverability of back-up data.” The other actual risk that an interviewee reported – the risk of losing bandwidth as a result of financial

---

97 Again, although the ARM journal literature does not specifically cite concerns about authenticity and integrity, these risks are discussed in the cloud computing literature already reviewed in the literature review. However, authors do not agree as to whether cloud computing adds to this risk or helps lower it.

98 This concern appears to refer to the risk of not being able to recover information in the event of a large data loss on the part of the service provider. This may be the same as the risk of not being able to retrieve one’s data during back-up restoration, although it was expressed in different terms by the interviewee.

99 This is an actual risk of recordkeeping in the cloud. Many cloud applications do not offer any recordkeeping functionality, and may in fact make recordkeeping more difficult due to the structure (or lack of structure) of the metadata.

100 The interviewee that expressed this risk cited the political cycle and the ability for funding to be cut on bandwidth, a risk which apparently turned into a reality for this interviewee’s cloud environment. However, the risk itself is not a recordkeeping risk of cloud computing. Rather it is a political risk that can influence successful technological operations under cloud computing.
impacts the political cycle – is an actual risk but it is neither a risk of cloud computing nor a recordkeeping risk. Rather, it is a political risk that can affect technical operations. While it could in some cases make cloud computing no longer a viable technological approach to information management and storage, it is a risk that could have multiple effects on a variety of different information technology structures.

Still, of the sixteen concerns cited by interviewees, only three of them were actual recordkeeping risks. The remaining apprehensions were concerns about changing organizational outcomes due to the altered power structures that the cloud computing implementation had produced.

4.6. Overall Findings

Although an interpretive study such as the one presented here makes no claims of generalizability, the themes and findings that are presented here can be used to create testable hypotheses for future studies. In addition, some of the findings presented here are consistent with theoretical frameworks from several archival and social scientific approaches. In particular, the findings provide empirical validation for theoretical approaches from Australian Continuum Theory, digital curation, and the property rights hypotheses of Brian Lavoie. They also support several social scientific theories, such as property rights theory in economics, the Psychological Ownership of Information Technology theory in organizational studies, and Resource Dependency theory in sociology. Because these theories do not represent mutually exclusive claims about the behavior exhibited in these cases, there is no inconsistency presented by an empirical case that supports several of them at once.
4.6.1. A variety of individuals from different occupational groups act as recordkeepers within their organization

These three cases showed that there is a great variety in recordkeeping roles and responsibilities across disperse occupational groups, in particular, records managers, archivists, information technologists, data practices liaisons, compliance personnel, clerks, and even epidemiologists. The tasks and functions shown in Table 10 are performed in a relatively distributed manner across the organizations studied.

Australian continuum theory asserts that the historical position taken by the records lifecycle model – that is, that records are “born,” exist in an active state for present transactions, eventually move to an “inactive” state and are then either preserved or destroyed (i.e., die) – is not an accurate depiction of the state of electronic records; in fact, the distinction frequently made between “active” and “inactive” records is not truly viable in electronic environments. Jay Atherton, “father” of the continuum theory, stated in 1985 that the “fluidity and continuity of the creation and re-creation of data”(47) and the public’s increased legal right to access has led to a situation in which “the formal differentiation between the active, dormant, and dead stages in the life of a record is becoming decidedly fuzzy” (47). Furthermore, the work of records managers and archivists are, according to Atherton, closely related and “intertwined” (47). As a result, he urges,

I believe we should replace the life cycle with a simpler, more unified model consisting of four rather than eight stages, and reflecting the pattern of a continuum, rather than a cycle. The first two stages would be the same as those in the traditional model: creation or receipt of the record and its classification within some predetermined system. I then suggest a significant change in the order. Scheduling of the information, joined with presumed later application of the schedules, becomes a separate third stage. The final element, then, is maintenance and use of the information - whether it be maintained in the creating office, an inactive storage area, or an archives. All four stages are interrelated, forming a continuum in which both records managers and archivists are involved, to varying degrees, in the ongoing management of recorded information (48).
Atherton also argued for a “combined records-management-archives goal” (51), that of the effective management of records through all stages of the continuum.

Continuum theorist Sue McKemmish asserts that the continuum model is, among other things, “a way of thinking about the integration of recordkeeping and archiving processes” (1997). The distinction between records managers and archivists that was debated in the North American arena for decades is artificial and disruptive, according to the continuum theorists. The titles “records manager” and “archivist” are somewhat arbitrary and do not represent true distinctions in their functional responsibilities. Rather, all [my emphasis] who engage in the “delivery of frameworks for accountable recordkeeping regimes that enable access to essential, useable evidence of social and business activity in the business, social and cultural domains” (1997) are recordkeepers.

This approach is also consistent with that of the digital curation theorists, whose approach is logically consistent with that of continuum theory. As discussed in Chapter 2, digital curation theorists treat curation as the active involvement of information professionals in the management of information, including its preservation. The difference between continuum theory and digital preservation theory is one of focus rather than kind. Whereas the continuum theorists have focused on the intertwined nature of records management and archival activities, digital curation theorists tend to focus on delineating the necessary technological, organizational, and management techniques needed to ensure that all information provides added value to the organization and to society.

McKemmish’s 1997 article reveals, however, that if many continuum theory articles focus on the notion of “recordkeeping” as opposed to “information management,” the spirit
of the continuum undertaking is essentially the same as the spirit of the digital curation undertaking:

Records continuum thinking is concerned about ideas about the role of recordkeeping which flow from this unifying concept - in five key areas. Firstly there is the role records play in governance, in regulating relationships between people and organisations, and as instruments of power and authority. Secondly, there is the nexus between recordkeeping and accountability in its broadest sense of accounting to each other for what we do to each other, encompassing corporate, social, cultural, and historical accountability. Thirdly, there is the role that recordkeeping plays in constituting corporate and collective memory, especially insofar as records capture experiential knowledge. Fourthly there is the way in which recordkeeping can be understood as a kind of witnessing, providing evidence of both personal and collective identity. And finally, there is the way records function as sources of value-added information and can be exploited as assets, with new records being created in the process.

Neither of these theoretical approaches focuses directly on the occupational roles and responsibilities of records managers or archivists. Rather they focus upon the records and information that must be selected and maintained for short- and long-term retention, even though that selection process will have to be revisited on a continuing basis over time.

Within the cases examined here, recordkeeping activities are highly distributed throughout the organization, and some functions may be comprised of task performance that spans several organizational and occupational groups. Portions of access provision, for example, may be distributed across an IT group, a cloud computing provider, and/or archives or records management groups. Ensuring authenticity may belong to the IT group while ensuring reliability may be shared between records management and agency personnel. Preservation of agency records is handled by both the archives (for a subset of the records) and the records management and IT groups (for a larger subset of records that are maintained for a very long period, if not permanently).
4.6.2. Thinking of stewardship as an organizational (as opposed to individual) responsibility may create disincentives for engaging in appropriate levels of long-term preservation.

All of the people interviewed in the BioSense 2.0 case thought of information stewardship in terms of organizational entities rather than individual occupational types or titles. As with interviews for the other cases, BioSense 2.0 respondents were not queried about particular occupations that were not initially included in their responses about stewardship, and none of these respondents offered occupational titles or descriptions when they identified recordkeeping stewards. Rather, they remarked that stewardship functions belong to organizational entities like “the CDC,” “jurisdictions,” or “Amazon.”

Several hypotheses arise for why the participants of the BioSense 2.0 project did not cite individuals or individual occupational types when talking about the stewardship of information. First, the personnel engaged in this project concern themselves with sharing epidemiological surveillance data. They are scientists or IT personnel who engage in collecting and creating data that can be used for population-level studies which are then reported at the population level. As a result, they may simply be accustomed to thinking about problems and data at the aggregate level and may not really tend to think about more granular data-related roles when asked about data stewardship.

Second, with one exception, the respondents were all either medical or IT professionals by training and practice.\footnote{One individual had been trained in informatics.} Individuals from occupations that are historically well-recognized as “professional” may not tend to turn their thoughts consciously toward professional roles and responsibilities as quickly as do those who belong to recordkeeping professions; for example, archives and records management have a relatively recent history
of struggling to gain the identity and status associated with professionalism. Thus, the ARM literature contains numerous cases of discussions about professionalism of the ARM occupations, as mentioned in Chapter 2. Whereas individuals from newly established professions may be more preoccupied with issues related to roles and responsibilities, those from longer- or more firmly established professions may require less conscious contemplation of roles and responsibilities.

A third possible hypothesis relates to the granularity of incentives to participate. The CDC offers competitive grants to jurisdictions who agree to participate in the BioSense 2.0 program. The money is disbursed to the jurisdiction, placing an emphasis not on internal jurisdictional or organizational structure but on the less granularly considered jurisdictional entity itself. Since most BioSense 2.0 participants are either in a position to offer grants or to receive a grant, if they are interested in joining the program they have a strong incentive to think of it as the source of competitively provided funding, and to think of other jurisdictions as potential competitors. This creates an incentive to think not about the individual organizational structures of other BioSense 2.0 participants but about the other jurisdictions as a whole as individual competitors.

Fourth, the participants explicitly distinguished between the primary data that is created at the local level and the secondary data that is fed into the BioSense 2.0 system. The respondents’ may have been more inclined to talk about specific data stewards (i.e., records managers, IT personnel, etc.) if they focused their thoughts upon the source data than when they focused their thoughts upon the secondary data that is fed to BioSense 2.0.

Finally, respondents in this case may consider the stewardship aspect of the program to be an entirely collaboratively-designed and distributed role. The initial identification of the
minimum required data elements was made by cooperative agreement between the jurisdictions, the CDC BioSense 2.0 management, and the Governance Committee, which includes members from all partner organizational types and jurisdictional levels of government. Because the initial determination of content and classification were determined by the collaborative’s representatives as a whole, and the custodianship of the data (i.e., the safe storage and the privacy and security requirements) are contracted to Amazon, using ASTHO as the manager, the representatives may actually think of stewardship itself as a joint effort.

Although the respondents themselves did not provide individual titles or occupations as stewards for the BioSense 2.0 project, the written process for onboarding jurisdictions explicitly and formally does state that stewardship is an entirely distributed role that is individually handled by each participating jurisdiction. Specifically, the formal onboarding document states that each jurisdiction will appoint two stewards – a Security Steward and a Data Steward. The Security steward assigns user access within his or her jurisdiction and holds authorized users to the “terms and conditions of data use and security standards” (BioSense 2.0 Redesign Team 2012, 2). The Data Steward controls the “level of data to be shared with other participating jurisdictions and…monitor[s] and conduct[s] ongoing quality assurance of data feeds” (2).

Each jurisdiction has its own storage location “within” the BioSense 2.0 repository; each jurisdiction ensures that its own security requirements are met; and each jurisdiction determines the level of sharing it will provide to other jurisdictions and makes sure that PII is
removed from view. Furthermore, data is only retained in Amazon AWS for two years.\textsuperscript{102} Amazon AWS acts as a custodian in the sense of providing a safe storage location that meets all separate participants’ individual security and privacy requirements (where proof of this is provided primarily through certifications). BioSense Redesign 2.0 can thus be considered a data repository only in the sense of providing a single logical “location” to which individual repositories can store and share their information. It is not an \textit{archival} repository in the sense of providing secure and safe long-term preservation of information.

In addition, no incentives (for any of the jurisdictions) exist to lead to additional scrutiny or concern about the handling of the BioSense 2.0 records outside of their own handling of this information because the data are not records in the legal definition of the term. Rather, the data is secondary information that is being shared according to a tightly drawn and secure contractual arrangement that explicitly recognizes that each organization that provides its data to the BioSense 2.0 program not only fully owns that information but is fully responsible for the accuracy, privacy, and security of the information at the time of transporting it to BioSense. Jurisdictions agree that if they are willing to place their information in the BioSense repository, they are taking on the responsibility of ensuring it is accurate and that they transport it in manners consistent with privacy and security requirements. In addition, in the event of breaches or issues on the part of the CDC, ASTHO, or Amazon, their recourse is also explicitly written into their contracts. Each jurisdiction’s

\textsuperscript{102} The respondents reported that they do not engage in long-term preservation of the data; it is only kept for two years in the repository and is then deleted. They have built authenticity and integrity checks into the technical systems and are relying upon certification and automated reports to determine that the data is authentic and exhibits integrity, on SLAs to ensure that reliability of access occurs, and on pre-negotiated forms and applications to ensure that access is offered at the “right” level of granularity and transformability.
contract is individually negotiated and represents that jurisdiction’s individual requirements for information management.

Setting up a collaborative relationship in a manner that requires the individual organizations to ensure the quality of the data they provide does not in itself put the data at risk. It allows each organization in the partnership to choose the level of control it wishes to assert over its information. From an economic point of view, this allows each organizational partner to assess the benefits and costs it receives from good quality data and thereby to privately attempt to maximize its positive goals, whether that be profit or some other mission-oriented set of goals. As will be seen in the section on ownership, this in itself creates disincentives for socially optimal levels of recordkeeping, but from a purely private view of information quality, it at first seems an efficient way of organizing recordkeeping activities.

However, what is optimal for each individual organization may not be optimal for the entire virtual organization comprising all BioSense 2.0 partners. In fact, given that a key goal of the BioSense 2.0 redesign is to facilitate information sharing, this approach is decidedly sub-optimal, if understandable. It is understandable because the BioSense 2.0 governance board recognizes that there are several competing goals they must follow to ensure that BioSense 2.0 is created effectively. Ensuring data quality is, of course, one of the goals. However, a crucial concern during the BioSense redesign has been that jurisdictions need to feel they own their own data. They also want to engage in data sharing only to the extent to which they feel comfortable. This places constraints on the ability for other members of the
BioSense 2.0 project to monitor the data quality of any particular jurisdiction. As a result, the governance board has determined that each jurisdiction needs to maintain its own data quality monitoring. However, this will reduce the incentives of each organization to maintain optimal levels of authenticity, reliability, and integrity. (Usability has been jointly determined by the BioSense 2.0 minimum set of data elements.) Benefits would accrue from a centralized records management function that would ensure each jurisdiction is, in fact, providing quality, authentic data. However, until jurisdictions are comfortable with sharing their data, such an arrangement is unlikely. This is a clear case of satisficing, as opposed to maximizing, recordkeeping activities with given cost constraints.

More discussion on the role of private versus social recordkeeping costs and benefits will occur in section 4.5.7.5.

4.6.3. Different occupational group members may have different notions of “stewardship.”

No records manager or archivist in Minnesota reported that the cloud vendor personnel act as recordkeeping stewards, although they clearly voiced their recognition that these individuals manage Minnesota records in the cloud system and may be the only individuals’ with access to perform certain records management functions that used to belong directly to the records management role. One individual did remark, however, that some of the activities previously belonging to records managers should now be considered under the ownership of IT, although this person did not refer to the cloud vendor as a recordkeeping steward.

In addition, monitoring data quality is a time- and resource-intensive activity and the BioSense 2.0 governing body may have determined that it is currently too expensive to engage in monitoring that would require additional redesign team personnel.
The MN.IT Services professionals, however, pointed to Microsoft’s cloud employees as responsible for helping them to engage in information stewardship. They did not specifically name the cloud employees as stewards but did clearly implicate them in stewardship activities, such as maintaining security and integrity of records in the Cloud. They stated that Microsoft can be a stewardship partner because of the security and legal safeguards that this vendor was required by the State to provide.

The difference in reports on information stewardship between the MN.IT Services personnel and the records managers may reflect differing characteristics of their respective professional identities or differing perceptions about the nature of stewardship of information. One perspective may be considered the “cultural value perspective” towards stewardship and the other may be considered the “information governance perspective.”

The notion of “stewardship” is found frequently in the relatively recent ARM literature (Stout 1995; Gingerich 1966; Lowell 1997; O'Meara and Tuomala 2012; Ericson and Ranger 1999; Gold 2007; Cloonan and Mahard 2010; Lazorchak 2011; Bastian, Cloonan, and Harvey 2011; Bradley 2007). This literature suggests that the notion of stewardship, having been used for a long time in the environmental community (Lazorchak 2011; Baker and Yarmey 2009), has been taken up by the library and archives professions and is used by them to represent an ethical notion of “broad cultural responsibility” (Bastian, Cloonan, and Harvey 2011). In the environmental sector, and in other areas such as religion, the notion of stewardship also represents a broad commitment and responsibility (Cloonan and Mahard 2010). Although often used interchangeably with the term “data curation,” the term “stewardship” is preferred by some ARM professionals because of its historical and ethical ties to “what we remember and forget” (Bradley 2007). These ties provide a continuity of
orientation to the ARM profession’s long-standing ties with the academic discipline of
history, in addition to newer links to the scientific data community. These two academically-
focused groups have been using the term “data steward” since the early 1990s (Committee on
Ensuring the Utility and Integrity of Research Data in a Digital Age 2009; Duerr et al. 2004).

The notion of stewardship as a broad ethical and cultural responsibility may be
termed the “cultural value perspective on stewardship.” This view is consistent with
historical ideas that archives and records management personnel are custodians of culture,
stewards of records, and keepers of cultural memory. Under this view, the notion of records
as evidence of activities and resources of continuity and of the ARM professions as
responsible for the “archival bond” hold as primary occupational identifiers. Historically,
those whose focus is on stewardship of this nature have felt that employees from other areas
within the organization do not truly understand the value of their stewardship roles because
those other employees focus directly upon the means-end relationship between the
organization’s output and input, in the form of efficient production of goods and services, as
opposed to the temporal means-end relationship between memory and current activities.

The term “stewardship” is also used frequently in organizations now in a manner
related to the popular trend toward information governance, a set of activities in which in
organizations attempt to align their business practices with strategic IT governance strategies
(Weber, Otto, and Österle 2009) rather than (purely) with the management and preservation
of information in and of itself for cultural and historical value. A data steward in a non-
archival organization is often defined by the organization to be someone who is responsible
for the content, context, and business rules of the data. Sometimes a newer category of
information worker, called a data custodian\textsuperscript{104}, is also present and is usually responsible for the safe custody, transport, storage, and business rule implementation of the data (Information Security Office 2009; ISACA 2013). In this latter situation, the stewards receive a professional status relatively equal to that of the Information Technology employees, with the delineation of stewardship duties often following lines of business. Peter Block, who first popularized the idea of stewardship (1993) in organizations, stressed the importance of creating a sense of partnership and equality within the organization and combined the notion of service with representations of a relatively flat organizational structure. In such an organization, the data steward would be recognized as having relatively equal power with other occupational groups that handle the information. They all work together as partners, according to Block’s recommendations.

In addition to focusing on content, some organizations make the data steward responsible for access decisions and security protection. They use the role of data custodian as one in which the data steward assigns specific data management responsibilities to the data custodian (NCSU 1990) and informally directs the custodian’s data management activities. For example, under the latter situation, the data steward would determine what privacy and integrity policies must be in place within his or her business area (e.g., agency), and would be responsible for making sure these rules are followed, but may assign other workers or groups as data custodians to manage the data storage and transport in a policy-compliant manner. In this example occupational status is more hierarchical than in the first information governance example. Such a stewardship structure tends to orient itself to functional and hierarchical views of information, and often provides data stewards the final authority over data, or at

\textsuperscript{104} Data stewards and data custodians are not typically job titles; they are typically roles and responsibilities assigned to particular individuals or groups within an organization.
least, gives them final accountability. Within such structures stewards assign data custodians to perform specific tasks related to the stewards’ needs. In such an environment IT may well be considered a custodian by whichever group is assigned stewardship. For example, records managers may be responsible for ensuring retention schedules are in place, but they must work with IT to ensure that the tasks that ensure compliance with the schedules are also put in place.

These two structures are common within organizations, and are often based upon the prevailing ideological orientation towards hierarchy. However, there is a wide variety of information governance models, each of which defines data stewardship in its own unique way (Thomas 2009, 15) and arranges recordkeeping structures in a variety of configurations. Noticeably, however, the two approaches do not necessarily have to be in conflict or competition. Depending upon how roles and responsibilities are assigned within an organization, one could peaceably unite both views.

The primary benefit of examining how organizational agents enact data stewardship is that one sees that the connotation of the term “steward” is quite different when viewed from the IT governance perspective than when looking at the cultural value perspective. From the information governance perspective, information is recognized as an organizational asset that provides value to the organization insofar as it is governed properly. The role of information steward is to ensure that access and use is sufficiently available to derive added value from the information. Rapid retrieval is essential and metadata should exist to ensure that the most valuable uses of the information can be made. Stewardship in this view is designed to ensure that workers can use the information to direct management’s focus toward meeting immediate and long-term organizational goals. The primary means for doing this is
through clear and consistent data definitions, efficient business process, and shared partnerships with clearly defined roles and responsibilities.

Under the cultural value approach, these goals sometimes reside in ethical tension with the stewardship goals of maintenance of cultural and memory, and provision of information as evidence. The strength of the focus on societal memory is directly related to the likelihood of tension between immediate efficiency and output goals and long-term evidential and social memory-maintenance goals.

Comparing the two approaches, the information governance perspective is a primarily future-focused intent to capture (economic) value, whereas the cultural value focus is a relatively historically-focused intent to capture context and understanding in order to add to “societal value.” It appears possible that the difference between IT personnel and records managers’ identification of data stewards in two of these cases, Minnesota and Kentucky, rests upon attitudes toward the nature of stewardship roles (i.e., “value-building partnership” versus “keeper of the memory”) and the role of information in society (i.e., “source of co-building value” versus “evidence of the past”). This difference in viewpoint toward stewardship may represent a distinction between ARM professionals and IT professionals in general with respect to information and its import, although this hypothesis requires testing with additional cases, an activity out of scope for this project.

The records managers from Minnesota appear to consider themselves as the rightful stewards in the hierarchical sense of the term stewardship, wherein they are the stewards of memory and responsible for making sure that recordkeeping activities provide long-term cultural and organizational value. They do not appear to see either IT or the cloud vendor as “inside” this stewardship domain. MN.IT Services, however, appears to consider it
appropriate to delegate responsibility for stewardship to other parties, who provide differing information management activities, as custodians.

Exacerbating the possible conflicts resulting from this difference in interpretation of the concept of “stewardship,” the records managers believe they have a lower status in these particular organizations. Whether the MN.IT Services people explicitly think they have higher status in the organization is not clear, but their comments and conversations suggest they think of themselves as service providers for their organization, fellow employees, other government organizations, and state citizens but the other employees as service providers only for their organizations. Regardless of their individual perceptions, however, in what is still a hierarchical information government structure the MN.IT Services personnel certainly do have higher status both in terms of having greater power over information resources and in terms of having greater authority over information management strategy.

Although it is beyond the scope of this research project, a key question that requires further research is whether one approach is superior in mitigating information and records-related risks and in ensuring that authenticity, reliability, integrity, and usability of information and records are ensured in an organization. A more important question, perhaps, is whether ARM professionals can maintain their hierarchically-oriented, functional approach to stewardship into the future, or whether there are social, technical, and organizational forces that are driving their mode of information stewardship out of practice entirely. If one accepts Giddens’ assertion that structure and individual identity/motivation co-evolve and co-create emergent structures in a reflexive and continuous manner, what impact will the introduction of the information governance view of stewardship and the growing trend toward large data environments have on the identity of ARM professionals and they operate
in environments requiring increasing knowledge of technologically sophisticated language and activities.

4.6.4. The distribution of information ownership and processing in a new cloud computing environment creates disincentives to engage in appropriate levels of recordkeeping.

Earlier in this chapter, it was noted that executive manager P-19 suggested that delivery stand and “real” requirements can be distinguished and furthermore, that the state can maintain ownership of information by identifying the “real” requirements. This viewpoint accepts a traditional definition of ownership that equates ownership with having control of the rights to assets and in particular, with maintaining the right to use these assets, whether tangible or intellectual, in a manner the asset owner sees fit (within boundaries defined by law and social norms) (Constant, Kiesler, and Sproull 1994; Jarvenpaa and Staples 2000, 2001). In other words, P-19 treats the information within the cloud system as a product or asset and separates the question of ownership from the question of how requirements are met and who meets them. Hence, ownership of the information, as defined by P-19, represents control over property, where the property in question is an intangible asset (i.e., information).

However, “control” is a slippery concept for intangibles such as information and may not easily be as easily distinguished from what P-19 thinks of as the delivery stand as P-19 suggests. Limited control is typically accepted as a component of ownership; certain restrictions over use are built into any concept of ownership, since ownership relies upon legal requirements for and constraints over the use of any property. That is, someone cannot do everything he or she may wish with his or her property because some uses could harm others. Further, within an organization, different agents are allowed different rights and
responsibilities with respect to the work processes and tasks involving the information. Thus, ownership itself is more complex than a simple possession and right-to-use of the (intangible) final product called “information.” It is also more complex than mere “possession.” If physical “possession” were sufficient for “ownership,” then if the state allows its data to be stored on externally owned cloud services, the cloud service provider has possession and therefore “ownership” of the bits and bytes that comprise and form the data. In fact, this case study presents a particularly complicated example of ownership, for what does it mean to say that the state owns the information within its information systems?

Several pre-existing, conceptual approaches to ownership attempt to explain how the complexities of information ownership affect information management. All of them point to a likely reduction in the quality of this management or to an increase in the risk of less effective information management under a third-party, distributed-storage and distributed-process environment such as cloud computing.

One common approach arises from traditional property rights theory in economics. Within traditional property rights theory, ownership of information systems provides the owner of those systems with a strong incentive to structure the information and its use in an efficient manner. For example, if an organization is the owner of an information system, and if the information is distributed across a number of different departments or units, the organization as a whole will be better off by standardizing data elements and structures in a manner that avoids redundancy. Rather than having the same information created and structured by a variety of departments, this view suggests that it is both more efficient and more effective to have a single source of data which all departments use for their own purposes. In fact, the organization as a whole would have very strong incentives to use its
information system in the most efficient manner because if it does so, it would profit fully from the gains derived from that use.\footnote{“Profit” is not meant here to represent the difference between revenues and costs in an economic sense but rather, by the general sense of receiving an advantage, benefit or gain. By no means does this conclusion result only for “for-profit” organizations.}

However, more recent versions of property rights theory recognize that the organization’s employees operate according to bounded rationality and further, that the “organization” (as a whole) does not make decisions – only subsets of employees within an organization have decision making power. As a result, the organization does not and cannot have complete knowledge of all the potential contingencies with respect to “its” property relations in the future. Because top management cannot formally cover every possible use of or control over the information asset in the future, the organization’s contracts related to that asset will necessarily be incomplete. Thus, there will always be residual rights to the property that are not covered by contract. This residual claim is vested in the owner, who has the right to determine how these uses will be exploited. For example, if a department within an organization owns a system that is used by other departments, and if the users pay for use of the system through some sort of lease contract (often called a “charge back mechanism”), the residual claimant is the department that owns the system. Any potential use of the system that is not specifically and formally written into the contract is assumed to belong to the department that has ownership.

This separation of ownership and use creates a number of inefficiencies of use that is well illustrated by Homburg and Bekker’s reference (2002) to the statement that “rental cars are driven less carefully than cars driven by their owners.” This inefficiency of use can also be depicted by the expected condition of a house that has been lived in by its owner as
opposed to a renter. Because the full benefits of ownership will be recouped at sale by the house owner, he or she has a stronger incentive to maintain the property in a manner that will maximize its expected profit at time of sale.

In the Minnesota case study, the IT consolidation led to a system in which all such purchases must be approved by MN.IT Services and in which all IT personnel report to MN.IT Services. Property rights theory might lead one to suggest that MN.IT Services is the residual claimant to all information systems and therefore the rightful “owner” of all information within those services. Because other units of government pay for the use (and maintenance) of those systems through lease contracts, one would expect less incentives on their part to manage the information within the systems thoroughly, ultimately resulting in lower quality information or more risk events. In many ways, this is similar to the “public goods” problem outlined by Lavoie (2003), who adopted a great deal of property rights economics in his own theoretical orientation to digital preservation. Because the individual units do not expect to gain the full benefits accruing from the use of an information system, they do not have a strong incentive to manage the information in a manner that will maximize its benefits to the organization as a whole. The attenuated sense of information ownership leads those who use the system to perform less careful actions upon it or reduce their willingness to deal with information management tasks of which they no longer see themselves as owners. In addition, because they no longer have clear accountability for the information they will not face the direct repercussion of errors of quality or omission.

Moreover, since any organization comprised of more than the barest minimum number of departments will in fact have a variety of different “mini-organizations” in the form of departments or functional units, other pressures also tend to lead to less than efficient
information management activities. The second commonly held theory about this comes from sociology - resource dependency theory – and explains this mechanism. Resource dependency theory suggests that organizations (or departments within an organization) reside in environmental “niches” which supply the resources necessary for the organizations’ existence. In order to increase the likelihood of survival, an organization will willingly share resources with others if it expects to receive needed resources from those other organizations in return. Each organization attempts to maximize the value it receives from these resources by minimizing its dependence on the other organization and maximizing the other organization’s dependence on its own resources.\textsuperscript{106} Thus, if information sharing provides value above and beyond its expected cost, the organization or department will agree to share. However, because resource sharing requires both parties to modify their practices - perhaps through data standardization, restrictive terms of use, or process handling modifications - the decision makers in each organization may be very hesitant to change the organization’s practices because frequently, such changes in practice shift balances of power and resource dependency in ways that may lead one of the organizations to gain at the expense of the other. Often the outcome is that the most powerful groups are able to exert greater influence over the terms of sharing than the less powerful units can and thereby may lead to contracts and agreements that reflect the language and culture of the more dominant group. This can lead to evolutionary changes in the resource management of the less powerful group, such that this group ceases to be viable. For example, when MN.IT Services consolidated IT in

\textsuperscript{106} The greater the reliance on another organization the greater is the likelihood that that organization will hold resources “hostage,” thereby demanding greater and greater concessions from the dependent parties. In a world of resource scarcity this will increase the dependent party’s overall resource costs and reduce the amount of resources it is able to purchase. Even if the other organizations never restrict supply in this manner, the knowledge that the risk is there will lead them to reduce resource usage in order to provide a safety net of resources in the event of such a possible restriction.
Minnesota, the individual IT departments within each agency ceased to exist. The personnel were still employed by the state, but they were removed from the control of the individual departments and no longer reported to those departments exclusively. As a result of this possibility, the management and employees of those departments that are less powerful in joint information sharing and work processes will often engage in behaviors that circumvent the agreement in ways that are geared to allow as much of their power to be retained as possible.\textsuperscript{107}

In fact, several (non-records management) interviewees in Minnesota voiced this problem, while not using the term “power shifts” explicitly. As mentioned earlier, these individuals referred to the changes in IT service provision as IT “sucking up” the personnel from other units, a phrase that suggests a sense of loss of control or autonomy.\textsuperscript{108}

The third theory, the psychological ownership of information technology theory (POIT) (Barki, Paré, and Sicotte 2008; Van Dyne and Pierce 2004), relies upon the recognition that not only actual ownership, but also perceptions of ownership (i.e., psychological ownership), create strong incentives for particular types of action upon information within organizations. Using both the idea of residual control over assets and the idea of distributed stakeholder control over organization, Carney, Anderson, and Place (2005), for example, suggest that there are different modes of ownership when dealing with intangible property such as information. For example, for digital information to be useful to

\textsuperscript{107} One way that an agency could do this, for example, is by contracting with a third-party IT person (perhaps calling this outside consultant something other than “IT,” in order to allow it to continue to engage in IT-types of activities “under the radar.” They may also attempt to contract for services that do not fit easily within predefined procurement categories. Cloud computing for example, could easily be such a service and could easily be used for such purposes.

\textsuperscript{108} This exact term was also used by an interviewee in Kentucky regarding the IT consolidation they recently underwent (P-6).
anyone, it requires hardware, software, business processes, users and other stakeholders of the entire information system. Because of distributed usage rights, individuals develop a sense of ownership over the information which is influenced by their particular rights and responsibilities over that information. Possession alone is not sufficient to identify information ownership, however. In addition, social norms, roles and responsibilities, perceptions, and cultural and legal factors unique to an organization contribute to what is treated within that organization as information ownership. Jarvenpaa and Staples (2001) agreed with this view in their study of how employee perceptions of joint ownership affect employees’ willingness to share information products and expertise within (and between) organizations. A wide variety of stakeholders hold different rights of use over information residing in organizational information systems, and these rights are defined by legal regulations, organizational mandates and customs, and social norms (Carney, Anderson, and Place 2005; Constant, Kiesler, and Sproull 1994; Jarvenpaa and Staples 2000, 2001), making ownership not merely a legal concept but also a highly normative and organization-specific concept.

For example, the issue of ownership in Kentucky was expressed quite differently than in Minnesota. Although a Kentucky interviewee (P-29) not only mentioned that cloud implementation occurred without her input into requirements but also that she did not even know there was going to be a cloud email adoption until after the adoption had occurred! Nonetheless, she expressed no particular rancor about this; neither did she express any particular concern about the recordkeeping outcomes, although she clearly knew about actual cloud computing risks to records (P-29). In addition, although the BioSense 2.0 participants reported a great deal of interest in the power distribution and the ownership of records, none
of that case’s respondents expressed concern or current worry about recordkeeping risks like security, privacy, and data quality.

Kentucky presents two characteristics that significantly change the initial power dynamics surrounding information and records management in comparison to Minnesota. In the first place, P-29 reported that within the local districts very little electronic records management occurs and to-date virtually no email archiving activities have occurred. In addition, P-29 also expressed the fact that KDE’s IT department KIDS has handled the management of email for so long that local district employees have begun to think of it explicitly as an IT task rather than a local task. Both of these factors have affected the recordkeeping stewards’ psychological perception of ownership. Specifically, because ARM personnel in the local districts do not yet engage in electronic recordkeeping to a great extent, they have not yet had the opportunity to develop a perception that they own the management of electronic records. In addition, KDE’s IT department has managed and controlled the records for so long that the localities have ceded (psychological) ownership “rights” to the IT department. Thus, one would expect a sense of stewardship over email records to occur only if the IT department exhibits that sense of stewardship, since the inclination toward stewardship will not be likely to generate from the local district employees and ARM professionals unless and until they receive some sort of accountability “kick in the pants.” That is, they would need to have the sense of information ownership and face the attendant penalties of non-action before they are likely to try to intervene in what they currently see as KIDS’ ownership responsibilities.

Discussions with an IT executive manager (P-20) suggest that this is unlikely at this point in time. When asked whether the records are sent to a preservation system after their
retention, P-20 responded, “If the district wants to keep a lot of that then they can put it on a thumb drive.” This individual also added, “I know all the lawyers try to scare you to death, yeah, ‘keep it for 30 decades’… I always thought of those things like fishing expeditions. They want to keep them a long time so they can fish and go find stuff.” Notwithstanding the potential truth of this perception, these two statements do express rather clearly the sense that this individual does not believe that the IT department has a particular responsibility towards its own role in retention. Rather, IT hands the responsibility to the agencies, who can keep information “on a thumb drive” and the agencies think the job belongs to IT. This creates a strong incentive for no stewardship activity on the records at all, since neither KIDS nor the local recordkeeping stewards have a perception that they own the recordkeeping responsibilities for electronic records.

Within BioSense 2.0, jurisdictions are assigned responsibilities for the quality of data sent to the system, but undergo no external monitoring of that quality. In addition, because the information is “only” secondary, the participants that send it to the BioSense 2.0 system have no responsibilities toward data preservation or planning for retention based upon the importance of information to the virtual organization. There are no dedicated records management personnel to investigate the value of maintaining any of the BioSense 2.0 administrative records or data for potential retention purposes, especially in light of the lack of regulatory mechanisms to deal with virtual organizations’ recordkeeping activities and accountability.

Organizational ownership of information is a form of social structure instantiated by the activities of the individuals representing the different occupations within the organization. When individuals perceive that they “own” their information and that the boundaries of this
ownership are clearly defined and uncontested, they are more willing (and more able) to engage in management of the information. For example, during the course of this project’s interviews, several records managers remarked that one impact of moving to a third-party cloud provider was that the records creators and records managers were no longer given access sufficient to allow them to make changes to or deletion of some information in the email system which had previously been their responsibility to handle; they had to cede that power to Microsoft, through requests to IT. In other words, from having originally had direct access, they were now two steps removed from that access (P-21; P-22; P-24; P-25). As a result, they began to suggest that the tasks dependent upon the access now “belonged” to IT rather than to the creators and records managers.

Because different (and sometimes conflicting) roles and responsibilities and different perceptions of ownership exist across occupational entities, one can see that speaking about state ownership of information may be perfectly appropriate in some contexts, but it may not be granular enough for other purposes. When P-19 suggested that ownership of the data has not been affected by allowing a third party vendor to take control of particular aspects of managing the information, she also implicitly suggested that IT executive management either speaks for the state or are the rightful representative of the state’s interest in its information. Given the state legislature’s willingness to give additional legislative power to MN.IT Services (by legislating MN.IT Services’ right to control all IT financial decisions made by agencies, and by legislating the consolidation of all IT services into MN.IT

---

109 I do not here mean to suggest that P-19 in any way deliberately meant to make this assertion. In all likelihood, this person simply took the most common language use of the term “ownership” and applied it to this situation. However, when one begins to see how complicated the notion of ownership truly is, especially when discussing intangibles such as information, one is led logically to questions of authority and control. Specifically, which types of authority and control are granted to which individuals and/or groups within the state? In this case, the interviewee seems to see IT as residual claimant and voice of the state in matters related to information management.
Services), this perception may be correct, from the point of view of those with the most power in the state. However, from the point of view of records managers, the impacts of that legislative decision have led to confusion over the records managers’ legitimate roles and responsibilities and have engendered a sense that although they *should* be able to act according to their occupational norms about records management roles and responsibilities, they cannot do this in all of their tasks. The ownership confusion is thus verbalized by voicing concern about the risk of losing ownership. According to property rights-related theories, this feeling of ownership reduction or loss will also lead to reduced incentives to engage in the optimal level of information management.

BioSense 2.0 reflects ownership as well, albeit in a different manner. Because ownership of information was a key factor influencing participation in the BioSense program, every jurisdiction is explicitly, legally contracted as an independent owner of that information. Rights and responsibilities are fully clarified in the contract,\(^{110}\) and a clearly outlined governance board and grievance process have been given to them. Participants appear to have no confusion about roles and responsibilities with regards to recordkeeping functions or about ownership of the information (by the time they have agreed to contract). The CDC is treated as a beneficiary (of surveillance information) and a funder in the arrangement. Although it inevitably is psychologically associated with the federal government, it is not contractually “in charge” of the program in any way. This lack of legal authority ameliorates two concerns participants might otherwise have: (1) ownership of information is clear and (2) fears of losing control to a more powerful entity are reduced significantly - jurisdictions can leave the program and take their data at any time they wish.

\(^{110}\) They are clarified as clearly as possible under bounded rationally, that is.
This particular aspect of the contractual arrangements provides incentives for appropriate levels of information management.

However, this case is also interesting because it presents several disincentives to engage in a socially appropriate level of ARM activities. Essentially, these disincentives result from similar factors to those that Lavoie voiced in his economic analysis of digital preservation incentives. That is, as a public good, accountable recordkeeping provides social value in addition to private value.

Decisions about undertaking activities occur by implicitly or explicitly taking into account the expectation of receiving benefits in return for incurring a cost. In the case of digital recordkeeping activities, the benefits accrue to participants who are not a party to the exchange. For example, with respect to BioSense 2.0, the jurisdictions expect to receive specific, private benefits: a safe and secure storage place for their epidemiological surveillance data, analytical tools to support their epidemiological research and situation awareness, greater data storage capacity, access to regional and national public health information, access to peers with whom they can share knowledge and best practice information (BioSense 2.0 Redesign Team 2013), and the possibility of receiving financial remuneration in the form of a grant from the CDC. In return, their “costs” are the costs of extracting and preparing the data, including security and privacy measures that must take place at the facility level. They must also consider risks of breach or misuse of their information, which from the point of view of a jurisdiction would be a potential unexpected cost that they must figure into their assessments. To the extent that the expected benefits to them exceed the expected costs, they will participate.
In addition, extra incentives to participate are provided to some facilities and local jurisdictions. Some states require that their local jurisdictions provide their surveillance data to the State’s Department of Public Health anyway. For jurisdictions, this means that engaging in the collection of data is a “sunk cost” when it comes to considering BioSense 2.0 participation. That is, they have to incur the cost of data preparation and transport anyway to meet their state mandated sharing, so there is no additional cost to them of sending their data to BioSense in addition, except with regards to their concerns about the risks of data sharing in general, i.e., the possibility of losing control of their “own” data and the possibility of breaches of security and privacy. The lack of additional costs to send data to BioSense increases the likelihood of their participation, while concerns about losing control of their data decreases this likelihood. If they anticipate they may receive a grant from the CDC for participation, this anticipated gain acts as an additional incentive to participate. What the final decision is for any given jurisdiction is an empirical question that depends on its cost-benefit assessment. Some jurisdictions have refused to participate on the basis of lack of resources and thus participation costs that are simply too high for them.

State Departments of Health go through the same type of benefit-cost analysis. With respect to North Carolina’s BioSense 2.0 participation, P-14 reported that since they do their own syndromic surveillance and rely primarily on telephone contact with nearby jurisdictions they don’t receive that great of a benefit in terms of the actual availability of data. However, they do receive a grant for participation and they also receive another benefit they consider substantial: by helping the CDC maintain this project, they receive goodwill from the CDC and they strengthen their ties with that entity. North Carolina is a state that already requires
local jurisdictions to send data to the State Department of Public Health, so the costs of data preparation are not significantly higher and the CDC grant largely covers those costs.

The CDC is required by federal law to engage in syndromic surveillance as part of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002, so it has strong incentives to attempt to engage as many participants as possible. This has led them to be willing to spend a great deal of time and effort educating their potential partners about the risks and benefits of cloud computing, to expend funds to hire an external consulting firm to ensure a safe and successful implementation, and to be willing to cede governance and ownership of the information sent to BioSense 2.0 to the jurisdictions.

All of the incentives cited thus far, however, are private incentives to the participants. They can be seen as private costs and benefits of the virtual organization and its members. But there are also benefits to society from this program. To the extent that a large number of jurisdictions send their data to BioSense and share that data with other jurisdictions, citizens and to medical researchers around the world benefit. The more information that is available, the more likely widespread trends in epidemiological outcomes (and causal factors) will be discovered and the quicker the response that can be expected in the event of an actual pandemic. This suggests that greater medical knowledge and improved responses to pandemic threats will accrue to society, leading to better health outcomes for citizens. In short, the greater the participation in BioSense 2.0, the greater the benefits would be even to those who play no role in the virtual organization at all. These benefits can be thought of as social benefits.

Moreover, society could achieve similarly positive benefits from the safe long-term archiving of data in the BioSense 2.0 system. The more data that is available, the greater is
the knowledge of how such an organization works, and thus, the greater are the benefits to researchers, private organizations, and government sponsors who want to understand more about collaborative enterprises (or about the role of government in health-related situations) such as this. Likewise, if there were ever a data breach, information would be readily available about the practices that occurred prior to that breach. This would be beneficial not only as a means for preparing for and handling potential legal issues; it would also be beneficial as a means to understand how to avoid such breaches in the future.

However, for society to accrue such benefits, it would require data in addition to just the epidemiological data itself. It would require management and archiving of all the official documents of the BioSense 2.0 virtual organization, which includes governance documents, records reflecting decision making activities, and other contextual information. As mentioned earlier, although the BioSense Redesign website publishes an “archive” of these documents, they are incomplete and the oldest do not appear to be available. It is not clear whether they are destroyed or deleted, or whether they reside on some server or servers but are simply no longer available for public access. In addition, the information showing all the details of administrative decision making and how the administration of the program occurs does not appear to be available. Without such documents there is no official “evidence” that BioSense operates as a virtual organization (although it clearly does). This evidence may or may not currently be kept within the boundaries of the CDC as a record of the CDC’s BioSense 2.0 program, but without transparency of such records, one cannot really know.

Certainly the lack of such records represents a loss in value to society. However, the key take-away here is that the incentives for participation on the part of contracted organizations are private and do not include assessments of the benefits to society as a whole.
Although the CDC may maintain information that is not currently accessible to the public, they do not maintain more than two years of data because the individual (private) participants have determined that the cost of doing so is too high to justify keeping information in a preservation archive.

To some extent, the lack of formal retention mechanisms occurs because, as a merely virtual organization, the BioSense 2.0 program does not face the same legal requirements that a formally instituted organization of its scope would face. This suggests a potential need in today’s collaborative environment to address the potential lack of regulatory requirements for the activities and outputs of virtual organizations.

Moreover, an underassessment of social cost is occurring in this situation and records management and archiving are probably not as prevalent as they should be if the social cost were included in benefits assessments.\(^{111}\) And in fact, the BioSense 2.0 project does not have records managers or archivists on its staff at all. Likewise, it does not engage in long-term preservation of its information at all. The lack of legal justification for preservation and the manner in which BioSense 2.0 participants customarily consider this data act as structures of signification and dominance; the individual partners do not consider the data they keep to be “records,” because it does not fit the definition that government gives to records and because the virtual organization falls outside the legal parameters for a formally instituted organization. Likewise, the partners have no compliance requirements for preserving the data over the long-term or even for considering retention requirements in the same manner as they would if this information were defined as “records” rather than “data.” Rather, they consider

\(^{111}\) It is recognized that some attempt to measure social costs and benefits may have been undertaken by the CDC personnel prior to implementing the BioSense 2.0 system. However, given that BioSense 2.0 is a direct result of the Bioterrorism Preparedness and Response Act of 2002, this need not necessarily be the case.
the information to be secondary data only, since their current conception of records suggests that all this information is duplicative - already present as records of the individual entities. However, in this case the type of information available as a whole has a value much greater than when the information is dispersed among a large number of entities and therefore not available in a single environment. To put it more simply: the BioSense 2.0 program does in fact constitute a “virtual organization” and therefore the working records that represent evidence of all the transactions (and output) of that organization are, in fact, the organization’s records, including data that could be dispersed as records among a large number of individual organizations. However, these records do not have a legal status accorded to them. Therefore, the incentives to engage in a socially optimal level of recordkeeping are not high enough to justify engaging in formal retention scheduling or preservation. If the “social value” were included in the cost-benefit assessment it is possible that the addition of this value to the private value would justify long-term preservation or greater attention to trustworthy recordkeeping responsibilities.

4.6.5. Cloud computing is used to help justify and smooth the changing power dynamics within organizations when they implement new technology perceived to represent this social structure.

The struggles for control over information point out that information ownership is one of several social structures (in Giddens’ sense of the term) that participants perceive to be a behavioral constraint in cloud computing implementations. The workers use terms such as “information,” “records,” or “data,” depending upon their occupational background, self-identity, and current position. For example, within Minnesota, the distinction between the terms “record” and “data” define not only who will set policies around retention and access, but also distinguishes the participants who talk about it. The IT workers did not use the term
“records” regularly, although they did use both “information” and “data” frequently. The IPAD respondent distinguished between “data” and “records” in order to point out her group’s responsibilities in contrast to that of the records managers. The fact that several records managers suggest that all data should now be considered “record” brings to mind the records managers’ concern that their own role of stewards of records may be threatened by the newly emerging social structure referred to as “cloud computing.” Rather than losing ground on the competition for control over “records,” as the term “record” begins to become more difficult to delineate in a rapidly evolving technological environment, records managers may choose to re-align themselves with their concept of “information” more generally (P-24).

Sue McKemmish mentions the role of information in her discussion of records more specifically. Discussing Giddens’ framework, she (1997) says,

In The Constitution of society, sociologist Anthony Giddens spoke of information as being both an allocative and an authoritative resource. As an allocative resource, it can be 'a feature of the environment, a means of production or a produced good'. As an authoritative resource, Giddens said information is 'a means of control or governance of social time-space', i.e. a way of governing and perpetuating relationships between people and organisations through time and across space. With reference to the above outline of the purposes of recordkeeping, records can also be usefully characterised in this way. As sources of value-added information, they function as an allocative resource; as evidence of activity and identity, as memory, and as instruments of power and authority, they function as an authoritative resource.

While agreeing with this characterization, this project suggests, however, that even our choice of terminology used to define our roles and responsibilities (i.e., the explicit expression of our knowledge of what a relevant information domain is), represents the contested nature of allocative and authoritative resources such as information. The confusion and frequent contradiction in usage of terms such as “records,” “knowledge,” “information,” and “data” may play a particular structural role in social systems. By blurring boundaries and
leading to confusion, they may help reduce direct confrontation over power relations, allowing room for conversation and communication between people with somewhat disparate backgrounds and identities to negotiate bilateral agreements. With “just enough” agreement in definition communication can occur; with “just enough” fuzziness, there is room for movement and ongoing negotiation.

In addition, with respect to occupational- and self-identity, the increasing difficulty with defining the term “record” in highly distributed information systems may be creating anxiety on the part of records managers, for whom the term signifies a domain of work, the rightful domination over a certain set of resources and the legitimization of their professional identity. As the term “record” begins to evolve into “data” and “information,” these individuals will be forced wonder just whose information this is. And when the answer to that question appears and points to other occupational members, the records managers feel increasing discomfort with their own identity in the organization. This confusion and discomfort may lead to the attempts of circumvention on the part of the less powerful group that were mentioned earlier in this chapter when discussing ownership.

It thus seems that otherwise vague terms such as “cloud computing,” “records,” “data,” and “information” represent loci of control and act as signifiers of meaning in Giddens’ sense of the term, both at the professional and at the organizational level. Because “information” has traditionally been considered the work of IT departments, the threat is even greater for records managers. “Information,” like “stewardship,” “cloud computing,” “records,” and “data” all appear vague and all represent contested ground within these organizations.
In addition, according to resource dependency theory, the feeling of “disenfranchisement” that records managers may feel could lead them to try to affect power relations by subverting the “rules.” Such an attempt was already mentioned earlier in the chapter with respect to Kentucky agency decisions to implement cloud services rather than use COT as a service provider. Engaging an external vendor to provide services creates a direct link with a service provider that is accountable to the agency alone. Such is not the case within consolidated IT department environments. The state IT department must act on behalf of “the state” or “the executive branch” (P-19). A cloud vendor that has contracted with the agency has to act on behalf of the agency-negotiated requirements. This allows agencies to bypass central IT decisions by simply engaging in their own contracts. Significantly in these cases, however, both states have recently consolidated their IT governance structure in order to save money for the state as a whole and to circumvent such attempts at “escape.” Cloud computing is not only an economic tool for participants; it is also a political tool.

4.6.6. The degree of worker mobility affects the participants’ reported levels of comfort with new cloud computing implementations. Another theme that arose in the three cases was that of losing control of one’s data to a more powerful entity. In all three cases the recordkeepers exhibit concerns about the impact of more powerful entities within their work environments. In fact, an additional factor influences this concern – in two of the three cases, some of the stewards have no ability to “drop out” of the cloud computing implementation. In Minnesota and Kentucky, prevailing organizational relationships are viewed by participants as constraints that records managers must accept. In short, the IT department in both states has a monopoly on IT implementation.
decisions and the organizational power structure makes it difficult for the records managers to exert a significant influence on those decisions. Thus, records managers report feeling that they must accept the decisions MN.IT Services or KIDS make. In the BioSense 2.0 program, however, the individual jurisdictions that steward the data are free to enter or leave the program as they wish. They are much more mobile with respect to the BioSense 2.0 program than are the records managers in the other two states. This helps to explain why very little frustration was expressed by BioSense 2.0 participants. As a result of this mobility, however, clearly designed contracts have been a necessity. These contracts act as a proxy for trust; by attempting to include participants’ concerns in contracts, the participants feel more inclined to participate fully. Their explicitly contracted responsibilities over the data quality also create some accountability. In the two Microsoft implementations, the contracts that influence accountability exist at the state level – with the IT departments contracting on behalf of the state with Microsoft. Microsoft is thus accountable to the state, or more specifically, to the IT Departments that manage the contracts. Because of this, one would expect that if the individual members of stewardship groups, such as records managers, do not receive the level or quality of recordkeeping services that they desire they must take their concerns to their IT departments and hope that they receive an appropriate response. The lack of formal assurances of an appropriate response (e.g., via contracts or MOUs) reduces trust and makes it less likely that actual recordkeeping problems will be resolved, rapidly or not. The ARM workers in these cloud environments do not have any contractual clout, because their service provider is a member of their own organization who has more power within that organization than they do. In short, the likelihood of reduced service levels or quantity of recordkeeping services is higher due a lack of clearly defined accountability mechanisms. In
the Minnesota and Kentucky cases, the combination of blurred (or even mistaken)
understandings of responsibilities and the break in the accountability chain between the cloud
vendor and some groups within the state organization make it more likely that errors or
omissions in recordkeeping will occur.

4.6.7. Presence or lack of trust between participants is a key factor in participants’
comfort with the implementation of cloud computing.

BioSense interviewees showed little concern about the actual stewardship risks that
cloud computing may embody. One respondent expressed her reason for having very little
concern by remarking that she had a long and successful relationship with the CDC and
considered them to be “trustful partners” (P-17). She relied on that trust when considering a
technological area about which she otherwise knew very little. She said that the CDC had
never caused a breach in any of the state’s previous data, giving her greater confidence that it
was unlikely to engage in risky behavior now. In addition, she relied upon the opinion of the
data analytic organization that her agency uses to clean and manage their surveillance data,
as well as the input from her agency’s other partners. Because those with greater knowledge
of the technology and with whom she has developed a trust relationship have given the
“green light” to the data security and other recordkeeping requirements, she is comfortable.
In addition, the BioSense Redesign team engage in a significant amount of outreach and
education about the risks of cloud computing and how these risks would be mitigated within
the program. The combination of trust and education gave her an overall comfort with cloud
computing itself and with the satisfaction of technical requirements in this particular project.

This respondent exhibits all of the types of knowledge that Duranti and Rogers (2012)
have identified as necessary for an individual to have trust in the records managed by others:
Traditionally, trust in records is based on four types of knowledge about their creator and/or their custodian: reputation, which results from an evaluation of the trustee’s past actions and conduct; performance, which is the relationship between the trustee’s present actions and the conduct required to fulfill his or her current responsibilities as specified by the truster; competence, which consists of having the knowledge, skills, talents, and traits required to be able to perform a task to any given standard; and confidence, which is an “assurance of expectation” of action and conduct the truster has in the trustee (522).

P-17 has had a long-lasting relationship with the CDC and during that time, they have performed in a manner consistent with expecting they will continue to perform in a trustworthy fashion. In fact, they have never breached confidence in their ability to manage P-17’s information. They have an excellent reputation as a foremost player in the area of epidemiological research and data analysis, according to P-17. P-17 knows researchers at the CDC and recognizes that they are competent. In fact, they are experts in their mutual field. Between the long-term relationship, the recognition of competence, and the past excellence in performance of data stewardship, P-17 has confidence that she can trust them to continue managing his agency’s data appropriately.

This facet of the BioSense 2.0 data raises the question of what types of mechanisms internal state projects and programs could use to increase the level of trust between less powerful workers such as records managers and their more powerful IT counterparts. If reputation, performance, competence, and confidence are necessary for a well-structured trust relationship, which of these characteristics are missing in the Minnesota and Kentucky cases and further, how could one stimulate the development of these trust characteristics? This question needs to be examined in much greater detail than has previously occurred in the ARM literature. Particularly, with respect to cloud computing, a key research agenda should include a detailed analysis of the nature of these characteristics, how they are subjectively perceived by different occupational groups and in particular, by archivists and
records managers, and what particular activities or outcomes information technology workers need to provide in order to increase trust and improve the incentive mechanisms that result from this trust.

4.7. Final Comments

The results presented here have highlighted a number of recurring themes that may prove important for understanding how new technological innovations such as cloud computing can and do affect recordkeeping practices in which state government plays a role. Among the findings, one appears conspicuous: when participants in the cloud implementation have discussed their concerns pre- and post-implementation, a majority of their concerns have focused not specifically on the new technology so much as on the effects the implementation has had on their positions within the implementing organizations. In fact, even those individuals who felt that the implementation was “transparent” or “highly successful” simultaneously acknowledge that the power structure had been, and in some cases continued to be, a major concern within the organizational structure. Thus, they implicitly recognized something that Giddens expressed in his definition of structure as “rules and resources, or sets of transformation relations, organized as properties of social systems” (1984, 23-24), where systems are “reproduced relations between actors or collectives, organized as regular social practices” (23). That is, for these individuals the implementation was in many respects less about the technical aspects of the systems than it

---

112 It is certainly possible that these themes may, in fact, also hold true for all organizations, or for a subset of organizations in which state government resides. However, the scope of the cases here is bounded by the inclusion of state government and thus any hypotheses resulting from this exploratory study will first examine state governmental participation. Only after becoming convinced that one can find theoretical constructs holding for that limited scope could one legitimately begin to examine situations that exclude state government in order to find out if the theories hold for a wider scope as well (Glaser and Strauss 2009).
was about the ways in which the implementation and technical aspects together would influence employees’ status within the organization and the ways in which employees would tend to consider these changes as new “constraints” on their roles and responsibilities.

That is not to say that cloud computing is not at all unique among the various types of new technologies that organizations implement. Cloud computing does indeed present some unique risks to recordkeepers, as discussed in Chapters 2 and 4. It also presents some unique affordances that influence incentives on behavior and thus the likely responses of employees to its implementation. For example, because a third party vendor becomes a highly visible partner in recordkeeping activities, recordkeeping stewards must grapple with how the entire structure will accommodate this new participant. Of course, this is also the case (to some extent) when any third party vendor enters into internal work processes. However, many third-party contracts involve tasks with well-defined boundaries, which “give” specific tasks to the vendor – tasks which are often performed without any noticeable introduction of the vendor into employee relations. For example, an organization might contract out paycheck processing as a discrete activity. This type of third-party contract creates a relatively easy-to-define boundary of responsibility and accountability that employees can use as information to re-align their own occupational areas of responsibility. In the case of cloud computing, that is not the case. Because discrete “bits and pieces” of the overall recordkeeping functions begin to reside within multiple occupational and organizational boundaries that did not perform them prior to the implementation, records managers experience confusion about their accountability and about the actual performance of tasks that they now perceive to belong to the new partner. Boundary confusion occurs and with it, anxiety about which tasks they will
be held accountable for. This confusion creates disincentives to participate in fully engaged recordkeeping.

Also, in the cases examined here, a variety of themes have been revealed, from distribution of information ownership and processing, breaks in lines of accountability, fears of power loss, participant mobility, and incentives to engage in socially optimal recordkeeping. One can see that various terms such as “records,” “data,” “information,” and “cloud computing” are quite fuzzy to most of the participants and they are defined different by the different participants. One hypothesis that has come out of this exploratory study is that this vagueness in terminology actually plays an important role in the power dynamics of organizations and in particular, in the ability for recordkeeping stewards to fulfill their responsibilities to their organizations, their professions, and their beliefs about their identity as stewards. That is, “fuzzy terms” may play an evolutionary role in smoothing the development of the emergent social structures, occupational identities, and personal identities that reflexively co-evolve as new technologies come into play in organizations.

Another hypothesis is that different linguistic use of terms such as “stewardship” may play a role in the ability for different recordkeeping occupational members to communicate with each other, and in fact, for one group to gain greater power within the organization when technological implementations occur. At the least, the realization that terms associated with an allocative and authoritative resource such as information may be divided along occupational boundaries leads one to question how a less powerful occupational group such as records management may begin to evolve as the common ways of using terms such as “stewardship,” “custodianship,” and other records- and information-related terms begin to reflect the dominant social usage more and more in the ARM professional realm of work.
Finally, in a cloud computing environment not only are recordkeeping incentives important, but the way in which cloud computing contracts affect the line of accountability within and between organizations may lead to disincentives to manage and preserve records in a socially optimal way. Thus, researchers need to begin building a greater understanding of the empirical relationship between property rights theories and actual recordkeeping environments and allocation of recordkeeping responsibilities. Which incentives play the greatest role in improving recordkeeping outcomes? Which incentives have the strongest influence on improving the coordination of diverse occupational groups’ recordkeeping activities? Because work, like all human activity, is situated and contextual, there may be no single answer to those questions, but the potentially multiple answers will nonetheless influence ARM workers’ ability to support and engage in trustworthy recordkeeping activities.
5. CONCLUSION

5.1. Recap

In this document we have examined the results of three case studies featuring cloud computing implementations which involve state government as partners in the cloud implementation. The different cases exhibited a variety of similarities and differences between themselves, as illustrated in Table 16.

These distinctions have been listed here because they may not only distinguish the organizations that undertook the cloud implementations, but could also potentially be factors that influence the outcomes of the cloud adoptions, either in terms of recordkeeper perceptions of the cloud impacts on their work or in terms of incentives to provide socially appropriate levels of recordkeeping stewardship, or both.

Two of the cases presented a “micro-level” view of recordkeeping stewards in that individuals representing the occupational groups directly involved in information stewardship discussed their perception of the implementation and its impacts on their work environment and roles. In the third case, however, a “macro-level” view was investigated, in which the participants interviewed were the decision-makers and consultants that decided to engage in the implementation and who spoke to the roles of the various organizations (rather than individuals) that act as information stewards. This presented more of a “birds’ eye” view but did provide key information regarding the role of mobility in creating incentives or disincentives for different types of recordkeeping stewardship. By examining participants who are “free” to choose to exit the cloud implementation, the visible power dynamics are
### Table 16 - Similarities and Differences between Cases

<table>
<thead>
<tr>
<th>Case Characteristic</th>
<th>MN Microsoft 365</th>
<th>KY Microsoft Live@Edu</th>
<th>BioSense 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-jurisdictional Implementation</td>
<td>---</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>State Level Entity as Cloud “Proposer”&lt;sup&gt;113&lt;/sup&gt;</td>
<td>✓</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Cross-Sector&lt;sup&gt;114&lt;/sup&gt;</td>
<td>---</td>
<td>---</td>
<td>✓</td>
</tr>
<tr>
<td>Coerced Participation of Stewards&lt;sup&gt;115&lt;/sup&gt;</td>
<td>✓</td>
<td>✓</td>
<td>---</td>
</tr>
<tr>
<td>Stewards Free to Leave Cloud After Adoption&lt;sup&gt;116&lt;/sup&gt;</td>
<td>---</td>
<td>---</td>
<td>✓</td>
</tr>
<tr>
<td>Archivists Participate in the Cloud Environment</td>
<td>---</td>
<td>✓</td>
<td>---</td>
</tr>
<tr>
<td>Records Managers Participate in the Cloud Environment</td>
<td>✓</td>
<td>✓</td>
<td>---</td>
</tr>
<tr>
<td>Contracts Drafted Between Internal Cloud Participants&lt;sup&gt;117&lt;/sup&gt;</td>
<td>---</td>
<td>---</td>
<td>✓</td>
</tr>
<tr>
<td>Archiving of Records Occurs Post-Implementation</td>
<td>✓</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Internal View of Cloud Tasks Available for Interviewer Discussion</td>
<td>✓</td>
<td>✓</td>
<td>---</td>
</tr>
<tr>
<td>All Occupational Types of Stewards Included in</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

<sup>113</sup> The term “proposer” here references the jurisdiction at which the implementation was initiated. “State Level” implies that the organization which implements the cloud is a state- rather than local- or federal entity.

<sup>114</sup> “Cross Sector” refers to all the parties undertaking the cloud implementation together; it does not reference the contracting of a private vendor.

<sup>115</sup> “Coerced” means that the recordkeeping stewards were required to participate in the cloud implementation or else leave their job.

<sup>116</sup> Again, the assumption here is that “freedom” implies an ability to exit the cloud environment without also having to terminate one’s current employment venue. This is true of BioSense 2.0 partners to the extent that the jurisdictions are accepted as being “stewards.” Within their own internal operations, once a facility makes the determination to join the BioSense 2.0 virtual organization, the archivists and records managers do not have the freedom to refuse to participate unless they choose terminate employment with the jurisdiction’s facility to which they belong.

<sup>117</sup> This excludes the organization’s contract with the cloud vendor.
quite different than within those cases in which a recordkeeping stewards must take part in the implementation even in the face of dramatic power imbalances.

Key themes that arose when discussing recordkeeping stewards’ perceptions of the cloud environment in which their organizations participated were:

- The report of pre-implementation fears on the part of one (or more) set(s) of stakeholders that they were at a power disadvantage and may lose some of their power to a more powerful entity in the implementation;

- The report of a set of stakeholders that post-implementation they felt they actually did lose power as a result of the implementation;

- For one of the cases, perceptions that roles and responsibilities had become less clear for records managers as a result of the cloud computing implementation;

- For two of the cases, at the same time or just prior to implementing the Cloud the states’ engaged in statewide IT consolidations;
  - For the case in which the implementation was an agency implementation, few impacts of that consolidation were perceived;

118 Minnesota did conduct some calculation of “Total Cost of Ownership” (TCO), but neither a complete Cost-Benefit Analysis nor a Business Case was drafted.
For the situation in which the implementation of the Cloud was statewide, respondents reported that they had difficulty distinguishing between the implementation and the consolidation. They described both as examples of IT “sucking up” their resources;

- Records managers reported that their fellow employees see them as primarily “paper-based” and that when recordkeeping involves electronic records, most of their colleagues immediately assume any question or issue is a question or issue for IT to resolve;

- The terms that represent key constructs in this study - “information,” “data,” “records,” “cloud computing” – seem very “fuzzy” to respondents, who generally defined these concepts in a variety of (often disparate) ways but nonetheless work together as if they all refer to the same thing; and,

- When discussing concerns about cloud adoptions, trust plays a key role for those who do not feel they have a clear technical knowledge of what “cloud computing” is. To the extent that stewards felt trust for the cloud implementer, the less concerned they were about risk.

5.2. Limitations

This research project was an exploratory study. It was oriented toward understanding better the a narrow aspect of cloud computing implementations – the decision making leading up to them, the perceptions of the participants, the ways in which the participants believe it has influenced their own roles and responsibilities or that of others, and their concerns about these influences. The study was undertaken to gain a granular enough view of the subject matter that one could realistically begin to create hypotheses and theoretical
constructs that are suitable for further research. For example, the participants all seemed to share a number of concerns about intra- and inter-institutional power. These concerns are consistent with the finding that power dynamics have a negative impact on work when participant mobility is low and a positive impact when mobility is high. Within the boundaries of these cases, however, the impacts appear to be tempered by the nature of the mechanisms that make mobility possible. For example, in order to ensure the greatest level of participation, private contracts were used between otherwise “equal” partners in the BioSense 2.0 system. We could see, however, that this type of private contract inherently leads to a common problem associated with social goods – unless society’s benefits of information stewardship can somehow be “privatized” so they are included in cost-benefit assessments about level and types of recordkeeping, the archives and records management services performed will be less than socially optimal.

It is important to keep in mind, however, that these findings provide some empirical support for a hypothesis; they are not “proof” of the overall theory directing that hypothesis. To begin assessing the findings and hypotheses given here, it is necessary for further study of more cloud computing implementations. Is it usually the case, for example, that records managers feel “disenfranchised” by such an implementation, or was this an anomaly that holds only for two implementations? Further, if one were to examine the BioSense 2.0 case at the “micro-level,” would we find as much satisfaction with the implementation among the individual recordkeeping occupations at the facility level? Or conversely, if we were to examine only the decision makers in statewide cloud implementations, would we learn that the satisfaction felt by the BioSense respondents was related more to their managerial capacity and decision making role than to the implementation itself? In other words, was
their assessment the result of a rational analysis of the outcomes of the implementation or was the assessment a form of self-congratulation or public relations? Looking at more decision makers could validate or dispute the notion that mobility is the prime factor here.

5.3. Import of Findings

This is the first in depth study of recordkeepers in organizations that adopt cloud computing. Of the few articles in the ARM literature that discuss cloud computing at all, none takes an empirical look at the actual environments in which the Cloud is adopted or at the actual recordkeepers who must engage in ARM practices before and after these implementations. The examination of the historical ARM literature with respect to recordkeeping functions, however, suggests that the academic literature presents somewhat unrealistic pictures of what information stewardship in a highly technological environment is like. In particular, it appears to suggest much greater power is held by archivists and records managers than is the case. In addition, it appears to suggest that the bulk of recordkeeping activities are performed or managed by records managers and archivists, when the “real world” cases suggest a much more distributed allocation of responsibilities. The continuum theorists have posited for some time that recordkeeping in organizations can no longer be considered the realm of archivists or records managers, but rather, that there are recordkeepers, that their roles and responsibilities are highly distributed, and that one must take this into account when examining ARM practices. This study has offered a view that supports that assertion. However, it also goes further than mere support – it presents some key factors that may influence recordkeepers’ ability and even willingness to engage in the quantity and quality of recordkeeping services to which they aspire as ARM professionals.
Within Lavoie’s publications on the economics of digital preservation, the discussion of incentives focused on incentives deriving from distinctions between funders, rights holders, archivists, and beneficiaries, leading to a conclusion that the structure of preservation arrangements can lead to sub-optimal levels of recordkeeping. This study goes further by showing that incentives exist such that all recordkeepers may choose to perform a less-than-socially optimal level of recordkeeping, *even while they ideologically and professionally support performing as much as possible*. This is an important possibility for ARM theorists to consider and one which has not yet been discussed in the literature. It is important because it points directly to the need to examine more closely internal, organizational structures and relationships in order to understand how best to create incentives in organizations for engaging in highly accountable and successful recordkeeping. This study also reveals the importance of including identity-related aspects of social structuration, such as how workers define their occupational and personal identity in changing technological systems, how the different ways in which they interpret “fuzzy” boundary terms influences their ability to collaborate to meet an organizational goal, and how the changing power relations affect recordkeeping incentives and expectations in new technological systems.

5.4. Further Research Requirements

Further empirical examination of the internal environment of organizations engaging in cloud computing are necessary. As Nicholas Carr predicted, “the Cloud” has become a widely used business and computing model for organizations in both the public and the private sector. In fact, as “big data” becomes used by more organizations, highly distributed technological architectures with highly centralized management will also become more
prevalent. It is important to understand how recordkeeping is affected both in theory and in practice in organizational settings such as this.

With the federal government’s focus on both cloud computing and “big data,” as well as its concern with improving records management within agencies, understanding the mechanisms that allow archival and records management functions to excel in cloud computing environments is essential. To do this requires, however, that more research examining the roles and responsibilities of the various information stewards occurs. Workers’ willingness and ability to perform recordkeeping activities may change for both technological reasons associated with new system implementations (e.g., changing affordances) and for motivational reasons that result from changing social and political structures within organizations. Currently, much recordkeeping research about activities within organizations is performed under the rubric of “information governance,” but most research on information governance is performed in IT or business management literature rather than in ARM literature. Further, most of this research focuses not on factors that affect employee motivations but instead on creating efficient information flows and developing policies and procedures to ensure such flows. While necessary to understand that procedural and structural aspect of recordkeeping, such understanding is not sufficient. Employees may react quite differently in two organizations facing a very similar technological environment because of historical, social, legal, and political differences operating within the two organizations.

Types of research projects that could move the information presented in this document forward are:
• conducting a longitudinal case study of a public sector entity as it begins to implement the Cloud, focusing on the roles and responsibilities of information stewards and the ways in which these roles and responsibilities change over time;

• conducting a longitudinal social network study that focuses on the interrelations between different recordkeeping occupational members over time;

• conducting a post-cloud implementation risk assessment of one or more public sector entities to determine the degree to which they are able to mitigate cloud risks;

• conducting an investigation of attitudes of both IT professionals and records managers to determine whether the degree of records management awareness is significantly different between the two occupations, as the records managers within the interviews discussed here have suggested; and

• conducting a study that examines the degree to which the affordances of cloud computing allow public sector organizations to meet a commonly accepted standard such as ISO 15489-1 or DOD 5015.02.
## APPENDIX A - DEFINITIONS OF CLOUD COMPUTING, BY SOURCE

<table>
<thead>
<tr>
<th>Definition</th>
<th>Source</th>
<th>Key Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>“...to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services...it is the sum of SaaS and Utility Computing, but does not include Private Clouds ... Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS), so we use that term. The datacenter hardware and software is what we will call a Cloud.”</td>
<td>Armbrust et al. (2009), p. 1.</td>
<td>SaaS  Utility computing model  Private clouds are excluded</td>
</tr>
<tr>
<td>“...a type of parallel and distributed system consisting of a collection of interconnected and virtualised computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements and established through negotiation between the service provider and consumers.”</td>
<td>Buyya et al. (2008), p. 2.  Buyya et al. (2009), p. 603.</td>
<td>Distributed  Virtualization  Dynamic provisioning (i.e., scalable on-demand)  SLAs</td>
</tr>
<tr>
<td>“...is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”</td>
<td>Cloud Computing Use Case Discussion Group. (2010), p. 6.  Wyld. (2009), p. 12.</td>
<td>Ubiquitous (i.e., one component of utility computing)  Distributed  Resource pooling  Dynamic provisioning</td>
</tr>
<tr>
<td>“...huge aggregates of various grids (academic, commercial), computing clusters and supercomputers.”</td>
<td>Delic &amp; Walker. (2008), p. 3.</td>
<td>Distributed</td>
</tr>
<tr>
<td>Definition</td>
<td>Source</td>
<td>Key Features</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>“…A large-scaled distributed computing paradigm that is drive by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms and services are delivered on demand to external consumers over the Internet.”</td>
<td>Foster et al. (2008), p. 1.</td>
<td>Distributed Economies of scale (i.e., candidate for utility computing) Abstraction Virtualization Scalable Centrally managed Dynamic provisioning Internet-oriented</td>
</tr>
<tr>
<td>“…the use of Internet-based technologies for the provision of services [1], originating from the Cloud as a metaphor for the Internet, based on depictions in computer network diagrams to abstract the complex infrastructure it conceals.”</td>
<td>Marinos &amp; Briscoe. (2009), p. 473.</td>
<td>Internet-oriented</td>
</tr>
<tr>
<td>“…a large pool of easily usable and accessible virtualized resources (such as hardware, development platforms and/or services). These resources can be dynamically reconfigured to adjust to a variable load (scale), allowing also for an optimum resource utilization. This pool of resources is typically exploited by a pay-per-use model in which guarantees are offered by the Infrastructure Provider by means of customized SLAs.”</td>
<td>Vaquero et al. (2009), p. 51.</td>
<td>Resource pooling Virtualization Dynamic provisioning Scalable Utility model SLAs</td>
</tr>
<tr>
<td>“…the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet)”</td>
<td>(Wikipedia.org 2011) <a href="http://en.wikipedia.org/wiki/Cloud_computing">http://en.wikipedia.org/wiki/Cloud_computing</a></td>
<td>Service, not product Resource pooling Utility model Network provision</td>
</tr>
<tr>
<td>Definition</td>
<td>Source</td>
<td>Features</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>&quot;...a hardware and software infrastructure that provides dependable,</td>
<td>Foster and Kesselman (1999), p. 18.</td>
<td>Dependable</td>
</tr>
<tr>
<td>consistent, pervasive, and inexpensive access to high-end computational</td>
<td></td>
<td>Consistent</td>
</tr>
<tr>
<td>capabilities.&quot;</td>
<td></td>
<td>Pervasive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inexpensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Access to high-end computers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;...an extension of the scalable computing concept: Internet-based</td>
<td>Chetty and Buyya (2002), p. 61.</td>
<td>Internet-based</td>
</tr>
<tr>
<td>networks of geographically distributed computing resources that scientists</td>
<td></td>
<td>Geographically distributed</td>
</tr>
<tr>
<td>can share, select from, and aggregate to solve large-scale problems.&quot;</td>
<td></td>
<td>Resource sharing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Large-scale problem solving</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;...a system that coordinates resources that are not subject to</td>
<td>Foster (2002), pp. 2-3.</td>
<td>Resource coordination</td>
</tr>
<tr>
<td>centralized control using standard, open, general-purpose protocols and</td>
<td></td>
<td>Non-centralized control</td>
</tr>
<tr>
<td>interfaces to deliver non-trivial qualities of service.&quot;</td>
<td></td>
<td>Standard, open, general-purpose protocols and interfaces</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-trivial QoS delivered</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;...a very large scale, generalized distributed NC [network computing]</td>
<td>Krauter, Buyya, and Maheswaran (2002), p. 135.</td>
<td>Large-scale</td>
</tr>
<tr>
<td>system that can scale to Internet-size environments with machines</td>
<td></td>
<td>Distributed</td>
</tr>
<tr>
<td>distributed across multiple organizations and administrative domains.&quot;</td>
<td></td>
<td>Scalable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Internet-based</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple organizations and administrative domains (i.e., non-centralized</td>
</tr>
<tr>
<td></td>
<td></td>
<td>control)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;...a large-scale geographically distributed hardware and software</td>
<td>Bote-Lorenzo, Dimitriadis, and Gómez-Sánchez (2004), pp. 295-296.</td>
<td>Large-scale</td>
</tr>
<tr>
<td>infrastructure composed of heterogeneous networked resources owned and</td>
<td></td>
<td>Geographically distributed</td>
</tr>
<tr>
<td>shared by multiple administrative organizations which are coordinated</td>
<td></td>
<td>Heterogeneous networked resources</td>
</tr>
<tr>
<td>to provide transparent, dependable, pervasive and consistent computing</td>
<td></td>
<td>Non-centralized control</td>
</tr>
<tr>
<td>support to a wide range of applications.&quot;</td>
<td></td>
<td>Resource coordination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple organizations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Transparent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dependable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pervasive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Consistent support</td>
</tr>
<tr>
<td>Definition</td>
<td>Source</td>
<td>Features</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>“...very large-scaled virtualized, distributed computing systems”</td>
<td>Delic and Walker (2008).</td>
<td>Large-scale, Virtualized, Distributed</td>
</tr>
<tr>
<td>“...a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed ‘autonomous’ resources dynamically at runtime depending on their availability, capability, performance, cost, and users’ quality-of-service requirements.”</td>
<td>Buyya et al. (2009), p. 603.</td>
<td>Parallel, Distributed, Resource sharing and selection, Geographical distributed resources</td>
</tr>
<tr>
<td>“...a form of distributed computing in which a virtual supercomputer is composed from a cluster of networked, loosely coupled computers, acting in concert to perform very large tasks.”</td>
<td>Marinos and Briscoe (2009), p. 5.</td>
<td>Distributed, “virtual supercomputer,” Cluster of networked, loosely coupled computers, Very-large task performance</td>
</tr>
</tbody>
</table>
APPENDIX C – MINNESOTA ARCHIVAL REQUIREMENTS FOR THE MICROSOFT
365 IMPLEMENTATION, JULY 2009

Enterprise Email – FINAL Archiving Requirements, July 30, 2009 (MN OET 2009)

Definitions – Project Scope

| Required: | The archiving solution MUST have this feature. |
| Desired – In Scope: | A desired feature that is IN scope of the email archiving project. |
| Desired – Out of Scope: | A feature that is OUT of scope for this project, but desired in future enhancements of an archiving solution. |

Definitions – Archiving Requirements

| Archived mail: | Mail that has been removed from an end user’s regular mail is stored in the archive for that end user. A stub may be used to replace the content in a user’s regular mailbox. |
| Confidential data: | As defined by the Minnesota Government Data Practices Act, it is data that is available to: Entities authorized by law; those who work requires access. |
| Content: | Email that may or may not have attachments (e.g. documents, images, audio files, etc.) |
| Deduplication: | The process of reducing content (e.g., documents, images, etc.) to a single instance. |
| Disclosure: | The rules associated with publishing, redaction, or the prevention of publication, depending on applicable statutes and/or rules associated with the definition of content information (e.g., public data, private data or confidential data as defined within the Minnesota Government Data Practices Act.) |
| e-Discovery: | Also known as “electronic Discovery” refers to any process in which electronic data is sought, located, secured, and searched with the intent of using it for Human Resource investigation or as evidence in a civil or criminal legal case. Also, electronic information is usually accompanied by metadata, which is not present in paper documents. Can span multiple mailboxes. |
| Government record: | A document described by the Minnesota Government Data Practices Act that includes public data or non-public data used for management, decisions, or operations of a government entity. |
| Importing: | The process of converting non-archived mail to archived mail. |
| Journaled mail: | A copy of mail that is created by the email infrastructure during message delivery for the purpose of capturing all communications (sent or received). Journaled mail contains all metadata from the original mail and is imported into the archive. |
| Legal Hold: | A hold placed on one or more messages, from one or more mailboxes for a set or unlimited time, that allows an agency to retain documents to support existing or potential litigation. Will override any other document retention rule. |
| Mailbox: | Identified by an email address, it is a collection of folders (e.g., “Inbox,” “Deleted Items,” “Calendar”) associated with a user, group or resource. Mailboxes are typically accessed using Microsoft Outlook or Outlook Web Access. |
| Data Practices Act: | The Data Practices Act establishes the data classification system as a cornerstone for the method used to regulate government data. This system, which must be extracted from the definitional section of the statute, is a logical way of classifying and labeling government data in terms of who is authorized to gain access to the data. |
| Metadata: | is generally defined as data about data. For our purposes, metadata means electronic data that describes such information as creation date, publication or send date, received data, creator/editor... |
user IDs or names, attached data, etc. It does not refer to text markup (such as underscore or highlights), or to internal document tracking data such as adds, changes, or deletions that might be found in a Microsoft Word document with change tracking.

**Native format:** Also known as “raw format,” is the term for data collected on the source which has not been subjected to processing or any other manipulation.

**Non-archived mail:** Mail that has not been removed from an end user’s *regular mail* and is not stored in the archive.

**Permanent retention:** Content that is retained indefinitely for various reasons.

**Private data:** As defined by the *Minnesota Government Data Practices Act*, it is data that is available to: those authorized by the data subject; entities authorized by law; those who work requires access; *Data Subject*.

**Protected data:** Sensitive data that includes personal and financial data (Social Security Number, credit card number, tax information, credit reports, etc.), Federal data (FERPA and HIPAA), State data (driver’s license number, security/access codes, passwords).

**Public data:** As defined by the *Minnesota Government Data Practices Act*, it is data that is available to anyone for any reason.

**Regular mail:** Mail that is stored in its entirety in the end user’s Outlook *mailbox*.

**Resource mailbox:** Resources are items that are available to an organization and require people to schedule or reserve the use of them. Some examples of resources include conference rooms, company cars, projectors and special equipment.

**Retention rule:** A rule/policy that specifies how long *content* should be kept in both/either the *journaled mail* or *archived mail*.

**Shared mailbox:** Are mailboxes that are used by several people for a single point of contact (e.g., project teams, groups of account or project managers).

**Single Instance Storage:** The primary copy (instance) of a mail document or mail document’s attachment.

**Spoliation:** is the intentional or negligent withholding, hiding, alteration or destruction of evidence relevant to a legal proceeding. The theory of the spoliation inference is that when a party destroys evidence, it may be reasonable to infer that the party had “consciousness of guilt” or other motivation to avoid the evidence.

**Stub:** A reduced entry (“shortcut”) in a user’s mailbox that contains the pertinent information about content that has been archived that can be used to access the full archived *content*.

### Business Requirements:

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The solution is transparent to an end-user using Outlook or Outlook Web Access (OWA) – <em>Non-archived mail</em> and <em>archived mail</em> that has been <em>stubbed</em> is accessible with a single interface (Outlook or OWA). An icon that shows the difference between <em>regular mail</em> and <em>archived mail</em> is acceptable.</td>
</tr>
<tr>
<td>2</td>
<td>Mail that is deleted from the user’s mailbox could still [be] accessible from the archive based on the retention rules.</td>
</tr>
<tr>
<td>3</td>
<td>The solution is capable of managing <em>retention rules</em> that span from 1 day to indefinitely.</td>
</tr>
<tr>
<td>4</td>
<td>Access from a Virtual Private Network (VPN) to the archiving solution using the Outlook client.</td>
</tr>
<tr>
<td>5</td>
<td>Within the solution, <em>archived mail</em> and <em>journaled mail</em> includes all <em>metadata</em> associated with the content, including user specified categories.</td>
</tr>
<tr>
<td>6</td>
<td>The solution provides the ability to create rules/policies that import <em>content</em> from <em>journaled mail</em> to <em>archived mail</em>. Triggers for this process can be based on content age, content size and/or</td>
</tr>
</tbody>
</table>
percent of mailbox quota used. Rules/policies should allow for *content* to be *stubbed* or removed from user’s view (but is still available in the archive for the duration specified by the *retention rule*).

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Within the solution, <em>archived mail</em> and <em>journaled mail</em> cannot be altered or allow for <em>spoliation</em></td>
</tr>
<tr>
<td>8</td>
<td>The solution is “Section 508” and ADA compliant – complies with accessibility laws from the Federal Government.</td>
</tr>
<tr>
<td>9</td>
<td>The solution is capable of importing Personal Store Files (.PST) into the archive.</td>
</tr>
<tr>
<td>10</td>
<td>The solution provides the ability to apply <em>retention rules</em> to specific groups of people (e.g., an agency).</td>
</tr>
<tr>
<td>11</td>
<td>The solution allows for tiered administration of <em>retention rules</em> for specific groups of people (e.g., an agency), while preventing access from other groups (e.g., a different agency).</td>
</tr>
<tr>
<td>12</td>
<td>The <em>stub of archived mail</em> references normal metadata (from/to/subject/dates/attachment names) and some of the message body.</td>
</tr>
<tr>
<td>13</td>
<td>Can the solution import messages into the archive based on the name of folder(s) within a user’s mailbox? Example: A user places <em>content</em> into pre-defined folders (e.g. “3-year retention,” “7-year retention”) for archiving. All other <em>content</em> will not be archived unless a <em>retention rule</em> specifies to do so.</td>
</tr>
<tr>
<td>14</td>
<td>The ability for administrators to modify which <em>retention rule</em> applies to a specific piece of <em>content</em>, if it is assigned an incorrect <em>retention rule</em>.</td>
</tr>
<tr>
<td>15</td>
<td>The solution provides an automated method to remove all <em>archived mail</em> from a mailbox and the archive based on agency defined <em>retention rules</em>, unless an administrative override is defined (e.g., <em>legal hold</em>).</td>
</tr>
<tr>
<td>16</td>
<td>The solution provides security of all <em>protected data</em> to authorized users.</td>
</tr>
<tr>
<td>17</td>
<td>The amount of time to access <em>archived mail</em> that has been <em>stubbed</em> should be similar to accessing <em>non-archived mail</em>, with slight delays being acceptable.</td>
</tr>
<tr>
<td>18</td>
<td>The solution allows users with the proper permissions to access <em>archived mail</em> for resource and shared mailboxes. Additionally, these users can manually select <em>content</em> to be <em>imported</em> into the archive.</td>
</tr>
<tr>
<td>19</td>
<td>The solution is capable of searching file shares for Personal Store Files (.PST), so that <em>content</em> can be <em>imported</em> into the archive.</td>
</tr>
<tr>
<td>20</td>
<td>The ability to publish search results to SharePoint for review by others (e.g., legal counsel, Human Resources)</td>
</tr>
<tr>
<td>21</td>
<td>Access to <em>stubbed archived mail</em> via BlackBerry and Windows Mobile devices.</td>
</tr>
<tr>
<td>22</td>
<td>End-users have the ability to access <em>content</em> that has been <em>stubbed</em> while in offline/cached mode in Outlook.</td>
</tr>
<tr>
<td>23</td>
<td>The ability for end-users to selectively archive content and select the appropriate <em>retention rule</em> (e.g., right-click a message and select “archive”).</td>
</tr>
<tr>
<td>24</td>
<td>The solution provides reports on the size and number of <em>archived mail</em> items an individual/department/agency has.</td>
</tr>
<tr>
<td>25</td>
<td>The solution provides a Microsoft Management Console (MMC) type administrative interface – Many users are familiar with this type of interface.</td>
</tr>
<tr>
<td>26</td>
<td>The solution provides a report or alert which details the <em>content</em> that is about to be deleted from <em>archive mail</em>.</td>
</tr>
<tr>
<td>27</td>
<td>The solution ensures <em>archive mail</em> is retained in the archive for the duration of the <em>retention rule</em>.</td>
</tr>
<tr>
<td>28</td>
<td>The solution ensures <em>archived mail</em> beyond its defined <em>retention rule</em> and <em>legal hold</em> is automatically removed from all occurrences in the archive.</td>
</tr>
<tr>
<td>29</td>
<td>Legal holds can have an indefinite/unlimited time duration.</td>
</tr>
<tr>
<td>30</td>
<td>Legal holds can span multiple mailboxes and include <em>journaled mail</em> and <em>archived mail</em>.</td>
</tr>
</tbody>
</table>
| 31 | The solution provides auditing capabilities to track who accessed the archive, their related
searches and the results returned.

| 32 | The solution contains robust e-Discovery features to search content and metadata, and place legal holds on the results for investigatory capabilities. |
| 33 | The solution provides the ability to define permissions for group(s) or individual(s) to execute search/e-Discovery investigations. |
| 34 | The solution provides the ability to export search/e-Discovery results in their native format for distribution to the requesting entity. |
| 35 | The solution has automatic classification/categorization features to index content based on pre-defined criteria (e.g., Social Security Numbers, key words). Additionally, the solution provides the ability to assign retention rules to these classifications/categorization. |
| 36 | When using the solution, all investigations contain a complete copy of content including both journaled mail and archived mail. |
| 37 | The solution provides “search within a search” functionality and can perform complex searches based on proximity and multiple criteria (e.g., content with attachments, to from, date). |
| 38 | The solution provides the ability to save search queries and their results for future review. |
| 39 | Vendor describes the solution’s ability to report, track and document agency-wide compliance with retention rules. |

**Technical Requirements**

| 1 | The solution is able to integrate with a Microsoft Exchange 2007 infrastructure in a Continuous Cluster Replication (CCR) configuration. |
| 2 | The solution is able to provide highly-available infrastructure, with multi-datacenter support for disaster recovery purposes. |
| 3 | The solution is able to deduplicate content to reduce storage space needs. |
| 4 | The solution supports the Microsoft Outlook 2003 with SP 2 and 2007 email client, as well as Outlook Web Access (OWA) from an Internet Explorer 7 browser or later. |
| 5 | The solution is able to scale with the needs of the business. |
| 6 | The solution supports a multi-forest Active Directory model for user authentication into the administrative and e-Discovery/search interfaces. |
| 7 | The solution can be monitored to ensure availability and overall system health. |
| 8 | The solution is capable of archiving individual mailboxes, entire mailbox information stores or both. |
| 9 | The solution provides granular reporting capabilities. |
| 10 | The solution is scalable for integration with other data sources (e.g., File Share, SharePoint). |
| 11 | The solution is able to deduplicate content across data sources (e.g., Email, file share, SharePoint). |
| 12 | The solution’s product roadmap includes future versions of Microsoft Exchange (e.g., Exchange 2010). |
| 13 | The solution utilizes 64-bit processes to maximize efficiency. |
| 14 | The solution connects to the existing email infrastructure via “RPC over HTTPS,” no MAPI requirements. |
| 15 | The solution supports a tiered/hierarchical storage management technology, which automatically moves content to other types of media (could be less expensive) based on defined criteria (e.g., content age). |
| 16 | To prevent access beyond an authorized time period, the solution provides the ability have permissions expire within the Administrative and e-Discovery/search consoles. |
| 17 | The solution provides the ability to document why a user was granted access to the archive.
(e.g., to document the authorizing user and date, any details about the investigation – case number).

18 In addition to deduplication, the solution has data compression methods.

19 The solution is capable of determining whether journaled mail or archived mail is considered a government record. If the solution cannot determine this, then all journaled and archived mail must be considered a government record.

20 The solution does not require any software to be deployed to end-user computers.

21 The solution provides the ability to separate/segment archived mail repositories based on agency/group (Note: may limit deduplication to the repository level).


23 Vendor: Describe the technology and method used to search journaled mail and archived mail. Describe the typical amount of them it takes to return search results.

Service Level Requirements (pending vendor selection and solution design)

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>In the event of a disaster, content that has ALREADY been archived will be accessible within 2 business days after the core Enterprise Email services have been restored.</td>
</tr>
<tr>
<td>2</td>
<td>In the event of a disaster, the content YET TO BE archived, will be imported within 3 business days after the core Enterprise Email services have been restored (read from Journaling mailbox by archiving system, or delivered to journaling mailbox).</td>
</tr>
<tr>
<td>3</td>
<td>When message journaling is enabled for a mailbox, content within a journal mailbox is to be archived within 4 hours.</td>
</tr>
<tr>
<td>4</td>
<td>In the event of a disaster, the archiving environment will be fully functional within 5 business days after the core Enterprise Email services have been restored.</td>
</tr>
</tbody>
</table>
Dear <name>,

I am writing to request your participation in a project titled “Evidence-as-a-Service: State Recordkeeping in the Cloud.” I am a doctoral candidate in the School of Information and Library Science at the University of North Carolina, Chapel Hill, examining the impact of cloud computing on recordkeeping activities in state government agencies and departments. In order to investigate the ways in which cloud computing affects recordkeeping in state government, I will engage in semi-structured interviews with state government professionals who engage in the management or creation of records or who have played a part in the adoption of cloud computing within their state. I hope that you can help me with this endeavor.

If you agree to participate, you will take part in a semi-structured interview about your perceptions and observations about the cloud computing implementation in your state and about its effects on your state’s abilities to meet its public records commitments. If you are unable to meet in person but would be willing to engage in a telephone interview, that can be arranged. I will audio-record the interview, and transcribe these notes after completion of the interview. Also after the interview, I will “clean up” my notes and send a copy to you for review, if you desire to do so. I will then analyze the interview notes to generate findings for this study. You can expect the interview to take about 45-60 minutes. If you agree, I may also contact you through email or telephone with brief follow-up questions to clarify or extend your comments from this interview. You do not have to agree to be contacted with follow-up questions in order to take part in the one hour semi-structured interview, however.

I have attached a copy of the consent form for your review. If you agree to be interviewed, I will also send you a copy of the interview questions, will send you further instructions about returning the consent form to me, and will schedule a mutually convenient date for the interview.

Thank you for considering to participate in this important study on public information governance. I hope you will be willing and able to speak with me!

If you have any questions prior to agreeing to participate, please contact me at 919-357-1363 or reply to this email (at lorraine.richards@unc.edu).

Regards,

Lorraine L. Richards, Doctoral Candidate
School of Information and Library Science
University of North Carolina, Chapel Hill
Consent to Participate in a Study

Evidence-as-a-Service: State Recordkeeping in the Cloud

**Investigator:** Lorraine (Lori) L. Richards, Doctoral Candidate, University of North Carolina School of Information and Library Science

**Description:** This research project investigates the impact of cloud computing implementations on recordkeeping in state government, focusing on the ways in which the requirements that inform the cloud computing implementation affect the ability of recordkeeping decision makers and records creators to comply with the recordkeeping requirements in their respective states. This study has implications for those involved in research, in information technology, and in archives and records management activities in the public information environments.

**Procedure:** If you agree to participate, you will take part in a semi-structured interview about the cloud computing implementation in your state and/or agency and about the ways in which your work and perceptions regarding recordkeeping activities have been affected by the implementation. I will take notes and, if you agree, make an audio recording of the interview. A transcriber will then create a transcript of everything said during the interview. I will analyze interview transcripts in order to generate findings for this study. You can expect the interview to take about 45-60 minutes. If you agree, I may also contact you through email with brief follow-up questions to clarify or extend your comments from this interview.

**Expected Benefits:** Although you may not receive direct benefit from your participation, others may ultimately benefit from the knowledge obtained in this study.

**Confidentiality:** In all reports and publications associated with this research, I will report data in a way (using participant numbers or pseudonyms) that does not disclose your identity. After completing the data analysis, I will erase the audio tape of this interview and any email messages you have sent me in response to follow-up questions. I will also dispose of any information that identifies you as an individual.

**Right to refuse:** Your participation is completely voluntary. You may skip questions that make you uncomfortable, and you are free to withdraw from participating at any point.

**Questions:** If you have any questions about this study, please feel free to contact:

**Principal Investigator:**
Lorraine (Lori) Richards, Doctoral Candidate
University of North Carolina, Chapel Hill
Dissertation Chair:
Dr. Christopher (Cal) Lee
University of North Carolina, Chapel Hill
216 Lenoir Drive,
CB #3360, 100 Manning Hall
Chapel Hill, NC  27599-3360
callee@email.unc.edu
919-962-7024

Should you have questions regarding your rights as a research participant, please contact the
Institutional Review Board at the Office of Human Research Ethics, CB 7097, Medical
School Building 52, 105 Mason Farm Road, Chapel Hill, NC  27599-7097, (919) 966-3113,
email: irb_questions@unc.edu.

Documentation of consent: One copy of this document will be kept with the research
records of this study. You will also be given a copy to keep.

Consent to Participate:
I understand and agree to all of the above. Lorraine (Lori) Richards has offered to answer any
questions I may have concerning the study.

________________________________________
Printed Name Consenting signature
DATE: ____________________________________

Audio Recording:
Please sign below if you are willing to have this interview recorded on audio tape. You may
still participate in this study if you are not willing to have the interview recorded.

________________________________________
Signature Date

Follow-up Email Correspondence:
Please sign below if you are willing to be contacted later with questions I might have about
your comments from this interview. You may still participate in this study if you are not
willing to engage in follow-up email correspondence.

________________________________________
Signature Date
Disclosure of Your Participation to Prospective Interview Participants:
In order to recruit additional interview participants for this study, it can be useful to inform them that someone they know has already taken part. Please sign below if you are willing to allow me to inform prospective interview participants that you have already participated in an interview. No reports or publications associated with this research would disclose your identity, and I will not provide any information to other participants about what you have said during the interview. You may still participate in this study if you are not willing to share your name with prospective participants.

_________________________________________
Signature Date
APPENDIX F – MINNESOTA RECORDKEEPING LAWS

The statutes most directly affecting information and records management and data practices in Minnesota are:

- Minnesota Statute 13, the “Data Practices Act”;
- Minnesota Statute 138.17, The “Records Management Act”;
- Minnesota Laws 1971, Chapter 529;
- Minnesota Statute 15.17, the “Official Records Act;” and
- Minnesota Statute 325L, the “Uniform Electronic Transactions Act.”

The Data Practices Act and the Records Management Act are often related. In general, the Records Management Act will require the creation of records that reflect the state’s ongoing business, whereas the Data Practices Act is related to the provision of access of those records to the public (IPAD 2000). In some cases, a relationship holds between the state’s data practices activities and the Minnesota Open Meeting Law, represented in Minnesota Statute 13D (P-16). For example, cases in which records which are classified as private by the Data Practices act may be considered public if they are discussed as a matter of procedure at an open meeting that is regulated by the Open Meeting Law (IPAD 1997). In some cases, portions of a record or information may be considered public and portions may be considered private, in which case someone redacts the record (IPAD 2000).

Data Practices Act (Minnesota Statute 13)

Minnesota’s Data Practices Law regulates access to government information. It attempts to balance the right to privacy with the responsibility for transparency in government action. The law states that all government information in the state is considered open to public inspection unless it has been explicitly classified as non-public by law or by temporary classification. The classification of non-public must be defined by legislative statute or by temporary classification. Although an agency can temporarily classify a piece of information as confidential (in the case of data about persons) or protected, non-public (in the case of data about non-person entities), this temporary classification must be approved by the Commissioner of Administration and by the legislature in order to remain non-public (Gemberling and Weissman 1982). If this permission is not forthcoming, it automatically becomes public after two years. The Data Practices Act operates in conjunction with the Records Management Act and in some cases can present seemingly inconsistent requirements for disclosure or non-disclosure of information. However, the primary point regarding data practices in Minnesota for the purposes of this dissertation is that all government information is considered public and therefore accessible by the public unless there is a specific law that classifies the information as non-public (or an agency temporarily classifies it as non-public).

Records Management Act (Minnesota Statute 138.17)

The Records Management Act gives the State Records Disposition Panel the power to direct the sale or destruction of records deemed not to be of permanent value and to direct the disposition (“by gift”) to the Minnesota Historical Society of records that are deemed to be of permanent value. It also specifies when agencies must submit records to the State Archives,
allows the State Archives to inspect records that are listed on a state records disposition schedule, and designates that all records in the State Archives are open to the public unless they are specifically classified as non-public by the archives for a variety of reasons similar to those specified in the Data Practices Act.

**Minnesota Laws 1971, Chapter 529**

Minnesota Laws 1971, Chapter 529 abolished the previously existent state archives commission and transferred records management oversight to the Commissioner of Administration. It empowered the Commissioner to establish standards, procedures, and techniques for effective management of public records, to make continuing surveys of paper work operations, and to recommend improvements in current records management practices including the use of space, equipment, and supplies employed in creating, maintaining, preserving and disposing of public records (Laws of Minnesota 1971, Chapter 529, 962).

It included Statute 138.17 within its text and specified that individual agencies, “when requested by the Commissioner,” must assist in creating an inventory of records in their possession and attaching to that inventory a records disposition schedule that “establishes a time period for the retention or disposal of each series of records” (Laws of Minnesota 1971, Chapter 529, 962). It also specified the provision of storage space for the MHS and gave MHS the right to adopt rules and regulations governing its own procedures and performance.

**Official Records Act (Minnesota Statute 15.17)**

Minnesota Statute 15.17, the “Official Records Act,” specifies that “all officers and agencies of the state, counties, cities, towns, school districts, municipal subdivisions or corporations, or other public authorities or political entities within the state … shall make and preserve all records necessary to a full and accurate knowledge of their official activities” (Minnesota Statutes 2012, Section 15.17). It also specifies that it is the duty of each agency and its chief administrative officer to carefully preserve all its records from “deterioration, mutilation, loss, or destruction” (Minnesota Statutes 2012, Section 15.17).

**Uniform Electronic Transactions Act (Minnesota Statute 325L)**

The Uniform Electronic Transactions Act states that if a law requires that a record be retained, the requirement is satisfied by retaining an electronic record of the information in the record. This includes retention of the front and back of checks. In essence, this law says that the storage medium of records is irrelevant for legal purposes. The law is considered a “uniform” law because it is binding in the same manner upon all states in the United States that adopt it and it has been adopted by most of those states.

**Open Meeting Law (Minnesota Statute 13D)**

The Minnesota Open Meeting Law, Statute 13D, specifies that all governmental meetings (with a very few exceptions) must be open to the public unless the public body has, prior to the meeting, stated on record that the meeting will be closed, what the grounds for closing the meeting are, and what the subject to be discussed will be. It also requires that the votes on any action taken must be recorded and that this record must be open to the public during normal business hours (Minnesota Statutes 2012, Section 13D).
APPENDIX G – LEGAL ENVIRONMENT AFFECTING BIOSENSE 2.0

Public Health Security and Bioterrorism (BT) Preparedness and Response Act of 2002

This Act does not specifically discuss recordkeeping for the BioSense 1.0 or BioSense 2.0 systems, but it does provide a mandate for the CDC to work together with other federal agencies to monitor records and information related to potential public health risks. It also authorized funding for this activity to be allocated to the CDC for this purpose, thereby providing the initial funding and mandate for BioSense 1.0.

Federal Information Security Management Act (FISMA)

Enacted in 2002 FISMA (“FISMA,” 44 U.S.C. § 3541 et seq.) is Title III of the E-Government Act (Public Law 107-347), passed by the 107th Congress and signed by President George W. Bush in December 2002 (section 3541 title 44). This act requires each federal agency to “develop, document, and implement an agency-wide program to provide information security for the information and information systems that support the operations and assets of the agency, including those provided or managed by another agency, contractor, or other source” (NIST 2002). The act defines information security to be “protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction” in order to provide integrity, confidentiality and availability (44 U.S.C. § 3541 et seq., 2002).

Close ties can be seen between FISMA compliance and traditional records management responsibilities by examining what is meant by “integrity,” and “availability” within the FISMA Act. Integrity refers to “guarding against improper information modification or destruction, and includes ensuring information nonrepudiation and authenticity” (44 U.S.C. §3542, 2002). Integrity and authenticity are two of the primary requirements for recordkeeping systems asserted in the ISO 15489-1 standard. Availability in FISMA refers to “ensuring timely and reliable access to and use of information” (§3542). Reliability and usability are inherent to FISMA and are also primary requirements for recordkeeping systems, according to ISO 15489-1. Thus, FISMA builds directly into its requirements the four primary ISO 15489-1 requirements for a recordkeeping system.

Although not referred to as such, FISMA is essentially a best practice “Records Management” Act focused specifically on successfully managing risks to information security. The CIO Council categorizes FISMA as a Cybersecurity / Information Assurance requirement while it categorizes ISO 15489-1 as an Information and Knowledge Management requirement. It asserts that both of these requirements represent fundamental areas of competency that a federal CIO must ensure are represented within the “knowledge, skills, and abilities” of the government workforce (CIO Council 2012).

BioSense 2.0 complies with FISMA and has also been through the Certification and Accreditation process performed by CDC security personnel. For example, it does not collect data from state and local health departments that contain personally identifiable information (PII). Rather, all data in the application is aggregated” (Gallagher 2012). In fact, through BioSense 2.0 “the CDC is the first government agency to complete all the necessary certification requirements for hosting health data in a public cloud” (Dublin 2012).
H.R. 1163 (Federal Information Security Amendments Act of 2013)

H.R. 1163 enhances FISMA by “improving the framework for securing federal information technology systems” (GOP.gov 2013) and by amending FISMA to reestablish oversight authority of the Director of the Office of Management and Budget (OMB) with respect to agency information and security policies and practices. In the Senate it was read twice and is currently under review by the Committee on Homeland Security and Governmental Affairs (Congress.gov 2013). Thus, like FISMA, it is related to records management activities and is a curation requirement insofar as it relates to maintaining information in a legal and secure manner.

Health Insurance Portability and Accountability Act (HIPAA)

HIPAA, enacted in 1996, attempts to combat fraud and abuse in health care (Stephens 2007). One of the means by which it does this is by placing stringent privacy regulations on medical records. Thus “managers of medical recordkeeping systems must ensure that patient-specific records are stored, maintained, transmitted, and accessed in a secure fashion, so as to protect the privacy of the individuals to which they relate” (99).

American Recovery Reinvestment Act (ARRA)

Public Law 111-5, the American Recovery Reinvestment Act (ARRA) of 2009 was designed to create and preserve jobs within the United States (US HHS 2013). The key relevancy of ARRA to recordkeeping is that it incorporates the Health Information Technology for Economic and Clinical health (HITECH) Act of 2009 within it.

Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009

This act increases the severity of monetary penalties for data breaches of health information (US HHS 2013; Horowitz 2007). In addition, it includes monetary incentives to eligible practitioners and eligible hospitals that meaningfully use Certified EHR (electronic health record) Technology” (HealthIT.gov 2013). The HITECH Act calls for the “development of a nationwide health information technology infrastructure that allows for the electronic use and exchange of information and that…promotes a more effective marketplace, greater competition...[and] increased consumer choice” (US HHS 2009, §3001(b)).

Individual State Laws

Although citing all relevant state laws is outside the scope of this document, two are provided here from North Carolina by way of example. The North Carolina Preparedness Response Budget of 2004, a byproduct of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (govtrack.us 2001-2002), provided funds to North Carolina’s DPH to develop a system that monitors emergency department visits for public health surveillance purposes. The DPH was already working with UNC to develop an emergency department database for the specific use of DPH and the University of North Carolina (UNC) to “emergency department administrative data” mainly (P-17), so this act allowed the DPH to expand its stakeholder base and information collection activities. With the additional funding, DPH collaborated with the North Carolina Hospital Association, which is a “really powerful organization in North Carolina” (D-17), to bring in more
hospitals on a voluntary basis. A year later, NC General Statue § 130A-480 turned the voluntary relationship into a legal requirement. North Carolina General Statute § 130A-480 requires that "for the purpose of ensuring the protection of the public health, the State Health Director shall develop a syndromic surveillance program for hospital emergency departments in order to detect and investigate public health threats that may result from (i) a terrorist incident using nuclear, biological, or chemical agents or (ii) an epidemic or infectious, communicable, or other disease" (NC DHHS 2012). Although NC DETECT originated as the result of a 1999 pilot project designed to demonstrate the value of collecting emergency department data for the purposes of surveillance and research (Castillo-Chávey 2001, 47), it now provides twice-daily feeds to BioSense 2.0 and has been provided the CDC BioSense data since 2007 (P-14). In 2005, NC General Statue § 130A-480 specified that the emergency department data must be submitted to the State Health Department. Not explicitly stated was the right of the DPH to forward the data to BioSense. In 2007 the additional verbiage was added that the DPH is explicitly allowed to share the de-identified data with the CDC (NC General Statue § 130A-480) (P-14; P-17).
APPENDIX H – KENTUCKY RECORDKEEPING LAWS

KRS 61.870 – 61.884, Kentucky Open Records Act
The Open Records Act establishes a right of access for citizens to public records, with several exemptions that specify which public records are not classified as “open,” and therefore not available for public access. The Kentucky Open Records Act provides the details which individuals must use to request a copy or a viewing of a public record, the methods which agencies and public officials must use to deny the request if they choose to deny it, and the recourse that the requestor can take if his or her request is denied. Because electronic records are also included in the definition of a public record, this law holds for both paper copy and electronic records.

KRS 171.410 – 740 (KY State Archives and Records Act)
KRS 171.410-740 specifies the rules and regulations associated with the archiving and records management activities under the purview of the KDLA. KRS 519.060 (Kentucky Public Records Statute).

KRS 171.223
This is a section of KRS 171 that spells out the responsibilities of the KDLA with respect to its actions under the Kentucky Open Records Act. Specifically, it requires the KDLA to provide, within 60 days, information about the proper retention and management of public records to the State Attorney General’s office whenever the Open Records Act is modified.

KRS 519.060 (Tampering with Public Records Statute)
This statute classifies “tampering with public records” as a Class D felony (Ky. Rev. Stat. 61:950(1992)).

725 KAR 1:010
This statute specifies the responsibilities of state agency records officers as noted above in “Recordkeeping Stakeholders” (Ky. Admin. Regs. 1:010 (2013)). Other portions of 725 KAR 1 relate to various activities associated with records management, for example, scheduling public records (KAR 1:030), creating specific procedures for an agency to collect and distribute public records (KAR 1:040). All of these administrative regulations refer specifically to the statutes list in KRS 171.

KRS 434.845 – 850 and KRS 434.855
These portions of KRS 434 relate specifically to unlawful access to computer information and misuse of computer information, respectively. Although not specific to records, they hold true for any electronic information and are therefore comprehensive.

KRS 15.257
This statute requires the State Attorney General to distribute explanatory materials to all public officials (explicitly including school superintendents) and to designated attorneys within 60 days of a modification to the Open Records Act. It also requires these officials to
provide the name and contact information of the attorney to which they wish these materials to be distributed.

**KRS 369.101 – 369.120 (Uniform Electronic Transactions Act)**

This law regulates the use of electronic records and electronic signatures in certain transactions not otherwise covered by rules about electronic records. It covers electronic records and electronic signatures related to business and government transactions. It does not require that electronic records or signatures be used within the course of these transactions but it does allow for them to be used with both parties’ agreement and it provides the rules and regulations regarding such use. Section 369.112 covers the requirements for retaining electronic records, noting that if an electronic record is an accurate reflection of the information set forth in the record, whether it was electronic or otherwise, that it is acceptable to use the electronic record for retention purposes.


———. “Who Are the Archivists and What Do They Do?” *American Archivist* 51, no. 1/2


370


Centers for Disease Control and Prevention. Key Findings from Public Health Preparedness: Mobilizing State by State. Washington, D.C.: Centers for Disease Control and


——. "Exploring Perceptions of Organizational Ownership of Information and Expertise." 

Jick, Todd D. "Mixing Qualitative and Quantitative Methods: Triangulation in Action." 


http://technology.ky.gov/about/Pages/HistoryofITOrganizationsinKY.aspx (accessed March 18, 2014).

———. “COT Background and Business Drivers of Technology.” Kentucky Commonwealth Office of Technology (COT) 2013b. 


403


