

SCHOLARS AND THEIR BLOGS: CHARACTERISTICS, PREFERENCES, AND  
PERCEPTIONS IMPACTING DIGITAL PRESERVATION

Carolyn F. Hank

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in  
partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of  
Information and Library Science.

Chapel Hill  
2011

Approved by:

Helen R. Tibbo

Lynn Silipigni Connaway

Jeffrey Pomerantz

Paul Jones

Richard Marciano

## **ABSTRACT**

**CAROLYN F. HANK: Scholars and their Blogs: Characteristics, Preferences, and Perceptions Impacting Digital Preservation**  
(Under the direction of Dr. Helen R. Tibbo)

This descriptive study investigated scholars who blog in the areas of history, economics, law, biology, chemistry and physics, as well as attributes of their respective blogs. It offers an examination of scholars' attitudes and perceptions of their blogs in relation to the system of scholarly communication and their preferences for digital preservation. Further, it investigates blog publishing behaviors and blog characteristics that influence preservation action. Findings are drawn from 153 questionnaires, 24 interviews, and content analysis of 93 blogs. Most feel their respective blogs are representative of their cumulative scholarly record. A majority see their blogging as benefitting several aspects of their scholarly lives, contributing to a sense of improvement in their teaching, writing, and research activities, as well as in communications with peers. It was found that scholars who blog are generally interested in blog preservation with a strong sense of personal responsibility. Most feel their blogs should be preserved for both personal and public access and use into the indefinite, rather than short-term, future. Scholars who blog identify themselves as most responsible for blog preservation. Concerning capability, scholars perceive blog service providers, hosts, and networks as most capable. National and institutional-based libraries and archives, as well as institutional IT departments, are perceived as least responsible and capable for preservation of scholars' respective blogs. Although over half of questionnaire respondents save their blog

content, in whole or in part, and many interviewees expressed a sophisticated understanding of issues of digital preservation, the findings also indicate that bloggers exhibit behaviors and preferences complicating digital preservation action, including issues related to rights and use, co-producer dependencies, and content integrity. For example, most use a blog publishing application and hosting service, many report editing and deleting of blog posts after publication, and less than half of blogs feature explicit rights and use statements.

To my Mom and Dad for supporting whatever it is that I dream up, and to my most important dreams come true: Maddy, Kennedy, Raegan, Izzy, Vin and Jackie.

## **ACKNOWLEDGEMENTS**

I have read a few dissertation acknowledgements that use the “it takes a village” metaphor to acknowledge those contributing to the respective authors’ success. In my case, I think it more apt to say it takes several “villages:” faculty, friends, and colleagues at the School of Information and Library Science (SILS); friends and colleagues at the University of North Carolina at Chapel Hill (UNC), particularly the University Libraries; friends and colleagues from other institutions in the field; and family and friends who have been supportive of me long before this journey toward a Ph.D. began.

A much deserved thank you to the members of my committee, Drs. Helen R. Tibbo (Chair), Lynn Silipigni Connaway, Jeffery Pomerantz, and Richard Marciano, and Paul Jones. They provided me with exceptional support throughout. In addition to serving on my committee, I also had the privilege of working with each of them on other projects throughout the years. My committee members not only provided excellent guidance in my dissertation completion, but also greatly contributed to my overall professional development.

Since this is Carolina, I think a basketball reference is appropriate. I was lucky to have the guidance of several other individuals who each served as a sort of “sixth man” on my dissertation committee, providing suggestions and guidance as I developed my research design and wrote my dissertation. Thank you to my “bench,” Drs. Christopher (Cal) Lee and Deborah Barreau, and Yvonne Belanger. Additionally, SILS faculty and staff, both current and former, were always available to listen, encourage and provide other support over the years. A thank you to the exceptional faculty and staff at SILS.

I'd like to make special mention of my relationship with Helen. I have worked closely with Helen since my first day in the Ph.D. program and on a range of projects. Helen not only provided me with many opportunities for professional development, she also encouraged and supported my own ideas. Our relationship extended beyond simply advisor-advisee; she made me feel much more like a trusted, respected colleague than simply an advisee, and her influence and impact is evident from any casual look at my CV.

Further, our close working relationship and the resulting opportunities for developing my research, teaching and service experience enabled me to cultivate relationships with many members of the campus community as well as researchers, educators, and practitioners at other institutions. While I have benefited professionally in many ways from the many contacts made through these opportunities, I do want to specifically acknowledge a few of those who offered support and encouragement specific to my dissertation pursuit. At the University Libraries at UNC, thank you to Dave Pcolar, Natasha Smith, Janis Holder, and members of the Digital Curation/Institutional Repository Committee and the Carolina Digital Repository program and Steering Committee. Also, thank you to Drs. Kathy and Bill Wisser, Simmons College; Richard Davis, University of London Computer Centre; Dr. Megan Winget, University of Texas at Austin; Dr. Nancy McGovern, Inter-University Consortium for Political and Social Research, University of Michigan; Dr. P. Bryan Heidorn, University of Arizona; and Dr. Seamus Ross, University of Toronto.

The research reported in this dissertation was informed by an earlier exploratory study on bloggers' perceptions on digital preservation. I would like to thank my collaborators on that study, Dr. Songphan Choemprayong and Laura Sheble, as well as my valued dissertation committee member, Paul Jones, who served as our study's faculty advisor. This

earlier work was extremely beneficial in helping me to design and implement my dissertation research.

My parents let me live out my long ago “only child” fantasy, cooking, cleaning, and otherwise providing for me during the three months I wrote this dissertation (and in the process, with photos to prove it, letting me take over two guest bedrooms, the dining room table, a coffee table in the den, and about one-fourth of their kitchen counter space). But this “only child” depiction is not entirely appropriate as my siblings, as well as friends, opened up their homes to me and my “mobile office” during this time. In particular, my sister Lara and her husband Mike let me take over their phone line for many hours over an extended period to conduct interviews, as did my dear friend Craig. As they always do, my family provided encouragement and support throughout my Ph.D. pursuit. I am happy to have only had a brief glimpse of life as the “only” child as I could not imagine a life without all of them. A much deserved thank you to my immediate family: my mom and dad; my brother Steve (and Kendra); my sister Lara (and Mike, Kennedy and Raegan); my brother Scott (and Maddy and Izzy); my brother Jeremy (and Sue, Vin, and Jackie); and my sister Katie.

I would like to acknowledge and thank my colleagues who assisted in pre-testing my instruments and otherwise shared ideas and suggestions, as well as other friends and members of my extended family for their kind words of encouragement throughout the dissertation process and, when needed, providing the requisite distractions (a.k.a., cocktails): Maureen and Josh Aronson; John Blythe; Jason Brabbs; Mike Brown and Liz Wing; Mike Canup and Daniela Vidal; Shawn Casey and Willis Arrington; Alex Chassanoff; Ken Chestnutt; Dr. Songphan Choemprayong; Tex Clark; Marc Conte; Megan Crouch; Richard Davis; Jackie Dean; Billy Doherty; Phil Edwards; Todd Gannon; Lisa and Danny Gregory;

Scott Grimsley; Samantha Guss; Gayle Gutter; Jackie Gutter and Zach Snyder; Dr. Pam Gutter and Tammy Moore; Christy Hank; Dana Hanson-Baldauf; Perla Innocenti; Dr. Jacob Kramer-Duffield; Craig Kullik; Irene Marik; Louise Mercurio; Michael Merrill; John and Bev Mielke; Lori Richards; Terrell Russell and Kelly Marks; Dave, Amy and Jerry Pcolar; Stephanie and Will Peterson; Emily Shreefter; Don Sizemore; Chelcy and Fred Stutzman; Dr. Cassidy Sugimoto, and Thomas; Frances Tibbo; Gardner Treneman; Amy Van Scoy; Jean Wagner; Jewel Ward; Dr. Meredith Weiss; Dr. Megan Winget; and Drs. Kathy and Bill Wisser.

I utilized the Academic Blog Portal (<http://www.academicblogs.org/>) in compiling my samples. Thank you to Dr. Henry Farrell, the portal's chief editor, and his senior editors. The portal was invaluable to me in conducting my research.

A grateful acknowledgement to Beta Phi Mu. My dissertation research was supported by a 2010 Eugene Garfield Doctoral Dissertation Fellowship awarded by Beta Phi Mu.

Finally, I would like to extend my sincere appreciation and gratitude to all of the scholars who participated in my study, taking time to complete the questionnaires, participate in interviews, and/or communicate with me in some other fashion. This dissertation was made possible through your cooperation. Thank you for helping me to achieve my research objectives and, ultimately, contribute to my achievement of a Ph.D.



## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	v
TABLE OF CONTENTS.....	ix
LIST OF TABLES .....	xviii
LIST OF FIGURES .....	xxv
Chapter	
I. INTRODUCTION .....	1
1.1. Motivators .....	3
1.2. Objectives.....	7
II. LITERATURE REVIEW .....	9
2.1. The System of Scholarly Communication .....	9
2.1.1. Foundations of Scholarly Communication .....	11
2.1.1.1. Crisis in Scholarly Communication .....	13
2.1.1.2. Ambiguity of Publication .....	15
2.1.2.Communication Aspects .....	17
2.1.2.1. Channels .....	17
2.1.2.2. Units .....	20
2.1.2.3. Actors .....	22
2.1.2.4. Functions .....	23
2.1.3. Evolution of the System.....	28

2.1.3.1. Evolving Actor Roles .....	32
2.1.3.2. Evolving Channels and Units .....	33
2.2. The Blogosphere .....	36
2.2.1. Bloggership, Blawgosphere and Blogademia.....	39
2.2.1.1. Instances .....	41
2.2.1.2. Scholars and their Blogs.....	44
2.3. Digital Preservation.....	54
2.3.1. Defining Digital Preservation .....	55
2.3.1.1. Key Concepts .....	57
2.3.1.2. Current State of Research.....	64
2.4. Blog Preservation .....	65
2.4.1. Blogger Characteristics.....	66
2.4.1.1. Basic Demographics.....	66
2.4.1.2. Style of Blogging .....	68
2.4.1.3. Topical Treatment .....	68
2.4.1.4. Identity and Access .....	68
2.4.1.5. Audience.....	69
2.4.1.6. Activity.....	70
2.4.1.7. Stability .....	71
2.4.1.8. Publishing Conventions.....	73
2.4.2. Blog Characteristics .....	73
2.4.3. Approaches .....	75
2.5. Conclusion .....	77

III. METHODS .....	79
3.1. Population .....	80
3.1.1. Sampling Frames .....	83
3.1.1.1. Inclusion Criteria.....	83
3.1.2. Blog Sample.....	87
3.1.3. Blogger Sample .....	89
3.2. Questionnaires.....	91
3.2.1. Instrument Design.....	91
3.2.2. Pre-Testing.....	93
3.2.3. Administration .....	94
3.2.4. Completed Sample.....	96
3.2.5. Data Analysis.....	96
3.3. Interviews.....	97
3.3.1. Interview Schedule and Protocol .....	97
3.3.2. Pre-Testing.....	98
3.3.3. Administration .....	99
3.3.4. Data Analysis.....	100
3.4. Blog Analysis .....	101
3.4.1. Coding System.....	101
3.4.2. Data Collection and Analysis .....	103
IV. RESULTS: QUESTIONNAIRES AND INTERVIEWS .....	104
4.1. Respondents and Interviewees .....	104
4.1.1. Questionnaire: Outcome Rates and Completed Sample.....	104

4.1.1.1. Response Rates .....	105
4.1.1.2. Cooperation Rates .....	106
4.1.1.3. Refusal Rate .....	106
4.1.1.4. Contact Rate .....	107
4.1.1.5. Completed Sample .....	107
4.1.2. Interview: Completed Sample .....	108
4.2. Blogger Profile .....	108
4.2.1. Field of Study .....	109
4.2.2. Highest Degree Earned .....	109
4.2.3. Professional Age, Age and Gender.....	110
4.2.4. Occupation .....	113
4.2.4.1. Faculty Bloggers .....	115
4.2.5. Motivations to Blog .....	117
4.3. Publishing to Other Blogs .....	120
4.3.1. Blog Deletion.....	122
4.4. Publication and Service.....	123
4.4.1. Publication History .....	127
4.4.2. Service History .....	128
4.5. Blogging and Professional Life.....	129
4.5.1. Scholarship Criteria .....	129
4.5.2. The Scholarly Record .....	131
4.5.3. Invitations to Publish, Present, Collaborate, and Serve .....	133
4.5.4. Impact on Scholarly Life .....	135

4.6. Blog Publishing Behaviors and Characteristics .....	138
4.6.1. Audience .....	139
4.6.1.2. Engagement via Commenting Systems .....	143
4.6.2. Availability .....	145
4.6.3. Host Location.....	145
4.6.4. Blog and Post Composition and Revision Activity .....	148
4.6.4.1. Blog Updating .....	148
4.6.4.2. Imported Content and Permissions .....	152
4.6.4.3. Post Editing .....	157
4.6.4.4. Post Deletion .....	158
4.6.5. Micro-blogging .....	160
4.6.5.1. Twitter and the Library of Congress .....	162
4.7. Blog Preservation Behaviors.....	164
4.7.1. Questionnaire Respondents .....	164
4.7.1.1. Single-Blogger Respondents .....	165
4.7.1.2. Co-Blogger Respondents.....	168
4.7.1.3. All Respondents .....	174
4.7.2. Interviewees .....	175
4.7.2.1. Networks .....	182
4.7.2.2. Co-Bloggers .....	183
4.8. Blog Preservation Preferences .....	184
4.8.1. Extent of Preservation for Private and Public Access and Use .....	184
4.8.1.1. Blogs and Preservation Prioritization.....	188

4.8.2. Responsibility and Capability for Preservation .....	196
4.8.2.1. Additional Perspectives on Responsibility and Capability .....	201
V. RESULTS: BLOG ANALYSIS.....	211
5.1. Completed Samples: Blogs and Interviews .....	211
5.2. Identity and Authorship .....	212
5.2.1. Access Points .....	212
5.2.2. Co-Authors: Extent.....	214
5.2.3. Single-Authors: Name and Location .....	214
5.3. Blog Elements and Features.....	216
5.3.1. Select Feature Set .....	216
5.3.2. Application Software .....	218
5.3.3. Network Affiliation .....	219
5.4. Post Elements and Features.....	219
5.5. Posting Activity.....	220
5.6. Post Content .....	222
5.6.1. Text .....	222
5.6.2. Other Content Types.....	224
5.6.3. Links .....	225
5.6.4. Comments .....	226
5.7. Rights, Policies, and Disclaimers.....	226
5.7.1. Rights and Use Statements .....	226
5.7.2. Disclaimers .....	228
5.7.3. Commenting Policy .....	230

5.7.4. Blog Editing Policy.....	231
5.8. Age, Access, and Authority .....	232
5.8.1. Blog Age.....	233
5.8.2. Blog and Page Views.....	234
5.8.3. Blog Access and Indexing .....	234
5.8.3.1. Google Blog Search .....	235
5.8.3.2. Technorati.....	235
5.8.3.3. Internet Archive Wayback Machine.....	235
5.8.3.4. Library of Congress Legal Blawgs Web Archive .....	236
5.8.4. Blog Authority .....	237
5.8.4.1. Technorati Authority Number and Top 100 Ranking .....	237
5.8.4.2. Blog Awards.....	238
VI. DISCUSSION.....	240
6.1. Scholars and their Blogs?.....	240
6.2. Blogs and the ‘Scholarly Record’.....	242
6.2.1. Characterizing the ‘Scholarly Blog’ .....	243
6.3. Perceptions and Preferences for Preservation .....	246
6.3.1. Personal Responsibility and Provider Capability .....	247
6.4. Characteristics and Behaviors Impacting Preservation.....	248
6.4.1. Publishing Behaviors and Preferences.....	248
6.4.2. Co-Producer Dependencies .....	250
6.4.3. Back-Up and Archiving Behaviors and Strategies .....	252
6.4.3.1. Known and Unknown Archiving .....	252

6.5. Implications .....	253
6.5.1. Back-up or Preservation? Or, Back-up is Preservation? .....	253
6.5.2. Preserve What? Issues of Representation, Selection, and Permissions .....	255
6.5.3. Perceptions of Actor Roles in Blog Preservation .....	258
6.6. Limitations .....	260
6.6.1. Potential Exclusion of Eligible Respondents.....	263
6.6.2. Participation .....	263
6.6.2.1. Ineligible Sample Members .....	264
6.6.2.2. Interview Recruitment.....	265
6.6.3. Instrument Design and Analysis .....	265
6.6.4. Researcher Bias .....	268
VII. CONCLUSIONS AND FUTURE DIRECTIONS .....	269
7.1. Recommendations to Facilitate Blog Preservation .....	272
7.2. Recommendations for Further Research.....	274
7.2.1. Further Analysis.....	275
7.2.2. Future Research .....	276
7.3. Let the Conversation Continue .....	279
Appendices	
A. SAMPLE CODING SYSTEM.....	281
B. DISQUALIFIED BLOGS .....	287
C. QUESTIONNAIRE I (SINGLE-BLOGGERS) .....	288
D. QUESTIONNAIRE II (CO-BLOGGERS) .....	307
E. QUESTIONNAIRE PRE-TEST CORRESPONDENCE.....	327



F. QUESTIONNAIRE RECRUITMENT CORRESPONDENCE.....	332
G. INTERVIEW CONSENT SCRIPT AND SCHEDULE.....	336
H. INTERVIEW DEBRIEFING SHEET .....	339
I. INTERVIEW RECRUITMENT CORRESPONDENCE.....	341
J. SINGLE- AND CO-BLOG CODING SYSTEM.....	344
REFERENCES .....	352

## LIST OF TABLES

Table 1. Informal and formal channels .....	18
Table 2. Informal and formal units .....	21
Table 3. Population count of blogs by domain and cluster .....	82
Table 4. Blog inclusion criteria assessment by cluster .....	88
Table 5. Count of blogs in blog sampling frame stratified by authorship .....	88
Table 6. Blog sample .....	89
Table 7. Co-blogger inclusion criteria assessment by cluster .....	90
Table 8. Single-blogger inclusion criteria assessment by cluster .....	90
Table 9. Blogger sample .....	91
Table 10. Completed sample.....	96
Table 11. Frequency distribution of interview times .....	100
Table 12. Frequency distribution of time to code .....	103
Table 13. Questionnaire final disposition codes, definitions, and results.....	105
Table 14. Response Rate 1 (RR1).....	106
Table 15. Response Rate 2 (RR2).....	106
Table 16. Cooperation Rate 1 (COOP1) .....	106
Table 17. Cooperation Rate 2 (COOP2) .....	106
Table 18. Refusal Rate 1 (REF2) .....	107
Table 19. Contact Rate 1 (CON2).....	107
Table 20. Completed sample for QI and QII .....	108
Table 21. Completed sample for interview stage.....	108
Table 22. Frequency of highest degree earned by authorship .....	110
Table 23. Frequency of professional age by authorship .....	111
Table 24. Frequency of professional age by cluster .....	111

Table 25. Frequency of age by authorship.....	112
Table 26. Frequency of age by cluster .....	112
Table 27. Frequency of gender by authorship .....	113
Table 28. Frequency of gender by cluster.....	113
Table 29. Frequency of employment type .....	113
Table 30. Frequency of post-secondary faculty, post-secondary non-faculty, and non-post-secondary employment by authorship.....	114
Table 31. Frequency of post-secondary faculty, post-secondary non-faculty, and non-post-secondary employment by cluster .....	114
Table 32. Frequency of job titles for post-secondary, non-faculty respondents by authorship .....	115
Table 33. Frequency of job titles for post-secondary faculty respondents by authorship ....	116
Table 34. Frequency of job titles for post-secondary faculty by cluster.....	116
Table 35. Frequency of tenure status for post-secondary faculty by authorship .....	117
Table 36. Frequency of tenure status for post-secondary faculty by cluster .....	117
Table 37. Frequency of tenure status for post-secondary faculty by job title.....	117
Table 38. Frequency of all blogs published to by authorship .....	120
Table 39. Frequency of all blogs published to by cluster .....	120
Table 40. Frequency of other single-blog instances by authorship .....	121
Table 41. Frequency of other co-blog instances by authorship .....	121
Table 42. Frequency of single- and co-blog instances by authorship.....	122
Table 43. Frequency of blog deletion by authorship .....	122
Table 44. Frequency of blogs intentionally deleted by authorship.....	122
Table 45. Frequency of reasons for intentional blog deletion .....	123
Table 46. Frequency of scholarly publishing history by unit-type of communication.....	127
Table 47. Frequency of scholarly publishing history across unit-types of communication .	128

Table 48. Frequency of professional service activity by authorship .....	129
Table 49. Frequency of single-blogger agreement or disagreement concerning criteria for scholarship .....	130
Table 50. Frequency of co-blogger agreement or disagreement concerning criteria for scholarship .....	130
Table 51. Frequency of all bloggers' agreement or disagreement concerning criteria for scholarship .....	130
Table 52. Frequency of all bloggers' agreement concerning criteria for scholarship by cluster.....	131
Table 53. Frequency of single-blogger agreement or disagreement concerning the scholarly record .....	132
Table 54. Frequency of co-blogger agreement or disagreement concerning the scholarly record .....	132
Table 55. Frequency of all bloggers' agreement or disagreement concerning the scholarly record .....	132
Table 56. Frequency of all bloggers' agreement concerning the scholarly record by cluster.....	133
Table 57. Frequency of invitations to publish by unit-type of communication.....	133
Table 58. Frequency of invitations to publish across unit-types of communication .....	134
Table 59. Frequency of invitations to provide professional service by authorship .....	135
Table 60. Frequency of single-blogger perceptions of impairment or improvement in their scholarly life.....	135
Table 61. Frequency of co-blogger perceptions of impairment or improvement in their scholarly life.....	137
Table 62. Frequency of all bloggers' perceptions of impairment or improvement in their scholarly life.....	137
Table 63. Frequency of all bloggers' perceptions of improvement in their scholarly life by cluster.....	138
Table 64. Frequency of all bloggers' perceptions of improvement in their scholarly life....	138
Table 65. Frequency of all bloggers' primary and secondary audience-types .....	140
Table 66. Frequency of publishing comments on other blogs by authorship .....	144

Table 67. Frequency of primary host location by authorship .....	146
Table 68. Frequency of primary host location by cluster .....	147
Table 69. Frequency of updating by authorship .....	148
Table 70. Frequency of updating by cluster.....	149
Table 71. Characterization of blogging over time by authorship .....	149
Table 72. Characterization of blogging over time by cluster.....	150
Table 73. Frequency of changing skin of blog .....	152
Table 74. Frequency of checking permissions for imported content by authorship .....	153
Table 75. Frequency of checking permissions for imported content by cluster .....	153
Table 76. Frequency of importance or unimportance for respondents to verify permission by authorship.....	154
Table 77. Frequency of importance or unimportance for respondents to verify permission by cluster.....	154
Table 78. Frequency of importance or unimportance for others to verify permission by authorship.....	155
Table 79. Frequency of importance or unimportance for others to verify permission by cluster.....	155
Table 80. Frequency of post editing by authorship.....	157
Table 81. Frequency of post editing in a three-month period by authorship.....	158
Table 82. Frequency of reasons for intentional editing of posts .....	158
Table 83. Frequency of post deletion by authorship.....	159
Table 84. Frequency of post deletion in a three-month period by authorship .....	159
Table 85. Frequency of reasons for intentional deleting of posts .....	160
Table 86. Frequency of association between respondents' blog and Twitter accounts .....	161
Table 87. Frequency of single-bloggers saving blog via an archiving service.....	165
Table 88. Frequency of single-bloggers saving blog independent of an archiving service..	166
Table 89. Frequency of single-blogger approaches for saving entire blog.....	167

Table 90. Frequency of blog components purposefully saved by single-bloggers.....	168
Table 91. Frequency of co-bloggers saving blog via an archiving service.....	169
Table 92. Frequency of co-bloggers saving blog independent of an archiving service.....	171
Table 93. Frequency of co-blogger approaches for saving entire blog.....	172
Table 94. Frequency of component types purposefully saved by co-bloggers.....	173
Table 95. Frequency of extent of component types purposefully saved by co-bloggers.....	173
Table 96. Frequency of co-blogger approaches for saving blog components .....	174
Table 97. Frequency of all bloggers purposefully saving blog.....	174
Table 98. Frequency of all bloggers' strategies for saving blog in whole or in part .....	175
Table 99. Frequency of single-blogger agreement or disagreement for blog preservation ..	186
Table 100. Frequency of co-blogger agreement or disagreement for blog preservation .....	186
Table 101. Frequency by all bloggers' agreement for blog preservation by cluster.....	188
Table 102. Frequency of all bloggers' agreement for blog preservation .....	188
Table 103. Frequency by all bloggers' ranking of responsibility for preservation.....	198
Table 104. Frequency by all bloggers' ranking of capability for preservation.....	199
Table 105. Frequency of actors perceived as most responsible and most capable for blog preservation .....	200
Table 106. Ranking of actors perceived as most responsible and most capable for blog preservation .....	200
Table 107. Blog sample .....	211
Table 108. Frequency of access points for information on bloggers' identity .....	213
Table 109. Descriptive measures of number of primary authors for co-blogs by cluster.....	214
Table 110. Frequency of name types .....	215
Table 111. Frequency of access points for single-bloggers' name.....	216
Table 112. Frequency of blog user interface elements and other features.....	217
Table 113. Frequency of blogroll links.....	218

Table 114. Frequency of blog application .....	218
Table 115. Frequency of blogs network affiliations .....	219
Table 116. Frequency of network affiliation for blogs by cluster .....	219
Table 117. Frequency of blogs post elements and features .....	220
Table 118. Frequency of recent posting.....	221
Table 119. Frequency of recent posting by cluster .....	221
Table 120. Descriptive measures of posts published to all blogs between August 1-31, 2010 by cluster .....	222
Table 121. Descriptive measures of total, quoted, and original word count for most recent blog posts .....	223
Table 122. Frequency of most recent posts' total, quoted, and original word count.....	223
Table 123. Means of total, quoted, and original words for most recent blog posts by cluster.....	224
Table 124. Frequency of content types in most recent blog post.....	225
Table 125. Frequency of text-based hyperlinks in most recent blog post .....	225
Table 126. Descriptive measures of internal and external links in most recent post.....	226
Table 127. Frequency of comments to most recent post .....	226
Table 128. Frequency of rights and use statements by type for all blogs.....	227
Table 129. Frequency of rights and use statements by cluster .....	227
Table 130. Frequency of rights statement type and location for all blogs.....	227
Table 131. Frequency of disclaimer statements for all blogs .....	229
Table 132. Frequency of disclaimer statements by cluster .....	229
Table 133. Frequency of commenting policies for all blogs .....	230
Table 134. Frequency of blog age by authorship.....	233
Table 135. Frequency of blog age by cluster.....	233
Table 136. Descriptive measures of blog visits and page views from Site Meter .....	234

Table 137. Frequency of blogs archived at the Internet Archive Wayback Machine by cluster .....	236
Table 138. Frequency of blogs indexed to Google, Technorati, Internet Archive and Library of Congress.....	236
Table 139. Frequency of Technorati Authority scores for blogs indexed to Technorati .....	238
Table 140. Characteristics of majority of sampled blogs and bloggers by cluster .....	241



## LIST OF FIGURES

Figure 1. Continuum of preservation priorities.....	192
Figure 2. Perceptions of blogs in relation to the scholarly record and scholarship criteria .....	242
Figure 3. Agreement for blog preservation.....	246
Figure 4. Perceptions on responsibility and capability for blog preservation.....	247

## **CHAPTER 1**

### **INTRODUCTION**

The opening title segment for this dissertation, “Scholars and their Blogs,” could also be recast, “Scholars and their Blogs?” With the emergence of the World Wide Web, or one’s “personal printing press,” as described by an interviewee in this study, the contemporary system of scholarly communication has been described as transformed or evolved. It might also be characterized as a bit of a fender bender, with traditional and untraditional units and channels of communication colliding in the infosphere.

A contemporary reflection in the academe of this transformation, evolution, fender-bender, or what have you, is the publication of blogs by scholars, faculty, and researchers. The neologisms, “bloggership” and “blogademia,” emerged in recent years, reflecting the adoption of blogs as channels for scholarly communication; the former in reference to legal scholarship blogs, or blawgs, and the latter to blogs across domains (e.g., Caron, 2006; Saper, 2006; Smith, 2006; Dayal, 2005a). Halavais (2006), in borrowing from Oldenburg (1999), describes the blogs of scholars as, “shaping a new ‘third’ place for academic discourse,” comparable to other informal channels of communication such as lab notebooks, coffee houses, and opinion pieces (p. 117).

The blogs of scholars pose a particular challenge when considered within the system of scholarly communication and, in particular, preservation, a primary function of the system. First, are blogs scholarship? Braxton, Luckey, and Helland (2002) include “unpublished scholarly outcomes and publications” in their definition of scholarship, continuing that,

“unpublished scholarly outcomes fully meet the definition of scholarship if they appear in a publicly observable form” (p. 141). The three parameters for this form are that, “it must be public, subject to critical review, and in a form that allows use and exchange by other members of the scholarly community” (Braxton et al., 2002, p. 141). The blogs of scholars may be assumed, to some degree, to satisfy these requirements. Blogs are typically published to the open Web, made available to peers and colleagues, and support use and exchange through commenting systems and other user interface elements. One blogger, as cited by Glenn (2003), thinks that blogs have, “some of the best aspects of peer review built in” (para. 11). However, beyond anecdotal reports in the literature, the extent to which scholars’ blogs satisfy these criteria is an under-examined area of inquiry.

Debate on the legitimacy, value and impact of blogs and other new or transformed modes of digital scholarship is a contemporary issue in the area of scholarly communication. It is succinctly commented upon by Borgman (2007):

While most of these new genres are too informal to have been considered publications in a print realm, they do contain important discussions, facts, and reports that are part of the scholarly discourse of a field. Furthermore, they can be captured because digital communications leave a trace. (p. 99).

These new genres may leave a trace, but when considering the nature of the blog form as well as the technical, regulatory and social frameworks in which blogging takes place, for how long? Rothenberg (1995), an early leader in advocating technical approaches for the preservation of digital records, cautioned that:

... the significance of many digital documents we consider too unimportant to archive become apparent only long after they become unreadable. Unfortunately, many of the traditional methods developed for printed matter are not applicable to electronic files. The content and historical value of thousands of records, databases and personal documents may be irretrievably lost to future generations if we do not take steps to preserve them now. (p. 42).

Without deliberate personal or programmatic approaches to the long-term stewardship of these digital communications, the scholar blogs of today may be unavailable into the future. The conversations informing this research – blogs, the scholars who publish them, and those interested, or not, in preserving them – is not new. The descriptive study presented within provides some evidence in response to these on-going conversations, informing emerging profiles of this particular class of content and creators, and current and future personal and programmatic digital preservation action.

### 1.1. Motivators

The blogs of scholars represent communications which may, in reference to the Association of Research Libraries' (1986) definition of the scholarly record, be reflective of scholarly works, knowledge and ideas, and representative of scholars' cumulative scholarly record. When framed within the functions of scholarly communication – summarized by Borgman (2007) as, “legitimization, dissemination, and access, preservation and curation,” and Roosendaal, Geurts, & van der Vet (2001), in reference to the value chain, as, “registration, certification, awareness, and archiving” – these communications may be seen as representative of products, or units, of communication, as well as processes for communication.

As potentially valuable additions to the human record, advocates have called for the stewardship and long-term preservation of blogs (e.g., Entlich, 2004; O'Sullivan, 2005; Paulus, 2006). Dempsey (2007) contributed his own call for stewardship for academic blogs, blogging (aptly):

While they [blogs] may be of enduring interest, little thought has probably been given to thinking about their longer term persistence ... Universities and university libraries are recognizing that they have some responsibility to the curation of the intellectual outputs of their academics and students ... What, if anything, should the Open

University or Harvard be doing to make sure that this valuable discourse is available to future readers as part of the scholarly record? (para. 7-8).

It is interesting to note that, two years after Dempsey's remarks, Harvard is doing something in regard to the blogs of their scholars. In a recent Harvard University Libraries Taskforce report, academic blogs are included in the plan to expand digital collections, along with other new, informal genres, including data sets, Websites, and email correspondence (Harvard University Library, 2009). Other institutions are "doing" as well. A prominent example of an active "blawg" archiving program is the Library of Congress' Legal Blawgs Web Archive, begun in 2007.<sup>1</sup>

In delineating benefits of long-term preservation, Hedstrom et al. (2003) ground their outlook on societal benefits of preservation through conditional statements, including "if unique information objects that are vulnerable and sensitive and therefore subject to risks can be preserved and protected," among others (p. 4). While there is support for the first part of this propositional statement – blogs are *potentially* unique, potentially since most definitely diverse – activities informing the latter part – that blogs can be preserved and protected – is a current and active area of investigation. While some institutions, such as the Library of Congress, have implemented blog archives, or, like the Internet Archive, have been collecting blogs as part of a broader Web archiving program, many issues of concern are still outstanding.

To highlight some select concerns here, consider the message as well as the medium. While the message of blogs may be similar to other communications (e.g., diaries, lab notes, coffee houses, salons, and opinion-type news pieces), the medium is distinct. Blogs are most definitely co-dependent, representing a mix of code, content and co-creators. As such, blogs

---

<sup>1</sup> Law Library of Congress Legal Blawg Archive: <http://www.loc.gov/law/find/web-archive/legal-blawgs.php>

have been described as a co-produced medium. Serfaty (2004) describes the blogosphere as a diverse, dynamic network of co-producers, comprised not only of bloggers, both the known and anonymous, but their readers and service providers. This multiplicity of producers complicates blog preservation actions, necessitating negotiations among a complex landscape of associated procedural, social, technical and regulatory issues.

Scholars who blog and the blogs of scholars pose particular challenges concerning trust, authenticity, reliability, and topicality. What distinguishes a scholarly blog from other blogs and, likewise, a scholar who blogs from other bloggers? Halavais (2006) distinctly summarizes this somewhat bipolar concern when he asks: "... what practices will help to differentiate 'scholars who blog' from 'scholarly blogging'?" (p. 123).

Further, blogs are a diverse – and pervasive – medium. The size of the blogosphere is overwhelming. As of October 6, 2010, Technorati.com, the largest blog directory in the world, lists 1,242,410 blogs to their blog index.<sup>2</sup> While this count is clearly demonstrative of the overall extent of the blogosphere, an exact count of all blog instances is unknowable, as blogs are continuously added, abandoned and deleted. A report from the Council on Library and Information Resources (2001), while not addressing blogs specifically, does provide some context in which to consider the complicated landscape for blog preservation:

With so much information coming at us, how do we distinguish between what is of long-term value and what is ephemeral? And of that ephemeral, what should be selected to be preserved to ensure for the future a rich record of the present" (p. 1).

So, the prospect of preserving scholar blogs is not only complicated by determinations of who is a scholar blogger and what is a scholarly blog, but also determinations of which blogs "merit" preservation. Developing selection criteria is an outstanding and critical step in any action toward preservation. As Lynch (2004) points out:

---

<sup>2</sup> Technorati Blog Directory: <http://technorati.com/blogs/directory/>

“There are undoubtedly a goodly number of blogs that deserve preservation, just as there are a goodly number of websites. Every website does not need to be preserved in every one of its versions, nor does every blog” (para. 30). Additionally, in recognition of blogs’ mix of co-produced code and content, this raises concerns as to what elements of blogs should be preserved, as well as what elements can actually be preserved when considering the technical and regulatory landscape in which blogs are published.

Selection considerations are particularly relevant when considering the need to negotiate between a climate of finite resources and the technical, organizational and resource infrastructures required to support digital preservation activities. The challenge of blog preservation is compounded by a myriad of factors in consideration of the goals for preservation: to get, describe, interpret, secure, authenticate, access and perform (Caplan, 2008). While these goals, respectively, necessitate a consideration of a triage of actions and approaches, when treated in the aggregate, it makes clear the need for an enhanced understanding of the range of behavioral, legal and technical issues impacting blog preservation.

Bloggers, in general, are not only active producers but also active blog consumers. Nine out of ten bloggers read other blogs (Lenhart & Fox, 2006). Bruns (2006) describes this dual role, reflecting the lack of exclusivity between production and consumption, as “produser.” Both the producer and consumer perspectives of bloggers would contribute to an enhanced understanding of selection decisions for blog preservation. That is, how might the preservation preferences of bloggers influence decisions on what, or what not, to preserve, both concerning their own blogs as well as the blogs of other scholars?

Scholars who blog can play a pivotal role in blog preservation in other ways as well. Garrett and Waters (1995) identify content creators as the “first line of defense” in digital preservation (p. 40). A deeper understanding of bloggers’ diverse publishing behaviors and practices would contribute to a deeper understanding of both preservation preferences as well as awareness. For example, bloggers in general, as reported by Viegas (2005), assume the persistency of their blogs unless deliberate action is taken by bloggers to remove their respective blog. Even if no longer maintained, it is assumed the blog will still be available via search engine caches. This assumes no other influences impacting service (e.g., discontinuation of service issues; changing terms of service from Web publishing application and hosting services).

## 1.2. Objectives

This descriptive study investigates the attitudes and perceptions of scholars who blog in relation to scholarly communication and digital preservation, as well as examines blogger behaviors and blog characteristics that may influence preservation action. Specifically, the study is designed to inform the following research questions:

- 1) How do scholars who blog perceive their blog in relation to their cumulative scholarly record?
- 2) How do scholars who blog perceive their blog in relation to long-term stewardship and, subsequently, who, if anyone, do scholars perceive as responsible as well as capable for blog preservation?
- 3) How do blog characteristics and blogger publishing behaviors and preferences impact preservation?



The research design is a triangulated approach, employing interviews, questionnaires and document analysis techniques. These methods were chosen to provide the most comprehensive picture of scholars' blogging and scholars' blogs to date, as well as an extensive look at the digital preservation preferences of this unique community of producers. The instrumentation gathered data on attitudes, opinions, practices, and preferences of scholars who blog, as well as characteristics of scholars' blogs.

This research is intended to benefit multiple stakeholder groups, including: 1) bloggers interested in personal preservation of their blog content; 2) bloggers interested in improved data management practices; 3) organizations with current, piloted or planned digital preservation initiatives who are considering the medium; 4) organizations without planned digital preservation initiatives, in order to inform future, strategic collection policy decisions; 5) researchers interested in the evolution of scholarly communication, particularly in relation to social media and public science; and 6) lastly, the digital preservation community in general. In their investigation of outstanding research areas, the DPE (2007) cautions that while research in the area of digital preservation has been ongoing for nearly twenty years, "our heritage may now be at greater risk because many in our community believe that we are making progress towards solving the preservation challenges" (p. 1). It is hoped that the findings presented within can contribute in some way to this ongoing need for continued progress.

## CHAPTER 2

### LITERATURE REVIEW

This literature review provides a framework from which to consider the underlying concepts informing this study. It begins with a treatment of the system of scholarly communication, providing context for situating blogs within this domain. Next, the emergence of blogging as a communication device is presented, with a particular emphasis on adoption by scholars. It concludes with a summary of the goals of digital preservation, in general, and the challenges of blog preservation, in particular.

#### 2.1. The System of Scholarly Communication

The Association of Research Libraries (ARL) describes scholarly communication as a system, emphasizing the social aspect, “whereby intellectual and creative activity is transmitted from one scholar to another (1986, p. 2). Further, ARL (1986) identifies actors, relationships, procedures and outputs comprising the system:

As a system, its major components are the creators or authors of scholarly works, the ideas and knowledge that they create or interpret, the means by which they initially and informally communicate these messages to their colleagues (conference papers, preprints), the means by which such information is formally reviewed and communicated (editorial boards and publishers), the means for distributing and providing access to journals (paper copy subscriptions, microforms, electronic access), research libraries, and finally, the consumers of this information (scholars and potential scholars). As the loop closes the process continues in a self-generating mode to the extent that research tends to be cumulative: discoveries, once known, stimulate new discoveries. (p. 2)

Per Borgman (1990), “a field’s interest in its own scholarly communication is a sign of its maturity” (p. 12). Made explicit in Borgman’s definition of scholarly communication

are the informal and formal channels of communication. Additionally, scholarly communication is considered from a research perspective:

By *scholarly communication* we mean the study of how scholars in any field (e.g. physical, biological, social, and behavioral sciences, humanities, technology) use and disseminate information through formal and informal channels. The study of scholarly communication includes the growth of scholarly information, the relationships among research areas and disciplines, the information needs and uses of individual user groups, and the relationships among formal and informal methods of communication. (Borgman, 1990, p. 13-14).

The system of scholarly communication is also characterized by the goals of various stakeholder groups within the system:

Each group shares this ultimate objective: Scholars want the best research to be widely publicized and shared, directors of scholarly presses and editors of scholarly journals want to publish works of highest quality and disseminate their publications widely, and librarians seek to develop comprehensive collections and make them easily accessible. (National Enquiry into Scholarly Communication, 1979, p. 6)

Borgman (2000) points out that, “new or evolving research areas are often characterized by debates over territory and terminology” (p. 413). A challenge in a review of literature in scholarly communication is how to distinguish between the concepts of scholarly communication and scholarly publishing. While it may appear the terms are used interchangeably, this may be a reflection of the confluence between the system of scholarly communication and the subsequent communicative actions of specific members within this system. Scholarly publishing is a sub-activity within the larger domain of scholarly communication. To borrow from Borgman (2007):

In the larger sphere of activities among scholars, ‘publication’ occurs when a document is “made public” with the intention that it be read by others. The term scholarly communication is used ... in the broader sense to include the formal and informal activities associated with the use and dissemination of information through public and private channels. (p. 48)

Synthesizing these and other definitions in the literature, scholarly communication is understood as a system that spans disciplines, comprised of multiple, interdependent groups

of stakeholders, including scholars, publishers, research libraries, and learned societies. It entails a range of communication behaviors and outputs, both formal and informal, resulting in a cyclic exchange of information. Further, this system is informed by external factors, notably funding agencies and information and communication technologies (e.g., Roosendaal, et al., 2001; NESC, 1979). Though understood as a system, it should not be viewed as static and inflexible, but as organic and evolving.

#### 2.1.1. Foundations of Scholarly Communication

While scholarly communication reemerged as a crucial area of investigation in the mid-1990s, on-going through present-day (Borgman, 2000), the system of scholarly communication has evolved over several hundred years. Johns (1998) looked at the intersection of the history of print and the history of science, referencing the “chronological skeleton, the joints of which are dates such as 1543, 1632, 1687, 1789, 1859 and 1905” (p. 42). These dates represent seminal scientific publications by Copernicus, Galileo, Newton, Lavoisier, Darwin and Einstein, respectively. The first peer reviewed serial publication, the *Philosophical Transactions of the Royal Society*, chartered in 1662, was issued in 1665 (Johns, 1998; Price, 1986; Solomon, 2002).

Price (1986) attributes the development of the scientific paper as a consequence of too many books. The intent was to save time, for both the author and reader, in the communication of information which was previously confined to, “the network of personal correspondence, private rumor, and browsing in Europe’s bookstores” (p. 57). These papers, ancestral forms of the contemporary journal article, fulfilled a social purpose – the what and who of research – and, per Price, served as a “social device” (p. 58).

Contemporary readings in foundational aspects of scholarly communication frequently reference leading theorists such as Kuhn, Licklider, Merton, and Price, in discussions of the principles underlying scientific communication (e.g., Borgman, 1990; Borgman, 2007; Cronin, 2005; Roosendaal, et al., 2001). Scholarly communications reflect aspects of the gift economy; that is, scholarship is owned by authors, their institutions, and funders. Hagstrom (1965) makes mention of the “gift” nature of scientific communication. Per Okerson (1992), scholarly communication models should be developed based on concepts of the gift economy (Okerson, 1992). Due to the systematic attributes of scholarly communication, it is viewed within a cultural context, further reflecting the theory of a gift exchange economy (Borgman, 2000, citing Lyman, 1999).

Borgman (2007) comments on additional influences on the understanding and evolvement of the system of scholarly communication, writing, “The system builds on long traditions in Western thought about open science and the social and economic benefits of the free flow of ideas” (p. 73-74). Further, while the system of scholarly communication is comprised of a range of actors, a dedicated body of literature emphasizes the scholar-actor role in particular.<sup>3</sup>

The NESC’s (1979) seven elements for an efficient system of scholarly communication are: access, entry, quality control, timeliness, coordination, adaptability and financial viability (p. 7). Regarding access, the emphasis is on availability to a broad collection of resources and reduction of barriers to access. The entry element addresses the requirement for multiple channels for communication, reflective of procedures in establishing scholarly merit. The quality control element extends this notion of merit through

---

<sup>3</sup> For a fuller treatment of scholars’ procedures and methods for scientific inquiry and subsequent communicative actions, see works by these seminal authors: Bazerman (1994, 2008), Ceccarelli (2001), Hagstrom (1965), Hyland (2000), Knorr-Cetina (1999), and Latour (1987).

establishing procedures for assessment, reflective of the respective mode for publication (i.e., journal article, technical report). Timeliness concerns efficiency in the time lapse between dates of submission and dates of publication. The coordination element addresses mechanisms for interaction among diverse actors in system. With the recognition of the dynamic nature of the system of scholarly communication, the element of adaptability is responsive to evolving methods for scholarly inquiry and communication. Lastly, financial viability requires coordination and agreement amongst various actors for sustaining the functions and channels within the system of scholarly communication.

In recent history, ARL (1986) describes the system of scholarly communication as relatively consistent through the previous two centuries, remaining stable even with the introduction of new journals, academic institutions and learned societies. Price (1989) however points out that the now recognized form of the journal article did not take shape until the past century,<sup>4</sup> and that, “as late as 1900, some of the most respected journals contained not one scientific paper of the present variety” (p. 58).

The stability referenced by ARL was shaken in the 1960s with what ARL (1986) describes as an information explosion in scholarly publishing, contributed to by an increase in the number of active scholars and the struggle by system stakeholders, notably publishers, their gatekeepers and libraries, to manage the growth in journal publications.

#### 2.1.1.1. Crisis in Scholarly Communication

This apprehension and instability is not new to the area of scholarly communication. Consider the foundations of scholarly journal publishing. Price (1989) characterized the climate in the late 1600s as “resistant,” and described the scientific paper as a “new and

---

<sup>4</sup> Hartley (2008) describes the generalized structure of a scientific article as an acronym, IMRAD: introduction, method, results, and discussion.

seemingly illicit practice” in comparison to “decent books” (p. 57). Johns (1998) provides further context for this new form of publication:

Since very few printing houses specialized in learned work ... knowledge remained lodged in the same premises as ephemera and pamphleteering. Learned scholars and gentlemen alike had to commit their tomes to be printed in the midst of almanacs, pamphlets, and (in the case of Newton’s *Principia*) pornography. (p. 100)

In modern times, the call of crises in scholarly communication has been ongoing over the past several decades (e.g., Day, 2003; NESC, 1979; Okerson, 1992; Van de Sompel et al., 2004). Descriptions of these crises are further qualified by attribute. In two recent examples, Day (2003) writes of the “serials pricing crisis,” and Van de Sompel, Payette, Erickson, Lagoze, and Warner (2004) qualify the crisis by “serials” and “permissions,” the latter resulting from “restrictions on use of publications once access has been obtained” (para. 1).

The extent of scholarly communication crises is aggravated when considering the diversity of the system’s stakeholders. To illustrate, consider the 1970s crisis and subsequent “endistic” predictions. A ripple effect among stakeholder groups emerged, highlighting their interdependent relationships. Publishers, including university presses, and journal editors, confronted declining sales; libraries negotiated between shrinking budgets and acquisitions; and scholars coped with delays in publication (NESC, 1979). Further, the NESC (1979) pointed to the simultaneous sense of financial crisis in higher education as an external factor contributing to perceptions of a crisis in scholarly communication. Concerning the research library in particular, new advances in technology and desktop computing contributed to predictions that the role of the library would be diminished (NESC, 1979).<sup>5</sup>

The NESC concludes that the 1970s crisis was not so much fatalistic as awakening, contributing to recommendations for supporting the evolution and sustainability of the

---

<sup>5</sup> See Brown and Duguid (2000) for reference to other “endistic” claims. For example, the end of the pencil, c. 1938; the end of the paper office, c. 1975; the end of paper copies, c. 1980; and the end of the fax, c. 1995.

traditional system of scholarly communication in light of changing circumstances, dominated by financial and technological influences. Describing the paradigm of traditional scholarly communication as unviable, Okerson (1992) furthered the call for transformation to the system of scholarly communication rather than its dismantling.

#### 2.1.1.2. Ambiguity of Publication

The channel of communication and the form of the communication also contributed to a “crisis of semantics:”

The crisis that libraries face first is more in the areas of unpublished and pre-published scholarship. The ability to acquire and store electronic information that can be changed and erased, for example, may also be a critical problem for scholarship of the future. The definition of ‘published’ is becoming unclear (ARL, 1986, p. 5).

Though written over two decades ago, the concern raised by ARL is a current and contested topic in the area of scholarly communication. Determinations on what constitutes “publication” can be informed by a consideration of the nature of the act. Borgman (2007) writes that, “in the larger sphere of activities among scholars, ‘publication’ occurs when a document is ‘made public’ with the intention it be read by others” (p. 48). This approach, however, is complicated by a dynamic, diverse, and constantly evolving information landscape, and associated tools and channels for communication, and is well reflected in Borgman’s use of quotes for the words ‘publication’ and ‘made public.’ For example, a scholar’s blog published to the open Web, within the parameter of publicity, would constitute a publication. However, that does not take into account degrees of formality, particularly in regard to the structuring of the reward system within scholarly communication (e.g., tenure and performance reviews). Borgman (2007) elaborates on other types of publications and tools for dissemination that complicate consensus on a definition of publication and, more crucially, determinations of value:



What has changed most since the days of print and post is the balance between public and private communication. Conversations that previously were oral are now conducted by e-mail or online discussion lists, sometimes leaving a public record for a long period of time. Presentations that would have been heard by a few people at a seminar are now widely available online via the posting of slides, speaking notes, and Webcasts. Manuscripts, preprints, technical reports and other written works that have circulated privately are now posted publicly. Online communication has accelerated the amount of informal communication among scholars and simplified the dissemination of formal products of scholarship. (p. 49).

As mentioned in the Introduction chapter, Braxton, Luckey, and Helland (2002) include “unpublished scholarly outcomes” in their definition of scholarship if, conditionally, these appear in a publicly observable form” (p. 141). These unpublished outcomes are considered publicly observable if made “public, subject to critical review, and in a form that allows use and exchange by other members of the scholarly community” (Braxton et al., 2002, citing Shulman & Hastings (1999), p. 141).

In further consideration of what constitutes a publication, consider the conventions governing academic writing. To borrow from the foreword to the Publication Manual of the American Psychological Association (APA, 2010), the intent of a style manual is, “to advance scholarship by setting sound and rigorous standards for scientific communication,” with “elements ... codified in the rules ... for clear communication” (p. xiii). The APA Style Manuals (2001, 2010) provide examples of the most common reference types in support of the objective for clear communication. The most recent edition presents several examples of references not presented in the previous edition.<sup>6</sup> These represent the content types referenced by Borgman (2007) above, including: 1) in-press article, pre-print archive; 2) blog post; and 3) video blog post. The addition of these content types in style manuals provides

---

<sup>6</sup> The 5th edition contained a heading for “Unpublished works and publications of limited circulation” (APA, 2001, p. 235). The 6th edition contains a similar heading, “Unpublished and informally published works” (APA, 2010, p. 196). However, the APA reorganized the section, moving references previously confined to the unpublished category to other categories representing “published works;” e.g., in-press work now appears under the heading, “Periodicals” (APA, 2010, p. 194).

some support for scholarly treatment of these resource types, through long-established practices for “clear communication,” regardless of whether deemed published or unpublished, or informal or formally communicated.

### 2.1.2.Communication Aspects

#### 2.1.2.1. Channels

Channels of communication may be formal or informal (e.g., ARL, 1986; Cronin, 2005; Garvey & Griffith, 1971; Roosendaal et al., 2001), and described as occurring along a continuum (Borgman, 2007). Garvey and Griffith (1971) constructed an oft-cited dissemination model for scientific communication reflecting this continuum. The authors’ model is illustrative of the processes from conceptualization through various stages of communication within the domain of psychology. Per Garvey and Griffith (1971), the transition from informal to formal occurs at the point of publication in a scholarly journal. An obvious example of an informal channel is communication with colleagues. In consideration of the current landscape, it is interesting to note that Garvey and Griffith, writing nearly forty years ago, qualified colleagues as “immediate;” that is, within an achievable physical proximity. The advent of wide-spread networked technologies in the current communication landscape removes barriers for physical proximity to support the exchange of texts and other media, extending scholars’ ability to interact informally with colleagues contingent simply on access to a browser.

The path from the formation of a specific scientific inquiry – the idea – through informal and formal communications is extensive. Garvey and Griffiths’ (1971) model is projected across a thirteen-year time span. Drawing from the work of Garvey and Griffith

(1971), Borgman (2007), Hagstrom (1965), and Price (1986), Table 1 distinguishes between examples of communication types by formal and informal channels.

**Table 1. Informal and formal channels**

<b>Informal</b>	Personal communication with immediate colleagues (e.g., scholars at home departments/schools/institutions)
	Personal communication with extended colleagues (e.g., scholars at other institutions via phone, email, letters).
	Communication at informal gatherings with restricted/invite-only audiences (e.g., staff meetings)
	Communication at formal events with unrestricted/open-registration audiences (e.g., professional conferences, symposia) <sup>7</sup>
	Conference proceedings and presentations
	Preliminary findings (e.g., technical report)
	Pre-prints
<b>Formal</b>	Journals
	Journal abstracts
	Monographs
	Citation indexes

While a focus on evidence of communication, particularly within the academes’ “publish or perish” culture, is achieved through formal channels, such as the publication of journal articles, it does not dispel the value of informal channels for communication. Garvey and Griffith (1971) make clear the importance of informal channels in the communication of science, writing in reference to an examination of over 200 research initiatives in the field of psychology, that:

... We found that ideas for less than one out of seven originated from sources such as journal articles, presentations at national meetings, etc. Instead, the scientist relies heavily on informal networks of information exchange to keep abreast of current activities and of the current views of the community on the value and relevance of specific research problems. (p. 354).

The value of informal communication is further evidenced by the concept of the “invisible college.” The invisible college, per Price (1986), Lievrouw (1990), and Borgman

<sup>7</sup> Table 1 represents a basic categorization and is not intended to be absolute. For example, while Garvey and Griffith (1971) place these events on the informal side of the continuum, Borgman (2007), in her reinterpretation of the model, places these events (i.e., “Special Group Meetings, National Convention, Invited Conference”) on the side of formal communication. In another example, proceedings may or may not be peer reviewed, which also contributes to confusion on where to place conference papers within the continuum of communication.

(2007), emerged in the 1600s, with the origin credited to London's Royal Society and, "at which time it described a group of scholars who were in geographic proximity and had common interests, but lacked a formal institution or college" (Borgman, 2007, p. 57). Lievrouw (1990) identifies it as the best known model of scientific communication. While the concept of the invisible college is centuries old, Lievrouw (1990) credits Price (1986) with re-introduction of the model in the contemporary landscape. Price (1986) provides the following illustration of the invisible college:

For each group there exists a sort of commuting circuit of institutions, research centers, and summer schools giving them an opportunity to meet piecemeal, so that over an interval of a few years everybody who is anybody has worked with everybody else in the same category. Such groups constitute an invisible college ... (p. 75-76).

Lievrouw's proposed definition of the invisible college de-emphasizes conditions for formal organizational structures and proximity: "An invisible college is a set of informal communication relations among scholars or researchers who share a specific common interest or goal" (p. 66).<sup>8</sup> Price (1986) employs the model of invisible colleges in his assessment of the transition from Little Science to Big Science over the previous three hundred years:

We tend now to communicate person to person instead of paper to paper. In the most active areas we diffuse knowledge through collaboration. Through select groups we seek prestige and recognition of ourselves by our peers as approved and worthy collaborating colleagues. We publish for the small group ... Only secondarily, with the inertia born of tradition, do we publish for the world at large. (p. 80).

---

<sup>8</sup> While a more thorough treatment is beyond the scope of this chapter, further insight on informal communications within the invisible college is informed by a consideration of social network theory and social exchange theory. Roloff (1981) describes social exchange theory in interpersonal communication as, "a symbolic process by which two people, bound together in a relationship, provide each other with resources or negotiate the exchange of resources." Social network theory maps social relationships, linking nodes, or actors, to ties, or the relationships among actors, within a system (Granovetter, 1973). The concern in social network theory is not on individual actor characteristics, but the ties between actors in the system.

Prince, writing in 1986, attributed this transition from formal to informal communication channels to developments in modern transportation and a rise in affluence among premier scholars. The growth of exchange is further evidenced two decades later when the consideration for “modern transportation” is extended to include the Web’s “information highway.” The concept of the invisible college is particularly compelling to today’s information landscape and the changing nature of interactions and “publications.” As Borgman (2007) succinctly illustrates, “in a few short years, we have gone from ‘logging on’ to specific tasks to ‘always on,’ where the Internet is the communication channel of first resort for a growing array of activities” (p. 1).

#### 2.1.2.2. Units

In borrowing from Roosendaal et al. (2001), unit is used here to describe an authored “unit of information, e.g., such as a scientific article or a set of data” (p. 17). Units of communication take many forms, including paper, digital, and microform, and are distributed via several modes, including subscription-based services. Units, in part, reflect the scholarly record, described as the aggregation of the, “message that is communicated by scholars,” which may be published or unpublished (ARL, 1986, p. 5). These communications may be reflective of scholarly works, knowledge or ideas (ARL, 1986). ARL (1986) qualifies conference proceedings, journal articles, published research reports and books as formal publications. The journal article is the prevailing, preferred unit of communication, though it is not without criticism (e.g., concerns for gate keeping procedures and lack of transparency; delay in time between submission of an article and subsequent publication, if accepted) (Van de Sompel et al., 2004).

Presented in Table 2 is a list of content types gleaned from surveys on institutional repository (IR) deployment activity. The development of IR programs in recent years has contributed to new opportunities for universities, colleges, and research centers in the stewardship of their digitally-produced, institutionally-based scholarship in traditional and non-traditional formats. A number of definitions of IRs are offered in the literature (e.g., Barton & Waters, 2004; Crow, 2002; Johnson, 2002; Jones, Andrew & MacColl, 2006; Lynch, 2003; Ware, 2004). Summarizing the essential, shared elements, an IR is a persistent, interoperable and open digital archive comprised of cumulative scholarly content, “representing an historical and tangible embodiment of the intellectual life and output of an institution (Crow, 2002, p. 16).

**Table 2. Informal and formal units**

Informal			Formal
<ul style="list-style-type: none"> <li>• Pre-prints</li> <li>• Working papers</li> <li>• Technical reports</li> <li>• Conference proceedings (if not peer reviewed)</li> <li>• Conference presentations</li> <li>• Datasets</li> <li>• Databases/spreadsheets</li> <li>• Interview transcripts</li> <li>• Laboratory protocol</li> <li>• Software/documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Learning Objects</li> <li>• Syllabi</li> <li>• Class notes</li> <li>• Exhibitions/performances</li> <li>• Image</li> <li>• Audio</li> <li>• Moving images</li> <li>• Digitized musical scores</li> <li>• Blogs</li> <li>• Newspapers/Newsletters</li> <li>• Web pages/sites</li> </ul>	<ul style="list-style-type: none"> <li>• e-Portfolios</li> <li>• University electronic records</li> <li>• Email</li> <li>• Institutional reports</li> <li>• Departmental reports</li> <li>• Committee reports</li> <li>• Meeting agenda/minutes</li> <li>• Maps/plans/blueprints</li> </ul>	<ul style="list-style-type: none"> <li>• ETDs</li> <li>• Postprints</li> <li>• Articles</li> <li>• Journals/eJournals</li> <li>• Manuscripts</li> <li>• Books/eBooks</li> <li>• Conference proceedings (if peer reviewed)</li> </ul>

As shown in Table 2, these content types are diverse, representing an aggregate view of content deposited to IRs, or targeted for deposit. In line with opinions expressed in the literature (e.g., ARL, 1986; Borgman, 2007; Garvey & Griffith, 1971; Hagstrom, 1965), these units are categorized as formal or informal. These distinctions are not intended to be absolute, but are simply presented to demonstrate the diversity in scholarly “publications.” Additionally, it is demonstrative of the challenges in making determinations of formality

based solely on categorizations as context for identification.<sup>9</sup> The resulting table reflects a large breadth of units of communication in the system of scholarly communication falling outside the bounds of formal publication.

#### 2.1.2.3. Actors

Recalling the definitional attributes of the system of scholarly communication and the attributes for efficacy proposed by the NESC (1979), several categories of actors are evident. These actors may fill multiple roles. Focusing on the scholar-actor in their consideration of the value chain of scholarly communication, Roosendaal et al. (2001) present an equation of the scientific communication market (p. 17):  $F(\text{scientific communication markets}) = F(\text{actor, content, accessibility, applicability})$ .

This equation identifies the “actor pair” factor as the “generic stakeholders,” representing scholars’ roles as authors and readers (p. 17). A more encompassing descriptor, in consideration of these two roles, is that of researcher. In addition, scholars also serve in other roles within the system of scholarly communication: administrator; collaborator; gatekeeper, including peer reviewer and editor; and mentor, including advisor and instructor (e.g., ARL, 1986; Borgman, 2007). Traditionally, the processes of information mediation, in support of the functions for legitimization and certification, is served by editorial boards and publishers. Generally, the functions of awareness and dissemination are the domain of research libraries, publishers, scholarly societies, and scholarly presses (e.g., ARL, 1986;

---

<sup>9</sup> In making determinations of what constitutes formal communications, it is useful to consider bibliometrics. Per Borgman (1990), citation analysis is applicable to the study of formal channels of scholarly communication, qualified as the written record of scholarship. New bibliometric strategies have emerged in the past decade, notably webometrics, to capture citation relationships for units of communication not included in journal citation indexes.

Borgman, 2007; NESC, 1979).<sup>10</sup> Additionally, research libraries participate in the archival and access, preservation and curation function.

This is a high-level treatment of these roles. A treatment of the evolution of roles and subsequent blurring of boundaries between these actor groups is presented later in this chapter. Considering these actors are aligned with specific organizational cultures and disciplines, the latter categorized broadly as the sciences, the social sciences and the arts and humanities, roles are further defined by practices and procedures within their respective organizations and disciplines. A simplified, traditional trajectory of scientific information communication is, “author to publisher to library to reader” (Roosendaal et al., 2001, p. 13).

#### 2.1.2.4. Functions

Borgman (2007) organizes the functional aspects of scholarly communication around three areas: legitimization; dissemination; and access, preservation and curation (p. 66). Roosendaal et al. (2001) provide a complementary framework of functionality: registration, archive, awareness, and certification. These functions are represented in the following equation:  $f(\text{scientific communication functions}) = f(\text{registration, archive, awareness, certification})$  (Roosendaal et al., 2001, p. 18).

While the authors concede that their equation may lack comprehensiveness, others have adopted these functional attributes in discussions of the transforming channels and units for scholarly communication (e.g., Crow, 2002; Van de Sompel et al., 2004). Further, Van de Sompel et al. (2004), in consideration of Roosendaal et al.’s (2001) functions, identify a fifth function – reward.

---

<sup>10</sup> Of note here is an issue discussed in the literature and succinctly raised by Okerson (1992); that is, the issue of institutions paying twice for scholarship – first, through the support of scholars and secondly, through the acquisition of publications in which their scholars’ work is published.



Roosendaal et al.'s registration function aligns with Borgman's (2007) legitimization function, enabling the identification and establishment of scholarly ideas. Borgman (2007) qualifies this function as a reflection of the "trustworthiness" of scholarly work, in addition to "establishing priority" (p. 66). Taking into account the varied channels of formal and informal communication, Borgman (2007) raises the issue: At what point does legitimization occur? If considered from the viewpoint of the ownership of ideas, ala the 1976 U.S. Copyright Act, registration and legitimization could be thought to occur at the stage in which the original work was fixed, whether distributed informally and formally.<sup>11</sup> This simplistic angling, however, does not reflect the evidence of trustworthiness, described by Borgman (2007) as, "a part of the social process to assure readers that the content meet community norms" (p. 66).

Subsequent to registration, Roosendaal et al.'s (2001) certification function is reflective of the NESC's (1986) objectives for quality and entry. Crow (2002) summarizes the certification function as the, "quality of the research and/or the validity of the claimed finding" (p. 7). Borgman (2007) identifies this as the legitimization function. Certification practices and procedures vary across disciplines, as do the outcomes. For example, the Health Sciences exhibit a lower rate of acceptance, reported as low as 10%, in comparison to other disciplines.<sup>12</sup> Ginsparg (2001) references his own field, physics, as having higher acceptance percentages.

---

<sup>11</sup> On a related note, Borgman (2007) applies Copyright Law to discussions on what constitutes a "publication" referring directly to the statute in consideration of what it means to publish, or disseminate, a work (p. 98).

<sup>12</sup> La Porte et al. (1995) report a 10% acceptance rate for several health sciences academic journals, including the BMJ, the Lancet and the New England Journal of Medicine.

Peer review is a prevalent procedure that falls within the functional attributes for certification and legitimization. Lamont (2009) provides a compelling description of the implications of peer review across the system of scholarly communication:

The Latin word *academia* refers to a community dedicated to higher learning. At its center are colleagues who are identified as ‘peers’ or ‘equals’ and whose opinions shape shared definitions of quality. In the omnipresent academic evaluation system known as peer review, peers pass judgment, usually confidentially, on the quality of the work of other community members. Thus they determine the allocation of scarce resources, whether these be prestige and honors, fellowships and grants to support research, tenured positions providing identifiable status and job security, or access to high-status publications. Peers monitor the flow of people and ideas through the various gates of the academic community. But because academia is not democratic, some peers are given more of a voice than others and serve as gatekeepers more often than others. Still, different people guard different gates, so gatekeepers are themselves subject to evaluation at various times. (p. 2).

While the processes and biases of peer review are well reported throughout the literature, a full examination is beyond the scope of this chapter. A few issues are highlighted here. For example, Ginsparg (2001) offers the following critique of the peer review system, commenting that:

... for the most part the current peer review system has itself escaped a systematic assessment. Despite its widespread use, and the widespread dependence on it both for publication and for grant allocation, much of the evidence for its efficacy remains largely anecdotal. (para. 20).

Ginsparg’s critique is provided to highlight just one of many concerns in regard to the current culture of peer review within the system of scholarly communication. Peer review is a time-intensive, resource-laden process for actors within the system, particularly, authors, reviewers and other actors in the gate-keeping process. Further, Lamont (2009), in her examination of peer review practices and perceptions across six disciplines, points to a lack of cross-disciplinary consensus on assessing publications.

Roosendaal et al.’s (2001) awareness function enables distribution and availability, reflective of Borgman’s (2007) dissemination function and the NESC’s (1979) access

objective. As illustration, Roosendaal et al. (2001) reference the historical development of the *Philosophical Transactions of the Royal Society*, with emphasis on the awareness function as a driver in establishing the journal. The awareness function is evident in definitions put forth by ARL (1986) in reference to the emphasis on discovery and by Borgman (1990), who emphasizes use and dissemination. Borgman (2007) further characterizes the dissemination function as interdependent to the legitimization function:

*Diffusion* and *publicity* capture the idea of awareness by others in the community. *Transparency* reflects scholars' responsibilities to register their work in the public sphere, both to legitimize it and enable others to assess it. *Discourse* emphasizes the role that publications (formal communication) play in the scholarly communication. (p. 67).

Roosendaal et al.'s (2001) fourth function, archiving, serves the purpose of maintaining and preserving scholarship over time so it is available for future use. This function is demonstrative of the definitional attribute of the cumulative nature of the scholarly record set forth by ARL (1986). Borgman (2007) expands on this functional attribute, identifying "access, preservation, and curation ... as one composite function of scholarly communication, reflecting the continuous availability of the scholarly record" (p. 68). Treating each composite aspect in turn, access is distinguished from the dissemination function in that it is reflective of further action, beyond simply making content available, to the need over-time that the content also be "useful and usable" (Borgman, 2007, p. 68). Preservation implies an organizational commitment to ensure access is maintained over time. Lastly, curation represents a life-cycle approach to content management, extending the attributes for access and preservation to include actions for "maintenance, value-added, trust and temporality" (Borgman, 2007, p. 67).

The reward function, per Van de Sompel et al. (2004), "rewards actors for their performance in the communication system based on metrics derived from that system (para.

10). For example, citation impact is an essential hallmark in scholarly communication, and is reflected in several functions – Borgman’s legitimatization and dissemination functions; Roosendaal et al.’s (2001) awareness and certification function; and Van de Sompel et al.’s (2004) reward function.

Hagstrom (1965) categorizes recognition as institutionalized or elementary. The former, per Hagstrom, “is given in formal channels of communication in science, whereas interpersonal approval and esteem, or ‘elementary recognition,’ is given in direct communication” (p. 23). Hagstrom provides further illustration to distinguish between these two types of recognition:

An individual establishes his status as a scientist by having his research contributions accepted by a reputable journal; he achieves prestige as a scientist by having his work cited and emulated by others; and he achieves elite status by receiving collective honors. (p. 65).

Murray and Moore (2006), in their structural model for academic writing, propose two types of rewards: intrinsic and extrinsic. Intrinsic rewards include, “satisfaction, insight learning, engagement, empowerment, self-efficacy, and curiosity,” and extrinsic rewards include, “promotion, tenure, external endorsement, developing profile, and recognition” (Murray & Moore, 2006, p. 179).<sup>13</sup> Some of these extrinsic and intrinsic awards are evidenced in La Porte et al.’s (1995) description of publication motives in the biomedical sciences:

We publish to exchange information and to archive our work with some degree of permanence so as to leave a paper trail of evidence for future scientific work. We also publish for currency, to obtain promotion, to obtain grant support, and to obtain accolades from our peers (para. 4)

---

<sup>13</sup> This model elaborates on intrinsic and extrinsic rewards from an earlier conceptual model tentatively proposed by Moore (2003).

The excerpt above, through reference to the notion of permanence of the research record, also is illustrative of Roosendaal et al.'s (2001) archiving function, and Borgman's (2007) function for access and preservation, minus curation.

### 2.1.3. Evolution of the System

As written earlier, Borgman (2000) described a renewed interest in scholarly communication research in the mid-1990s. This renewed interest is influenced by a number of factors, both internal and external to the system, including: 1) ubiquity of networked, digital technologies; 2) evolving scholar expectations and behaviors of access and dissemination; 3) evolving roles and practices of stakeholder groups; 4) evolving mechanisms for content creation and communication; and 5) evolving capacity for hardware and storage (e.g., Borgman, 2007; Carr & Harnad, 2005; Cronin, 2007; Crow, 2002; Maron & Smith, 2008; Roosendaal et al., 2001; Van de Sompel et al., 2004; Willinsky, 2000).

Multiple phrases have been used to describe the evolution of media, tools, and procedures impacting scholarly communication practices and products; for example, information explosion, information overload, information revolution, information highway ... (e.g., 1986; CLIR, 2001; La Porte, 1995). Garvey and Griffith described the "information crisis" in scholarly communication in 1971; this concern continues thirty years hence. Roosendaal et al. (2001) comment, "indeed we seem to suffer from abundance, rather than from a scarcity of information and data. This leads us to the issue of complexity and above all the seeming abundance of scientific information" (p. 14). CLIR (2001) succinctly framed the implications of the digital environment for the contemporary academe:

... the introduction of digital technology onto campuses and into research libraries has fundamentally altered the information landscape and created problems for scholars and students that have potentially serious ramifications. The creation and dissemination of digital resources are creating new models of service and access, such

as licensing rather than owning essential intellectual assets. The mutability of digital documents is redefining what constitutes a text. (p. 3)

The contemporary landscape in which the system of scholarly communication takes place has contributed to a range of perspectives and recommendations in regard to roles and functions of the traditional system and the ancillary activity of scholarly publishing. Though the emergence in recent decades of networked information and communication technologies contributes to what some have identified as new or novel means and modes of communication, “renewed” may be a more appropriate term. The emergence of Web 2.0 technologies has produced new channels for communication and dissemination. These technologies are inherently social. Per Price (1986), the earliest manifestations of scientific papers, as mentioned earlier in this chapter, were a social device. While delivery mechanisms have evolved across both formal and informal communication channels, the essential functions of scientific communication are, per Roosendaal et al. (2001), “intrinsically invariant or – otherwise stated – are independent of changes in parameters such as information technology” (14). For example, the value of peer review is persistent. La Porte et al. (1995), in proposing an author submitted electronic archive for a global health information server, recommends, among other things, a feature supporting community review of submitted texts by permitting ranking and commenting.

In considering the contemporary, evolving nature of the system of scholarly communication, it is useful to look from the lens of both producer and consumer. Thirty years ago, the NESC (1979) reported the scholar-consumer as generally *satisfied* with the scholarly communication system, but the scholar-producer less so. Further, the NESC reported that other actors in the system – libraries, publishers and journal editors –

demonstrated more interest in the issues influencing the current state of scholarly communication than scholars in the humanities and social sciences.

The system has evolved in the three decades since the NESC report. The information world of today, per Van de Sompel (2000), “runs on Internet time” (p. 9). The current climate of scientific communication is in a state of flux, per Roosendaal et al. (2001). As mentioned previously, the literature provides examples of dissatisfaction with the current system of peer review (e.g., Ginsparg, 2001; Lamont, 2009). Awareness and dissemination are other cited areas of dissatisfaction. In advocating institutional repositories, Crow (2002) writes that, “the current system of scholarly communication limits, rather than expands, the readership and availability of most scholarly research (while also obscuring its institutional origins)” (p. 7).

Van de Sompel et al. (2004) attribute dissatisfaction with the traditional system to escalating subscription prices, issues of copyright, and, in consideration of traditional rewards for publication, concerns on what constitutes appropriate units of communication and channels for dissemination. The authors also note dissatisfaction with delays between the time something is submitted for journal publication until it is actually disseminated, a concern echoed by Roosendaal et al. (2001) and reflected in the NESC’s (1979) objective for “timeliness.” In the biomedicine domain, for example, La Porte et al. (1995) report the typical lapse of time between submission of an article and publication, if accepted, as between twelve to eighteen months.

Digital technologies have enabled broader access to scholarship previously confined to print-based modes of delivery.<sup>14</sup> These developments have contributed to speculation on the imminent demise of the print medium and the scholarly journal. La Porte et al. (1995)

---

<sup>14</sup> See Kling and McKim (1999) for an oft-cited work on the evolution from paper-based to digital journal publication.

severely advise, “Thus the terms ‘article,’ ‘paper,’ and ‘publication’ should die” (para. 14), while Odlyzko (1995) predicted, “Traditional scholarly journals will likely disappear within 10 to 20 years and the electronic alternatives will be different from current periodicals, even though they may carry the same titles.”<sup>15</sup> Though scholarly journals are still around, fifteen years after Odlyzko’s (1995) statement, both statements do provide indications for transformation rather than substitution. While the production of digital surrogates of previously print-exclusive publications provides enhanced access to scholarship, this approach does not leverage the extent of evolutionary potential possible in the digital realm. Further, per Parks (2002), there is a lack of incentive for scholars to use alternate publishing mechanisms, hence dependence on traditional journals will persist. While technological developments and contemporary imperatives, such as the Open Access movement, have contributed to new forms and behaviors in scholarly communication, print publications and the scholarly journal article remain a vital and consistent component of the system (e.g., Borgman, 2000; Day, 2003; Van de Sompel et al., 2004)

Roosendaal et al. (2001) propose a transformative approach to the system of scholarly communication as opposed to approaches that seek to substitute or replace nodes on the typical information chain. In advocating new opportunities, Ginsparg (2001) comments that:

Rather than ‘electronic publishing’ which connotes a rather straightforward cloning of the paper methodology to the electronic network, many researchers would prefer to see the new technology lead to some form of global ‘knowledge network,’ and sooner rather than later (para. 2).

La Porte et al. (1995) call for a redesign of the traditional journal article, recommending the term “research communication” in place of “article or paper” since these

---

<sup>15</sup> While La Porte et al. (1995) and Odlyzko (1995) point to the emergence of new technologies as contributory factors to the decline of the print journal, La Porte et al. (1995) also mentions the positive environmental impact once journals in print form are discarded.



two terms convey permanence in a two-dimensional format, while new technologies support enhanced communication and adaption beyond simply static transmission of stable texts.

#### 2.1.3.1. Evolving Actor Roles

Roosendaal et al. (2001) note the change in the customary chain of communication. In regard to actors within the system, this chain typically followed a path from creator to publisher to disseminator to end-consumer. Aberration from this simplified chain of explicit actor-delineated roles is contributing to shifts in actors' involvement in the functional aspects of scholarly communication. Typically, the functions for access, preservation and curation were seen within the domain of librarians and archivists (Borgman, 2007). Scholars are taking more active roles in these areas, and introducing themselves at all stages of the information chain – creation, publication, and dissemination.

In addition to new roles, the contemporary landscape also supports a return of functionality to certain actors. Ginsparg (2006) comments that, “it is also possible that the technology of the 21st century will allow the traditional players from a century ago, namely the professional societies and institutional libraries, to return to their dominant role in support of the research enterprise” (para. 10). In addressing libraries specifically, Van de Sompel (2000) highlights the value of libraries and their role in satisfying functional elements in the system of scholarly communication. Registration is enabled due to proximity to authors, while archiving and awareness is supported through long-standing service models. However, Van de Sompel (2000) also points to characteristics of libraries which may impede participation, describing libraries as “slow movers, hosted by slowly moving institutions,” and are in turn, “slow to recognize the fact that a new technology may allow [or beg] for a new mode of operation” (p. 9).

### 2.1.3.2. Evolving Channels and Units

Investigations into scholars' changing information seeking and use behaviors demonstrates the trend in scholars' transition from print to digital resources, and emerging preference for digital surrogates due to ease of access and enhanced functionality, such as full-text searching. Ginsparg (2006) elaborates on this digital dependence and the contemporary need for open access:

Because the current generation of undergraduates, and the next generation of researchers, already takes for granted that such materials should be readily accessible from anywhere, it is more than likely that this percentage will only increase over time and that the publishing community will need to adapt to the reality of some form of open access, regardless of the outcome of the government mandate debate. (para. 8).

Open access is an active area of activity in scholarly communication research, emerging in the mid-1990s (Borgman, 2007). Examples of seminal initiatives in support of open access to scholarship include the Budapest Open Access Initiative (BOAI) and the Open Archives Initiative, the latter resulting in tools for federated discovery of open access content<sup>16</sup> (Day, 2003). Cronin (2005), in a discussion of the influences on traditional bibliometric procedures, comments:

The web has challenged, and may revolutionize, many of the assumptions that have underpinned the established scholarly communication system. Radical proposals for open access and self-archiving have been put forward by a growing band of scholars, scientists, librarians, research administrators and others ... (p. 177).

Contemporary scholars have several channels for which to disseminate their scholarly work, outside the bounds of the traditional avenues (e.g., conferences, journals, monographs), including digital repositories, personal Websites and blogs, among others. However, as Jones et al. (2006) caution, "as personal web pages tend to be ephemeral, the long-term preservation of the research materials held on them is extremely doubtful" (p. 167). While

---

<sup>16</sup> For more information, see the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and the Open Archives Initiative Object Reuse and Exchange (OAI-ORE) at <http://www.openarchives.org/>.

organization-wide administration of back-up services is a typical benefit in an academic setting, this strategy does not guarantee long-term preservation.

Prepublication, when seen in context of the emergence of digital disciplinary and institutional repositories over the past two decades, is considered a recent dissemination development among scholars. While these mode of delivery are new, the practice is not. Garvey and Griffith (1971) reference dissemination of pre-prints among scholars in the 1960s. The disciplinary digital repository, arXiv.org, is an early example of a managed digital utility supporting the prepublication of scholarly research. Jones et al. (2006) propose a positive correlation between disciplines with a trend for faculty dissemination of materials online and disciplinary digital repositories. Self-archiving is a contemporary term used to describe these prepublication activities (e.g., Carr & Harnad, 2005; Borgman, 2007; Cronin, 2005). Trends in self-archiving vary among disciplines and are also influenced by trends within institutional units (e.g., universities, schools, departments).

These prepublication activities lead to considerations of the long-standing tradition of peer review in the vetting and subsequent dissemination of scholarship, in addition to a number of other concerns. How does prepublication impact subsequent formal publication? Is there a devaluing of published scholarship if disseminated in a pre-publication form? Or, does prepublication enhance impact? For example, facilitating perceptions of transparency through intermittent releases of new and novel research. Further, when considered from the consumer perspective, how do multiple access points to multiple versions of a unit of communication impact understanding within the research community? From the production angle, how do multiple copies, whether of different versions, or the same versions in multiple locations, impact an author's scholarly record and evaluation of impact factors? While these

issues are outside the scope of this chapter, examination of prepublication of scholarship is an active and timely area of research.

A look at the content of digital repositories provides some insight into not only the changing nature of communication units, but also evolved channels for dissemination. ARL (2009) identified nine types of digital repository collections, spanning a range of content types and reflective of the diverse units of communication as were presented in Table 2. These nine collection types are: 1) published faculty research archived for institutional purposes; 2) unpublished text material from faculty; 3) research data in various numeric and image formats; 4) administrative records; 5) primary source documents from libraries and research centers; 6) digitized book, journal and image collections; 7) instructional materials and courseware; 8) platforms for publishing journals; and 9) software (ARL, 2009, p. 5).

Writing over fifteen years ago, La Porte et al. (1995) asked:

How will the world of health information look when every original paper, letter of criticism, and review article, as well as every form, chart, and database in the computerised world, is accessible with a couple of dozen clicks of a mouse? We envision new forms of transmission of research communications. (para. 4).

Findings from ARL (2009) demonstrate movement toward La Porte et al.'s vision. Maron and Smith (2008) provide further insight into the evolution of units of communication. In a study of faculty perceptions on the value of digital scholarly works, eight types of valued scholarly resources emerged: 1) electronic-only journals; 2) reviews; 3) pre-prints and working papers; 4) encyclopedias, dictionaries, and annotated content; 5) data; 6) blogs; 7) discussion forums; and 8) professional and scholarly hubs (p. 7). In describing their study approach, Maron and Smith (2008) used the ARL parameter of "original and scholarly works" to evaluate the resources above, concluding these eight content types qualified as "scholarly resources" (p. 7). In a study of IR deployment, Chan adopts a proponent viewpoint

on the role of novel, non-traditional forms within the scholarly communication milieu, writing in the conclusion that:

By preserving and making accessible academic digital objects, datasets, and analytic tools that exist outside of the traditional scholarly publishing system, IRs also represent a recognition of the importance of the broader range of scholarly material that is now part of the scholarly communication process and record. (para. 74).

Borgman includes blogs as well as discussion lists, “chats, RSS feeds, and others soon to be invented” in describing the contemporary, “complex ... array of informal scholarly communication genres” (p. 99). The next section addresses one particular genre – blogs – within the system of scholarly communication.

## 2.2. The Blogosphere

Before considering blogs within the system of scholarly communication, it is necessary to address the pervasiveness of the medium in general. Since their introduction in the mid-1990s, blogs are now a ubiquitous component of online life and culture (e.g., Rainie, 2005; Gillmor, 2004). Sifry (2004) describes the blogosphere as doubling in size at least every five months. The growth in blogging is attributed to many catalysts. Notably, the availability of free, easy to use blog publishing and hosting services in the late 1990s lowered barriers for participation (e.g., Lenhart & Fox, 2006). The current extent of the blogosphere – the diverse heterogeneous aggregation of blog content and code – is immense, representing networks of co-producers comprised of bloggers, their readers, and service providers (Serfaty, 2005). Investigations on extent typically employ a survey research design.<sup>17</sup> The Pew Internet & American Life Project (PIALP) has conducted three survey studies to

---

<sup>17</sup> For example, Lenhart and Fox (2006) used a sample survey research design conducted via a random digital dial technique, targeting American adults identifying themselves as bloggers in earlier surveys of Internet users conducted by the PIALP between 2004 and 2005. White and Winn (2009) also used a sample survey research design, drawing from a sampling frame of 1.2 million bloggers registered at Technorati.com.

identify and characterize bloggers and blog readers (Lenhart, Fallows, & Horrigan, 2004; Lenhart & Fox, 2006; Rainie, 2005). Periodic reports on the characteristics of blogs – and their respective bloggers – registered at Technorati have been published since 2004 (Sifry, 2007; White & Winn, 2009; Olsen, Berlin, Olsen, McLean & Sussman, 2009).

An exercise in “counting” the blogosphere entails a look, in whole or in part, at instances, producers, consumers, and co-producers. For example, as of October 6, 2010, Technorati, the largest blog directory available, lists 1,242,410 blogs to their directory.<sup>18</sup> White and Winn (2009) report that 133 million blog records are indexed at Technorati. Blogpulse, a blog trend discovery service, identified over 71 million blogs as of January 2008; as of October 2009 that number has grown to over 107 million blogs.<sup>19</sup> However, something to bear in mind is that only a fraction of blogs, per Herring et al. (2005), are considered A-list; that is, widely read, linked to and referenced.

Lenhart and Fox (2006) report that 12 million U.S. adult Internet users – approximately 8% of all U.S. Internet users – have a blog. Over 1.2 million bloggers are registered at Technorati (White & Winn, 2009). Readership is extensive. About 57 million adult Internet users in the U.S. (39% of all users) read blogs (Lenhart & Fox, 2006). Bloggers are not only active producers, but also active consumers; nine out of ten read other blogs (Lenhart & Fox, 2006). Further, about eight out of ten bloggers contribute to other blogs by posting comments.

Provided below is a listing of popular blog service providers, reported in blog survey research findings from Lenhart and Fox (2006) and Hank, Sheble and Choemprayong (2007),

---

<sup>18</sup> Technorati Blog Directory: <http://technorati.com/blogs/directory/>

<sup>19</sup> See “Blogpulse Stats,” available at: <http://www.blogpulse.com> (accessed January 20, 2008 and then October 18, 2009).

as well as other reports from the literature (e.g., Gillmor, 2004; Rosenberg, 2009; Serfaty, 2005). This list represents a range of service options, from application-only tools to application and hosting services, either free or fee-based. Further, blog software programs may be provided as an exclusive application or in addition to a suite of applications.

- 20six (<http://www.20six.co.uk/>)
- Blog-City (<http://www.blog-city.com/community/>)
- Blogger (<https://www.blogger.com/start>)
- DiaryLand (<http://members.diaryland.com/edit/welcome.phtml>)
- Live Journal (<http://www.livejournal.com/>)
- Movable type (<http://www.movabletype.com/>)
- MySpace (<http://www.myspace.com/>)
- Radio UserLand (<http://radio.userland.com/>)
- Spaces –Windows Live (<http://home.spaces.live.com/>)
- Squarespace (<http://www.squarespace.com/>)
- TypePad (<http://www.pitas.com/>)
- Xanga (<http://www.xanga.com/>)
- WordPress.com (<http://wordpress.com/>)
- WordPress.org (<http://wordpress.org>)

Most bloggers elect to use a blog service provider; only a small minority (2%) report building their own blogging applications (Lenhart & Fox, 2006).

Investigations on the blogosphere take place across multiple disciplines and in response to a variety of research themes. A comprehensive account of the range of blog-related research is beyond the scope of this chapter, though a few examples are provided here to demonstrate the variety of investigations and objectives. For example, Miura and Yamashita (2007) and Nardi, Schiano, Gumbrecht, and Swartz (2004) investigated the socially situated actions of bloggers. Gurak, Antonijevic, Johnson, Ratliff, and Reyman (2004) conducted an in-depth examination of the blog as “rhetorical artifact.” Viegas (2007) and Qian and Scott (2007) investigated issues of identity, privacy and self-disclosure. Farrell and Drezner (2008) explored the role of blogs in political discourse.

### 2.2.1. Bloggership, Blawgosphere and Blogademia

While blogs are ubiquitous, there is a polarity of opinion on blogs, particularly in regard to the information value of the communication as well as endurance as a communication channel. Lenhart and Fox (2006) comment on this polarity of opinion in the context of news media and politics, writing that:

Some observers have suggested that blogging is nothing more than the next step in a burgeoning culture of narcissism and exhibitionism spurred by reality TV and other elements of the modern media environment. But others contend that blogging promises a democratization of voices that can now bypass the institutional gatekeepers of mainstream media. This democratization is thought to have implications for the practice and business of journalism as well as the future of civic and political discourse.” (p. 1).

Lynch (2004) addresses the potential value of blogging in the context of journalism and scholarship:

Blogs are important in the same way that websites are. Blogs really do range from the sublime to the ridiculous. They can be wonderful sources of thinking, analysis, pointers of interest. There is a growing fuss over the extent to which the blog is an essential form of legitimate journalism or scholarly communication. Blogs succeed or fail on the quality of the content. They can be a significant form of communication. (para 30).

The adoption of blogs as channels for communication by faculty, researchers and academics has led to the creation of three neologisms. First, the term “bloggership” emerged from the legal scholarship domain, reflecting the network of legal scholar bloggers, their blogs and, subsequently, their readers. Caron (2006) credits the coining of the term to Rafael Gely, founder of the Workplace Prof Blog.<sup>20</sup> The term evolved during discussions in preparation of the symposium, *Bloggership: How Blogs are Transforming Legal*

---

<sup>20</sup> Workplace Prof Blog: [http://lawprofessors.typepad.com/laborprof\\_blog/](http://lawprofessors.typepad.com/laborprof_blog/)



*Scholarship*.<sup>21</sup> Caron (2006) described Gely as a “recovering blogger,” since, after establishing the Workplace Prof Blog, he subsequently, “decided that blogging was not his cup of tea” (p. 1025). However, Gely’s Workplace Prof Blog is still active, as of October 7, 2010, edited and published to by a cohort of law school professors.

Smith (2006), a co-founder of The Conglomerate Bloggers,<sup>22</sup> describes and distinguishes bloggership as:

... a useful neologism because it distinguishes this sort of scholarship from the traditional, long-form scholarship that appears in law reviews and scholarly journals and because it distinguishes blogging that has scholarly aspirations from other forms of blogging (para. 8).

Additionally, the neologism “blawgosphere” has also been used, described by Dayal (2005a) as, “blogs by law school profs, students and grads,” and representing, “one of the most organized and lively pockets of online academic discourse” (para. 4).

Another term derived to broadly represent scholarship in the blogosphere is blogademia, with references emerging from the journal literature (Saper, 2006), as well as the popular press (Dayal, 2005a, 2005b; Halford, 2009). Rice (2009) credits Saper with the introduction of the, “neologism ‘blogademia’ to describe the emergence of the academic blog as a genre” (p. 304). Saper writes on the relation between blogs in the academe and current – and future – perceptions of scholarship:

The challenge of blogademia is to focus on this translation process of scholarship and knowledge into the currently disparaged and debased sociopoetic form of blogs. Beyond apprehending the issues at stake in using this form, one can begin to articulate the advantage of research that uses the blog, not as an object of study, but as a vehicle to comprehend mood, atmosphere, personal sensibility, and the possibilities

---

<sup>21</sup> The symposium was held at Harvard Law School on April 28, 2006. For more information, see: <http://cyber.law.harvard.edu/node/2457#>).

<sup>22</sup> The Conglomerate is a collaboratively-authored blog of law school professors writing on the topics of business, law, economics and society (see <http://www.theconglomerate.org/>). It was identified as one of the 100 best Websites by lawyers, for lawyers, as chosen by the editors of the ABA Journal (see [http://www.abajournal.com/blawgs/blawg100\\_2008/professors](http://www.abajournal.com/blawgs/blawg100_2008/professors)).

of knowledge outside the ego's conscious thought. The blog, podcast and wikis may hint, fleetingly, at the future tools of academia (para. 20).

There are many first-hand accounts, both informally and formally communicated, of scholars' perceptions of blogging in relation to scholarly communication (e.g., Althouse, 2006; Bearman, 2005; Caron, 2006; Glenn, 2003; Halavais, 2006; Jensen, 2007; Oi, 2006; Podgor, 2006; Quiggin, 2006; Smith, 2006; Solum, 2006a; Walker, 2006). However, research on scholars' perceptions of the blog in relation to scholarly communication and/or digital preservation is limited. See, for example, Hendricks (2010), Primary Research Group (2009), and Kjellberg (2009). However, no studies of scholars who blog, scholar blogs, and digital preservation, in combination, are identified in the literature.

#### 2.2.1.1. Instances

No exhaustive, complete inventory of scholar blogs in the blogosphere exists. Blogging takes place in a dynamic, shifting landscape. Further, it is unknown – and potentially, unknowable – what percent of scholars publish to blogs since it would require knowledge of how many scholars there are, both in the blogosphere and the real-world. Many issues complicate an attempt to identify all instances of scholars' blogs and scholars who blog, including: 1) the nature of the format (i.e., low barrier to create, as well as delete, a blog; 2) the publishing practices of bloggers (i.e., access controls, pseudonymous authorship); and 3) definitional ambiguity on who is a scholar blogger and, in turn, what is a scholarly blog (i.e., issues related to credibility and authority, as well as style and topical treatment). However, activity shows that scholar blogs are available across a range of disciplines. For example, Caron (2006), in citing Best (2006) and Solove (2006), reports 600 law-related blogs and 235 law professor bloggers (p. 1025). Further, Saper (2006) writes, “Academics, especially in the arts and humanities, have taken to blogs like ducks to water”

(para. 1), while Glenn (2003) notes, “Scholarly blogs are sprouting like mushrooms” (para. 6).

Though a complete inventory is lacking, there are a variety of strategies for identifying scholar blogs and, in turn, scholars who blog. While blogs and bloggers may be identified through simply ‘convenience’ (i.e., personal communication, links to scholar Websites; searches on the open Web), there are several types of sources available for improved identification: 1) references in the literature; 2) networks of scholar blogs; and 3) directory listings.

Articles on the role of blogging in the academe typically list author-identified instances of scholar blogs (e.g., Dayal, 2005a, 2005b; Ewins, 2005; Glenn, 2003; Mortensen & Walker, 2002). Dayal (2005a, 2005b) described and cited 13 scholarly blogs. When considering both the ephemeral nature of blogs and issues of access, 12 of these cited blogs are still publicly available, five years after publication of Dayal’s articles.<sup>23</sup> Eight are characterized as active, with new post(s) published within the previous month.<sup>24</sup> The other four are still available online.<sup>25</sup> Of these, two provide information in their most recent posts on the discontinuation of publishing to the blog. One attributed cessation to the project-specific timeline of the blog; the other attributed the reason for cessation to multiple factors, including the demand on time, technical maintenance, and shifting of professional responsibilities.

---

<sup>23</sup> The status of the thirteenth blog is unknown. The blogger name, as identified at the referenced blog URL, did not match the name of the blogger identified in the article.

<sup>24</sup> The URLs were visited January 24, 2010. See: <http://rushkoff.com/>; <http://pajamasmedia.com/instapundit/>; <http://drezner.foreignpolicy.com/>; <http://k-punk.abstractdynamics.org/>; <http://wayneandwax.com/>; <http://crookedtimber.org/>; <http://volokh.com/>; and <http://blogs.discovermagazine.com/cosmicvariance/>.

<sup>25</sup> URLs were visited January 24, 2010: <http://journalism.nyu.edu/pubzone/weblogs/pressthink/>; <http://www.sivacracy.net/>; <http://www.lessig.org/blog/>; and <http://www.zmag.org/blog/noamchomsky>).

Scholar blogs may also be identified by networks. Networks are distinguished as aggregates of blogs with a shared host location, typically organized by topical treatment or institutional affiliation, and are typically sponsored by an organization, such as a publisher or a university. Law Professor Blogs (<http://www.lawprofessorblogs.com/>) is an example of a network of legal scholarship blogs. Established in 2004, it is self-described as:

... a network of web logs ("blogs") designed from the ground-up to assist law professors in their scholarship and teaching. Each site focuses on a particular area of law and combines both (1) regularly-updated permanent resources and links, and (2) daily news and information of interest to law professors. Our editors are leading scholars and teachers who are committed to providing *the* web destination for law professors in their fields. (Caron & Hodnicki, n.d., para.1).

Another example of a topical network of blogs is ScienceBlogs, managed by Seed Media Group, and described as "... a digital science salon featuring the leading bloggers from a wide array of scientific disciplines."<sup>26</sup> The blogs are organized by topical treatment: Life Science, Physical Science, Environment, Humanities, Education, Politics, Medicine, Brain and Behavior, Technology, and Information Science. The Berkeley Blog is an example of an institutional network, aggregating and providing access to blogs from, "more than 150 UC Berkeley professors and scholars."<sup>27</sup>

Another source is directories of scholar blogs.<sup>28</sup> For example, the Academic Blog Portal (<http://www.academicblogs.org/>) is a resource, "for (a) academic bloggers, and (b) people who want to read academic blogs. It is a first step towards Scott McLemee's suggestion of an aggregating site to provide the academic blogosphere with a greater degree

---

<sup>26</sup> ScienceBlogs: <http://scienceblogs.com/channel/about.php>. Additionally, for other science-related blog networks, see: Nature.com Blogs (<http://blogs.nature.com/>) and Discover Blogs (<http://blogs.discovermagazine.com/>);

<sup>27</sup> The Berkeley Blog: See <http://blogs.berkeley.edu/about/>. For another example of an institution-based network, see: Kellogg Faculty Blogs (<http://www.kellogg.northwestern.edu/faculty/blogroll.aspx>).

<sup>28</sup> For example, see BlogScholar: Academic Blogging Portal (<http://blogscholar.com/>).

of coherence” (Academic Blog Portal, 2006, para. 1). Using a wiki format, the directory is maintained by a group of editors, but open for contributors to add new listings based on parameters presented by the site’s managing editors. These parameters are informative in regard to definitional ambiguities on who is a scholar:

... the author should be either a member of a third level institution's faculty (i.e. community college, college, university, technical institute or whatever), or pursuing a doctoral degree, or employed by a third level institution to do academically relevant work (such as working as a university librarian). (Academic Blog Portal, 2006, para. 6).

Another type of source is publicly published listings of scholars’ blogs compiled for various reasons. For example, Cline (2006) provides a listing of “Professors Who Blog.” This list, while no longer maintained, is still accessible. Lists may be more directed, with a topical focus, such as Worthen’s (2009) a list of medievalist Weblogs, or represent blogs attaining a formal recognition of merit, such as the Third Annual ABA Journal Blawg 100 (McDonough & Randag, 2009), the Research Blogging Awards 100,<sup>29</sup> or the Wall Street Journal’s Top 25 Economics Blogs.<sup>30</sup>

#### 2.2.1.2. Scholars and their Blogs

In general, bloggers choose to establish and maintain a blog for a variety of reasons. The list below represents motivations as described in findings from several researchers, including Lenhart and Fox (2006), Miura and Yamashita (2007), Nardi et al. (2004), and White & Winn (2009):

- Creative self-expression;
- Documenting and sharing personal experiences and emotions;
- Staying in touch with friends and family;
- Sharing practical knowledge, skills, and expertise;

---

<sup>29</sup> Research Blogging Awards 2010: <http://researchblogging.org/static/index/page/awards>.

<sup>30</sup> Top 25 Economics Blogs: <http://online.wsj.com/article/SB124768581740247061.html>

- Sharing viewpoints, opinion and commentary;
- Advocating action;
- Entertainment;
- Storing important information;
- Influencing thoughts and perceptions;
- Networking, meeting new people, and establishing and maintaining community, and;
- Generating revenue

Most bloggers, per Lenhart and Fox (2006), blog for creative expression and for documenting and sharing personal experiences. White and Winn (2009) also found self-expression as a primary motivator, along with sharing of expertise.

While there is limited empirical research into scholars' motivations for blogging, there is some anecdotal reporting on why scholars choose to blog. In academia, motivations for blogging are equally diverse. Althouse (2006), a legal scholar "blawger," describes her motivations for blogging, reflective of the "creative self-expression" factor:

The blog unlocks my thinking and creates a state of flow ... I get my momentum in part from the freedom not to connect every paragraph to the next. In this freedom not to connect things in the conventional way of the written page, I find new connections. Without the need to follow a theme, to produce a big idea, according to the regimentation of law review writing, I have a chance to see something new, to think something new. (para. 2).

Althouse's (2006) description is also inwardly (self) reflective of the factor for "influencing thoughts and perceptions." In returning to motives for bloggers in general, Lenhart and Fox (2006) identified a small group of bloggers for whom their activities were motivated as a public endeavor. Consider Borgman's (2007) position that the system of scholarly communication, "builds on long traditions in Western thought about open science and the social and economic benefits of the free flow of ideas" (p. 73-74). Blogging by scholars may be viewed as a "public endeavor," in support of the advancement of scholarship. Courant (2008) writes, "fundamentally, scholarship is about learning, describing,

and explaining what one has learned and how one has learned” (p. 204). The next section looks at how these essential elements are reflected in scholars’ blog publishing and blog readership.

In reference to scholarly blogging, Glenn (2003) asked, “is this a revolution in academic discourse, or is it CB radio?” (para. 1). Considering blogging in the context of scholarly communication, the process may mimic several behaviors and attributes of scholarly communication, reflective of the channels and functions described by Garvey and Griffith (1971), Price (1986), Roosendaal et al. (2001), Borgman (2007), and Van de Sompel et al. (2004). Presented here is a brief treatment of these issues. Further research on the perceptions and practices of scholars who blog would provide empirical evidence in support of this relationship.

In considering the foundations of the system of scholarly communication, Price (1986) describes the scientific paper, introduced in the 17th century, as a social device. Likewise, blogs are an interactive social device, exemplifying an open, shared community space for scholarly dissemination and collaboration. For example, Schmidt (2007) applied social interaction theory, and Marlow (2004) and Blanchard (2004) social network theory in their respective works examining relationships among communities of co-producers.

Blog commenting systems support interaction between bloggers and their readers, providing a convenient mechanism to post comments to specific posts. Such commenting systems facilitate an exchange of ideas and sharing of feedback (Lenhart and Fox, 2006). Gate-keeping, including peer review and editorializing, is a long-standing practice in establishing quality within the system of scholarly communication, typically manifested in formal publication in scholarly journals. While commenting on a blog does not constitute the

formality or rigor of peer review in journal publications, it may contribute (somewhat) to establishing quality and credibility. Additionally, a quick examination of collaboratively-authored blogs in the Law Professor Blog network, as well as a handful of other scholar blogs, provides examples of blogs identifying “editors” on their front page.

The social, communicative and collaborative nature of blogging places it well within the model of the invisible college, defined by Lievrouw (1990) as, “a set of informal communication relations among scholars or researchers who share a specific common interest or goal” (p. 66). Scholarly blogging supports the development of new ideas, providing an avenue for exploration with colleagues and other interested parties not confined by physical proximity. As Althouse (2006) writes, “If you write a specialized legal blog, it might occur to people to check-in with you if they happen to notice that there is a significant development in the field they remember is yours ...” (para. 16). Further, blogs are demonstrative of collaboration amongst scholars. Scholarly blogs, as with general blogs, may be singularly or collaboratively authored. This collaboration may transcend institutional affiliations. Two well-known examples of collaboratively authored blogs, Balkanization (<http://balkin.blogspot.com/>) and SCOTUSblog (<http://www.scotusblog.com/wp/>) represent inter-institutional collaborations among law faculty, legal scholars, and practicing attorneys.

Blogs may also be seen to serve Roosendaal et al.’s (2001) registration, certification and awareness functions, and Borgman’s (2007) legitimization and dissemination functions. For example, the blog medium may be used in the pre-release or drafting of a scholarly idea preceding or complimentary to other channels of communication. As such, the stewardship of blogs of this type may serve as an interesting complement to subsequent products. Further, there are examples in the literature of blog content being transferred to another media, in



whole or in part. For example, ideas and discussion initially posted to a group blog, upon invitation, were transformed into a paper symposium issue (Althouse, 2006). This example is further reflective of the reward function proposed by Van de Sompel et al. (2004). In their survey of general bloggers, White and Winn (2009) reported on the benefits of blogging. Approximately one out of four bloggers attributed invitations to events and/or speaking engagements to their blogging activities, and for about 20%, their blogs contributed to subsequent contributions to traditional print media.<sup>31</sup> Future investigations focused specifically on the rewards of blogging amongst scholars, both intrinsic and extrinsic (Murray & Moore, 2006), would serve to inform the ongoing dialogue regarding blogs within the system of scholarly communication.

Blogs support a range of content types (i.e., images, audio, video). For scholarship dependent on multimedia representations and interactive visualizations, blogs may provide a supplementary avenue for dissemination, in addition to other digital resources. Blogs also serve the role as a synthesized accounting for information traditionally published in other forms. Althouse points specifically to the long-worded law review article, noting that they “have become terribly bloated over the years” (para. 35). Blogs fill a role in that a concise synthesis of such works as law review articles may be more accessible to consumers, both in terms of physical access, but also in terms of enabling a quicker dissection and pointed understanding of the informational content.

Havalais (2006) and Mortensen and Walker (2002) contextualize scholar blogs within other information channels and units, including the salon, the coffee house, the lab notebook, and op-eds. From these metaphors, blogging can be seen as an alternative means for

---

<sup>31</sup> These events include: interest group meetings, blogger roundtables, and industry events. White and Winn’s (2009) categorization of events, while not grounded in the system of scholarly communication, could be seen to be reflective of informal and formal channels described by Garvey and Griffith (1971) and Borgman (2007).

discourse (i.e., discussing current issues, thoughts on new approaches, or emerging research agendas). Blogs provide an avenue for exploration and sharing of ideas in real-time as opposed to reliance on formal publications, in which the time between development of an idea to formal dissemination of findings may elapse over several months or years.

Discussions on blogs as a type of scholarship attribute the rise in academics' blogging to the need for quicker and broader dissemination of scholarship at multiple stages of development. The conditions under which to fulfill these needs, it could be seen, are manifested in the blog platform (e.g., Althouse, 2006; Borgman, 2007; Caron, 2006; Saper, 2006). While blogs are not replacements of traditional scholarly vehicles, they can benefit traditional publications by promoting more transparency of research and scholarship through public dissemination at multiple stages in the research process, from conceptualizing research questions through investigation, data collection, and analysis.

While the examples above provide a snapshot of viewpoints in which to consider blogs as part of the system of scholarly communication, reflective of channels and functions, there are several viewpoints critical of blogs as a component of this system. Althouse (2006) writes, "I have to confess that I balk at other kinds of writing. I feel so much resistance" (para. 6). While such sentiment could be used to dismiss academic blogging as simple resistance and disagreement to the process of more formal writing, it does point to a gap. That is, the difference in preparing something to be formally disseminated through traditional channels (e.g., peer reviewed journals) and preparing something for consideration with fewer restrictions placed on production. However, as made clear by Griffith and Garvey (1971), scientific communication is comprised of several channels in which scholars' exchange "unpublished" units of scholarship.

The “resistant” academic proponent of blogging described above recognizes the potentially capricious nature of blogs and issues of perception in the academe. As Althouse (2006) puts it:

When I first started to blog, it felt like a clandestine activity. I had no idea of what my colleagues would think of what I was doing ... But maybe you ought to think of ways to protect yourself. You might tick off some colleagues if you write about politics. People who don’t quite get blogging might decide you’re a lightweight if you talk about television. And, most disturbing for the scholar, you could lose the feeling for how to compose things on a grand scale. We do still need to write law review articles, don’t we? (para 10-11).

Several observations regarding blogs and scholarship are made from this statement. One, blogs are not a replacement of traditional models of publication. While blogs may (or may not) be viewed as a complement, blogs are not a substitute for traditional, peer reviewed journals and other publications. Saper (2006) writes, “most academics, including bloggers, do not think these blogs have any impact on scholarship, and the production of knowledge, precisely because they have no editorial review process” (para. 8). Podgor (2006), writing from the perspective of a legal scholar, sees blogs as more appropriate for assessing and recognizing professional service as opposed to scholarly publication. This was evidenced in Hendricks (2010) study:

... at this point in time, most academic library promotion and tenure committees do not weigh publishing a blog the same as publishing a peer-reviewed article. Some recognize it as service toward the profession ... (para. 30).

The preceding comments of Althouse (2006) and Saper (2006) are reflective of the fuzzy boundaries that occur when writing to the open Web. Saper (2006) comments on the view of blogs as channels for gossip, a view he supports but which he believes may be of some benefit to the scholarly enterprise:

Blogs are about gossip, but when academics gossip they may also lead to a new form of knowledge. When bloggers discuss infrastructure (who in your department is an idiot or a psychopath; how poorly the administration of your University functions, or

simply the trials and travails of the tenure-track publish or perish mill), they also discuss, unwittingly, the social processes of knowledge production, what counts as scholarship and discipline formation. (para. 12).

It is interesting to note that half of bloggers, as reported by White and Winn (2009), identify their predominant “styles” as sincere, conversational, and/or humorous; snarky and confessional were the least reported styles.

Scholars’ blogs may not be exclusive to scholarly topics. The format itself, as described by Althouse (2006), “perfectly accommodates moving from one thing to another, in and out of professional expertise, and shifting from humorous to serious” (para. 23). The recognition of the humorous factor of blogs provides support for those who do not identify blogs as potential channels and units of scholarly communication since humorous, typically, is not a hallmark of scholarly writing.<sup>32</sup> Some, such as blogs affiliated with the Law Professor Blogs network, make explicit their focus on specific topical treatments. The following excerpt describes the academic and teaching focus of the network:

Our blogs are not a collection of personal ruminations about the Presidential campaign, the war in Iraq, or what the editor had for dinner last night. Neither do our editors offer their personal views on every policy issue in the news or every new court decision. We leave that terrain to the many existing blogs with that mission. Instead, our editors focus their efforts, in both the permanent resources & links and daily news & information, on the scholarly and teaching needs of law professors. Our hope is that law professors will visit the Law Professor Blog in their area (or areas) as part of their daily routine. (Caron & Hodnicki, n.d., para. 5).

Some scholars’ blogs, such as the blog of Althouse (2006), cross boundaries between scholarly posts and posts resulting from simply everyday life. This is reflective of White and Winn’s (2009) finding that approaches to blogging are not exclusive; bloggers tend to identify their blogs as both personal and professional. This leads to the quandary of whole or in part; that is, do we identify a true scholarly blog as one that conveys research-related ideas

---

<sup>32</sup> Though, for general bloggers, humor is a popular style, as reported by White and Winn (2009).

exclusively, or do we distinguish at the item-level – scholarly posts among all posts? Further, what benefits would such distinctions provide? If we consider the scholarly blog as a candidate for programmatic digital preservation, we will have to consider how to address the confluence of themes, scholarly or not. Althouse (2006) does note the utility of her blog, self-reporting that, “it has been clearly expressed to me in a number of ways that the people at the school appreciate my blogging and see the twenty percent of it that deals with law as part of my work, an intriguing hybrid of scholarship and service” (para. 22).

Perceptions of how scholars’ perceive their blogs in relation to scholarly communication is an outstanding area of research, though there are numerous reports in the literature (and blogosphere) of individual opinions. Some see a clear relationship between scholarly blogging and scholarly communication (e.g., Caron, 2006; Smith, 2006; Althouse, 2006; Havalais, 2006; Solum, 2006a; Walker, 2006). Althouse (2006) sees blogs as a “competing model of scholarly expression,” though she notes that “blogs won’t replace more sustained, detailed scholarly projects,” but, in reference to a discipline-specific publication type, could contribute to, “raising the standards for law reviews” (para. 38).

Althouse (2006) cites an interactive exchange via a blog on this very question, posed by Bearman (2005), titled, “How might we improve blogs as an academic medium.”

Bearman (2005) provides some insight into his query, writing that:

With this question, I mean to pick up some themes that Ron Wright raised in this recent post about blogs in the academy. Ron suggested, *inter alia*, that it may become important for legal scholars “to show how the blog format serves as a *supplement* to their other academic work, and not an efficient *replacement* for much of that work.” This insight raises the broader question of whether and how the blog format — either through improved technological infrastructure or through the development of professional norms about the medium — can and will evolve to become a more respected and trusted academic medium for law professors.

Of course, a lot of law professors may not want blogs to become a more respected and trusted academic medium, perhaps for fear that blogging will become just another

professional obligation. But I seriously doubt blogs will ever evolve into a *required* medium for law professors: I suspect tenure committees in the future are unlikely to be concerned that a junior prof has not done enough blogging, just as tenure committees now are unlikely to be concerned that a junior prof has not written enough op-eds.

I am concerned, however, that blogs may never evolve into a truly *respected* academic medium for law professors, even though I think a case might be made, especially with the recent forums sponsored by SCOTUSblog, that we are starting to see more and more first-rate scholarly ideas appearing on the blogosphere well before they surface in more traditional outlets. Thus my question for those interested in having (at least some) blogs further evolve into a respected and trusted academic medium for law professors: how might we help make that happen? (para. 3-5).

In a summary of the exchanges taking place in the comments section to Bearman's post, Althouse (2006) notes the dichotomy between work and hobby, coming down on the side of hobby (para. 28). A further look at the comments and trackbacks show several themes emerging, both on the side for play and the side for work. This is reflective of Lenhart and Fox's (2006) finding that bloggers, in general, view blogging a hobby. While the sense of blogging as hobby may diminish the value and import of blogging as both a product and process in the system of scholarly communication, this viewpoint may not be so simple to decipher. For example, the boundaries between professional, academic, and personal life may be difficult to distinguish because these aspects may be intertwined in the lives of scholars.

Some recognize the boundaries between topicality and generalizations may place scholars' blogs as just another blip in the blogosphere, and not easily identified, or understood, as an extension of scholarly thought and writing. Wegner (2005) notes that, "the inclusion of off-topic posts probably ceases to be a helpful branding tool and becomes a serious branding liability." Further, in response to a comment to his post on how the blog, as an academic medium, might be improved, Bearman (2005) writes that, "My sense is that, despite the genius of google, the blogosphere currently lacks means to ensure quality scholarly posts are easy to find, access and utilize." While Bearman is addressing issues of

immediate accessibility and utility of these self-described “scholarly posts,” the challenge is further exacerbated when considering issues of access and use into the future.

Future work investigating motivations for blogging as well as perceptions on how the blog fits within the scholarly communication milieu would be a valuable contribution to current understandings of scholarship in the digital age. This is particularly telling since academics’ motivations for blogging and perceptions of impact and reward, both personal and professional, may be diverse. An author’s intent of publishing to a peer reviewed journal, it may be assumed, is evident – to disseminate accepted and vetted scholarship in an accepted form – though underlying intents may be varied (i.e., in compliance with funding requirements; in pursuit of tenure, etc.). What, then, is achieved through blogging?

### 2.3. Digital Preservation

While preservation is a function of the system of scholarly communication (Borgman, 2007; Roosendaal, Guerts, & van der Vet., 2001), the digital landscape in which these communications take place complicates approaches toward achieving this function. Contemporary calls for digital preservation usually adopt doomsday-like forecasts. Tibbo (2003) frames the approach to this problem succinctly: “The challenge, if today’s cultural and scientific heritage is to be preserved, is to build the technologies and infrastructure to preserve digital assets, not simply try to rescue them after near disaster” (p. 13). The CLIR report (2001), while primarily concerned with preservation of physical formats, does address implications for digital preservation, “grounded in recognition that without preservation today, there will be no access tomorrow” (p. 4). CLIR cautions that, “If libraries, archives, or historical societies do not collect instances of recorded information, then the chance of their survival is slim. Loss is inevitable” (p. 3). It is interesting to note that Harvey (2008), in the

role of devil's advocate, calls for actual reports of data loss when making the case for digital preservation, as opposed to the prevalent practice of supporting calls through reports of data recovery.

Hedstrom et al. (2003) delineate genre- and format-agnostic benefits of long-term digital preservation, including: "protection and conservation of cultural memory; global access to open knowledge and support for cross-disciplinary collaboration; preservation for accountability; reduction of costs by information re-use; foundation of a knowledge economy; and development of digital libraries" (pp. 4-5). CLIR (2001) succinctly describes the intent, or "art of preservation," regardless of format, "is to minimize the risk of loss while continuing to keep collections usable for researchers" (p. 87). It involves both "preventive" and "corrective" measures to "stabilize items" (CLIR, 2001, p. 90).

### 2.3.1. Defining Digital Preservation

Many definitions of digital preservation are presented in the literature. Digital information spans disciplines and stakeholder groups. This diversity of communities involved in digital preservation necessitates negotiating amongst differences in terminology. The Center for Research Libraries and OCLC (2007) note, "Digital preservation interests a range of different communities, each with distinct vocabulary and local definitions for key terms" (p. 8). The provision of glossaries in key documentation in support of digital preservation program development and research provides some evidence of efforts for consistency across key terms (e.g., CRL & OCLC (2007), CCSDS (2002), and DCC & DPE (2007)).

However, DigitalPreservationEurope (2007) points to the need for a consensual definition of digital preservation as a contributing factor in the lack of success in current



digital preservation research activities. This lack of consensus contributed to the recent drafting of definitions by the American Library Association's Association for Library Collections and Technical Services' (ALCTS) Preservation and Reformatting Section (PARS) Working Group on Defining Digital Preservation (WGDDP). The intent was to contribute to shared terminology and clarity on fundamental concepts. The PARS approach resulted in three definitions. The "short" definition is the most succinct: "Digital preservation combines policies, strategies and actions that ensure access to digital content over time" (WGDDP, 2007, para. 7). The "medium" definition extends to material types, challenges and goals:

Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time. (WGDDP, 2007, para. 8)

Evident in these definitions is the element of extent. That is, for how long should digital material be preserved? In the Open Archival Information System (OAIS) Reference Model, long-term is defined as:

A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future. (CCSDS, 2002, p. 1-11).

Further, an exercise in defining digital preservation is not complete without the recognition of other terms treated in the literature as synonyms. The term, archiving, has been co-opted to reference digital preservation actions. As Tibbo (2003) makes explicit: "Traditionally, *preservation* and *archiving* have had two distinct definitions with preservation being a necessary component of, but not equivalent to, the totality of archiving" (p. 11). Hedstrom (1991) points out that, "the word archive has lost much of its traditional meaning and associations," and that, "archivists have literally lost control over the definition of

archive” (p. 336). CLIR (2001) cautions that, “despite the recent appropriation of ‘archive’ as a verb to mean ‘store’ or ‘to preserve,’ the traditional meaning of archives as a noun is narrower” (p. 85). Tibbo (2003) provides further clarification in terminology, emphasizing the goals of the processes of archiving:

For archivists, archiving is a complex process that can range the lifecycle of information and that involves an array of different functions including selecting, acquiring, arranging, and describing, preserving (physical preservation), and providing access. Archiving always has the goals of *preserving* and *making accessible* documents, records, and other data of *enduring value*. Enduring value stems from a document or record’s intrinsic attributes, the contextual documentation that surrounds it, its relationship to other records and entities, and assurance of its authenticity and reliability. (p. 8).

The emphasis on an information object’s life cycle, a feature of the excerpt above, is further demonstrated through the recent emergence of the term “digital curation.”<sup>33</sup> Pennock (2007) provides the following definition of digital curation:

Digital curation, broadly interpreted, is about maintaining and adding value to a trusted body of digital information for both current and future use: in other words, it is the active management and appraisal of digital information over its entire life cycle. (para. 2).

The Digital Curation Lifecycle Model provides a high level overview of the steps involved in the curation of information objects, beginning at the point conception (Higgins, 2007). The model prescribes a path for data, whether digital objects or databases, through three categories of action: full-lifecycle, sequential, and occasional.

#### 2.3.1.1. Key Concepts

The seminal publications of Garrett and Waters (1996) and Rothenberg (1995) are frequently referenced as milestones in the establishment of digital preservation as an active

---

<sup>33</sup> The term digital curation emerged in 2003 in a report to the Joint Information Systems Committee (JISC) in support of the establishment of the Digital Curation Centre. For more background on these activities and the emergence of the term digital curation, see “What is Digital Curation” at the Digital Curation Centre Website: <http://www.dcc.ac.uk/about/what/>

practice area. However, the 1962 establishment of the Survey Data Archive at the Inter-university Consortium for Political Research<sup>34</sup> preceded these respective publications by three decades. In writing on the impetus and development of the archive, Converse (1964) states, “the great flexibilities of duplication and transmission of information permitted by the new technology can now justify the systematic archival accumulation of voting statistics in permanent and machine-manipulable form ...” (p. 274). Lynch (2003) stressed in his definition of an IR that it is an organizational commitment; this characteristic can be seen to extend beyond an IR to any systematic response to digital preservation, including data archives and other types of digital repositories.

The contemporary terms, digital archive and digital repository, are often used interchangeably, though Ware (2004) comments “that ‘repository’ has become the preferred term rather than ‘archive’ because the latter implies the panoply of stewardship and long-term preservation that may not in fact exist” (p. 5). Peters (2002) distinguishes five kinds of digital repositories: individual, disciplinary, institutional, consortial, and national. Heery and Anderson (2005) expand on these categories in their definition of a repository: “a managed storage system with content deposited on a personal, departmental, institutional, national, regional, or consortial basis, providing services to designated communities, with content drawn from the range of digital resources that support learning, teaching and research (p. 3). Further, Heery and Anderson (2005) offer a typology of digital repositories, categorized around content type, coverage, functionality, and designated communities. While the term may be applied across a range of contexts and uses, Rusbridge, as cited by the DCC and DPE

---

<sup>34</sup> The organization is now known as the Inter-university Consortium for Political and Social Research; ‘Social’ was added in 1975, per organization’s Website; see: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/org/timeline-more.jsp#1962c>.

(2007), makes clear that, “few apply it only to collections of digital materials that are intended to survive in an understandable way for very long periods into the future” (p. 10).

Garret and Water’s (1996) clarify the multiple actors and their roles – creators, archivists, users and society as a whole – in making explicit a fundamental principle in digital preservation: “The first line of defense against loss of valuable digital information rests with the creators, providers and owners of digital information” (p. 40). Hedstrom et al. (2003) proposed ten principles and assumptions for digital preservation research and projects, summarized here as entailing: 1) a long-term perspective, from point of creation forward, reflective of the life-cycle of digital objects and reflective of a dynamic, iterative process utilizing multiple approaches to preservation; 2) requirements for authenticity and integrity of objects, and scalable processes; and 3) prioritization of preservation research under the condition it not be done in isolation of practice and reflect shared responsibilities among a community of multi-disciplinary researchers (p. 5-9). Moore (2008) characterizes the underlying concepts of digital preservation in three parts: 1) as “communication with the future;” 2) as “validation with communication from the past;” and 3) the “ability to verify properties of the preservation environment ... are working correctly” (paras. 1-3).

Framing outcomes of digital preservation in reference to outcomes for traditional, analog-based preservation, CLIR (2001) points out that, “preservation of the digital artifact it is successful to the degree that it maintains over time the chief distinctions of that object as digital – its functionality, its formatting or whatever is important about the digital object for a particular use and a particular user” (p. 50).<sup>35</sup> Caplan (2008) succinctly describes the goals of

---

<sup>35</sup> Additionally, in comparing principles for non-digital and digital materials, CLIR (2001) finds that “the fundamental principles of preservation apply to all formats ... but the techniques and costs of the various treatments differ dramatically” (p. 87).

digital preservation as: to get, describe, interpret, secure, authenticate, access and perform.

These outcomes are supported by what Garret and Waters (1996) refer to as essential operations for a digital preservation archive: appraisal and selection, accessioning, or ingest, storage, access, and systems engineering.

Key concepts are evident in the WGDDP's (2007) long-form definition for digital preservation, and the full lifecycle, sequential and occasional actions mapped in the Digital Curation Lifecycle Model (Higgins, 2007). In addition to the elements for organizational commitment and goals of digital preservation provided in the short and medium definitions, the WGDDP (2007) elaborates on approaches for content creation, integrity and maintenance. The definition is reflective of fundamental concepts and best practices offered in seminal publications from leaders in digital preservation:<sup>36</sup>

Content creation includes: 1) clear and complete technical specifications; 2) production of reliable master files; 3) sufficient descriptive, administrative and structural metadata to ensure future access; and 4) detailed quality control of processes.

Content integrity includes: 1) documentation of all policies, strategies and procedures; 2) use of persistent identifiers; 3) recorded provenance and change history for all objects; 4) verification mechanisms; 5) attention to security requirements; and 6) routine audits.

Content maintenance includes: 1) a robust computing and networking infrastructure; 2) storage and synchronization of files at multiple sites; 3) continuous monitoring and management of files; 4) programs for refreshing, migration and emulation; 5) creation and testing of disaster prevention and recovery plans; and 6) periodic review and updating of policies and procedures. (WGDDP, 2007).

This definition reflects the two most common approaches to preservation: migration – both format and media migration – and emulation. Caplan (2008) frames these and other

---

<sup>36</sup> The literature is rich with compelling publications on conceptualizing digital preservation and, subsequently, strategies toward digital preservation action. There are several leading publications that merit particular note due to their impact on the contemporary digital preservation research environment: Garret and Waters (1996); Harvey (2005); Hedstrom (1998); Hedstrom et al. (2003); Lavoie and Dempsey (2004); Lynch (2002, 2003); NSF and LOC (2003); Ross (2000); Rothenberg (1995); and Tibbo (2003).

approaches as passive or active. Passive approaches are media refreshment, media migration and maintenance. Active approaches entail emulation, format migration, format normalization, computer museum approach, universal virtual machine, and universal virtual computer. Other approaches to preservation as described in the literature include what CLIR (2001) describes as, “preservation through handling approach” (p. 52). Illustrating how this somewhat coincidental approach to preservation may be approached systematically, CLIR (2001) writes that, “digital information that is frequently used by patrons stands a better chance of being migrated and refreshed, and therefore is more likely to continue to be available in future generations” (p. 52). The principle behind the Lots of Copies Keeps Stuff Safe (LOCKSS)<sup>37</sup> method for preservation is described by CLIR (2001) as, “preservation through proliferation” (p. 52). The lots of copies viewpoint is echoed by Crow (2002) in his description of the interoperability attribute for IRs. Per Crow (2002), “digital archiving best practice suggests that multiple mirrored and distributed repositories, varying in location and formats, contribute to a sound preservation policy” (p. 10).

The concepts of selection and appraisal are also key to understanding the principles and practices of digital preservation. Waters and Garrett (1996) succinctly capture the dilemma of the long-term availability of digital information, writing:

Rapid changes in the means of recording information, in the formats for storage, and in the technologies for use threaten to render the life of information in the digital age as, to borrow a phrase from Hobbes, ‘nasty, brutish and short.’ (p. 2).

They go on to write that, “Some information no doubt deserves such a fate, but numerous examples illustrate the danger of losing valuable cultural memories that may appear in digital form” (p. 2). In response to the latter statement, decisions about what digital information may be left to such fates is an active area of inquiry.

---

<sup>37</sup> See the LOCKSS website for more information: <http://www.lockss.org/lockss/Home>

The diversity and changing nature of the contemporary information landscape complicate selection and appraisal decisions. In drawing on work in other disciplines, the DPE (2007) makes mention of memory, writing:

In human psychology, forgetting is not just a failure of memory, but an essential mechanism for clearing “out” irrelevant facts and preserving the efficiency of recall. Problems here are the identification of what to forget, how to deal with related knowledge and the actual implementation of “forgetting.” (para. 51).

The DPE (2007) composed their research roadmap in support of the “concept of a Collective Memory of the future” (para. 41). In relating several of their research recommendations, the DPE, among many recommendations in support of the Collective Memory of tomorrow, makes explicit the need for selection.

Decisions on selecting materials for preservation are a compelling practice area in digital preservation. Hedstrom (1991), in laying out a research agenda for the preservation of electronic records, notes the benefit of electronic records research for archivists, heeding archivists to “account for the social, economic, and political aspects of organizational life that mediate how information technologies are adopted and used by organizations (p. 337). The author continues, writing that, “electronic record keeping is both a technological and a sociological phenomenon, and solutions lie in both the technical and administrative realm” (p. 337-338). Though appraisal is a critical underlying concept in digital preservation approaches, it does not translate to an integrated approach. For example, consider the role of archivists in IR programs. Jones et al. (2006) make explicit the necessity of an archival function, taking into consideration not just the goal of preservation, but the dual need to appraise the content in determination, “as to ... future value for the institution” (p. 15). However, Markey, Rieh, St. Jean, Kim, and Yakel (2007) describe, “the marginalization of the archivist with respect to IRs” (p. 79). The authors reference Crow’s (2002) viewpoint as

IRs in competition with archives as a potential contributing factor to the diminished role of the archivist in IR initiatives.

Blouin (1999) observed, “There is a heightened interest in the implications of selection and what constitutes archives” (p. 103), and that, “though we have procedures for the selection and appraisal of our collections and record groups, we are used to making these decisions in comparative isolation” (p. 110). While appraisal is the domain of archival professionals, there is recognition of the need to involve producers and consumers. This is reflective of Garrett and Water’s (1996) stress on the role of the creator in digital preservation activities. For example, CLIR (2001) writes that:

The act of choosing what gets treatment and deciding why is based on a constantly changing evaluation of the value, use, and condition of an item, and all of these factors are constrained by limited resources and inadequacies of technology. While the library and its personnel can be relied on to provide the expertise to assess condition, decide on treatments, track use, and so forth, the question of the changing valuation of the significance of artifacts for research and teaching is a matter that needs to be informed by scholars and teachers. (p. 90)

Further, CLIR (2001) makes note of the challenge of multiple opinions in making selection decisions, illustrating the tension by comparing two camps, “those who think we should collect as broadly as possible to expand our research base and those who think too much information can impede one’s ability to find meaningful information” (p. 2).

Elements to be considered in preserving physical resources, distinguished as “cardinal features” resulting in “best practice,” are: 1) age; 2) evidential value; 3) aesthetic value; 4) scarcity; 5) associational value; 6) market value; and 7) exhibition value (CLIR, 2001, p. 9). These accepted best practice approaches introduce challenges in how to translate these elements for print, analog materials to born digital objects. Further complicating selection decisions for digital materials, as pointed out by CLIR (2001), is the challenge of translating



elements for assessing the intrinsic value of print-based resources – e.g., originality, faithfulness, fixity and stability – to born digital resources.

#### 2.3.1.2. Current State of Research

Numerous projects and activities have been ongoing over the past several decades, resulting in scores of agencies contributing to hundreds of research initiatives, programs, standards, tools, and systems.<sup>38</sup> However, work in the area of digital preservation may still be considered developmental. As Heery and Anderson (2005) point out, “the effective curation and preservation for the long term is a complex process and the necessary methods and skilled staff are only now starting to emerge” (p. 21). For example, in their census of IRs at American universities and colleges, Markey et al. (2007) found little evidence of preservation functionality beyond guaranteeing long-term preservation for PDF files. The authors concluded that the lack of preservation functionality is a pressing concern in overall IR functionality, remarking, “at the least, such an improvement fulfills a key reason for the very existence of IRs” (Market et al., 2007, p. 83).

Restoration approaches, per the DPE (2007), are an outstanding research area in digital preservation, writing that, “rendering ... objects and making their content known remains a significant challenge for digital preservation and still requires much research into the reconstruction of digital objects” (para. 31). Lynch (2005), speaking on this issue, states that:

This is a hard area to do compelling research in: without the digital analogs of physical accelerated ageing test beds, most research is either about tools, about identifying approaches that don’t work, or is highly speculative in nature – how do you prove that your approach in fact will preserve data for a thousand years without

---

<sup>38</sup> For a listing of projects and initiatives, see: [http://ils.unc.edu/~hcarolyn/inls752\\_supplement\\_fall2009.pdf](http://ils.unc.edu/~hcarolyn/inls752_supplement_fall2009.pdf). This list is derived from compilations originally prepared by Helen R. Tibbo as supplement to the syllabus for INLS 752: Digital Preservation and Access, School of Information and Library Science, University of North Carolina at Chapel Hill, Fall Semester, 2005.

having to wait that long? (You can prove it will fail in much less time, of course). (para. 14).

The DPE (2007) points to several factors contributing to the lack of success in realizing sound digital preservation resolutions despite nearly two decades of active research in the area. Lack of consensus in defining digital preservation is one contributing factor. Others include: 1) loss of focus due to a broadening rather than narrowing of the research agenda; 2) lack of practical experience due to few instances of proposed approaches transitioning from the research environment to real-world, practice settings; 3) fragmentation, brought about due to partial, as opposed to holistic, approaches to research problems; 4) frictional issues and losses that arise from competing, as opposed to collaborating, organizational research initiatives; 5) issues of intellectual property rights; and 6) lack of training opportunities for promoting new, and needed, professional skills for implementing digital preservation activities in practice settings (para. 17).

#### 2.4. Blog Preservation

As potentially valuable additions to the human record, advocates have called for the stewardship and long-term preservation of blogs, both in general as well as specific to the blogs of scholars (e.g., Dempsey, 2007; Entlich, 2004; O'Sullivan, 2005; Paulus, 2006). But the gap between interest and capability for doing so is evident. Entlich (2004) illustrates the challenges for blog preservation to Web archiving in general, noting strategies for the latter are still in the early stages of development. The author identifies six common issues: 1) copyright; 2) robot exclusion; 3) dynamic content; 4) inaccessibility due to password protection; 5) unique file formats, and 6) miscoded material. Further, Entlich (2004) points to two issues specific to blog preservation: 1) blogs that are off the grid; that is, not hosted by

typical service providers; and 2) complications for interactivity features, namely commenting and in-bound and out-bound linkages.

Lynch (2004) makes the case that preservation of blogs may be easier to accomplish than that of Websites, drawing comparisons between the nature of Websites, in general, with that of a typical blog:

In another sense, however, blogs are much easier to handle than typical websites. Most blogs that I have seen are in essence “grow-only” with older material rolled off to an archive perhaps. On web sites, when you update a page you replace it. So it’s much more likely that when you “harvest” or “crawl” a blog you are going to get its entire history up to the point you are visiting it. For most web sites, you have to visit over and over again to get that history, which is represented as a series of versions. So I’m not sure that technically blogs are worse than websites broadly. (para. 31).

The challenge of blog preservation is compounded by a myriad of factors in consideration of the goals for preservation: to get, describe, interpret, secure, authenticate, access and perform (Caplan, 2008). While these goals, respectively, necessitate a consideration of a triage of actions and approaches, treated in the aggregate makes clear the need for enhanced understanding of a range of behavioral, legal and technical issues. Blogs are dynamic and mutable. In their exploratory research of general bloggers’ perceptions on digital preservation, Hank et al. (2007) report that 71% of bloggers support preservation of their own blogs, with 36% supporting preservation of all blogs. However, the diversity and extent of blog instances, the nature of the blog format, and the diverse publishing practices and preferences of bloggers complicates digital preservation. The next section highlights some of these challenges, drawing on survey research of bloggers in general.

#### 2.4.1. Blogger Characteristics

##### 2.4.1.1. Basic Demographics

Most bloggers, regardless of type or topical treatment, are young. Lenhart and Fix (2006) found that about 50% are below 30 years old, and 30% between the ages of 30 and 50

years old, while White and Winn (2009) report similar findings, identifying half of bloggers as aged 18 to 34 years old. Lenhart and Fox (2006) report no gender difference, with blogging equally prevalent among females and males. In comparison, White and Winn (2009) found a slight difference, with male bloggers representing 57% of all U.S.-based bloggers. While most bloggers are white (66%), race is more diverse among bloggers than in comparison to all Internet users (Lenhart & Fox, 2006).

Bloggers are more “educated, affluent, and influential” than the general Internet population (White & Winn, 2009).<sup>39</sup> Three in four are college graduates, and over four out of ten have attended graduate school (White & Winn, 2009). Lenhart and Fox (2006) found a lower incidence of college graduates among respondents to their study: 37% had a college degree compared to 27% of all American adults. Bloggers tend to work in knowledge-based professions (38%); for American adults in general, knowledge-based professional workers account for only 13% of the work-force (Lenhart & Fox, 2006).

Blog to blogger is not a simple one-to one ratio. While about 50% of bloggers publish to one blog, about 20% publish to two, and about 10% publish to three or more (Lenhart & Fox, 2009). Further, blogs may be singularly- or collaboratively-authored. For bloggers with two or more blogs, about 30% report that their primary blog is collaboratively-authored (Lenhart & Fox, 2006). White and Winn (2009) report that experienced bloggers – 51% of their respondents – contribute to an average of four blogs. Multiple instances and collaborative authorship raises critical considerations since blogger behaviors, practices, and

---

<sup>39</sup> As an example of “influence,” consider RSS. Bloggers have been described as “pioneers of RSS feeds, streamlining users’ experience by allowing them to interact with fresh content in one central clearinghouse instead of having to visit blog after blog” (Lenhart & Fox, 2006, p. 21). It is noted, however, that RSS feeds have been reported as a less popular element when compared to other blog elements; only 18% of bloggers utilize an RSS service for disseminating their content (Lenhart & Fox, 2006).

preferences, as well as blog and audience attributes, may be different depending on which instance is under consideration.

#### 2.4.1.2. Style of Blogging

Herring et al. (2005), in citing Blood, identify three primary types of blogs: the filter blog, the personal blog, and the knowledge blog. White and Winn (2009) also identify three primary types of blogs, including the personal blog as well as the professional and the corporate blog. While about eight out of ten bloggers responding to White and Winn's survey identify themselves as personal bloggers, the authors make clear that these categorizations are not mutually exclusive. Nearly half of respondents identify themselves as professional bloggers; for bloggers in this category, 59% also identify themselves as personal bloggers (White & Winn, 2009).

#### 2.4.1.3. Topical Treatment

Blog topical treatments vary widely. Lenhart and Fox (2006) report that the majority of blogs are personal journals, with nearly four in ten respondents citing their "life and experiences" as the main topic of their respective blogs. About one-third concentrate on a single topic, while two-thirds blog across a variety of topics. White and Winn (2009) found similar results. Approximately 75% of bloggers in their survey cover three or more topical areas, with an average of about five topical themes per blog (White & Winn, 2009).

#### 2.4.1.4. Identity and Access

Bloggers self-identify. As such, they may choose to blog under their real names or use a pseudonym. Per Lenhart and Fox (2006), over half (55%) blog under a pseudonym, with 46% reporting that s/he identify their own names on their blogs. White and Winn (2009) found that two-thirds of bloggers make their identity known, while about one-third think it

important to conceal their identity. Reasons for employing a pseudonym, as reported by respondents to Lenhart and Fox's (2006) survey, are to protect personal privacy and for establishing barriers between one's life in the blogosphere and life elsewhere.

Though two-thirds of White and Winn's (2009) respondents "openly exposed their identities on their blogs," it may not necessarily reflect that real names were provided. Further, use of pseudonyms does not necessarily mean that bloggers are anonymous. Identity could also be assumed through other identifiers published to a pseudonymous blogger's blog (i.e., occupation, affiliations, etc.). Additionally, while most blogs are characterized as open, bloggers may limit access to all or parts of their blog. These attributes complicate preservation action, including issues of authorship, credibility, permission, and authentication.

#### 2.4.1.5. Audience

An understanding of a blog's audience provides some information on content consumers. It may have potential implications for selection decisions by demonstrating audience types which may be interested in blogs' long-term persistence and access. Lenhart and Fox (2006) report that about half of all bloggers predominantly blog for themselves rather than others, while about three out of ten blog with an audience in mind.

Lenhart and Fox (2006) describe two types of measures, off-blog and on-blog, for evaluating readership. Examples of on-blog measures for deriving estimation of blog consumption are blog traffic logs, comments and tagboards, and blogrolls.<sup>40</sup> Off-blog measures suggestive of readership are informal, including conversations via external communications (e.g., email, chat, IM). In addition to these, White and Winn (2009)

---

<sup>40</sup> Per Lenhart and Fox (2009), about four out of ten bloggers make use of a blogroll or similar linking device. Further, about 40% of bloggers list under ten blogs to their respective blog rolls, with almost 30% listing between 10 and 49, and nearly 20% listing 50 or more.

reference other “measures” for determining bloggers’ personal perceptions of success, including their respective Technorati Authority number.<sup>41</sup>

In describing the use of blogrolls for assessing readership, Lenhart and Fox (2009) point out that, “readership may be suggested – though not necessarily guaranteed” (p. 17). This idea of suggestibility rather than accuracy could be applied across the measures described immediately above since there is variety in the types of blog elements adopted by bloggers and, hence, conclusive assessments of readership across the blogosphere in general lack rigor. Lenhart and Fox (2006) found that almost half of bloggers do not know the traffic statistics for their respective blogs. For those that know, a minority (13%) report 100 or more hits a days. This may reflect a low percent of blogs that have a somewhat identifiable, and somewhat substantial, designated community of consumers.

#### 2.4.1.6. Activity

Lenhart and Fox (2006) identify two primary types of bloggers. “Hobbyists” (84%) view blogging as a peripheral, “extracurricular” activity and typically spend less than five hours per week on their blogs. “Dedicated bloggers” view blogging as an important component of their lives and invest more time into what is viewed as a public endeavor. Ten percent of respondents to Lenhart and Fox’s (2006) survey spend more than ten hours per week on their blog, with the majority (59%) spending only one-to-two hours per week. Respondents to White and Winn’s (2009) survey invested more time in their blogging activities. About 25% spend ten or more hours per week blogging, while 19% blog between one-to-three hours per week.

Respondents to White and Winn’s (2009) study have been actively blogging, on average, for three years. Lenhart and Fox (2006) found that most of their respondents (81%)

---

<sup>41</sup>Technorati: <http://technorati.com/what-is-technorati-authority>

have been blogging for three years or less. Blogging is perceived as a persistent and continuing activity, with 82% anticipating that s/he will continue blogging one year from the time the question was posed, particularly among bloggers who identify writing for an audience as opposed to themselves (Lenhart & Fox, 2006). However, it should also be noted that this expressed sentiment for persistence may not necessarily be reflective of activity. Lenhart and Fox's (2006) survey, characterized as a call-back since using samples from earlier studies, resulted in a 71% responses rate, possibly reflecting a change in blog practices among earlier participants – that is, bloggers responding to the earlier surveys were no longer blogging. Of course, it can also mean that previous participants simply no longer chose to participate in the study.

#### 2.4.1.7. Stability

Blogs are updated or added to iteratively. For most bloggers, as reported by Lenhart and Fox (2006), adding new content is not a regularly scheduled activity (22%), but something done when inspiration strikes (70%). Though adding new content may occur irregularly, for most bloggers, it is a weekly activity: 13% add new content at least once a day, and about 40% add content at least once per week. Bloggers may choose among several techniques for composing and publishing to his/her respective blog, including a blog's internal editor feature, or methods external to the blog, such as word processing programs, email composers, or desk-top plug-ins.

Bloggers may easily update, modify and alter their blogs, both in terms of the look, or layout of the blog, as well as the content published to the blog. Nearly four out of ten bloggers post corrections to their blogs often or sometimes (Lenhart & Fox, 2006). Hank et al. (2007) found that nearly all respondents to their survey (96%) report editing posts after



initial publication to their respective blogs. These changes are typically of a presentation nature. For example, to correct spelling and grammatical errors, as was reported, in frequency, as often by 49% of the study's respondents. Such activity supports the notion of the ephemeral, changing nature of blogs, not only in contributing new content, but in the alteration and modification of existing content.

Hank et al. (2007) found that nearly four out of ten bloggers (39%) responding to their survey reported deleting a post after publication. Reasons for deletion are typically more content-oriented, as opposed to presentation oriented. Just as we speak of the ephemeral nature of blogs, this behavior is demonstrative of the ephemeral nature of human beings, and the capacity to change opinions and thoughts. About 26% percent of those deleting posts report doing so because they no longer held the opinion expressed or no longer believed the information correct (Hank et al., 2007).

The incidence of blog deletion provides additional context. Hank et al. (2007) report that 20% of respondents report intentional deletion of an entire blog. Reasons given include: 1) migration to another blog platform; 2) loss of interest in subject matter; 3) too time demanding; 4) privacy concerns; and 5) too much spam. The ease in which content may be modified or deleted complicates decisions for blog preservation. Preservation action is not intended to be ephemeral but, in borrowing from the language of the OAIS Reference Model (CCSDS, 2002), intended for, "a period into the indefinite future."

Not all blog deletion is the result of personal preference. Hank et al. (2007) report that 3% of respondents to their survey have had a blog unintentionally deleted.<sup>42</sup> The

---

<sup>42</sup> In a well-publicized example, the Google Blog was "accidentally deleted" and then recovered once the error was detected. For more information, see a post at the Official Google Blog, "And We're Back" (March 27, 2006), available at: <http://googleblog.blogspot.com/2006/03/and-wereback.html>.

interdependent relationship among blog co-producers may come into play when considering unintended blog deletion. For example, deletion may be the result of blog service providers discontinuing services or enforcing conditions of their terms of service agreements, such as those stipulating that blogs in violation of the terms be deleted.

#### 2.4.1.8. Publishing Conventions

Lenhart and Fox (2006) found that over half (56%) of bloggers responding to their survey tend to verify factual information contained in their posts. About 57% provide links to original source materials. Almost eight out of ten bloggers (77%) publish works of their own creation to their blogs, while nearly half (44%) remix and republish existing, external online content. However, Lenhart and Fox (2006) found that few of their respondents get permission to post copyrighted material. Only 20% report getting permission often or sometimes, while 50% never get permission.

#### 2.4.2. Blog Characteristics

Blogs support a range of objects and formats, including text, images, audio and video.<sup>43</sup> Text is the most frequently published content class, followed by photos (Lenhart & Fox, 2006). Posts are typically displayed in reverse chronological order. Content may be original to the blog or represent work originating elsewhere, imported, embedded or accessed via hyperlinks. Blogs support two-way communication channels through commenting system features, and also support authentication and access controls. Such features enable a variety of interaction scenarios, dependent on authorship (i.e., single or collaborative) and limitations imposed on audience (i.e., password protection).

---

<sup>43</sup> While blogs may support the display of a variety of content classes, blogs may also be limited to or focused on a specific representation of content, contributing to categorization by format. For examples, the Vlog, of videolog, is a blog predominantly featuring video files (Lenhart & Fox, 2006).

Blogs support a variety of user interface elements. White and Winn (2009) identified 13 common blog tools for facilitating content dissemination and interaction, reporting that bloggers use, on average, seven of these tools. The most common tools, used by over 80% of respondents, are: 1) commenting systems; 2) archiving posts by date or category; and 3) built-in syndication. Interactivity is a hallmark of blogs. Lenhart and Fox (2006) report that nine in ten bloggers in their survey allow readers to post comment. Other common tools, as reported by White and Winn (2009), are: search utilities; widgets; assigning posts to more than one category; track-backing; video hosting; photo hosting; collaborative authorship; Twitter or other real-time update applications;<sup>44</sup> photoblog capabilities, and mobile updating tools. (White & Winn, 2009). Additional tools that facilitate interactivity are: a) blogrolls, described as “lists of links to other blogs generally found in a blog’s sidebar,” and also referred to as “Friends” or Subscriptions;” and b) tagboards, a “general space for commenting on the entire blog or website” (Lenhart & Fox, 2006, p. 17).

Most blogs are published via a blog publishing and hosting service (Lenhart & Fox, 2006; Hank et al., 2007). As such, these service providers have terms of use which impact the stewardship of the blog in regard to control of and access to the content published by bloggers, as well as context, representing the tools and features described in the preceding paragraph. Without these essential features, it may undermine the informational and referential value of blogs.

---

<sup>44</sup> White and Winn (2009) report that about 31% of bloggers responding to their survey use Twitter or a similar real-time updating application. These microblogging services are beyond the scope of this section. However, it is noted that many in the blogging community view these tools as valuable additions to the blogosphere. Michael Arrington, founder of TechCrunch, writes: “Microblogging platforms like Twitter and Friendfeed are the fast food equivalents of the blogging world, and continue to gain popularity because they let people update multiple times per day with 140 characters or less ...” (White & Winn, 2009).

Both where the blog is hosted and the application on which the blog is built have implications for digital preservation. Resulting loss of personal online content is usually due to issues of service discontinuation or conflicts with service provider policies rather than technical failure. Marshall, McCown, and Nelson (2007) found that people rely heavily on service providers, often taking no additional actions toward the persistence of their personal digital assets. Preservation action necessitates careful consideration of these co-producer dependencies, and service agreements in particular. Such considerations are compounded by the number of service providers represented in the blogosphere, with examples provided earlier in this chapter (See section 2.2. for a listing of select blog publishing and hosting services).

#### 2.4.3. Approaches

Current approaches to blog preservation may be categorized as either personal or programmatic. In terms of responsibility, Entlich writes, “Thus, for the most part, it is up to individual bloggers to maintain copies of their creation or else rely on the Internet Archive to do the job (para. 27).” Marshall, Bly, and Brun-Cottan (2006) used field methods to explore five basic personal Web archiving strategies: 1) using system backups; 2) moving whole files to another computer; 3) replicating specific files on removable media; 4) using email attachments; and 5) retaining old computers. Other approaches, as found by Hank et al. (2007), include subscription to Web-based archiving services and, while it may seem an oxymoron, printing hard copies. Meyer (2005) commented on adopting the latter approach in a blog post, and gave two reasons for this print approach: 1) audience preference for paper copy, and 2) personal choice.

For the most part, programmatic blog preservation programs take either a deposit or harvesting approach (e.g., O’Sullivan, 2005; Paulus, 2006; Jyn & Leng, 2005). For example, the Internet Archive Wayback Machine takes a harvesting approach.<sup>45</sup> Additionally, the Internet Archive introduced Archive-It (<http://www.archive-it.org>), a subscription service, in 2005. Subscribers to the service select content for harvest to create their own Web archive collections. The Library of Congress Law Library uses Archive-It in their legal blog preservation program.<sup>46</sup> A current blog archiving project at the University of London Computer Centre and the British Library Digital Preservation Department takes a slightly different approach. Rather than strictly a Web-crawler or deposit approach, the ArchivePress project is facilitating harvest of blogs, including comments, embedded content and metadata, through RSS feeds and blog APIs. Further, WordPress applications are used to stored and maintain content with the intent of capturing the look and feel of the blog in terms of structure and relationships.<sup>47</sup>

Another programmatic approach is blog deposit to IRs or other digital repositories or archives. The National Library of Australia’s Pandora project collects blogs as part of their digital preservation program, accepting deposits of blogs that support academic publications collected by the Library (Entlich, 2004).<sup>48</sup> In regard to IRs, Markey et al. (2007) included blogs as a content type for planned or actual IR deposits in their extensive census of IR deployment in the US. Few IRs reported accepting blogs for deposit; only eight operational IRs and one planned pilot instance identified blogs as a deposited content type. Other IR

---

<sup>45</sup> Internet Archive Wayback Machine: <http://www.archive.org/web/web.php>

<sup>46</sup> Library of Congress Legal Blawg Archive: <http://www.loc.gov/law/find/web-archive/legal-blawgs.php>

<sup>47</sup> ArchivePress: <http://archivepress.ulcc.ac.uk/>

<sup>48</sup> Pandora: See <http://www.pandora.nla.gov.au>

deployment surveys did not include blogs as a researcher-supplied category in consideration of content types accepted or planned for deposit (Bailey, 2006; Lynch & Lippincott, 2005; Kennan & Kingsley, 2009; Shearer, 2006). However, the use of an “other” category was used, with the option to report content types not included in the researcher-supplied listing. Only Lynch and Lippincott (2005) specifically mention a blog as an “other” type, supplied by one just one respondent to their survey.

## 2.5. Conclusion

In the preface to her book on the system of scholarly communication in a networked, digital landscape, Borgman (2007) signs off with, “Let the conversation begin” (p. xix). The literature presented in this chapter is intended to contribute to this conversation, with the overall intent of contextualizing a specific type of content within the system of scholarly communication and implications for preservation. Presented was a framework of functions in which to consider blogs as units and channels of scholarly communication through scholars’ actor-roles as author and reader. Further, the considerations presented illustrate the range of issues that impact the preservability of scholars’ blogs and the critical decisions necessary for planning and implementing preservation action. Such decisions will be enhanced through an informed understanding of the perceptions, characteristics and behaviors of this particular class of bloggers and blogs.

The intent of this research, with the methodology described in the chapter immediately following, is to benefit personal data management and preservation activities, as well as planning activities and practices in a variety of information contexts, including archives, libraries and other information organizations that support services to identify, collect, and protect our collective memories represented in the blogosphere. Further the

emphasis on scholarly blogs and scholars who blog will contribute to improved understanding of the evolving system of scholarly communication, with preservation of scholars' digital assets, manifested in the blog, contributing to Borgman's scholarly communication function for access, preservation and curation, and Roosendaal et al.'s (2001) archive function.

## **CHAPTER 3**

### **METHODS**

This descriptive study investigates the attitudes and perceptions of scholars who blog in relation to scholarly communication and digital preservation, as well as behaviors and blog characteristics that influence preservation action. Specifically, the study is designed to inform the following research questions:

- 1) How do scholars who blog perceive their blog in relation to their cumulative scholarly record?
- 2) How do scholars who blog perceive their blog in relation to long-term stewardship and, subsequently, who, if anyone, do scholars perceive as responsible as well as capable for blog preservation?
- 3) How do blog characteristics and blogger publishing behaviors and preferences impact preservation?

Applying survey research and content analysis methods, this study takes a triangulated approach to support investigation of two units of analysis: scholars who blog in the areas of history, law, economics, physics, biology and chemistry, and the blogs of scholars in these areas.

Analysis is derived from three data sources: questionnaires, interviews, and publicly available blogs. Self-administered, Web-based questionnaires are selected to capture information only known or better known by the blogger, including preferences, perceptions, and behaviors. Blog analysis is selected to capture information more conveniently known to



the researcher. Additionally, it may be more accurately known. Dillman (2007) recommends questionnaires be designed that do not overtax respondents ability for recall, and support ease and accuracy in completion. Data on particular blog characteristics, such as word count of post, commenting history, and extent of contextualizing content, like number of blogs listed to a blogroll, is better treated through a content analysis than a survey approach. Interviews are selected to provide members of the sample additional opportunity to share more extensive insight on their attitudes and opinions on blogging, scholarly communication, and digital preservation than might be captured in a questionnaire format.

### 3.1. Population

Probability sampling techniques are characterized as more scientific, precise and cost-effective (e.g., Dattalo, 2008; Neuman, 2006; Powell & Connaway, 2004). However, nonprobability sampling is appropriate when probability sampling is not viable or the population is unknown (e.g., Babbie, 1990; Berg, 2004; Powell & Connaway, 2004; Wildemuth, 2009). Due to the nature of this inquiry and the units of analysis under investigation, the population is not simply unknown; it is unknowable. There is no viable strategy for comprehensively identifying all instances of scholars who blog or scholarly blogs in the blogosphere. Further, the population under study is somewhat ambiguous. What attributes define a scholar blogger, and what distinguishes a blog as scholarly?

In response to these population issues, purposive sampling is applied. Sample sources were carefully considered, including: 1) blog networks (e.g., Law Professor Blogs; George Mason University's History News Network Blogs; ScienceBlogs); 2) blog directories and listings (e.g., Academic Blog Portal; BlogScholar Academic Blogging Portal; Medievalist

Weblogs; Rhetorica: Professors Who Blog), including listings of award winning blogs;<sup>49</sup> and 3) blogs and bloggers identified and aggregated by this researcher from multiple sources, including blogrolls, references in the literature, alert services and blog indexing services.

Ultimately, the population for this study is drawn from a single directory source, the Academic Blog Portal (Portal).<sup>50</sup> This source was selected for several reasons. It is an extensive aggregation of scholar blogs, maintained and managed by a group of editors and made publicly available in a wiki format. It is well-organized, with blogs categorized by discipline within four broad domains: the Humanities, Sciences, Social Sciences, and Professions and Useful Arts. Descriptive information is provided for some of the blogs listed, including information on topical treatment, age of blog, and blogger's name, contact information and institutional affiliation. The blog listings are compiled by the editors as well as registered community contributors, with clear instructions for adding, deleting or modifying listings. Excerpted below are explicit eligibility criteria for blogs listed to the Portal:

...the blog has to be written by an academic. That is to say, the author should be either a member of a third level institution's faculty (i.e. community college, college, university, technical institute or whatever), or pursuing a doctoral degree, or employed by a third level institution to do academically relevant work (such as working as a university librarian)" (Academic Blog Portal, 2006, para. 6).

The editors do address some of the ambiguity in discerning what is and what is not a scholar blog, and provide a vetting procedure by which to nominate blogs for inclusion published by non-academics. The following excerpt describes this procedure:

---

<sup>49</sup> For example, see the Research Blogging Awards 2010, available at: <http://www.researchblogging.org/static/index/page/awards>

<sup>50</sup> Academic Blog Portal: [http://www.academicblogs.org/wiki/index.php/Main\\_Page](http://www.academicblogs.org/wiki/index.php/Main_Page)

Very rarely, blogs written by non-academics may qualify. If you think a blog by a non-academic qualifies on its merits, you should suggest it to one of the Senior Editors, with supporting evidence. The Senior Editor will then decide whether to nominate it. Otherwise it is liable to be deleted summarily. If you are a non-academic and you nominate your own blog for consideration, don't be offended if it isn't accepted - only very exceptional blogs will be included” (Academic Blog Portal, 2006, para. 6).

In consideration of the dynamic nature of the blogosphere, the Portal also addresses issues of currency and activity. Blogs with no new posts published within three months are subject to deletion from the Portal.

Following source selection, the population under investigation is further refined through cluster sampling. Selection of the clusters was based on: 1) disciplines with the most blogs within each of the four domains; and 2) similarity in number of blogs listed for each discipline. Due to fewer blogs across all disciplinary listings in the Sciences domain, when compared to the other domains, a composite cluster was constructed from the disciplines of biology, chemistry, and physics, abbreviated as BioChemPhys throughout the remainder of this chapter and subsequent chapters. The selected clusters for the remaining three domains are history, economics, and law. After removing duplicate blogs (n=5), 644 total blogs were identified at the source. Table 3 shows the total count of blogs identified in the population, organized by domain and respective cluster.

**Table 3. Population count of blogs by domain and cluster**

Domain	Cluster	Blogs Listed at Source	Duplicates	Total Blogs
Humanities	History	190	1	189
Social Sciences	Economics	192	0	192
Professions & Useful Arts	Law	120	1	119
Sciences	BioChemPhys	147	3	144
<b>All Domains</b>	<b>All Clusters</b>	<b>649</b>	<b>5</b>	<b>644</b>

*Note.* For total blogs within the Sciences cluster, BioChemPhys (N=144), sub-fields were represented as follows: Biology blogs, 39% (n=56); Chemistry blogs, 15% (n=21); and Physics blogs, 47% (n=67).

### 3.1.1. Sampling Frames

Since the study examines two units of analysis – blogs and bloggers – two sampling frames were constructed. The blog sampling frame is used to generate a sample for the blog analysis stage of this study. The blogger sampling frame, derived from the former, is used to identify participants for the study’s questionnaire and interview stages. The procedures for constructing the sampling frames are detailed immediately below, followed by a description of the resulting samples.

#### 3.1.1.1. Inclusion Criteria

As mentioned earlier, the Portal has explicit inclusion criteria for blogs listed to the directory. However, a cursory examination at the source evaluation stage revealed some non-compliant instances. Researcher-derived eligibility criteria were developed and applied to allow improved control over the sample and increased homogeneity among the units of analysis. These criteria were created in response to issues of blog access, language, currency, and blogger identity.<sup>51</sup> Further, while there are many hallmark characteristics of blogs, such as archiving systems and post time-stamping, these are not requirements for publishing a blog. These are, however, requirements for inclusion in this study. Hence, the inclusion criteria were also developed in consideration of the diversity of blog features in order to facilitate efficient and effective data collection for the blog analysis portion of this study.

The inclusion criteria and order of assessment are summarized below. The complete coding system for assessment is provided in Appendix A. Criteria one through six establish a blog’s eligibility, while criteria seven through nine establish the respective blog author’s

---

<sup>51</sup> As mentioned earlier, some descriptive information was provided at the source through internally linked “about” pages or provided next to the blog’s entry in the directory. While this supplemental information may be helpful in assessing blogs and, subsequently, bloggers’ eligibility for inclusion in the study, it was not considered sufficient for verification purposes for this study. Further, for the 644 blogs identified in the population (see Table 3), only 80 (12%) had links to internal scope notes.

eligibility. Collaboratively-authored blogs present particular challenges in regards to authorship. These blogs, abbreviated as co-blogs throughout, are distinguished as blogs published to by two or more primary bloggers. Issues related to co-blogs are addressed further on in this section.

Returning to the criteria, first, the blog must be publicly accessible. To assess availability, the URL posted at the source listing was accessed. For example, password-protected blogs are disqualified from the blog sample. Further, if a URL is a dud – for example, resulting in a 404 error message or “domain for sale” message – the blog is excluded. No further actions to identify alternative URLs are taken.

Second, the blog is published in English. To assess language, posts published to the first page of the blog were quickly reviewed. Blogs with posts published exclusively in another language are disqualified. Further, blogs with posts published in English and one or more other languages are also excluded.

Third, the blog is a knowledge or personal blog. As mentioned in the preceding chapter, Herring et al. (2005), in citing Blood, identify three primary types of blogs: the filter blog, the personal blog, and the knowledge blog. Blog types, per White and Winn (2009), are not necessarily exclusive; bloggers tend to identify their blogs as both personal and professional. Due to the potential ambiguity in characterizing blogs by type, careful consideration was applied at this stage of inclusion assessment. Those blogs found to be predominantly filter blogs, characterized as, “primarily containing observations and evaluations of external, typically public events” are excluded from the study (Herring et al., 2005, p. 147). Serfaty (2004) took a similar approach in compiling the sample for her study analyzing personal online diaries and blogs, excluding those that published, “mere lists of

annotated links to other websites” (p. 16). Blogs that serve primarily as Twitter aggregators or other news aggregators are also excluded, as are instances that, while powered by a blog software application, do not follow a conventional blog format, such as chronologically listed posts, and more closely resemble a general Website. Additionally, class blogs, used to facilitate a specific, named course, were excluded.

Fourth, the blog posts are time-stamped. To assess time-stamping, the most recent post was reviewed. If no time-stamping showing, at minimum, the date on which the post was published, then the respective blog was excluded. Time-stamping is necessary to assess currency of the blog and blogger publishing activity, contributing to the next criteria.

Fifth, the blog is actively published to, characterized as the most recent post published to the blog within 14 days from the date the blog was assessed for sampling.

Sixth, the blog is at least one year old. To assess age of blog, the blog’s archive was reviewed. Blogs that do not employ an “archives” feature, accessible from either the side bar or “archives” page, are disqualified. No other actions to assess age were taken (i.e., navigating through “older posts” link at bottom of the page).

Seventh, the blog has identifying information on authorship available from the first page of the blog, including: “about” information posted to header, title tag, or sidebar; internal link to “about” or similar page; or external link to “about” page, personal Website, or similar page. Blogs that do not contain any information on authorship found at these locations are disqualified.

Eighth, the blog’s primary author or authors (for co-blogs) are scholars.<sup>52</sup> Primary denotes that the blogger is a regular, current contributor to the blog. Authors may be

---

<sup>52</sup> Throughout the remainder of this chapter and subsequent chapters, collaboratively authored blogs are referred to as *co-blogs*; singularly-authored blogs, distinguished as blogs published by only one primary author, are

described in author-identification fields as blogger, author, editor, co-editor or contributor. Excluded are bloggers identified as guest editors, emeriti bloggers, or other categories that denote either a former or temporary relationship to the blog. This assessment of authorship is made from a review of “about” information published to the blog, or linked to the blog from an external source, as described in the preceding criterion. While the sub-criteria for assessing scholar status of single-bloggers and co-bloggers is the same, procedures to assess the criteria are slightly different.

For single-bloggers, a blogger is deemed a “scholar” if one of the following four sub-criteria is met:

- a) One or more of the following keywords or phrases are used to describe the blogger:  
Ph.D., Dr., Professor, Reader, Lecturer, Doctoral Student, or Doctoral Candidate;
- b) One or more of the following keywords or phrases are used to describe the blogger:  
Scholar, Academic, Researcher, Research Director, Fellow, and/or Other Role Identifier (i.e., biologist); **and** institutional affiliation is provided, either specifically by name or through a general description denoting a post-secondary education or research setting;
- c) A link to blogger’s curriculum vitae (CV), either full or partial, or other listing, either full or partial, of blogger’s publications is accessible, **with** at least one citation to a journal article published or co-published by the blogger, or;
- d) The blogger is described as a graduate student **and** explicit reference to area of study or pursuant degree is provided.

---

referred to as *single-blogs*. Likewise, authors of c-blogs are referred to throughout as *co-bloggers*, and single-blog authors as *single-bloggers*.

Those bloggers not satisfying sub-criteria a, b, c, or d are excluded from the sample, as are their respective blogs.

For co-blogs, determination of current authors precedes assessment of “scholar” criteria. Only co-bloggers publishing to their respective blogs within one month from the date of assessment (approximately 31 days) are considered for inclusion. Following assessment of posting activity, co-bloggers are then assessed for scholar status across the four sub-criteria, as described immediately above. Co-bloggers not satisfying sub-criteria a, b, c, or d are disqualified. Further, if none of a co-blog’s primary authors are deemed eligible, the blog is also disqualified.

Ninth, contact email addresses for qualifying single- and co-bloggers are publicly available from the first page of the blog or at one of these other locations: an internal link to “about” or similar page; external link (e.g., to personal Website, CV, or Blogger profile); or through a simple Google search when blogger’s full name and institutional affiliation made known in “about” information posted to blog. If no email contact information is available through any of these means, the respective blogger is disqualified, though the blog remains eligible for inclusion in the blog sample.

### 3.1.2. Blog Sample

The 644 blogs listed at the source were assessed for eligibility between July 30 and August 4, 2010. Table 4 shows the count and percentage of blogs, by cluster, meeting the inclusion criteria (criteria 1-8).<sup>53</sup> The table is arranged in the order by which the criteria were assessed. Overall, 188 blogs qualified for the study, representing 29% of blogs identified at the source listing (N=644). While this study presents findings for only those blogs eligible

---

<sup>53</sup> Criteria 8 is a cross-over criterion, as it is used in construction of the blog sample and blogger sample. Blogs with no blog authors qualifying as “scholars” are excluded from both samples.



for sampling, the disqualification of blogs (n=456) contributes to some insight into issues of persistence, authority, identity, and blog characteristics. A summary table on the blogs disqualified is provided in Appendix B.

**Table 4. Blog inclusion criteria assessment by cluster**

Criterion	History Freq (%)	Economics Freq (%)	Law Freq (%)	BioChemPhys Freq (%)
<b>One: Blog is publicly available</b>	168 (90%)	163 (85%)	113 (95%)	126 (88%)
<b>Two: Blog is published in English</b>	159 (84%)	151 (79%)	111 (93%)	123 (85%)
<b>Three: Blog is a knowledge or personal blog</b>	146 (77%)	140 (73%)	93 (78%)	119 (83%)
<b>Four: Blog posts are time-stamped</b>	145 (77%)	140 (73%)	93 (78%)	118 (82%)
<b>Five: Blog is actively published to</b>	68 (36%)	83 (43%)	58 (49%)	62 (43%)
<b>Six: Blog is at least one year old</b>	58 (31%)	66 (34%)	53 (45%)	54 (38%)
<b>Seven: Blog contains personal identifiers in regard to authorship</b>	53 (28%)	59 (31%)	48 (40%)	48 (33%)
<b>Eight: Blog is authored by one or more bloggers meeting scholar parameters</b>	46 (24%)	51 (27%)	47 (40%)	44 (31%)

Table 5 shows the number of blogs comprising the blog sampling frame. Each cluster, as shown in the table, is stratified by authorship (single or collaborative).

**Table 5. Count of blogs in blog sampling frame stratified by authorship**

Clusters	Single-Blogs	Co-Blogs	Total Blogs
History	32	14	<b>46</b>
Economics	34	17	<b>51</b>
Law	22	25	<b>47</b>
BioChemPhys	37	7	<b>44</b>
<b>All Clusters</b>	<b>125</b>	<b>63</b>	<b>188</b>

*Note.* For blogs in the BioChemPhys cluster, disciplines were represented as follows: Single-blogs: biology 43% (n=16), chemistry 14% (n=5), and physics 43% (n=16); and Co-blogs: biology 0%, chemistry 29% (n=2), and physics 71% (n=5).

The blog sample is derived from random sampling at a 50% sampling ratio. The resulting sample, as shown in Table 6, reflects overall frequency of blogs by cluster and authorship. In drawing the sample, each cluster in the sampling frame was sorted alphabetically, and then, using the Research Randomizer (<http://www.randomizer.org>), a random number set was generated from which to select blogs for the sample.

**Table 6. Blog sample**

<b>Clusters</b>	<b>Single-Blogs Count</b>	<b>Co-blogs Count</b>	<b>Total Blogs Count</b>
History	16	7	<b>23</b>
Economics	17	9	<b>26</b>
Law	11	13	<b>24</b>
BioChemPhys	17	4	<b>21</b>
<b>All Clusters</b>	<b>61</b>	<b>33</b>	<b>94</b>

### 3.1.3. Blogger Sample

Bloggers contributing to the 188 blogs comprising the blog sampling frame were assessed for eligibility simultaneous to blog assessment, from July 30 to August 4, 2010. Blogger sampling presented two particular challenges when assessing eligibility: duplicate bloggers and, as mentioned in Section 3.1.1.1., issues related to co-authorship.

In regard to duplicates, bloggers may publish to two or more blogs listed to the blog sampling frame. In line with the questionnaire and interview data collection portions of this study, detailed later in this chapter, bloggers are invited to participate in the study based on a particular blog. Therefore, a protocol was established for dealing with duplicate bloggers to ensure sample members did not receive multiple invitations to participate if s/he publish to two or more blogs in the blog sampling frame. This protocol is bulleted below:

- 1) For a blogger contributing to two or more single-blogs listed to the same cluster, the blog listed earliest in the blog sampling frame was selected. For single-blogs across clusters, the blog listed to the cluster with the fewest single blogs was selected.
- 2) For a blogger contributing to both single- and co-blogs, the single-blog was selected.
- 3) For a blogger contributing to two or more co-blogs listed to the same cluster, the co-blog with the fewest qualifying co-bloggers was selected. If this condition was not applicable, then the blog listed earliest in the blog sampling frame was selected.

Across all co-blogs comprising the blog sampling frame (N=63), 583 bloggers were identified. After assessing bloggers in line with the criterion for active posting – evidenced by publishing to their respective blogs within the previous month from the date of assessment – 241 co-bloggers qualified for further assessment in regard to criteria 8 and 9, scholar status and contact information, respectively. See Table 7 for assessment of inclusion criteria for co-bloggers identified in the blog sampling frame. Table 8 shows the assessment of single-bloggers. For single-bloggers, no duplicate bloggers were identified. Additionally, there was no need to further consider criterion 8 (“blogger meets scholar parameters”) as this was completed in construction of the blog sampling frame.

**Table 7. Co-blogger inclusion criteria assessment by cluster**

<b>Criteria</b>	<b>History (N=151) Freq (%)</b>	<b>Economics (N=155) Freq (%)</b>	<b>Law (N=228) Freq (%)</b>	<b>BioChemPhys (N=49) Freq (%)</b>
<b>Eight (Special Condition): Blogger published within previous month</b>	43 (29%)	65 (42%)	114 (50%)	19 (39%)
<b>Eight: Blogger meets scholar parameters</b>	31 (21%)	58 (37%)	107 (47%)	16 (33%)
<b>Nine: Blogger contact information available</b>	27 (18%)	56 (36%)	102 (45%)	15 (31%)
<b>Protocol: Revised count and percentage after removal of duplicate listings</b>	23 (15%)	53 (34%)	99 (43%)	15 (31%)

*Note.* Count of duplicate co-bloggers by cluster: History (4); Economics (3); Law (3); and BioChemPhys (0).

**Table 8. Single-blogger inclusion criteria assessment by cluster**

<b>Criterion</b>	<b>History (N=32) Freq (%)</b>	<b>Economics (N=34) Freq (%)</b>	<b>Law (N=22) Freq (%)</b>	<b>BioChemPhys (N=37) Freq (%)</b>
<b>Nine: Blogger contact information available</b>	27 (84%)	32 (94%)	21 (96%)	28 (76%)

Ultimately, for single-bloggers across all clusters, 108 qualified for the study, representing 86% of all single-bloggers identified (N=125). For co-bloggers, 190 qualified for the study, representing 33% of all bloggers identified (N=583), and 79% publishing to the

blog within 31 days of the assessment (n=241). Table 9 shows the counts of bloggers, by authorship and cluster, comprising the blogger sample.

**Table 9. Blogger sample**

Clusters	Single-Bloggers Count	Co-Bloggers Count	Total Bloggers Count
History	27	23	50
Economics	32	53	85
Law	21	99	120
BioChemPhys	28	15	43
All Clusters	108	190	298

### 3.2. Questionnaires

#### 3.2.1. Instrument Design

Due to differences in authorship, two questionnaires were designed and administered: questionnaire I for single-bloggers, referred to throughout as QI, and questionnaire II, or QII, for co-bloggers. Due to differences in authorship, wording of question stems and response categories is slightly different. Additionally, QII has four additional questions concerning issues of co-authorship. Otherwise, QI and QII are nearly identical.

QI and QII are comprised of primarily closed-ended questions. Both were constructed using Qualtrics (<http://www.qualtrics.com/>), a Web-based survey application made available by the Odum Institute for Research in Social Science.<sup>54</sup> The questionnaires, in part, were derived from the instrumentation for other survey research on blogs and bloggers, and scholars and scholarly communication, including Lenhart and Fox (2006); Hank et al. (2007); Sheble et al. (2007); Morton and Price (1999); Olsen, Berlin, Olsen, McLean, and Sussman (2009); Rainie (2005); and White and Winn (2009). The design was further informed by recommendations from Dillman, Smyth, and Christian (2009), Czaja and Blair (2005), and Punch (2003).

---

<sup>54</sup> Web Survey Data Collection service at The Odum Institute for Research in the Social Science: [http://www.irss.unc.edu/odum/jsp/content\\_node.jsp?nodeid=362](http://www.irss.unc.edu/odum/jsp/content_node.jsp?nodeid=362)

Depending on respondents' answers, single-bloggers were presented between 41 and 58 questions in QI, and co-bloggers were presented between 41 and 62 questions in QII. The questionnaires were organized into nine sections: 1) background; 2) blogging and scholarly communication; 3) blogging activity, identity, and audience; 4) blog publishing behaviors and preferences; 5) blog revision history; 6) blog preservation behavior; 7) blog preservation preference; 8) other blogging activities; and 9) demographics. See Appendix C for QI and Appendix D for QII. Included in these appendices are the welcome screen, incorporating the consent script. Due to the extensive skip logic built-in to QI and QII, questions were not numbered. Rather, a progress bar, visible in the footer on each screen, alerted respondents to the percentage of the questionnaire completed.

Bloggers may publish to more than one blog. A major consideration in both questionnaire design and administration was the provision of mechanisms by which bloggers would respond based on the particular instance identified during sample frame construction. This instruction was made explicit in the recruitment materials inviting bloggers to participate as well as re-emphasized in directions provided in QI and QII. Bloggers were also assigned a unique, four-digit personal identification number (PIN). The PIN was required to proceed to the first question of both QI and QII.<sup>55</sup> This is elaborated on in the recruitment subsection to follow, as it also served to facilitate reminder invitations. Additionally, in regard to co-authorship, specific instruction was given in QII to respond based on respondents' own perceptions, preferences and experiences rather than respondents' impressions on how her/his co-bloggers might respond.

---

<sup>55</sup> Providing a PIN was the only item in QI and QII that required a response. For all other questions, per the instructions given in the Welcome screen, respondents were, "free to answer or not answer any question."

### 3.2.2. Pre-Testing

Prior to administration, QI and QII were pre-tested, following recommendations from Dillman (2007), Czaja and Blair (2005), Fowler (2002), Powell and Connaway (2004), and Punch (2003). Fifteen people were identified and invited to take part in the pre-test.<sup>56</sup> Due to the nature of the inquiry, the testers represented several groups: single-bloggers, co-bloggers, and practitioners and researchers in scholarly communication, scholarly publishing, and/or digital preservation. These groupings were not necessarily mutually exclusive. Several bloggers invited to participate in the pre-test were also experienced researchers in digital preservation and/or scholarly communication. Additionally, while the questionnaires are limited to bloggers who publish in English, based on the sampling frame inclusion criteria, there was no exclusion criterion related to bloggers' geographic location. The pre-test group also included a UK-based tester. This was intentional in order to ensure clarity in terminology on both sides of the proverbial pond.

The pre-test reflected the conditions under which QI and QII would be administered to study participants, with the exception of the inclusion of an introductory note which preceded the recruitment correspondence. See Appendix E for pre-test email correspondence. This introductory section included specific considerations for the testers when completing and evaluating the questionnaires, and instructions for providing feedback. Testers were provided with a PIN and the questionnaires were administered via Qualtrics. Eight testers were invited to complete QI and seven were invited to complete QII. The pre-test was

---

<sup>56</sup> In addition to the pre-test, three other activities are worth noting in regard to the testing of the questionnaires. First, the questionnaires were drafted, vetted and revised over an extended period among members of my dissertation committee. Second, results and feedback from an earlier survey on which I served as PI, *Blogger Perceptions on Digital Preservation*, were also reviewed (See Hank et al., 2007; Hank, Choemprayong, & Sheble, 2007; and Sheble et al., 2007). And third, simultaneous to the pre-test period, myself and another colleague reviewed QI and QII in Qualtrics to test all possible skip logic patterns.

administered over a one-week period. Invitations to the pre-testers were sent on July 28, 2010. Ultimately, ten testers participated. Seven completed QI and three completed QII, as well as provided additional feedback by email. Completed questionnaires and emailed feedback were reviewed. Overall, feedback was positive. A few minor revisions were made to re-work wording for enhanced clarity in terminology and correct for minor typographical errors in both the questionnaires and recruitment materials.<sup>57</sup>

### 3.2.3. Administration

Sample members were invited to participate through email, with all communications sent individually rather than to groups of bloggers. All contacts with the sample were personalized. Each invitation began with a personalized salutation based on contact information identified at the sampling stage. For subjects self-identifying as professors, the email salutation was Dear Professor [Surname]. For those not identifying as a Professor, but as a Ph.D., then the salutation was Dear Dr. [Surname]. For those who qualified as scholars, but did not identify as either a professor or a Ph.D., then the salutation was Dear Mr./Ms. [Surname]. For those blogging with a pseudonym, even if real name also identifiable, pseudonym was used. This decision was made to make my familiarity with the blog clear to the blogger.

Further, sample members' blog and blog URL were provided, in addition to a unique PIN to be used when accessing the survey. The decision to require a PIN for access was made for several reasons: 1) to reinforce instruction for bloggers' to reply based on the specific instance identified in the emailed invitation; 2) to ensure that only sample members

---

<sup>57</sup> The two questions which underwent the most extensive revision following the pre-test – extensive characterized as the addition of a new response category – were a question on tenure status and a question on bloggers' Twitter activities. Additionally, concerns about the time to complete the questionnaires were also raised, with concern that study participants would drop out. The pre-testers completed the questionnaires in between 18 and 36 minutes.

completed the questionnaire, preventing a snowball effect if bloggers were to share the QI or QII URL with others; 3) to assist in tracking QI and QII completion; and 4) to facilitate sending reminder invitations to non-respondents.

No tangible inducements for participation were provided in any recruitment materials. However, it should be noted that a final report will be made available to those responding to QI or QII and indicating interest.<sup>58</sup>

The questionnaires were available to participants for three weeks. Data collection commenced August 5, 2010, and closed August 28, 2010.<sup>59</sup> Initial invitations to participate were emailed to all sample members – 108 single-bloggers and 190 co-bloggers – between 7:00 am and 10:30 am (EDT) on Thursday, August 5, 2010. Due to mode of delivery, the timing during which the initial invite was sent was a deliberate strategy. As Dillman et al. (2010) note, “there is some indication that e-mail invitations are most successful if they are delivered to recipients’ inboxes early in the morning” (p. 280). For efficiency in administration, all bloggers were sent invites in a dedicated block of time. However, due to differences in geographic locations, time sent did not necessarily reflect time received by invitees.

The first reminder invitation, sent approximately one week after the initial invite, was emailed to 67 single-bloggers and 139 co-bloggers between 10:30 am and 1:00 pm (EDT) on Thursday, August 12, 2010. The second and final reminder was emailed to 47 single-

---

<sup>58</sup> The final question in QI and QII asked if respondents would be interested in receiving a final report from the study, to be made available in January 2011. For all respondents returning a complete questionnaire (N=153), 135 (88%) indicated interest in receiving the report.

<sup>59</sup> There was an error in the correspondence which was not caught during testing. Bloggers were informed that the questionnaire, “closed at midnight (EDT) on 27 August 2010.” More accurately, this should have been restated as 11:59 pm (EDT) on 27 August 2010. Due to any errors in interpretation this may have caused, the survey remained open until 11:59 pm (EDT) on 28 August 2010. It should be noted: only one questionnaire was returned on August 27, 2010; zero questionnaires were returned on August 28, 2010.



bloggers between 1:00 and 2:00 pm (EDT), and 103 co-bloggers between 9:30 and 11:00 am (EDT) on August 23, 2010. See Appendix F for QI and QII recruitment correspondence.

#### 3.2.4. Completed Sample

Across all invited participants (N=298), 159 total questionnaires were returned. Only complete questionnaires, defined as 80% or more of all applicable questions completed, were accepted for analysis. Partial and break-off questionnaires (n=6) were excluded from analysis. Overall, the completed sample is comprised of 153 respondents, defined by Dillman et al. (2009), citing the AAPOR, as, “all of the units that complete the questionnaire” (p. 43). The completed sample (N=153) represents all questionnaires accepted for analysis. Table 10 shows the counts of bloggers, by authorship and cluster, comprising the blogger sample.

**Table 10. Completed sample**

<b>Clusters</b>	<b>Single-Bloggers Count</b>	<b>Co-Bloggers Count</b>	<b>Total Bloggers Count</b>
History	16	15	<b>31</b>
Economics	18	21	<b>39</b>
Law	15	38	<b>53</b>
BioChemPhys	18	12	<b>30</b>
<b>All Clusters</b>	<b>67</b>	<b>86</b>	<b>153</b>

QI and QII outcome rates, including measures of response, cooperation, refusal, and contact, are detailed in Chapter 4, Section 4.1.1. (Questionnaire: Outcome Rates and Completed Sample).

#### 3.2.5. Data Analysis

Questionnaire data was exported from Qualtrics to Excel to assess completeness, clean-up data, normalize responses from open-ended response categories, and remove any personal identifiers. Open-ended responses were analyzed in Excel. Quantitative results were preliminary analyzed in Excel and then exported to IBM SPSS Statistics 19 for final analysis. Additionally, select results were also re-verified in Excel. The overall goal of this study is to

describe attributes, perceptions, preferences and practices of a particular community of blogging practice, identified via purposive sampling. As such, data analysis was limited to descriptive measures. Quantitative data were analyzed using univariate and bivariate descriptive statistics. The results, as reported in Chapter Four, primarily report frequency distributions.<sup>60</sup>

In assessing and preparing data for analysis, three specific actions merit noting. First, PINs entered into the survey were cross-referenced with the sample. Second, one respondent submitted a partial questionnaire followed by a complete questionnaire on a subsequent day. Since the PIN allowed for identification of one sample member but for two submissions, the partial questionnaire was deleted and not considered in the calculation of any outcome rates or for any further analysis. Third, a completed questionnaire was revised at the request of a respondent. Changes were made to a three-question segment in regard to blog deletion, communicated by respondent via a telephone conversation.

### 3.3. Interviews

#### 3.3.1. Interview Schedule and Protocol

The semi-structured interview schedule was designed to clarify initial findings from the questionnaire phase of the study and to further explore respondents' perceptions on blogging, scholarly communication and digital preservation. Additionally, the schedule was designed to reflect characteristics specific to particular interviewees in regard to authorship status, preservation preference and practices, and network affiliations. The schedule contains 14 questions. Depending on characteristics identified via sampling and preliminary review of questionnaire data, interviewees were asked between 11 and 14 questions. See Appendix G

---

<sup>60</sup> Other descriptive statistical measures examined include central tendency measures, measures of variability and correlation statistics. These results are considered preliminary at this stage, and will be further reviewed and revised, as needed, in preparation for future publications post-dissertation.

for the interview script, including informed consent statement and schedule. Additionally, an interview debriefing sheet was devised. A debriefing sheet was completed after each interview. It provided supporting documentation of each interview and aided in identification of problems and issues, if any, impacting data collection. See Appendix H for the interview debriefing sheet.

All interviews were conducted by telephone via a landline and digitally recorded using a Sony ICD-UX71 Stereo Recorder and a Radio Shack Multi-Phone Recording Control. Specific steps taken during and post-interview, informed by recommendations from Berg (2004) and Czaja and Blair (2005), include:

- 1) Extensive note-taking during interview, capturing key points and other areas of interest;
- 2) Completion of interview debriefing sheet;
- 3) Listening to playback of recording immediately after interview, resulting in additional note-taking, and;
- 4) Listening to recording again, three to five days post-interview, followed by a review of all notes.

### 3.3.2. Pre-Testing

Prior to administration, the interview schedule was pre-tested. This also allowed testing of the digital recording devices and telephone connection. Interviews with two testers were conducted under the same conditions in which interviews were held, including completion of the interview debriefing sheet. Following the test interviews, feedback was solicited from the testers and the taped interview was played to assess sound quality. The testing resulted in minor revisions to improve wording of select questions in the interview schedule. No changes to equipment or other steps in the protocol were made.<sup>61</sup>

---

<sup>61</sup> In addition to the two pre-test interviews, it should be noted that the schedule was drafted, vetted and revised over an extended period among members of my dissertation committee.

### 3.3.3. Administration

Interview participants were identified from respondents' returning completed questionnaires. At the end of both QI and QII, respondents were asked to indicate whether s/he were interested in taking part in a follow-up phone interview. Invitations to participate in an interview were sent to 51 bloggers (24 single- and 27 co-bloggers) between August 13 and August 26, 2010. Bloggers were asked to respond to the email, if interested in continued participation, with preferred dates and times for the interview, as well as a preferred contact number. No invitation reminder emails were sent to the invited respondents. For those responding affirmatively, a confirmation email was sent within one to two days to confirm interview date, time, and contact information. A reminder email was sent to all interviewees one day before the scheduled interview.

Additionally, since interview recruitment occurred simultaneous to questionnaire data collection, emails were sent to all respondents indicating interest after the initial invitations were sent. The intent of this correspondence was to thank respondents for their interest and to inform them that interviews were no longer being scheduled. See Appendix I for all correspondence with respondents, including invitation, confirmation, reminder, and decline emails.

In the original research design, 15 to 20 interviews were planned. Due to interest among respondents and diversity in discipline, authorship, and network affiliations, 24 interviews were ultimately scheduled and completed.

All interviews took place between August 18 and September 3, 2010. While interviews were intended to last between 15 to 25 minutes, 12 of the interviews ran longer. Interviewees were told when the 25 minute time limit was approaching, and instructed s/he

could end the interview or continue past the allotted time. All interviewees expressed interest in continuing, and permission was given to extend the interview. See Table 11 for frequency of interview times. These times reflect the point from when the first question was asked until the interview was concluded. Taking into account the interaction in its totality, including obtaining informed consent and permission to record, the entire duration in which the interviewer was engaged in the telephone conversation would be extended by about one-to-two minutes.

**Table 11. Frequency distribution of interview times**

<b>Time in Minutes</b>	<b>Frequency (Percentage)</b>
<b>≤ 14</b>	1 (42%)
<b>15 to 19</b>	3 (13%)
<b>20 to 24</b>	8 (33%)
<b>25 to 29</b>	4 (17%)
<b>30 to 34</b>	2 (8%)
<b>35 to 39</b>	4 (17%)
<b>≥ 40</b>	2 (8%)

All interviewees gave permission for digital recordings to be made, though only 23 interviews were recorded due to an error in setting up the recording equipment for one of the interviews. This interview was still included in analysis as notes were made throughout the interview and, in-line with interview protocol, extensive notes were taken immediately following the interview.

#### 3.3.4. Data Analysis

Interview notes were entered into Excel. These notes were supplemented by partial transcripts of the recordings, created in MSWord and then exported into Excel. Any personal identifying information captured during the note-taking and transcription was de-identified, in compliance with this study's Institutional Review Board approval.<sup>62</sup>

---

<sup>62</sup>It merits noting that two interviewees, in the course of their respective interviews, gave permission to be identified by name. Both the enthusiasm for the study and willingness to explicitly and publicly share their views is much appreciated. However, due to the study's design and in consultation with members of the

The data were organized around concepts established at the schedule design stage: perceptions on scholarly communication; the role of blogging as a unit and process of scholarly communication; perceptions on preservation, both on this specific unit of communication and within a continuum of scholarly communication units, as well as considerations of personal and organizational responsibility and capability for preservation; and consideration of the role and influence of co-producers, including co-authors and networks. In addition to the notes and partial transcripts, each recording was played back a minimum of three times to allow sufficient opportunity for the capture of direct quotations to illustrate key concepts and emergent themes.

### 3.4. Blog Analysis

#### 3.4.1. Coding System

Concerning development of the coding system, some of the categories, particularly in regard to blog structure and author elements, were derived from the code book devised by Herring et al. (2005). This included: blog author elements; technical features; history and activity level; post-specific elements; commenting system-specific elements; and overall blog content, including widgets and badges posted to sidebars. Other categories in the blog coding system were developed using a grounded approach, with categories emerging from preliminary analyses of blogs identified in the sampling frame. It was also informed from an earlier investigation of blogs hosted at ScienceBlogs (Hank, 2009). This grounded approach resulted in both on-site and off-site coding categories, including: additional author information concerning education and employer; additional user interface features; content related to use, including rights and disclaimer statements; elements related to authority and

---

dissertation committee, no personal identifiers, including name of interviewee, blog, or institution, is disclosed, regardless of permission to do so.

audience, including blog statistics, awards, network affiliations, and indexing to blog directories; and content types, including embedded or imported content published to posts.

Due to differences in authorship, two coding systems were ultimately devised. The coding system for co-blogs captures 57 data points across six categories: 1) blog elements and features; 2) rights and disclaimers; 3) authority and audience; 4) blog publishing activity; 5) post features; and 6) archiving. The coding system for single-blogs is the same as the co-blog coding system except for the addition of one other category – authorship. The coding system for single-blogs captures 63 data points across these seven categories. Additionally, each coding system has a data management section, capturing 5 data points to facilitate data collection. See Appendix J for the blog coding system.

Prior to commencing data collection, the blog coding systems were evaluated. First, a random sample of ten unsampled blogs from the blog sampling frame was drawn. These blogs were coded, contributing to minor revision of the coding systems. Second, a coder was trained in the use of the coding systems. Ten blogs from the blog sample were randomly selected to be used in inter-rater reliability testing. After each of the first three blogs were coded by the coder and then me, results were compared. This could be better described as extended training as opposed to genuine inter-rater reliability testing. Just as the adage goes that no two snowflakes are alike, so the same might be said of blogs. This applies to the diversity in layout and format, including “structural” clues on where content on the blog may reside, if at all. The coding system was further revised after the third blog was coded to improve instructions for coding. In total, training took about four hours. The remaining seven blogs were each coded by the tester and me, and then the results were compared. Bauer (2000) categorizes reliability as very high (> 90%); high (>80%); and acceptable (66 to

79%). Comparisons between the tester's coding and mine fell in the high to very high categories, ranging from 82 to 94% across those seven blogs.

### 3.4.2. Data Collection and Analysis

All 94 blogs in the blog sample were coded and entered into an Excel workbook from September 16-21, 2010.<sup>63</sup> Ultimately, coding was completed for 93 blogs.<sup>64</sup> Completion times varied due to diversity in blog layout and user elements. See Table 12 for frequency of time to code for all blogs in the sample.

**Table 12. Frequency distribution of time to code**

<b>Time in Minutes</b>	<b>Single-Blog Frequency (%)</b>	<b>Co-Blog Count Frequency (%)</b>
<b>≤ 9</b>	17 (28%)	5 (15%)
<b>10 to 19</b>	32 (52%)	24 (73%)
<b>20 to 29</b>	9 (15%)	2 (6%)
<b>30 to 39</b>	2 (3%)	1 (6%)
<b>≥ 40</b>	1 (2%)	-

The overall goal of this phase of the analysis is to describe attributes of a particular group of blogs identified via purposive sampling. As such, data analysis was limited to descriptive measures. Quantitative results were analyzed in Excel. Open-text entered for select coding categories was normalized, categories were derived, and these were subsequently analyzed in Excel. In reporting findings, select open-text responses are also used to illustrate examples of specific blog attributes.

---

<sup>63</sup> A major portion of the sample (n=78) had initially been coded for analysis beginning August 20, 2010. However, some data was lost due to an error on my part. A decision to recode all blogs was made due to issues of currency and for homogeneity among all blogs in the sample.

<sup>64</sup> There were issues in coding one co-blog assigned to the economics cluster. The page kept timing out, so that not all categories could be coded. Ultimately, this blog was excluded from the sample.



## **CHAPTER 4**

### **RESULTS: QUESTIONNAIRES AND INTERVIEWS**

This chapter presents results from the questionnaire data collection phase of the study. Reported are quantitative results from both single- and co-bloggers completing QI and QII, respectively. Additionally, select results from the interview portion of the study are also provided here to illustrate the quantitative results from the questionnaires. Before turning to the results, outcome rates for the questionnaire portion of the study and the completed sample for the interview portion of the study are provided immediately below.

#### **4.1. Respondents and Interviewees**

##### **4.1.1. Questionnaire: Outcome Rates and Completed Sample**

QI and QII completion and communications with sample members was carefully monitored during data collection. Outcome rates were calculated based on disposition codes and calculations for Internet surveys of specifically named persons from the American Association for Public Opinion Research (AAPOR, 2009). See Table 13 for definitions of terms and counts used in outcome calculations.

Four sample members self-identified as ineligible in email responses to invitations to participate and were subsequently classified as such, resulting in a revised sample size (N=294). Based on the revised sample, four outcomes are reported in this section: response rate, cooperation rate, refusal rate, and contact rate.

**Table 13. Questionnaire final disposition codes, definitions, and results**

Term	Definition	Count of Sampled Cases	
		QI (N=108)	Q II (N=190)
Complete Questionnaire (I)	Count of respondents returning questionnaires with 80% or more of all applicable questions completed.	67	86
Partial Questionnaire (P)	Count of respondents returning questionnaires 50 to 79% complete. Partial responses not considered for analysis.	-	4
Refusal and Break-off (R/O)	Count of respondents returning questionnaires less than 50% complete. (R)	1	1
	Count of respondents declining to participate via an email reply to an invitation email. (R)	2	3
	Count of respondents who emailed and acknowledged receipt of invitation but did not complete questionnaire. (O)	3	5
	Count of respondents who accessed questionnaire, but selected “No Thanks” from the Welcome Screen when prompted to continue on to the first question. (R)	2	2
	Count of respondents who accessed questionnaire but failed to complete any items after selecting “Yes” when prompted to continue on to the first question. (R)	-	1
Non-Contact (NC)	Count of respondents for which invitations were replied to with an “out of office” message. (NC)	1	10
Unknown (UO/UH)	Count of respondents for which emailed invitations were returned undelivered. (UO)	2	2
	Count of respondents for which no information on outcome of invitation is known. (UH)	29	73
Ineligible	Count of respondents who self-identified as ineligible in email reply to an invitation. Reasons provided include: non-scholar status; non-academic or scholarly blog; or inactive blogger.	1	3

#### 4.1.1.1. Response Rates

Two response rates are provided. Response Rate 1 (RR1), or the minimum response rate, is calculated as:  $I / (I + P) + (R + NC + O) + (UH + UO)$ . It is defined as, “the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews ... plus all cases of unknown eligibility” (AAPOR, 2009, p. 35). Response Rate 2 (RR2) counts partial interviews as respondents, and is calculated as:  $(I + P) / (I + P) + (R + NC + O) + (UH + UO)$ . RR1 for QI, QII, and QI & QII are shown in Table 14, and RR2 is provided in Table 15.

**Table 14. Response Rate 1 (RR1)**

Questionnaire	RR1 [Equation]
QI: Single-bloggers	<b>63%</b> [= 67 / (67+0) + (5+1+3) + (29+2)]
QII: Co-bloggers	<b>46%</b> [= 86 / (86+4) + (7+10+5) + (73+2)]
QI & QII: All bloggers	<b>52%</b> [= 153 / (153+4) + (12+11+8) + (102+4)]

**Table 15. Response Rate 2 (RR2)**

Questionnaire	RR1 [Equation]
QI: Single-bloggers	<b>63%</b> [= (67+0) / (67+0) + (5+1+3) + (29+2)]
QII: Co-bloggers	<b>48%</b> [= (86+4) / (86+4) + (7+10+5) + (73+2)]
QI & QII: All bloggers	<b>53%</b> [= (153+4) / (153+4) + (12+11+8) + (102+4)]

#### 4.1.1.2. Cooperation Rates

A cooperation rate is the, “proportion of all cases interviewed of all eligible units ever contacted” (AAPOR, 2009, p. 37). Two cooperation rates are provided. Cooperation Rate 1 (COOP1), or the minimum cooperation rate, is calculated as:  $I / (I + P) + R + O$ . It is defined as, “the number of interviews (complete or partial) plus the number of non-interviews that involve the identification of and contact with an eligible respondent” (AAPOR, 2009, p. 37). Cooperation Rate 2 (COOP2) counts partial interviews as respondents in the calculation, and is calculated as:  $(I + P) / (I + P) + R + O$ . COOP1 for QI, QII, and QI & II is shown in Table 16, and COOP2 is provided in Table 17.

**Table 16. Cooperation Rate 1 (COOP1)**

Questionnaire	COOP1 [Equation]
QI: Single-bloggers	<b>89%</b> [= 67 / (67+0) + 5 + 3]
QII: Co-bloggers	<b>84%</b> [= 86 / (86+4) + 7 + 5]
QI & QII: All bloggers	<b>86%</b> [= 153 / (153+4) + 12 + 8]

**Table 17. Cooperation Rate 2 (COOP2)**

Questionnaire	COOP2 [Equation]
QI: Single-bloggers	<b>89%</b> [= (67+0) / (67+0) + 5 + 3]
QII: Co-bloggers	<b>88%</b> [= (86+4) / (86+4) + 7 + 5]
QI & QII: All bloggers	<b>89%</b> [= (153+4) / (153+4) + 12 + 8]

#### 4.1.1.3. Refusal Rate

A refusal rate is, “the proportion of all cases in which a ... respondent refuses to do an interview, or breaks-off an interview, of all potentially eligible cases” (AAPOR, 2009, p. 38). Refusal Rate 1 (REF1) is calculated as:  $R / (I + P) + (R + NC + O) + (UH + UO)$ . It is

defined as, “the number of refusals divided by the interviews (complete plus partial) plus the non-respondents (refusals, non-contacts, and others) plus all cases of unknown eligibility”

(AAPOR, 2009, p. 38). See Table 18 for REF1 for QI, QII, and QI & QII.

**Table 18. Refusal Rate 1 (REF2)**

Questionnaire	REF1 [Equation]
QI: Single-bloggers	<b>5%</b> [= 5 / (67+0) + (5+1+3) + (29+2)]
QII: Co-bloggers	<b>4%</b> [= 7 / (86+4) + (7+10+5) + (73+2)]
QI & QII: All bloggers	<b>4%</b> [= 12 / (153+4) + (12+11+8) + (102+4)]

#### 4.1.1.4. Contact Rate

The AAPOR defines contact rates at the household level as, “the proportion of all cases in which some responsible member of the housing unit was reached by the survey” (2009, p. 39). Using only information from known respondents, the Contact Rate 1 (CON1) for QI and QII are calculated based on respondent-level contacts, in compliance with AAPOR guidelines. CON1, shown in Table 19, is calculated as:  $(I + P) + R + O / (I + P) + R + O + NC + (UH + UO)$ .

**Table 19. Contact Rate 1 (CON2)**

Questionnaire	CON1 [Equation]
QI: Single bloggers	<b>70%</b> [= (67+0) + 5 + 3 / (67 + 0) + 5 + 3 + 1 + (29 +2)]
QII: Co-bloggers	<b>55%</b> [= 86+4) + 7 + 5 / (86+4) + 7 + 5 + 10 + (73+2)]
QI & QII: All bloggers	<b>60%</b> [= (153+4) + 12 + 8 / (153+4) + 12 + 8 + 11 + (102+4)]

#### 4.1.1.5. Completed Sample

The completed sample, originally shown in Table 10 in Section 3.2.4., is reproduced in this section as Table 20. It shows the counts of respondents, by authorship and cluster, comprising the completed sample for the questionnaire portion of the study. As reported in the preceding chapter, only complete questionnaires, defined as 80% or more of all applicable questions completed, were accepted for analysis. Partial and break-off questionnaires (n=6) were excluded from analysis. Overall, the completed sample is comprised of 153 respondents.

**Table 20. Completed sample for QI and QII**

Cluster	QI: Single-Bloggers Count	QII: Co-Bloggers Count	QI & QII: All Bloggers Count
History	16	15	31
Economics	18	21	39
Law	15	38	53
BioChemPhys	18	12	30
<b>All Clusters</b>	<b>67</b>	<b>86</b>	<b>153</b>

#### 4.1.2. Interview: Completed Sample

At the end of both QI and QII, respondents were asked to indicate whether s/he were interested in taking part in a follow-up phone interview. From the 153 complete questionnaires returned by respondents, 72 (47%) indicated interest in participating in a follow-up telephone interview (37 single- and 35 co-bloggers). Overall, 24 interviews were scheduled and completed. Table 21 shows the count of interviewees, by authorship and cluster, comprising the completed sample for the interview portion of the study.

**Table 21. Completed sample for interview stage**

Cluster	Single-Bloggers Count	Co-Bloggers Count	All Bloggers Count
History	2	4	6
Economics	3	4	7
Law	4	3	7
BioChemPhys	4	0	4
<b>All Clusters</b>	<b>13</b>	<b>11</b>	<b>24</b>

#### 4.2. Blogger Profile

This section reports results from questions designed to inform a profile of bloggers responding to the questionnaires. Basic demographic information was captured and is reported here, including field of study, education, age and gender, professional age, and occupation. The latter includes information on tenure-status for those respondents identifying as post-secondary faculty. Also reported is information on respondents' overall blog publishing history, excluding respondents' respective blogs as identified at sampling.

Additionally, findings related to motivations for blogging, as revealed during the interview portion of the study, are also provided.

#### 4.2.1. Field of Study

Respondents were asked to identify their primary field of study from researcher-derived categories based on categorization of respondents' blogs, respectively, as listed at the sampling frame source: history, economics, law, biology, chemistry, and physics. Across all respondents, 142 (93%) identified with these fields. Seven percent did not: five single-blogger respondents and six co-bloggers.

For respondents not identifying with one of these fields, an text-box was provided to record their primary fields of study. Two respondents assigned to the history cluster described their fields as interdisciplinary: women's studies, and history and creative writing, respectively. Eight respondents listed to the composite sciences cluster for biology, chemistry, and physics (BioChemPhys), identified with other fields: seven in the domain of science (i.e., computer science, materials science, and mathematics) and one in philosophy. One respondent assigned to the law cluster reported the field as political science.

While these 11 respondents identified other fields as their primary field of study in the questionnaires, respondents' results are still reported based on assignment to their respective clusters at sampling.

#### 4.2.2. Highest Degree Earned

Respondents were asked to enter their highest degree earned in a text box. A text-box, as opposed to a drop-down menu of response choices, was used due to diversity of degree types across disciplines as well as across geographic areas. Table 22 shows frequency of degree types for single- and co-blogger respondents across all clusters. Two categories of

“other” degree types were derived to represent those degrees reported with less frequency: master’s level degrees (i.e., A.M., M.Litt., Laurea, and A.B.D. – all but the dissertation), and doctorate-level degrees (i.e., J.S.D., and combined J.D. and Ph.D.). All respondents have earned an advanced degree, with 96 (63%) holding a doctorate-level degree.

**Table 22. Frequency of highest degree earned by authorship**

<b>Degree</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>M.A.</b>	3 (4%)	3 (3%)	<b>6 (4%)</b>
<b>M.S.</b>	2 (3%)	1 (1%)	<b>3 (2%)</b>
<b>J.D.</b>	9 (13%)	29 (34%)	<b>38 (25%)</b>
<b>LL.M.</b>	2 (3%)	4 (5%)	<b>6 (4%)</b>
<b>Ph.D.</b>	48 (72%)	45 (52%)	<b>93 (61%)</b>
<b>Other: Master’s</b>	1 (1%)	3 (3%)	<b>4 (3%)</b>
<b>Other: Doctorate</b>	2 (3%)	1 (1%)	<b>3 (2%)</b>

As shown in Table 22, a higher percentage of single-blogger respondents (72%) hold a doctorate degree in comparison to co-blogger respondents (52%). This is attributed to the prevalence of law cluster co-bloggers in the completed sample (n=38), with a clear majority – 33 respondents (87%) – holding a J.D., J.S.D., or LL.M. For co-blogger respondents assigned to the history, economics and BioChemPhys clusters (n=41), 85% hold a Ph.D. For single-blogger respondents assigned to these same clusters (n=52), 88% hold a Ph.D.

#### 4.2.3. Professional Age, Age and Gender

Respondents were asked to report the year in which their respective highest degree earned was awarded. Professional age is calculated based on Braxton et al.’s (2002) equation: the number of years between year highest degree earned and present year. For example, if a blogger’s Ph.D. was earned in 1995 and present year is 2010, then professional age is calculated as 15 years. Table 22 shows frequency of professional age in five-year intervals for bloggers across all clusters.

**Table 23. Frequency of professional age by authorship**

<b>Years</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>≤ 4</b>	11 (18%)	13 (16%)	<b>24 (17%)</b>
<b>5-9</b>	7 (11%)	18 (22%)	<b>25 (17%)</b>
<b>10-14</b>	11 (18%)	14 (17%)	<b>25 (17%)</b>
<b>15-19</b>	11 (18%)	15 (18%)	<b>26 (18%)</b>
<b>20-24</b>	9 (15%)	12 (14%)	<b>21 (14%)</b>
<b>25-29</b>	6 (10%)	7 (8%)	<b>13 (9%)</b>
<b>≥ 30</b>	7 (11%)	4 (5%)	<b>11 (8%)</b>

Across all respondents, professional age ranged from 0 to 39 years. The mean professional age is 15 years. Single-blogger respondents, with a mean professional age of 16 years, are only slightly “older” than co-blogger respondents, with a mean professional age of 14 years. Representation across age groups, as shown in Table 23, is fairly dispersed, with all age groups comprising only between 18 and 8% of respondents. The most represented professional age group is 15 to 19 years (18%). This is closely followed by three age groups each representing 17% of respondents: 10 to 14 years; 5 to 9 years; and 4 years or younger. Thirty-one percent of respondents have professional ages between 20 and 30 or more years.

When looking at professional age by cluster, respondents in the history and BioChemPhys clusters tend to be “younger” than respondents assigned to the economics and law clusters. The mean professional age for respondents in the history cluster is 10.5 years; mean professional age for respondents in the BioChemPhys cluster is 12.5 years. The mean professional age for respondents in the law cluster is 16 years, and 18 years for respondents in the economics cluster. Table 24 shows frequency of professional age by cluster.

**Table 24. Frequency of professional age by cluster**

<b>Years</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>≤ 4</b>	10 (37%)	6 (16%)	2 (4%)	6 (21%)
<b>5-9</b>	4 (15%)	4 (11%)	12 (23%)	5 (18%)
<b>10-14</b>	4 (15%)	7 (19%)	10 (19%)	4 (14%)
<b>15-19</b>	4 (15%)	4 (11%)	12 (23%)	6 (21%)
<b>20-24</b>	3 (11%)	5 (14%)	8 (15%)	5 (18%)
<b>25-29</b>	2 (7%)	5 (14%)	5 (9%)	1 (4%)
<b>≥ 30</b>	-	6 (16%)	4 (8%)	1 (4%)



Information on real age was also collected. See Table 25 for the frequency of respondents' real age presented in five-year intervals. Across all respondents, the range of age is between 25 and 70 years old. The mean age is 45 years old. The most represented age group among all respondents is 40 to 44 years old (21%), closely followed by those aged 35 to 39 (20%). More than seven out of ten (73%) are between 35 and 54 years old.

**Table 25. Frequency of age by authorship**

<b>Years</b>	<b>Single-Bloggers Frequency</b>	<b>Co-Bloggers Frequency</b>	<b>All Bloggers Frequency (%)</b>
<b>25-29</b>	4 (7%)	1 (1%)	<b>5 (4%)</b>
<b>30-34</b>	3 (5%)	6 (9%)	<b>9 (7%)</b>
<b>35-39</b>	9 (16%)	15 (22%)	<b>24 (20%)</b>
<b>40-44</b>	10 (18%)	16 (24%)	<b>26 (21%)</b>
<b>45-49</b>	11 (20%)	9 (13%)	<b>20 (16%)</b>
<b>50-54</b>	10 (18%)	10 (15%)	<b>20 (16%)</b>
<b>55-59</b>	3 (5%)	8 (12%)	<b>11 (9%)</b>
<b>60-64</b>	3 (5%)	2 (3%)	<b>5 (4%)</b>
<b>≥ 65</b>	3 (5%)	-	<b>3 (2%)</b>

Frequency of real age in five-year intervals is also presented by cluster in Table 26. On average, respondents in the BioChemPhys cluster are the youngest, with a mean age of 41 years old, with respondents in the law cluster comprising the next oldest group, with a mean age of 44. Respondents in the economics and history clusters tend to be slightly older, with a shared average age of 48 years.

**Table 26. Frequency of age by cluster**

<b>Years</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>25-29</b>	-	2 (7%)	-	3 (12%)
<b>30-34</b>	-	2 (7%)	3 (7%)	4 (15%)
<b>35-39</b>	6 (25%)	1 (4%)	13 (29%)	4 (15%)
<b>40-44</b>	3 (13%)	4 (14%)	12 (27%)	7 (27%)
<b>45-49</b>	6 (25%)	6 (21%)	4 (9%)	4 (15%)
<b>50-54</b>	5 (21%)	5 (18%)	8 (18%)	2 (8%)
<b>55-59</b>	2 (8%)	4 (14%)	5 (11%)	-
<b>60-64</b>	1 (4%)	2 (7%)	-	2 (8%)
<b>≥ 65</b>	1 (4%)	2 (7%)	-	-

Concerning gender, almost eight out of ten respondents (78%) are male. Table 27 shows frequency of gender by authorship and Table 28 shows frequency by cluster. The law

cluster has the highest representation of female respondents (30%), followed by the history cluster (29%). The economics cluster has the highest representation of male bloggers (95%). Data on gender was not collected at sampling, so it is not known how these gender groupings represent the overall frequency of gender in the sample.

**Table 27. Frequency of gender by authorship**

<b>Gender</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>Female</b>	12 (18%)	21 (24%)	<b>33 (22%)</b>
<b>Male</b>	55 (82%)	65 (76%)	<b>120 (78%)</b>

**Table 28. Frequency of gender by cluster**

<b>Gender</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>Female</b>	9 (29%)	2 (5%)	16 (30%)	6 (20%)
<b>Male</b>	22 (71%)	37 (95%)	37 (70%)	24 (80%)

#### 4.2.4. Occupation

Respondents were asked to reply to a series of questions related to their current, primary employment. Employment type was broadly categorized as either self-employed or other-employed, with the latter representing employment by any type of organization. Table 29 shows the frequency of respondents, by authorship, identifying as self- and other-employed. Those identifying as self-employed represent only 3% of all respondents; nearly all bloggers (97%) are employed by an organization.

**Table 29. Frequency of employment type**

<b>Employer</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>Self</b>	3 (4%)	2 (2%)	<b>5 (3%)</b>
<b>Other</b>	64 (96%)	84 (98%)	<b>148 (97%)</b>

The 148 respondents not identifying as self-employed were asked to enter his/her job title in a text box. From these titles, three broad categories to characterize nature of employment were derived: 1) post-secondary faculty; 2) post-secondary non-faculty; and 3) non-post-secondary. The first category includes respondents who reported their primary job as instructional and in a post-secondary education setting. The second category includes

those employed by a post-secondary education setting in another capacity. The third category includes employment in any sector other than a post-secondary education setting, including, in the case of one respondent, an unidentified sector. See Table 30 for the frequency of respondents assigned to these categories by authorship. Nearly all (94%) are employed by a post-secondary institution.

**Table 30. Frequency of post-secondary faculty, post-secondary non-faculty, and non-post-secondary employment by authorship**

Category	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Post-Secondary Faculty</b>	47 (77%)	66 (85%)	<b>113 (81%)</b>
<b>Post-Secondary Non-Faculty</b>	8 (13%)	10 (13%)	<b>18 (13%)</b>
<b>Non-Post-Secondary</b>	6 (10%)	2 (3%)	<b>8 (6%)</b>

Table 31 shows frequency of bloggers assigned to these categories by cluster. Nearly all bloggers in the law (98%) and economics clusters (94%) are categorized as post-secondary faculty. Faculty are treated more extensively in Section 4.2.4.1.

**Table 31. Frequency of post-secondary faculty, post-secondary non-faculty, and non-post-secondary employment by cluster**

Category	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Post-Secondary Faculty</b>	17 (61%)	31 (94%)	49 (98%)	16 (57%)
<b>Post-Secondary Non-Faculty</b>	8 (29%)	2 (6%)	1 (2%)	7 (25%)
<b>Non-Post-Secondary</b>	3 (11%)	-	-	5 (18%)

For respondents' categorized as post-secondary non-faculty, five sub-categories were derived to represent job titles: Post-Doc, Librarian, Researcher, Student, and Other. "Post-Doc" includes those respondents identifying as post-doctoral fellows, researcher or scholars. "Researcher" comprises those respondents identifying a research-oriented title (i.e., senior researcher, research assistant). "Student" comprises those identifying as a student. Included in this sub-category are respondents identifying as Ph.D. candidates as well as those providing other descriptive information in addition to their student designation (i.e., teaching fellow, lecturer). Lastly, "Other" represents titles not falling within these subcategories. These include: director/associate director, managing attorney, managing director, and

systems administrator. Table 32 shows frequency of respondents assigned to these five sub-categories by authorship.

**Table 32. Frequency of job titles for post-secondary, non-faculty respondents by authorship**

Sub-Category	Single-Bloggers Frequency	Co-Bloggers Frequency	All Bloggers Frequency (%)
Post-Doc	-	3	3 (18%)
Librarian	1	1	2 (12%)
Researcher	-	2	2 (12%)
Student	4	2	6 (35%)
Other	2	2	4 (24%)

For the eight respondents' categorized as non-post-secondary, five sub-categories representing industry or type of employment were identified. Three respondents are employed in the chemical and pharmaceutical industries; one by a research institute; two work in publishing; and one works in accounting. For the eighth and remaining respondent, identifying as, "director of a center," industry type could not be determined.

#### 4.2.4.1. Faculty Bloggers

For the 113 respondents categorized as post-secondary faculty, eight position categories were derived from respondents' open-text responses. These categorizations are presented in Tables 33 and 34, with the former showing frequency of faculty titles by authorship, and the latter showing frequency by cluster. Several of these categories merit further explanation. Faculty identifying as distinguished professors, endowed chairs, and chairs are reported under the sub-category *Distinguished/Chaired Professor*. Respondents identifying as professor or full professor are reported in the *Professor/Full Professor* sub-category. Faculty identifying two positions (e.g., professor and library director; professor and museum director; associate dean and professor) are reported here as *Professor/Full Professor+*. Those identifying as a research professor, associate research professor, or assistant research professor are assigned to the sub-category, *Research Professor*. Faculty with under-represented titles (e.g., emeritus professor, economist, senior lecturer, instructor,

and adjunct professor) are reported under the sub-category, *Faculty – Other*. Further, one respondent identified as a Visiting Assistant Professor; this blogger is reported under the sub-category, Assistant Professor.

**Table 33. Frequency of job titles for post-secondary faculty respondents by authorship**

Category	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Distinguished/Chaired Professor</b>	2 (4%)	3 (5%)	<b>5 (4%)</b>
<b>Professor/Full Professor +</b>	-	3 (5%)	<b>3 (3%)</b>
<b>Professor/Full Professor</b>	24 (51%)	26 (39%)	<b>50 (44%)</b>
<b>Associate Professor</b>	9 (19%)	22 (33%)	<b>31 (27%)</b>
<b>Assistant Professor</b>	6 (13%)	5 (8%)	<b>11 (10%)</b>
<b>Research Professor</b>	1 (2%)	3 (5%)	<b>4 (4%)</b>
<b>Reader</b>	2 (4%)	1 (2%)	<b>3 (3%)</b>
<b>Faculty – Other</b>	3 (6%)	3 (5%)	<b>6 (3%)</b>

**Table 34. Frequency of job titles for post-secondary faculty by cluster**

Category	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Distinguished/Chaired Professor</b>	1 (6%)	-	4 (8%)	-
<b>Professor/Full Professor +</b>	1 (6%)	-	2 (4%)	-
<b>Professor/Full Professor</b>	6 (35%)	15 (48%)	22 (45%)	7 (44%)
<b>Associate Professor</b>	4 (24%)	7 (23%)	17 (35%)	3 (19%)
<b>Assistant Professor</b>	3 (18%)	4 (13%)	1 (2%)	3 (19%)
<b>Research Professor</b>	-	3 (10%)	-	1 (6%)
<b>Reader</b>	-	-	1 (2%)	2 (13%)
<b>Faculty – Other</b>	2 (12%)	2 (6%)	2 (4%)	-

Data on tenure status was also collected. Due to respondents' variety in specific occupation, variations in the practice of tenure across institutions, both nationally and abroad, and the potentially sensitive nature of the question, respondents were presented with multiple response stems concerning the question on tenure status: tenured; non-tenured in tenure track position; not applicable (not in a tenure track position); and prefer not to answer. For display in the tables to follow, these response categories are abbreviated as follows: tenured, tenure track, not applicable, and not reported. Table 35 shows frequency of tenure status for responding post-secondary faculty by authorship, and Table 36 by cluster. A majority (78%) are tenured.

**Table 35. Frequency of tenure status for post-secondary faculty by authorship**

Status	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Tenured</b>	38 (81%)	50 (76%)	<b>88 (78%)</b>
<b>Tenure track</b>	7 (15%)	13 (20%)	<b>20 (18%)</b>
<b>Not applicable</b>	1 (2%)	3 (5%)	<b>4 (4%)</b>
<b>Not reported</b>	1 (2%)	-	<b>1 (1%)</b>

**Table 36. Frequency of tenure status for post-secondary faculty by cluster**

Status	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Tenured</b>	13 (76%)	24 (77%)	39 (80%)	12 (75%)
<b>Tenure track</b>	2 (12%)	6 (19%)	9 (18%)	3 (19%)
<b>Not applicable</b>	2 (12%)	-	1 (2%)	1 (6%)
<b>Not reported</b>	-	1 (3%)	-	-

Tenure-status is also shown for job titles of all respondents categorized as post-secondary faculty. See Table 37 for count of tenure status for the eight position types reported in Tables 33 and 34. These types are abbreviated in Table 37; additionally the *Professor/Full Professor* and *Professor/Full Professor+* categories are combined.

**Table 37. Frequency of tenure status for post-secondary faculty by job title**

Status	Dist./ Chaired Prof Freq (%)	Prof/ Full Freq (%)	Assoc. Prof Freq (%)	Asst. Prof Freq (%)	Research Prof Freq (%)	Reader Freq (%)	Other Faculty Freq (%)
<b>Tenured</b>	5 (100%)	53 (100%)	21 (68%)	-	3 (75%)	3 (100%)	3 (50%)
<b>Tenure-Track</b>	-	-	10 (32%)	9 (82%)	1 (25%)	-	-
<b>Not Applicable</b>	-	-	-	2 (18%)	-	-	2 (33%)
<b>Not Reported</b>	-	-	-	-	-	-	1 (17%)

#### 4.2.5. Motivations to Blog

While not a direct question in the interview schedule, eight interviewees disclosed motivations for blogging during the course of their respective interviews. Findings from these interviewees are presented here to complement other respondent attributes presented in the preceding sub-sections, and to contribute to a richer profile of scholars who blog.

A tenured law professor described starting his blog, begun about five years ago, as, “a lark,” and that he thought the, “new technology was interesting and had no idea if it would succeed or not.”

The other interviewees had more explicit objectives for launching a blog, including to bring attention to particular areas of study; for example, by creating a “presence ... on the Internet.” Another reason was for advocacy purposes, “to try and influence public debate.” Concerning the former, one tenured associate professor of law remarked, “I started the blog to signal to people that this [particular area of legal scholarship] is a discipline now.” Concerning the latter, one tenured history professor remarked, “I have a specific agenda. I do it for advocacy purposes. And I also do it to demonstrate my strengths. And I do it from an academic point of view.”

For another interviewee, a tenured professor assigned to the BioChemPhys cluster, motivation was not simply to share information about her research, but to share information about the process of doing research. She describes her blog as a “research blog” and not a “scientific blog,” and remarked that there are, “still very few people who blog about the process of research and the scientific content of their research.” Further, she provided a very detailed account of the catalyst for her decision to start her research blog four years ago:

[There was a] posting that somebody else had made about how it was a shame that there wasn't more information available. That what went on in researchers' minds and in researchers' labs wasn't publicly available. That the ordinary man in the street has no way to find out what scientists actually do and what they think, except for sort of fancy, formal interviews ... so when I read that I said, 'duh, maybe you should start a blog.'

Other interviewees commented on blogs as tools and avenues for writing, thinking and sharing ideas, with one interviewee remarking that, “that space ... is very important in my writing and my scholarly work,” and blogging allows her to, “be a little freer.” A Ph.D. working in the chemical industry described his motivation and the subsequent effect in this way:

The intersection of people who do my kind of science with the set of people who can talk intelligibly about my kind of science is not a large one ... So that was the attraction of the blog to me. I thought well, now I can sit down and write about what I do because I like doing the science and I also like talking about it and I can finally combine the two.

Another interviewee, a tenured professor in economics, described two primary objectives for blogging, touching on both the attraction of the form for writing and the benefit for communicating with the broader community:

... first of all, I hope it'll spark some thinking by me. If I write a couple hundred words on something, I'm a little bit smarter about it than when I started ... that's goal number one. And goal number two, we all think of ourselves in a community and we all think of ourselves as naturally sharing views and opinions, and I think people do that in blogs and all these review sites ... even Wikipedia. There are people stepping forward for all kinds of reasons and saying, 'here's what I think,' without necessarily getting traditional remuneration or traditional recognition or anything like that.

A tenured history professor equated his blog not so much as a tool, but as a "weapon." He had a very specific motivation. And, in consideration of other interviewees' comments, a very unique and deliberate objective. He related a personal story in describing his motivation, with an excerpt from his remarks provided below:

I got into blogging as a counterweight and as an explicit resistance to persecution of me ... The blog was clearly a weapon, it was a type of armor ... it was part of my preservation of myself. I did mine to save my job and preserve my honor ... this was a way of preserving my very essence as a person and defending my rights within academia so it had that tremendous sense of passion and self-preservation and anger and conviction to it.

He goes on to comment that, "in the beginning it was clearly a crusade of rebellion," but as time has passed, "I don't quite have that fire." While interviewees commented on how their respective blogs have or have not influenced opportunities for promotion and tenure, shared in a later section of this chapter, this is a unique example, and one absent in the literature as far as I have found, of an academic choosing to blog to keep one's job.



While interviewees may have casual or very explicit objectives for starting their respective blogs, these objectives are not necessarily satisfied. As commented on by a distinguished professor of law:

When I started it [the blog], I actually thought I was going to join a great invisible college of communication ... I discovered very, very quickly that that absolutely did not work for me because the level of precision and detail it takes for me to do scholarly work and put it into a state where I want to put my name to it is just as high, would be just as high, for blogging as it would be for any other communication.

#### 4.3. Publishing to Other Blogs

While the majority of findings reported in this chapter concern questionnaire respondents' respective blogs as identified at sampling, a series of questions gathered information on other blogs published to by respondents. Respondents were asked if the blog identified at sampling was the first blog s/he published. The blog identified at sampling was the first blog published to by 84 (55%) of respondents.

Respondents also were asked if s/he publish to blogs other than the blog identified at sampling. Sixty-four respondents (42%) publish to two or more blogs. Table 38 shows the frequency of respondents publishing to only one blog and those publishing to two or more blogs, by authorship, and Table 39 show this by cluster. More respondents in the history (65%) and law clusters (42%) reported publishing to more than one blog than those in the economics (31%) and BioChemPhys clusters (33%).

**Table 38. Frequency of all blogs published to by authorship**

<b>Blogs</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>1</b>	36 (54%)	53 (62%)	<b>89 (58%)</b>
<b>2 or more</b>	31 (46%)	33 (38%)	<b>64 (42%)</b>

**Table 39. Frequency of all blogs published to by cluster**

<b>Blogs</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>1</b>	11 (35%)	27 (69%)	31 (58%)	20 (67%)
<b>2 or more</b>	20 (65%)	12 (31%)	22 (42%)	10 (33%)

For those publishing to two or more blogs, respondents were asked to provide a count of how many other single- and co-blogs they publish, excluding the blog identified at sampling. These distinctions were clarified for bloggers in the questionnaires, with a single-blog characterized as a blog for which the respondent is the only author, and a co-blog characterized as a blog for which the respondent is one of two or more authors. Table 40 shows the frequency of single-blogs published to by single- and co-bloggers, and Table 41 shows the frequency of co-blogs published to by single- and co-bloggers.

**Table 40. Frequency of other single-blog instances by authorship**

<b>Blog Count</b>	<b>Single-Bloggers Frequency</b>	<b>Co-Bloggers Frequency</b>	<b>All Bloggers Frequency (%)</b>
<b>1 blog</b>	13 (68%)	14 (70%)	<b>27 (69%)</b>
<b>2 blogs</b>	2 (11%)	2 (10%)	<b>4 (10%)</b>
<b>3 blogs</b>	2 (11%)	2 (10%)	<b>4 (10%)</b>
<b>4 blogs</b>	2 (11%)	-	<b>2 (5%)</b>
<b>5 blogs</b>	-	-	<b>-</b>
<b>6 or more</b>	-	2 (10%)	<b>2 (5%)</b>

**Table 41. Frequency of other co-blog instances by authorship**

<b>Blog Count</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>1 blog</b>	12 (52%)	8 (40%)	<b>20 (47%)</b>
<b>2 blogs</b>	4 (17%)	6 (30%)	<b>10 (23%)</b>
<b>3 blogs</b>	2 (9%)	2 (10%)	<b>4 (9%)</b>
<b>4 blogs</b>	1 (4%)	1 (5%)	<b>2 (5%)</b>
<b>5 blogs</b>	1 (4%)	-	<b>1 (2%)</b>
<b>6 or more</b>	3 (13%)	3 (15%)	<b>6 (14%)</b>

Additionally, Table 42 shows frequency of both single- and co-blogs, combined, excluding blogs identified at sample. For those respondents publishing to other blogs, most (44%) publish to only one other blog, representing 18% of the completed sample (n=153). The next largest segment of respondents (22%) publishes to two other blogs, representing 9% of the completed sample. While the third largest group publishes to 6 or more blogs (14%), these respondents only represent 6% of the completed sample.

**Table 42. Frequency of single- and co-blog instances by authorship**

<b>Blog Count</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>1 blog</b>	14 (45%)	14 (44%)	<b>28 (44%)</b>
<b>2 blogs</b>	5 (16%)	9 (28%)	<b>14 (22%)</b>
<b>3 blogs</b>	3 (10%)	5 (16%)	<b>8 (13%)</b>
<b>4 blogs</b>	1 (3%)		<b>1 (2%)</b>
<b>5 blogs</b>	3 (10%)	-	<b>3 (5%)</b>
<b>6 or more blogs</b>	5 (16%)	4 (13%)	<b>9 (14%)</b>

#### 4.3.1. Blog Deletion

Across their overall blog publishing history, respondents were asked if s/he had ever accidentally or purposefully deleted an entire blog. Only 24 bloggers (16%) have reported accidental or purposeful deletion of a blog. Table 43 shows frequency of blog deletion by authorship.

**Table 43. Frequency of blog deletion by authorship**

<b>Blog count</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>0</b>	55 (83%)	73 (85%)	<b>128 (84%)</b>
<b>1 or more</b>	11 (17%)	13 (15%)	<b>24 (16%)</b>

These respondents were asked to report how many blogs s/he had purposefully or accidentally deleted. Only one blog, as reported by one respondent, was identified as accidentally deleted. All other occurrences of blog deletion were reported as intentional. Table 44 shows the frequency of the number of blogs reported as purposefully deleted. For those reporting intentional deletion (n=24), most (79%) reported deleting either one or two blogs.

**Table 44. Frequency of blogs intentionally deleted by authorship**

<b>Measure</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>1 blog</b>	5 (45%)	4 (31%)	<b>9 (37%)</b>
<b>2 blogs</b>	3 (27%)	7 (54%)	<b>10 (42%)</b>
<b>3 blogs</b>	3 (27%)	1 (8%)	<b>4 (17%)</b>
<b>4 blogs</b>	-	1 (8%)	<b>1 (4%)</b>

These 24 respondents were presented with a listing of eleven reasons for blog deletion and asked to identify all of the reasons which contributed to their decisions to delete

their respective blogs. Table 45 shows frequency of reason selected for single- and co-bloggers in the aggregate. The most cited reason for deletion is due to topical treatment: 13 respondents (54%) reported loss of interest in the blog’s subject matter. The second most cited reason for deletion is life-cycle related. That is, the blog was only intended to be available for a specific period. Twelve bloggers (50%) cited the blogs’ project-based or time-limited nature as cause for deletion. Other reasons most cited include time required to maintain the blog (38%) and “blog regret,” (33%), characterized as just not wanting the blog “out there anymore.”

**Table 45. Frequency of reasons for intentional blog deletion**

<b>Reason</b>	<b>Frequency</b>	<b>Percentage</b>
Because it was too much time required to maintain blog	9	38%
Because it was too expensive to maintain blog	1	4%
Blog was unpopular or under-read	3	13%
Because of dissatisfaction with blog publishing application	4	17%
Because of dissatisfaction with blog host service or location	1	4%
Because of loss of interest in subject matter of blog	13	54%
In response to a request from someone	-	-
Because it was a project-based or time-limited blog	12	50%
Because it was a duplicate blog	1	4%
Because of “blog regret;” I just didn’t want it out there anymore	8	33%
Because blog received too much spam	1	4%

While only 16% of the completed sample report purposeful deletion of blogs, their reasons point to the ephemeral nature of both the blog form and respondents’ personal publishing preferences.

#### 4.4. Publication and Service

This section reports questionnaire results from questions designed to capture respondents’ scholarly publishing and service history, apart from their respective blogs. Results on scholars’ perceptions of their respective blogs in their professional lives is reported in section 4.5. Before turning to these results, perceptions on scholarly communication gathered during the interview portion of the study are presented to provide

insight and context as to how scholars' characterize scholarly communication. Interviewees were asked to define the system of scholarly communication in their own words.

Interviewees' responses are reflective of Borgman's (1990) definition of scholarly communication, particularly the aspect that, "By *scholarly communication* we mean the study of how scholars in any field (e.g. physical, biological, social, and behavioral sciences, humanities, technology) use and disseminate information through formal and informal channels" (p.13).

Overall, interviewees' definitions of scholarly communication included reference to both informal and formal communications. These two aspects were succinctly described by a tenured professor of economics in his definition of scholarly communication: "A combination of oral and written discourse, through personal interactions and writing and reading journal articles." A Ph.D. in economics defined scholarly communication as informal and formal, but also as short- and long-term: "People talking over lunch, in the short-term; in the long-term, what people end up citing."

An interviewee assigned to the BioChemPhys cluster used an onion metaphor to describe the relationship between informal communications and the "core to the ecosystem ... published peer review literature," noting that there, "are layers surrounding the paper ... like an onion around it." These layers include papers, comments, notes, supplemental data, media coverage, other reading levels, newspapers, blogs, and popular science magazines.

A lot of things, in the words of a tenured associate professor of law, fall under the umbrella term of scholarly communication: "It is not just the printed words that actually get published. I think the conversations that go around the publications, that lead up to the

publications, are all part of it.” This viewpoint is echoed by a Ph.D. in economics, who commented:

On the one hand, most of the work and thinking occurs in the informal sphere and on the other hand, arguments are built up ... in the formal documented or archives sphere. It’s a building mound of authority ... We think informally and go to authority in terms of formal archives.

Adjectives applied to formal channels of communication by interviewees include “traditional,” “standard,” “well-defined,” and “accepted.” Traditional, established unit types referenced include journals, scholarly articles, books, and monographs. Less formal communications identified by interviewees include: drafts, working papers or other works-in-progress; presentations at a variety of venues, including conferences, faculty forums, seminars, and other meetings; letters, email, blogs and face-to-face conversations; and teaching and course materials.

Peer-review was another common attribute in definitions shared by interviewees. One tenure-track assistant professor of economics and business commented: “Refereed journal articles are the biggest and most important thing in economics. Reading and commenting on journal articles and reading and commenting on books is the main way people communicate.” As commented on by a Ph.D. working in the chemical industry, most scholars in his area of specialty, “have a pretty clear idea of what the hierarchy of these journals are, where the good places to get published, where the okay, and where the not so good places are.” However, in consideration of legal scholarship, as remarked by one tenured associate professor of law, “peer review journals are not the norm in legal scholarship,” a point echoed by several interviews assigned to the law cluster.

Interviewees also defined scholarly communication in consideration of intent and objective. In regard to more formal communications, these include the intent of

dissemination and the resulting recognition and rewards that come from publication, with many interviewees' specifically mentioning tenure. Feedback was frequently mentioned as both an outcome and benefit of more informal communications, such as conference presentations and working papers. A few interviewees also made mention of actors within the system of scholarly communication. In addition to peers, in the roles of both producers and consumers, other actor-roles specifically mentioned include scholarly societies and academic publishers.

Several interviewees commented on the need for a re-evaluation of the definition of scholarship. As one tenured humanities professor remarked in regard to his field of study:

The definition of scholarship, the character of scholarship, has got to evolve some. That sticking with the monograph and the scholarly journal article in the humanities isn't going to be sufficient any longer in the new media environment that we're in to define what makes someone an expert or a scholar in a particular subject.

Some interviewees responded by calling into question the wording of the interview question itself; that is, the notion of a *system* of scholarly communication. A couple of interviewees did not support the depiction of scholarly communication as necessarily *systematic*. As one tenured law professor commented: "I don't think it is very systematic. I read things, talk to people, check my mail, go to conferences ... I consider all that part of the system.

Another tenured humanities professor described the system as corrupt, commenting that, "we should not have the notion that only peer-review publications are acceptable ... so the fact that there is a system leads to the corruption." He elaborated on his viewpoint in regard to scholarly discourse, noting that it is, "either totally corrupt, because it's ideologically controlled by relatively few people, or it is restricted to such an arcane, small audience that it is virtually, totally irrelevant."

A Ph.D. in economics critiqued the current nature of scholarly communication by reference to the model of supply and demand. He commented that, with the volume of journals in scholars' respective fields in contemporary times:

... fewer and fewer people read each paper and there's less and less communication occurring through academic papers and publications. If you go back to an era where there were, say, 50 journals, it was pretty clear everyone in the business was going to read your article if it appeared in the flagship ... now it's not even close to being that way because either there's too many journals or people are concentrating too much on getting publications out for their own career.

He continues by describing the outcome of this development, remarking: "So you have a bunch of people talking past each other, or rather publishing past each other, and scholarly communication is much less robust, either in terms of a common knowledge or in terms of critical debate."

#### 4.4.1. Publication History

Returning to the questionnaire portion of the study, questionnaire respondents were provided with a list of ten types of units of communication and asked to identify which, if any of these units, s/he has ever published. Table 46 shows the frequency of respondents reporting at least one publication within each unit- type. No information on number of units published per unit-type was collected.

**Table 46. Frequency of scholarly publishing history by unit-type of communication**

<b>Unit-Type</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
Textbook	18 (27%)	20 (23%)	<b>38 (25%)</b>
Scholarly book or monograph (editor or co-editor)	17 (25%)	33 (38%)	<b>50 (33%)</b>
Scholarly book or monograph (author or co-author)	31 (46%)	44 (51%)	<b>75 (49%)</b>
Scholarly book chapter	46 (69%)	66 (77%)	<b>112 (73%)</b>
Peer-reviewed journal article	63 (94%)	73 (85%)	<b>136 (89%)</b>
Comment in a peer-reviewed journal	32 (48%)	42 (49%)	<b>74 (48%)</b>
Book review in a scholarly journal	46 (69%)	55 (64%)	<b>101 (66%)</b>
Book review for a newspaper or a magazine	33 (49%)	26 (30%)	<b>59 (39%)</b>
Scholarly paper in a conference proceedings	57 (85%)	74 (86%)	<b>131 (86%)</b>
Non-fiction essay, article, or op-ed piece in a newspaper or magazine	55 (82%)	57 (66%)	<b>112 (73%)</b>



Most respondents (89%) have published a peer reviewed journal article. Additionally, most have published a paper in a conference proceedings (86%), a scholarly book chapter (73%), and a non-fiction essay, article or editorial in a newspaper or magazine (73%).

Based on respondents' answers, a listing of how many of these respective unit-types s/he had ever published to was derived. Again, this does not represent total units published by respondents. It shows how many of these particular unit-types to which respondents have published. Table 47 shows the frequency of total unit-types published to by respondents.

**Table 47. Frequency of scholarly publishing history across unit-types of communication**

<b>Count of Unit-Types</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>0 units</b>	-	-	
<b>1 unit</b>	3 (4%)	4 (5%)	<b>7 (5%)</b>
<b>2 units</b>	3 (4%)	7 (8%)	<b>10 (7%)</b>
<b>3 units</b>	4 (6%)	9 (10%)	<b>13 (8%)</b>
<b>4 units</b>	7 (10%)	12 (14%)	<b>19 (12%)</b>
<b>5 units</b>	9 (13%)	5 (6%)	<b>14 (9%)</b>
<b>6 units</b>	15 (22%)	12 (14%)	<b>27 (18%)</b>
<b>7 units</b>	8 (12%)	16 (19%)	<b>24 (16%)</b>
<b>8 units</b>	7 (10%)	8 (9%)	<b>15 (10%)</b>
<b>9 units</b>	8 (12%)	6 (7%)	<b>14 (9%)</b>
<b>10 units</b>	3 (4%)	7 (8%)	<b>10 (7%)</b>

While all respondents have published to at least one of the unit-types, as shown in Table 46, publishing experience ranges across all respondents. Twenty percent of respondents have published to between one to three of these unit-types, while 39% have published to between four to six of these unit-types. Nearly half (49%) have published to seven or more of these unit-types.

#### 4.4.2. Service History

To gauge professional service history in scholarly publishing, questionnaire respondents were asked if s/he had ever performed three specific service activities. These categories are in no way exhaustive, nor intended to be. The categories of service were selected to highlight only a few service-oriented activities. Table 48 shows the frequency of

all respondents' scholarly publishing-related service by activity type. Across all respondents, 139 (91%) have performed at least one of these service activities: 62 (93%) of single-bloggers and 77 (90%) of co-bloggers.

**Table 48. Frequency of professional service activity by authorship**

Service Activity	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
Served as a referee for a scholarly journal article	52 (78%)	69 (80%)	<b>121 (79%)</b>
Evaluated a scholarly book or manuscript for a publisher	47 (70%)	60 (70%)	<b>107 (70%)</b>
Served as an editor or associate editor of a scholarly journal	24 (36%)	34 (40%)	<b>58 (38%)</b>

Overall, questionnaire respondents are active producers, disseminating scholarship through formal and informal channels of communication, and most are active and experienced evaluators of other scholars' publications.

#### 4.5. Blogging and Professional Life

This section reports results from questions designed to capture respondents' perceptions of their respective blogs, as identified at sampling, in relation to scholarly communication, including publishing, service, teaching, research, promotion and collaboration.

##### 4.5.1. Scholarship Criteria

Braxton, Luckey, and Helland (2002) include "unpublished scholarly outcomes and publications" in their definition of scholarship, writing that, "unpublished scholarly outcomes fully meet the definition of scholarship if they appear in a publicly observable form" (p. 141). The three parameters for this form are that, "it must be public, subject to critical review, and in a form that allows use and exchange by other members of the scholarly community" (Braxton et al., 2002, citing Shulman & Hastings (1999), p. 141). Questionnaire respondents were presented with these three criteria and asked, to what extent, s/he agreed or disagreed that their blog, as identified at sampling, satisfied these criteria.

Results are presented in several tables below. Table 49 shows the frequency to which single-bloggers agree or disagree with the criteria, while co-blogger agreement or disagreement is shown in Table 50. Agreement or disagreement for all respondents is presented in Table 51.

**Table 49. Frequency of single-blogger agreement or disagreement concerning criteria for scholarship**

<b>Criterion</b>	<b>Strongly Agree Freq (%)</b>	<b>Somewhat Agree Freq (%)</b>	<b>Somewhat Disagree Freq (%)</b>	<b>Strongly Disagree Freq (%)</b>
My blog is public	66 (99%)	1 (1%)	-	-
My blog is subject to critical review	16 (24%)	28 (42%)	14 (21%)	8 (12%)
My blog is in a form that allows use and exchange by other members of the scholarly community	39 (58%)	20 (30%)	6 (9%)	2 (3%)

**Table 50. Frequency of co-blogger agreement or disagreement concerning criteria for scholarship**

<b>Criterion</b>	<b>Strongly Agree Freq (%)</b>	<b>Somewhat Agree Freq (%)</b>	<b>Somewhat Disagree Freq (%)</b>	<b>Strongly Disagree Freq (%)</b>
My blog is public	85 (99%)	1 (1%)	-	-
My blog is subject to critical review	22 (26%)	38 (44%)	12 (14%)	13 (15%)
My blog is in a form that allows use and exchange by other members of the scholarly community	54 (63%)	30 (35%)	2 (2%)	-

**Table 51. Frequency of all bloggers' agreement or disagreement concerning criteria for scholarship**

<b>Criterion</b>	<b>Strongly Agree Freq (%)</b>	<b>Somewhat Agree Freq (%)</b>	<b>Somewhat Disagree Freq (%)</b>	<b>Strongly Disagree Freq (%)</b>
My blog is public	151 (99%)	2 (1%)	-	-
My blog is subject to critical review	38 (25%)	66 (43%)	26 (17%)	21 (14%)
My blog is in a form that allows use and exchange by other members of the scholarly community	93 (61%)	50 (33%)	8 (5%)	2 (1%)

In consideration of the first criterion, all respondents agree, either strongly or somewhat, that their respective blogs are public. In consideration of the third criterion, nearly all respondents (94%) agree, strongly or somewhat, that their respective blogs are in a form that allows use and exchange by members of the scholarly community. For the second criterion – that respondents' respective blogs are subject to critical review – fewer respondents (68%) either strongly or somewhat agree. For all criteria, this criterion generated

the most disagreement: 17% somewhat disagree and 14% strongly disagree that their respective blogs are subject to critical review. Agreement is also presented by cluster, as shown in Table 52 below. More respondents in the law (79%) and BioChemPhys (73%) clusters either strongly or somewhat agree that their respective blogs are subject to critical review than those in the history (52%) and economics (62%) clusters.

Across all respondents, 101 (66%) strongly or somewhat agree that their blog satisfies all three criteria; of these respondents, 32 (21%) strongly agree for all three criteria. Over half of respondents in the law, economics and BioChemPhys clusters agree that their respective blogs satisfied these criteria. Comparing agreement between the clusters, the law cluster comprised the largest percentage of respondents (79%) to strongly or somewhat agree that their respective blogs satisfy all three criteria. The next largest cluster was the BioChemPhys cluster (70%), followed by the economics cluster (59%). Respondents in the history cluster agree the least; less than half (48%) strongly or somewhat agree.

**Table 52. Frequency of all bloggers' agreement concerning criteria for scholarship by cluster**

<b>Criteria</b>	<b>History Freq (%)</b>	<b>Economics Freq (%)</b>	<b>Law Freq (%)</b>	<b>BioChemPhys Freq (%)</b>
My blog is public	31 (100%)	39 (100%)	53 (100%)	30 (100%)
My blog is subject to critical review	16 (52%)	24 (62%)	42 (79%)	22 (73%)
My blog is in a form that allows use and exchange by other members of the scholarly community	27 (87%)	37 (95%)	50 (94%)	29 (97%)

#### 4.5.2. The Scholarly Record

Next, respondents were presented with the following definition of the scholarly record: “The Association of Research Libraries defines the scholarly record as the message communicated by scholars, which may be published or unpublished and hence informally or formally communicated. These communications may be reflective of scholarly works, knowledge, or ideas.” Respondents were asked to what extent they agreed or disagreed that their blog, as identified at sampling, was part of their respective scholarly record. Table 53

shows frequency of agreement or disagreement by single-blogger respondents and Table 54 by co-blogger respondents. Table 55 presents frequency of agreement or disagreement for all respondents.

**Table 53. Frequency of single-blogger agreement or disagreement concerning the scholarly record**

Characteristic	Strongly Agree Freq (%)	Somewhat Agree Freq (%)	Somewhat Disagree Freq (%)	Strongly Disagree Freq (%)
My blog is part of my scholarly record	28 (42%)	23 (34%)	14 (21%)	2 (3%)

**Table 54. Frequency of co-blogger agreement or disagreement concerning the scholarly record**

Characteristic	Strongly Agree Freq (%)	Somewhat Agree Freq (%)	Somewhat Disagree Freq (%)	Strongly Disagree Freq (%)
My blog is part of my scholarly record	30 (35%)	41 (48%)	10 (12%)	4 (5%)

**Table 55. Frequency of all bloggers' agreement or disagreement concerning the scholarly record**

Characteristic	Strongly Agree Freq (%)	Somewhat Agree Freq (%)	Somewhat Disagree Freq (%)	Strongly Disagree Freq (%)
My blog is part of my scholarly record	58 (38%)	64 (42%)	24 (16%)	6 (4%)

Across all respondents, eight out of ten agree, strongly or somewhat, that their blogs are part of their respective scholarly records. In terms of authorship, this represents 76% of single-blogger respondents and 83% of co-blogger respondents.

Agreement is also presented by cluster, as shown in Table 56 below. As with agreement to criteria for scholarship as presented in Table 52, the law cluster comprised the largest percentage of respondents (83%) to strongly or somewhat agree that their respective blogs are part of their scholarly record. This distinction is shared with respondents in the history cluster. While the history cluster comprised the lowest percentage of bloggers agreeing with the three criteria of scholarship, 25 respondents (83%) agreed that their blogs are part of their respective scholarly records. Further, while the BioChemPhys cluster comprised the second largest percentage of respondents in agreement with the three criteria for scholarship, respondents in this cluster represented the lowest percentage of respondents agreeing their blogs are part of their scholarly record: 22 respondents (73%) in the

BioChemPhys cluster strongly or somewhat agree that their blogs are part of their scholarly record.

**Table 56. Frequency of all bloggers' agreement concerning the scholarly record by cluster**

Characteristic	History Freq (%)	Economics Freq (%)	Law Freq (%)	BioChemPhys Freq (%)
My blog is part of my scholarly record	25 (83%)	31 (79%)	44 (83%)	22 (73%)

#### 4.5.3. Invitations to Publish, Present, Collaborate, and Serve

A series of questions were asked to assess questionnaire respondents' perceptions on whether their blog, as identified at sampling, had led to opportunities to publish, present, provide professional service, or collaborate. For publications, respondents were asked if they felt their blog had led to an invitation to publish across ten unit-types of communication. Table 57 shows the frequency of respondents replying that their blog led to an invitation to publish by unit-type. While all unit-types were represented in responses, the most frequently cited are invitations to publish non-fiction essays, articles or op-ed pieces (59%), and conference papers (63%).

**Table 57. Frequency of invitations to publish by unit-type of communication**

Unit-Type	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
Textbook (publish or co-publish)	15 (22%)	10 (12%)	<b>25 (16%)</b>
Scholarly book or monograph (edit or co-edit)	14 (21%)	14 (16%)	<b>28 (18%)</b>
Scholarly book or monograph (author or co-author)	23 (34%)	20 (23%)	<b>43 (28%)</b>
Scholarly book chapter (author or co-author)	27 (40%)	31 (36%)	<b>58 (38%)</b>
Peer-reviewed journal article (author or co-author)	26 (39%)	30 (35%)	<b>56 (37%)</b>
Comment in a peer-reviewed journal	25 (37%)	21 (24%)	<b>46 (30%)</b>
Book review in a scholarly journal	27 (40%)	32 (37%)	<b>59 (39%)</b>
Book review for a newspaper or magazine	30 (45%)	23 (27%)	<b>53 (35%)</b>
Scholarly paper for a conference	37 (55%)	60 (70%)	<b>97 (63%)</b>
Non-fiction essay, article, or op-ed piece in a newspaper or magazine	44 (66%)	46 (53%)	<b>90 (59%)</b>

Based on respondents' answers, a listing of how many of these respective unit-types s/he had ever received an invitation to publish to was derived. This does not represent total units for which respondents received an invitation to publish. It shows how many of these

particular unit-types to which respondents received invitations. Table 58 shows the frequency of total unit-types for which respondents received an invitation to publish.

**Table 58. Frequency of invitations to publish across unit-types of communication**

<b>Count of Unit-Types</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>0 units</b>	11 (16%)	16 (19%)	<b>27 (18%)</b>
<b>1 unit</b>	2 (3%)	15 (17%)	<b>17 (11%)</b>
<b>2 units</b>	10 (15%)	7 (8%)	<b>17 (11%)</b>
<b>3 units</b>	9 (13%)	10 (12%)	<b>19 (12%)</b>
<b>4 units</b>	5 (7%)	6 (7%)	<b>11 (7%)</b>
<b>5 units</b>	7 (10%)	12 (14%)	<b>19 (12%)</b>
<b>6 units</b>	10 (15%)	7 (8%)	<b>17 (11%)</b>
<b>7 units</b>	4 (6%)	8 (9%)	<b>12 (8%)</b>
<b>8 units</b>	5 (7%)	2 (2%)	<b>7 (5%)</b>
<b>9 units</b>	4 (6%)	-	<b>4 (3%)</b>
<b>10 units</b>	-	3 (3%)	<b>3 (2%)</b>

Across all respondents, over eight out of ten (82%) report that his/her blog led to at least one invitation to publish. Thirty-four percent of respondents reported invitations to publish for between one to three of these units-types, while 30% report invitations for between four to six of these unit-types. About two out of ten (18%) have received invitations for seven or more of these unit-types.

Questionnaire respondents were also asked if they felt their blog had ever led to an invitation to serve as a speaker or panelist at a conference. Over three out of four respondents replied in the affirmative: 50 single-bloggers (75%) and 67 co-bloggers (78%) felt that their blogs had led to invitations to speak. Additionally, respondents were asked if their blog had ever led to an invitation to collaborate on a research project. About half replied in the affirmative: 29 single-bloggers (43%) and 48 co-bloggers (56%) felt their blogs had led to invitations to collaborate.

Next, respondents were asked if s/he felt their blog had ever led to an invitation to perform three service related activities. These include two specific service activities (“evaluate a scholarly book or manuscript for a publisher,” and “serve as an editor or

associate editor for a scholarly journal”), and one broad category of service (“serve your scholarly or professional society in some way”). About seven out of ten of respondents (n=109) reported invitations to perform at least one of these three categories of service. Table 59 shows type of service activity and the frequency of respondents replying that their blog led to an invitation to perform service.

**Table 59. Frequency of invitations to provide professional service by authorship**

<b>Service Activity</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
Evaluate a scholarly book or manuscript for a publisher	30 (45%)	42 (49%)	<b>72 (47%)</b>
Serve as an editor or associate editor of a scholarly journal	13 (19%)	5 (6%)	<b>18 (12%)</b>
Serve your scholarly or professional society in some capacity (e.g., conference planning committee, working group)	34 (51%)	49 (57%)	<b>83 (54%)</b>

#### 4.5.4. Impact on Scholarly Life

Questionnaire respondents were asked to what extent s/he felt their blog had improved or impaired ten aspects of their respective scholarly lives. These aspects relate to research, writing, teaching, and intrinsic and extrinsic rewards. Due to variations in occupational roles, respondents were instructed to skip any aspect they felt did not apply in their own scholarly life. Table 60 shows frequency of perceptions of improvement or impairment for single-blogger respondents, and Table 61 for co-blogger respondents. Table 62 shows results for all respondents.

Across all respondents, only a minority feel their blog has impaired certain aspects of their respective scholarly lives. Research productivity was reported as somewhat impaired by 21 (14%) respondents, and opportunities for institutional promotion reported as somewhat or strongly impaired by 13 (8%) respondents. Concerning the latter aspect, a majority (66%) report no improvement or impairment for opportunities for promotion.



Only two other aspects were reported to be strongly impaired, and represent less than 1% of all respondents: quality of writing (n=1) and writing efficiency (n=2). Otherwise, the other aspects found to be somewhat impairing represent 3% or less of all respondents: overall quality of research (n=4); quality of writing (n=2); writing efficiency (n=5); quality of teaching (n=1); and ability to share pre-publication materials (n=1). For three of these aspects – research productivity, overall research quality, and quality of teaching – about four out of ten of all respondents report no improvement or impairment. Further, about three out of ten report no improvement or impairment concerning the quality of their writing and their writing efficiency.

**Table 60. Frequency of single-blogger perceptions of impairment or improvement in their scholarly life**

Aspect	Greatly Improved Freq (%)	Somewhat Improved Freq (%)	Neither Improved nor Impaired Freq (%)	Somewhat Impaired Freq (%)	Strongly Impaired Freq (%)
My research productivity	8 (12%)	24 (36%)	26 (39%)	9 (13%)	-
My creativity in examining research in new ways	22 (33%)	33 (49%)	12 (18%)	-	-
The overall quality of my research	10 (15%)	31 (46%)	24 (36%)	2 (3%)	-
The quality of my writing	20 (30%)	33 (49%)	14 (21%)	-	-
My writing efficiency	22 (33%)	27 (40%)	17 (25%)	1 (1%)	-
The quality of my teaching	15 (23%)	28 (42%)	22 (33%)	1 (2%)	-
My ability to share pre-publication material with colleagues	28 (42%)	23 (35%)	15 (23%)	-	-
My enjoyment of my work as a scholar	31 (46%)	33 (49%)	3 (4%)	-	-
My greater visibility as a scholar	41 (61%)	20 (30%)	6 (9%)	-	-
Opportunities for promotion at my institution	4 (6%)	14 (21%)	41 (62%)	4 (6%)	3 (5%)

**Table 61. Frequency of co-blogger perceptions of impairment or improvement in their scholarly life**

Aspect	Greatly Improved Freq (%)	Somewhat Improved Freq (%)	Neither Improved nor Impaired Freq (%)	Somewhat Impaired Freq (%)	Strongly Impaired Freq (%)
My research productivity	6 (7%)	34 (40%)	34 (40%)	12 (14%)	-
My creativity in examining research in new ways	20 (24%)	49 (58%)	16 (19%)	-	-
The overall quality of my research	8 (9%)	43 (50%)	33 (38%)	2 (2%)	-
The quality of my writing	20 (23%)	33 (38%)	30 (35%)	2 (2%)	1 (1%)
My writing efficiency	15 (17%)	37 (43%)	28 (33%)	4 (5%)	2 (2%)
The quality of my teaching	13 (15%)	35 (41%)	37 (43%)	1 (1%)	-
My ability to share pre-publication material with colleagues	38 (44%)	29 (34%)	18 (21%)	1 (1%)	-
My enjoyment of my work as a scholar	31 (36%)	42 (49%)	13 (15%)	-	-
My greater visibility as a scholar	53 (62%)	32 (37%)	1 (1%)	-	-
Opportunities for promotion at my institution	5 (6%)	16 (19%)	59 (69%)	4 (5%)	2 (2%)

**Table 62. Frequency of all bloggers' perceptions of impairment or improvement in their scholarly life**

Aspect	Greatly Improved Freq (%)	Somewhat Improved Freq (%)	Neither Improved nor Impaired Freq (%)	Somewhat Impaired Freq (%)	Strongly Impaired Freq (%)
My research productivity	14 (9%)	58 (38%)	60 (39%)	21 (14%)	-
My creativity in examining research in new ways	42 (28%)	82 (54%)	28 (18%)	-	-
The overall quality of my research	18 (12%)	74 (48%)	57 (37%)	4 (3%)	-
The quality of my writing	40 (26%)	66 (43%)	44 (29%)	2 (1%)	1 (1%)
My writing efficiency	37 (24%)	64 (42%)	45 (29%)	5 (3%)	2 (1%)
The quality of my teaching	28 (18%)	63 (41%)	59 (39%)	2 (1%)	-
My ability to share pre-publication material with colleagues	66 (43%)	52 (34%)	33 (22%)	1 (1%)	-
My enjoyment of my work as a scholar	62 (41%)	75 (49%)	16 (10%)	-	-
My greater visibility as a scholar	94 (61%)	52 (34%)	7 (5%)	-	-
Opportunities for promotion at my institution	9 (6%)	30 (20%)	100 (66%)	8 (5%)	5 (3%)

Concerning perceptions of improvement, Table 63 shows the frequency of improvement, either greatly or somewhat improved, for all respondents by cluster. Additionally, Table 64 shows frequency of improvement for all bloggers regardless of cluster. Nearly all respondents feel their blogs have improved their visibility as a scholar (95%) and their enjoyment of their work as a scholar (90%). Most feel their blog has improved their creativity for examining research in new ways (82%) and their ability to share

pre-publications materials with colleagues (78%). A majority feel their blog has improved four other aspects of their scholarly lives: the quality of respondents' writing (69%); respondents' writing efficiency (66%); the quality of respondents' teaching (60%); and the overall quality of respondents' research (60%). Almost half (47%) feel their blog has improved their research productivity. Only about a quarter of respondents (26%) feel their blog has improved opportunities for promotion at their respective institutions.

**Table 63. Frequency of all bloggers' perceptions of improvement in their scholarly life by cluster**

Aspect	History Freq (%)	Economics Freq (%)	Law Freq (%)	BioChemPhys Freq (%)
My research productivity	18 (58%)	14 (36%)	28 (53%)	12 (40%)
My creativity in examining research in new ways	28 (90%)	27 (71%)	44 (83%)	25 (83%)
The overall quality of my research	22 (71%)	21 (54%)	36 (68%)	13 (43%)
The quality of my writing	23 (74%)	25 (64%)	31 (58%)	27 (90%)
My writing efficiency	19 (61%)	28 (72%)	30 (57%)	24 (80%)
The quality of my teaching	15 (50%)	28 (72%)	33 (62%)	15 (50%)
My ability to share pre-publication material with colleagues	26 (84%)	31 (82%)	43 (81%)	18 (60%)
My enjoyment of my work as a scholar	30 (97%)	36 (92%)	45 (85%)	26 (87%)
My greater visibility as a scholar	28 (90%)	35 (90%)	53 (100%)	30 (100%)
Opportunities for promotion at my institution	4 (13%)	10 (26%)	16 (30%)	9 (30%)

**Table 64. Frequency of all bloggers' perceptions of improvement in their scholarly life**

Aspect	All Bloggers	
	Frequency	Percentage
My research productivity	72	47%
My creativity in examining research in new ways	124	82%
The overall quality of my research	92	60%
The quality of my writing	106	69%
My writing efficiency	101	66%
The quality of my teaching	91	60%
My ability to share pre-publication material with colleagues	118	78%
My enjoyment of my work as a scholar	137	90%
My greater visibility as a scholar	146	95%
Opportunities for promotion at my institution	39	26%

#### 4.6. Blog Publishing Behaviors and Characteristics

Respondents were asked a series of questions concerning their perceptions of their respective blogs' audience, their blog and post publishing behaviors, their micro-blogging

activities, if any, and other characteristics of their blog, including availability and host location. The results from these questions are presented in this section. The majority of results are presented across all respondents, regardless of cluster-assignment. Additionally, select questionnaire results are supplemented in this section with findings from the interview portion of the study.

#### 4.6.1. Audience

From a researcher-derived list, respondents were asked to identify audience-types they felt best represents the primary and the secondary audience for their respective blogs. Five audience-types were provided: colleagues and professional peers; family and friends; the general public; students; and the respondent, presented as “myself” in the questionnaires. Concerning secondary audience, for those only identifying with a primary audience, respondents were able to select an option reflecting this (“I only blog for my primary audience”). For those responding (N=152), eight respondents (5%) reported blogging only for their respective primary audiences.

Additionally, respondents’ were provided with a text box to enter audiences not represented by these five audience-types. These open-text responses were reviewed and, where applicable, assigned to one of the five audience-type categories. For example, in reply to primary audience, one respondent applied a qualifier to the audience-type, the *general public*, writing, “scientifically inquisitive, general public.” This response was included in the count for *general public*. In reply to secondary audience, one respondent utilized the text-box for “none of the above” to elaborate on motivation for blogging for their self, writing “myself, but in a way that I hope improves my writing ability --- not just to tell people about my thoughts or my work or myself.” This response was included in the count for “Myself.”

Those open-text responses (n=10) that did not fall within the five audience-types after review were assigned to the category, *Other*. This includes responses reflective of an “all of the above” approach, as well as identification of specific occupational roles, including journalists, judges, lawyers, government officials, and politicians, and specific-topical interests, including “foodies,” “those interested in sports business,” and “economics autodidacts.”

Frequency of primary and secondary audience types for all respondents is presented in Table 65. Two audience-types account for 93% of respondents’ perceived primary audiences: colleagues and professional peers (61%) and the general public (32%). These two audience-types also represent the most frequently identified secondary audiences, reported by 54% of all bloggers. The third most frequently cited secondary audience type is students (30%).

**Table 65. Frequency of all bloggers’ primary and secondary audience-types**

<b>Audience Type</b>	<b>Primary Audience Frequency (%)</b>	<b>Secondary Audience Frequency (%)</b>
<b>Colleagues and professional peers</b>	92 (61%)	34 (22%)
<b>General public</b>	48 (32%)	52 (34%)
<b>Students</b>	2 (1%)	45 (30%)
<b>Family and friends</b>	3 (2%)	3 (2%)
<b>Myself</b>	1 (1%)	6 (4%)
<b>Other</b>	6 (4%)	4 (3%)

Additional perspectives on audience were gathered during discussions with interviewees. While there was no specific question in the schedule on interviewees’ perceived audience, seven interviewees commented on audience when responding to other questions. These comments, for the most part, confirmed perceptions of colleagues, professional peers and the general public as primary audience-types. Concerning colleagues and professional peers, a tenured associate professor of law commented that her co-blog is intended for a very narrow audience comprised of scholars in her specific area of study.

Others viewed audience more broadly. One single-blogger assigned to the BioChemPhys cluster commented on the notion of audience extending to the general public, stating that:

Very few bloggers think of their blogging as communications between scholars working on related projects. They much more think about them as communications between scholars and the rest of the world, whether the rest of the world is Joe Blow clicking ‘next blog’ on Blogger or whether it’s a community of people interested in science.

Several interviewees’ comments reflected an intersection between contributing to the public good and pursuit of a general audience, resulting in benefits for academia. A tenured professor assigned to the history cluster noted , “... I think blogging can be a very constructive means for academicians to share their knowledge with a broader audience which strikes me as being very much in the public interest.” A single-blogger in the BioChemPhys cluster remarked on the benefit of blogging in changing misconceptions of scholars as isolated and removed from society in general:

I would hope that people within academia would be a little faster in appreciating the importance of blogs in engaging the audience which, until now, has been pretty much shut out of the ivory tower. It is good for the scientists themselves, and for the science, for the general audience to stop with the stereotypes of isolated, socially inept people who have, you know, magic powers and are concocting something in their labs.

In consideration of the predominant audience types reported by questionnaire respondents – colleagues, peers, and the general public – the networked nature of the blog form and the diversity in topical treatment may contribute to readership that is both unpredictable as well as unplanned. As commented on by a tenured professor in history and American studies:

Over time you get linked [from other sources] and you build a following that’s actually much more diverse than the following you would get just from people who read academic blogs.

Even in consideration of the public nature of blogging and opportunity for exposure to a broad audience, actual readership may be more limited. A tenured professor blogging in the area of biology commented that, “the main regular people who follow it [the blog] are the other people in my research group.” Further, extent of readership may not be perceived as important. One tenured associate law professor relayed a story of a former co-blogger with the opposite intent of her and her co-blog’s other authors, illustrating the divergent intents of blogging for status versus blogging for service:

We don’t care how many people read this blog. We don’t care how many hits we get ... He couldn’t get that we didn’t care if we had a narrow but loyal audience. ... We just want to provide our service to our academic community.

Returning to the questionnaire, only seven respondents identified themselves, represented as “myself” in the questionnaire, as a primary or secondary audience. However, two interviewees did make comments in-line with the notion of themselves as an audience type for their respective blogs, with their blogs serving as vital memory devices. One single-blogger assigned to the BioChemPhys cluster, commenting in regard to a question about preferences and practices in personal blog preservation, noted her dependency on her blog, saying: “... it works like a searchable lab notebook almost ... it is only because I can search the past of my blog that I can find these ideas.” A co-blogger in the area of law also echoed this sentiment of the blog as a memory device, remarking:

...so the 2010 [me] can remember what the 2005 [me] was thinking about ... I almost use my blog to talk to myself in the future ... and I do this with other people’s writings, with my co-bloggers, and other people that I sort of follow.

One Ph.D. in economics, however, had a somewhat opposite viewpoint of his blog as a sort of storehouse of past thinking, commenting, “very few people are out there looking at stuff I wrote in 2007; even I’m not there.”

#### 4.6.1.2. Engagement via Commenting Systems

In reflection of issues related to audience, questionnaire respondents were asked two questions. The first was to capture information on respondents' engagement with their respective blog readers through commenting systems. The second was to capture information on respondents' roles not as blog authors but as contributing blog readers, with contributions demonstrated through commenting.

Respondents were asked if s/he has ever responded to comments posted to their respective blogs. Since blogs may not support commenting directly to posts, respondents were also provided a response category to indicate if their respective blog did not support a post commenting feature. For all respondents, three (2%) did not have a post commenting system. This does not necessarily mean that communication with readers is not supported by other means, as illustrated by an open-text response from one of these respondents: "I invite email and will respond to those and post them on blog even though comments per se are closed." For all other respondents, 146 (96%) have responded to comments posted to their respective blogs, while two respondents (1%) have not.

Next, respondents were asked to estimate how often s/he has contributed comments to someone else's blog during the past year. Eight measures of frequency were provided (i.e., daily, weekly, and monthly measures), including a measure of "not at all." Table 66 shows frequency of commenting to other blogs by measure. Across all bloggers, only three respondents (2%) report not having contributed comments during the past year. Publishing comments to others' blogs is a somewhat regular activity for 40% of respondents, characterized as at least once a week. For 17 respondents (11%), commenting to others' blogs is a daily activity, while for 44 (29%) it is a weekly activity. About six out of ten



respondents (58%) publish comments less frequently: 39 (25%) publish comments between one to three times a month, and 50 (33%) a few times over the past year.

**Table 66. Frequency of publishing comments on other blogs by authorship**

<b>Measure</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>More than once a day</b>	8 (12%)	4 (5%)	<b>12 (8%)</b>
<b>About once a day</b>	3 (4%)	2 (2%)	<b>5 (3%)</b>
<b>More than once a week</b>	16 (24%)	10 (12%)	<b>26 (17%)</b>
<b>About once a week</b>	10 (15%)	8 (9%)	<b>18 (12%)</b>
<b>2 to 3 times a month</b>	6 (9%)	14 (16%)	<b>20 (13%)</b>
<b>About once a month</b>	8 (12%)	11 (13%)	<b>19 (12%)</b>
<b>A few times</b>	15 (22%)	35 (41%)	<b>50 (33%)</b>
<b>Not at all</b>	1 (1%)	2 (2%)	<b>3 (2%)</b>

While not a direct question in the interview schedule, four interviewees provided insight on commenting during the course of their respective interviews. Interviewees emphasized how communication is facilitated through a blog’s commenting system and through personal emails. One tenure-track co-blogger in the history cluster commented on the benefit of feedback:

When I do get feedback in the form of comments on the blog or private emails, it’s very valuable to me. And it really helps shape my more formal forms of scholarship in terms of journal articles and the books I’m writing.

Another interviewee, however, remarked on the disadvantage of commenting systems, pointing specifically to issues of spam and comments that are not, “terribly useful.” He commented on his particular approach in regard to comments posted to his blog and referenced problems encountered by his peers:

I very rarely respond to comments, particularly from people who post anonymously and obviously, at that point, the cost of being a jerk in an anonymous blog comment is relatively low. I’ve had some friends who’ve had some real trouble maintaining their civil discourse in their blog comment threads ... [resulting, ultimately, in friends’ blocking several ISPs].

#### 4.6.2. Availability

Questionnaire respondents were asked if all posts s/he publish to their blogs are publicly available. For co-bloggers, this concerns only posts published by the respondent, as opposed to all posts, including those of their co-authors, published to their blogs. Across all respondents, 152 (99%) make all of their posts publicly available. As one interviewee commented, in reference to an interview question about a doomsday scenario where his blog was lost, “a blog that’s on my hard drive [as back-up] is not useful; [a blog] is only useful if people can get to it.” For all but one respondent to the questionnaire, people can get to all posts published to the respondents’ respective blogs. For the sole single-blogger who does not make all of their posts publicly available, this applies to only a minority, characterized as between 1 to 19% of all the respondent’s posts.

#### 4.6.3. Host Location

Questionnaire respondents were asked to identify where their blog is currently hosted. Six researcher-derived host locations were provided: 1) a free, third party blog publishing and hosting service (e.g., Blogger, Wordpress.com); 2) a paid, third party blog hosting service (e.g., IPOWERWEB, Go Daddy); 3) a managed blog network (e.g., ScienceBlogs, Law Professor Blogs); 4) a personal server; 5) a work server; and 6) a commercial server. Respondents were also provided with two other options: “None of these” and “I do not know.” If none of the above was selected, respondents were prompted to describe the host location in his or her own words in an open text box. Since blogs may be hosted at more than one location, respondents were instructed to select the location that s/he identify as the blog’s primary host.

For the four respondents (3%) replying *None of these*, their open-text responses were reviewed and, where applicable, assigned to one of the six host location categories. Two of these responses – “university computer club server” and “English dept server” – are included in the count for the host-location category, *Work Server*. Those open-text responses (n=2) that did not fall within the six host location categories were assigned to the category, *Other*. The two host locations identified by these respondents are, respectively, a foundation Website and a colleague’s server.

Six respondents (4%), all co-bloggers, did not know where their respective blogs are currently hosted. Table 67 show frequency of primary host location for the other 147 respondents, reflecting the six host location categories and the “Other” category, as described in the preceding two paragraphs. A little over half of respondents (51%) use a free blog publishing and hosting service, while 15% subscribe to a paid blog hosting service. The third most frequently cited host location is a managed blog network, reported by 12% of respondents. The least reported of the six host locations is a personal server, with only 7 respondents (5%) selecting this as their respective blogs’ primary host location.

**Table 67. Frequency of primary host location by authorship**

Host	Single-Bloggers Freq (%)	Co-Bloggers Freq (%)	All Bloggers Freq (%)
<b>Free blog publishing and hosting service</b>	32 (48%)	43 (54%)	<b>75 (51%)</b>
<b>Paid blog hosting service</b>	10 (15%)	12 (15%)	<b>22 (15%)</b>
<b>Managed blog network</b>	7 (10%)	11 (14%)	<b>18 (12%)</b>
<b>Personal server</b>	4 (6%)	3 (4%)	<b>7 (5%)</b>
<b>Work server</b>	6 (9%)	4 (5%)	<b>10 (7%)</b>
<b>Commercial server</b>	8 (12%)	5 (6%)	<b>13 (9%)</b>
<b>Other</b>	-	2 (3%)	<b>2 (1%)</b>

Frequency of primary host location type is also presented by cluster, as shown in Table 68. A free blog publishing and hosting service is the most cited primary host location among all clusters, though frequency varies. A majority of respondents in the history (61%) and economics (65%) clusters use a free blog publishing and hosting service, while this

primary host location is identified by a little less than half of respondents in the law cluster (45%), and only 33% of respondents in the BioChemPhys cluster. No respondents assigned to the economics cluster identified a managed blog network as a primary host. While this host category was selected by eighteen bloggers in the other three clusters, more respondents in the law (n=11) and (n=5) BioChemPhys clusters selected a managed blog network as a primary host than respondents in the history cluster (n=2).

**Table 68. Frequency of primary host location by cluster**

<b>Host</b>	<b>History Freq (%)</b>	<b>Economics Freq (%)</b>	<b>Law Freq (%)</b>	<b>BioChemPhys Freq (%)</b>
<b>Free blog publishing and hosting service</b>	19 (61%)	24 (65%)	22 (45%)	10 (33%)
<b>Paid blog hosting service</b>	3 (10%)	7 (19%)	11 (22%)	1 (3%)
<b>Managed blog network</b>	2 (6%)	-	11 (22%)	5 (17%)
<b>Personal server</b>	2 (6%)	3 (8%)	-	2 (7%)
<b>Work server</b>	3 (10%)	-	2 (4%)	5 (17%)
<b>Commercial server</b>	2 (6%)	2 (5%)	3 (6%)	6 (20%)
<b>Other</b>	-	1 (3%)	-	1 (3%)

Information on which specific blog publishing and/or hosting services and applications to which respondents' subscribe was not collected in the questionnaires.

Information on frequency of blog application is presented in the next chapter.

Respondents were also asked if s/he had ever migrated their respective blogs from one hosted location to another. Ninety-seven respondents (65%) responded no. Additionally, eleven respondents (7%) – four single and seven co-bloggers – did not personally migrate their respective blogs, though their service provider, network or, in the case of co-bloggers, co-author(s) did. For the 42 respondents (28%) reporting that s/he migrated their blog from one hosted location to another, 25 (17%) have done so just once, and 17 (11%) have done so more than once.

#### 4.6.4. Blog and Post Composition and Revision Activity

This section reports results from questions designed to inform a profile of respondents' blog and post composition and revision behaviors and practices as applied to respondents' respective blogs as identified at sampling.

##### 4.6.4.1. Blog Updating

All respondents in the completed sample were characterized as active at the sampling stage, as detailed in the Methods Chapter. Results from the questionnaires, for the most part, confirmed current and active publishing to respondents' respective blogs. Respondents were asked to characterize, on average, how often s/he update their respective blog. Updating was defined in the questionnaire as, “ adding new posts, modifying existing posts, responding to comments, modifying links, and other activities where [respondents] create new content or modify content published to [their] blog.” Table 69 shows frequency of updating across seven measures of time for all bloggers.

**Table 69. Frequency of updating by authorship**

<b>Measure</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>3 or more times a day</b>	11 (16%)	5 (6%)	<b>16 (10%)</b>
<b>1 or 2 times a day</b>	14 (21%)	12 (14%)	<b>26 (17%)</b>
<b>A few times a week</b>	18 (27%)	25 (29%)	<b>43 (28%)</b>
<b>At least once a week</b>	16 (24%)	17 (20%)	<b>33 (22%)</b>
<b>Once every few weeks</b>	5 (7%)	17 (20%)	<b>22 (14%)</b>
<b>Once a month</b>	3 (4%)	6 (7%)	<b>9 (6%)</b>
<b>Less than once a month</b>	-	4 (5%)	<b>4 (3%)</b>

Across all respondents, 42 (27%) update their blogs daily, characterized as at least one update a day. A larger percentage of single-bloggers update daily (37%) when compared to co-bloggers (20%). Across all respondents, 77 (50%) update their blogs weekly, characterized as once or a few times a week, representing 34 (51%) single-bloggers and 42 (49%) co-bloggers. Only eight-single bloggers (11%) report updating once every few weeks

or once a month (11%). For co-bloggers, only four (3%) update less than once a month, with 31 (21%) update either once every few weeks or once a month.

Results on frequency of blog updating are also provided by cluster, as shown in Table 70. In this table, the seven measures reported in Table 69 are condensed to represent daily, weekly, monthly, and less than monthly updating.

**Table 70. Frequency of updating by cluster**

Measure	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Daily</b>	5 (16%)	13 (33%)	16 (30%)	8 (27%)
<b>Weekly</b>	18 (58%)	17 (44%)	27 (51%)	14 (47%)
<b>Monthly</b>	8 (26%)	8 (21%)	8 (15%)	7 (23%)
<b>Less than monthly</b>	-	1 (3%)	2 (4%)	1 (3%)

Respondents were also asked to characterize if and how their updating activity has changed over time. Presented with five statements, respondents were asked to select which one best describes how frequently s/he update their respective blog now compared to when s/he first began to publish the blog. These statements, as provided in the questionnaires, are: 1) I blog a lot more frequently; 2) I blog somewhat more; 3) I blog about the same amount; 4) I blog somewhat less; and 5) I blog a lot less. Table 71 presents the frequency of respondents' characterization of their blog updating over time.

**Table 71. Characterization of blogging over time by authorship**

Time Spent Blogging	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>A lot more</b>	7 (10%)	5 (6%)	<b>12 (8%)</b>
<b>Somewhat more</b>	12 (18%)	10 (12%)	<b>22 (14%)</b>
<b>About the same</b>	19 (28%)	35 (41%)	<b>54 (35%)</b>
<b>Somewhat less</b>	17 (25%)	22 (26%)	<b>39 (25%)</b>
<b>A lot less</b>	12 (18%)	14 (16%)	<b>26 (17%)</b>

Respondents tend to blog about the same amount of time or less in comparison to when s/he first began to publish to their respective blogs. Across all respondents, 65 (42%) blog less than when s/he first began to blog, while 54 (35%) report blogging about the same amount. Thirty-four (22%) blog either somewhat more or a lot more.

Frequency of characterization of blog updating over time is also presented by clusters, as shown in Table 72. Respondents in the economics cluster (31%) tend to blog slightly more now in comparison to respondents assigned to the law (20%), history (20%), and BioChemPhys (17%) clusters. Half of respondents assigned to the BioChemPhys cluster blog somewhat less or a lot less, followed by nearly half of respondents in the history cluster (48%) and over four in ten (42%) in the law cluster. About three in ten respondents in the economics cluster (33%) blog somewhat or a lot less.

**Table 72. Characterization of blogging over time by cluster**

<b>Time Spent Blogging</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>A lot more</b>	3 (10%)	2 (5%)	5 (9%)	2 (7%)
<b>Somewhat more</b>	3 (10%)	10 (26%)	6 (11%)	3 (10%)
<b>About the same</b>	10 (32%)	14 (36%)	20 (38%)	10 (33%)
<b>Somewhat less</b>	11 (35%)	6 (15%)	10 (19%)	12 (40%)
<b>A lot less</b>	4 (13%)	7 (18%)	12 (23%)	3 (10%)

While not an explicit question in the interview schedule, six interviewees commented on their time spent blogging, providing further insight on how their blogging has changed over time and implications for blogging into the future. A tenured history professor remarked that when he launched his blog in 2005 he was trying to, “get [it] established and so ... was putting a lot more effort into it in the beginning.” Not so much now. He commented on factors that contributed to his decision to reduce the time spent blogging:

I was feeling absolutely enslaved by it [blogging]. It was becoming a psychological problem ... it was controlling me and it's not now ... so I'm blogging more for myself, I guess, than I am for others ... I definitely cut down because it was just too much.

Another blogger, also a tenured professor assigned to the history cluster, also commented on the approach to write for himself:

The pact I made with myself was I would write something when I wanted to ... I'm not attempting to create a readership that is waiting for me every day or every week. And part of keeping the pact is not investing too much effort in thinking about its future. It is what it is for as long as it is and then, when it isn't, it can go off to the great blog heaven in the sky.

Another blogger, a tenured associate professor of law, offered similar sentiments in regard to investment of time in blogging. Speaking for herself and her co-bloggers, she commented that they, “tried hard from the beginning to say we are not living for this thing,” and they, “blog when we can and when we want to.” She further remarked on the intent to keep her co-blog going, for now, but if she and her co-authors, “made it a pressure point, we would probably decide not to do it.”

Additionally, several co-bloggers commented on the implications of co-authorship in terms of regular updating and participation. A tenure-track associate professor of economics remarked:

And as long as it's fun for all of us to do and I think as long as at least someone, at least a subset of us, are participating enough so that it doesn't go six months without an update, it's open-ended and very, very fluid and just kind of spontaneous, I think.

His response implies that, while co-blogs may be currently published to, they are not necessarily currently published to by all co-authors. A tenure-track, history co-blogger remarked on the extent of his co-authors' participation and, in response, the potential for new opportunities to emerge from the co-blog:

Of the 14 of us on the blog, probably only half regularly blog. The other half want to be a part of [the] association, not necessarily the blog, so we're thinking about splitting that apart. It's all very new to us.”

Returning to the results from the questionnaire, in other considerations of the changing content of blogs, respondents were asked if either s/he or their service provider, network or, for co-bloggers, co-authors, have ever changed the skin of their blog. “Skin” was described as a blog's “layout and design.” For the 67 single-bloggers responding, 38 (57%) have changed the skin of their respective blogs. Six (9%) report that they have not, but that their service provider or network has. For the 83 co-bloggers responding, 18 (22%) have changed the skin of their respective blog, while 28 (34%) report that s/he have not but their



service provider, network or co-author did. For all bloggers responding (n=150), sixty (40%) have not changed the skin of their blog, while 37% report personally changing the skin and 23% report that other co-producers have done so.

The 56 respondents who have personally changed the skin of their blog were asked to approximate how often s/he did so. Respondents were provided with five interval measures: 1) at least once a month; 2) every few months; 3) a couple of times a year; 4) about once a year; and 5) less than once a year. Only two of the five intervals presented to respondents were selected. Table 73 shows the frequency of how often respondents change their respective blog's skin. Most (88%) change the skin of their blogs less than once a year.

**Table 73. Frequency of changing skin of blog**

Measure	Frequency	Percentage
About once a year	7	12.5%
Less than once a year	49	87.5%

#### 4.6.4.2. Imported Content and Permissions

Questionnaire respondents were asked to respond to a question concerning posts published to their blogs during the past month. Respondents were asked if s/he had imported and posted any content originating from someone other than themselves. This content, per further explanation provided in the questionnaires, includes text, images, video and audio. Over three in four respondents (77%) replied in the affirmative, while 35 (23%) reported no importing of content originating somewhere else in the past month. Cluster representation for those replying in the affirmative, by count and percentage, are: 22 respondents (71%) in the history cluster; 34 respondents (87%) in the economics cluster; 41 respondents (77%) in the law cluster; and 21 respondents (70%) in the BioChemPhys cluster. No information on whether or not respondents had links to content originating elsewhere was collected.

However, post-based links were analyzed in the blog analysis portion of the study. Findings from this analysis are presented in the next chapter.

Returning to imported content, a follow-up question concerning permissions was presented to those 118 respondents importing and publishing external content to their posts during the previous month. Respondents were asked how often s/he check whether or not there are any rights and use statements associated with the content before publishing it, such as a copyright statement or a Creative Commons license. Five measures of frequency were provided: always, usually, about half the time, seldom, and never. Results are presented in Table 74 by authorship and Table 75 by cluster.

**Table 74. Frequency of checking permissions for imported content by authorship**

Measure	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Always</b>	11 (23%)	14 (20%)	<b>25 (21%)</b>
<b>Usually</b>	12 (25%)	15 (21%)	<b>27 (23%)</b>
<b>About half the time</b>	3 (6%)	4 (6%)	<b>7 (6%)</b>
<b>Seldom</b>	20 (42%)	13 (19%)	<b>33 (28%)</b>
<b>Never</b>	2 (4%)	24 (34%)	<b>26 (22%)</b>

**Table 75. Frequency of checking permissions for imported content by cluster**

Measure	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Always</b>	6 (27%)	4 (12%)	7 (17%)	8 (38%)
<b>Usually</b>	7 (32%)	8 (24%)	6 (15%)	6 (29%)
<b>About half the time</b>	-	1 (3%)	3 (7%)	3 (14%)
<b>Seldom</b>	7 (32%)	10 (29%)	13 (32%)	3 (14%)
<b>Never</b>	2 (9%)	11 (32%)	12 (29%)	1 (5%)

Fifty-eight respondents (50%) check if rights or use statements are associated with imported content at least half of the time or more. Twenty-six respondents (22%) report that s/he never check. Those respondents assigned to the BioChemPhys cluster check for rights or use statements more often than respondents assigned to other clusters, with seventeen (67%) responding that s/he check more than half of the time, characterized as always or usually. The second most represented cluster reporting checking more than half the time are respondents in the history cluster (59%). Respondents in the economics and law clusters

check for permissions less frequently. Sixty-one percent of respondents in both the economics and law clusters seldom or never check, compared to 40% in the history cluster and 19% in the BioChemPhys cluster.

All respondents, regardless of recent experience in posting imported content, were asked two questions concerning permissions. First, respondents were asked how important they feel it is to verify permission to post content s/he believe may be protected by copyright before publishing the content to a post in their respective blogs. For further explanation, verifying permission was described in the questionnaire to include, “multiple means, including directly contacting authors or complying with copyright or use statements applied to the content at the source location.” Table 76 shows the frequency for which those bloggers responding (N=151) feel it important or unimportant to verify permission. Table 77 presents these results by cluster.

While half of respondents (50%) check permissions for content originating elsewhere, at least half of the time, as shown in Table 74, a majority of respondents (69%) feel it is important to verify permissions, with 57 (68%) feeling it very important and 47 (31%) feeling it somewhat important. Twenty-three respondents (15%) feel it not at all important.

**Table 76. Frequency of importance or unimportance for respondents to verify permission by authorship**

Degree of Importance	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Very Important</b>	30 (45%)	27 (32%)	<b>57 (38%)</b>
<b>Somewhat important</b>	20 (30%)	27 (32%)	<b>47 (31%)</b>
<b>Somewhat unimportant</b>	10 (15%)	14 (17%)	<b>24 (16%)</b>
<b>Not at all important</b>	7 (10%)	16 (19%)	<b>23 (15%)</b>

**Table 77. Frequency of importance or unimportance for respondents to verify permission by cluster**

Degree of Importance	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Very important</b>	13 (42%)	11 (28%)	18 (35%)	15 (50%)
<b>Somewhat important</b>	8 (26%)	13 (33%)	17 (33%)	9 (30%)
<b>Somewhat unimportant</b>	6 (19%)	6 (15%)	8 (16%)	4 (13%)
<b>Not at all important</b>	4 (13%)	9 (23%)	8 (16%)	2 (7%)

Second, respondents were asked how important they feel it is for others to verify permission to share content originating in a post published by respondents to their own blogs. Table 78 shows the frequency for which those bloggers responding (N=152) feel it important or unimportant for others to verify permission. Table 79 presents these results by cluster.

While 60% feel it important for themselves to verify permission to use content originating elsewhere, fewer respondents feel it important for others to verify permission to share content originating in respondents' own posts. Across all respondents, 75 (49%) feel it important, characterized as very or somewhat, for others to verify permission. Additionally, while 15% of all respondents feel it not at all important to verify permission to use content originating elsewhere in their own posts, more respondents feel it unimportant for others to verify permission to use content originating in respondents' own posts. Across all respondents, 45 (30%) feel it is not at all important.

**Table 78. Frequency of importance or unimportance for others to verify permission by authorship**

Degree of Importance	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>Very Important</b>	16 (24%)	13 (15%)	29 (19%)
<b>Somewhat important</b>	20 (30%)	26 (31%)	46 (30%)
<b>Somewhat unimportant</b>	15 (22%)	17 (20%)	32 (21%)
<b>Not at all important</b>	16 (24%)	29 (34%)	45 (30%)

**Table 79. Frequency of importance or unimportance for others to verify permission by cluster**

Degree of Importance	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
<b>Very important</b>	6 (19%)	4 (10%)	11 (21%)	8 (27%)
<b>Somewhat important</b>	10 (32%)	8 (21%)	16 (31%)	12 (40%)
<b>Somewhat unimportant</b>	6 (19%)	12 (31%)	9 (17%)	5 (17%)
<b>Not at all important</b>	9 (29%)	15 (38%)	16 (31%)	5 (17%)

Complementing these questionnaire results are comments shared by four interviewees concerning permissions to republish content originating in their respective blogs, as well as comments on their use of Creative Commons licenses. A tenure-track assistant professor of economics and business who publishes to multiple blogs remarked on how he likes blogging

for one particular, institute-affiliated blog because, “everything they do is under a Creative Commons license,” and that he, “doesn’t really care if people repost [his] stuff in different places or borrow from it liberally or what have you.” He also characterized Creative Commons licenses as time-saving devices, remarking that it is, “kind of nice because I have to answer fewer ‘may I republish your article, may I republish your blog post’ kind of emails.”

A tenured distinguished professor of law also referenced his use of Creative Commons licenses on his blog, and their benefit for facilitating re-use. He also commented on his reason for selecting a particular license type:

As an academic, I am really thrilled that people want to re-use my work as long as they’re not making money off it. To the extent that they’re being educated or find it useful, that’s what it’s for. To the extent that they want to monetize, maybe I want a cut, so that’s the license I pick.

Two other interviewees also remarked on permissions-related issues from a financial perspective. A tenured professor in history and American studies commented on payment for use of her blog content, drawing from her own experiences, and distinguished between uses where permission (or payment) is appropriate and uses where permission is unnecessary:

My stuff gets picked up time to time by [identified various sources, including journals, magazines, professional societies] ... If any of those places plant a piece in its entirety, they ask my permission and pay me a small amount of money ... Where a lot of places [other online resources] just get their content is by linking ... so they will just quote a paragraph of what you’re doing [from her blog post] and say, ‘everybody should read this for this reason, here’s the link, go read it.’ And you don’t get any money for that and they don’t have to ask your permission to use it. But what’s great about that is it brings readers to blogs.

Another interviewee, a tenured economics professor, also raised concerns for “real financial copyright” interests, illustrated by a particular scenario and reference to standard conventions for citing and acknowledging others’ work:

We allow anyone to reproduce it [blog content] as long as they acknowledge the author and the original source ... But other situations are not so simple. Some of my colleagues write essays that appear in two places at once, or almost simultaneously on the blog and at other major media outlets. If the New York Times is involved or Forbes magazine ... then you've got sort of real financial copyright interests that need to be understood and dealt with.

#### 4.6.4.3. Post Editing

As reported earlier in Table 69 (Section 4.6.4.1.), a majority of respondents (77%) update their blog, at minimum, weekly. This updating involves adding and revising content posted to respondents' respective blogs, including publishing new posts and modifying existing posts. Concerning the latter, respondents were asked two series of questions in regard to post editing and post deletion. This section reports on the former. Respondents were asked if s/he have ever intentionally edited a blog post after it was published to their respective blog. Most respondents have; only 7 (5%) report no editing of blog posts after initial publication. Table 80 show the frequency of those respondents having ever edited a post by authorship.

**Table 80. Frequency of post editing by authorship**

Characteristic	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
Editing posts	65 (97%)	81 (94%)	<b>146 (95%)</b>

These 146 respondents were asked to estimate the percentage of posts they had intentionally edited in consideration of all posts they published in the past three months. Table 81 shows the frequency of post editing during this period across seven measures of extent. The majority of respondents (57%) edit a minority of posts, characterized as between 0 to 19% of all posts published during the three-month period. Only 21 respondents (11%) report editing 60% or more of posts published during that period.

**Table 81. Frequency of post editing in a three-month period by authorship**

Percentage of Posts	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
<b>100%</b>	1 (2%)	5 (6%)	<b>6 (4%)</b>
<b>80-99%</b>	1 (2%)	4 (5%)	<b>5 (3%)</b>
<b>60-79%</b>	2 (3%)	3 (4%)	<b>5 (3%)</b>
<b>40-59%</b>	5 (8%)	9 (11%)	<b>14 (10%)</b>
<b>20-39%</b>	19 (29%)	14 (17%)	<b>33 (23%)</b>
<b>1-19%</b>	35 (54%)	45 (56%)	<b>80 (55%)</b>
<b>0%</b>	2 (3%)	1 (1%)	<b>3 (2%)</b>

Next, these 146 respondents were presented with a listing of seven reasons for post editing. Respondents were asked to identify all of the reasons which contributed to their decisions to edit posts. Table 82 shows frequency of reasons for post editing for all respondents. The most cited reasons for editing are to make corrections (99%), such as to correct spelling, grammar or formatting, and to make revisions (90%), such as to rephrase existing content, add new content, or update links. Over half of respondents (53%) edit posts because of premature publication, and 52% edit posts to revise information no longer believed correct. While a majority of respondents edit posts because of incorrect information, fewer respondents (12%) edit posts because s/he no longer believe in the opinion expressed. Nearly half (47%) have edited posts in response to a request from someone, though fewer (18%) do so due to concerns over sensitive or revealing content.

**Table 82. Frequency of reasons for intentional editing of posts**

Reason	Frequency	Percentage
To make corrections (e.g., spelling, grammatical, or formatting errors)	145	99%
To make revisions (e.g., rework phrasing, add new information, update links)	131	90%
To revise information no longer believed correct	76	52%
To revise opinions no longer believed	17	12%
In response to a request from someone	69	47%
To remove sensitive or revealing content	26	18%
Because accidentally published before post was complete	77	53%

#### 4.6.4.4. Post Deletion

Next, respondents were asked if s/he have ever intentionally deleted a blog post after it was published to their respective blog. While intentional editing of posts is a common

activity, as reported by 95% of respondents, intentional post deletion is not. Across all respondents, 109 (71%) report never having deleted a blog post after publication. Table 83 show the frequency of those respondents who report having ever deleted a post. Cluster representation for these 44 respondents, by count and percentage, is: 11 respondents (35%) in the history cluster; 9 respondents (23%) in the economics cluster; 19 respondents (36%) in the law cluster; and 5 respondents (17%) in the BioChemPhys cluster.

**Table 83. Frequency of post deletion by authorship**

Characteristic	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
Deleting posts	22 (33%)	22 (26%)	44 (29%)

These 44 respondents were then asked to estimate the percentage of posts s/he has intentionally deleted in consideration of all posts published in the past three months. Table 84 shows the frequency of post deletion during this period for all respondents across three measures of extent. Respondents were presented with the same percentage intervals as those presented to respondents for estimating percent of posts edited, as shown in Table 81. No respondents reported deletion of 20% or more of posts in the three-month period. As reflected in extent of post editing, a majority of respondents (59%) delete a minority of posts, characterized as between 1 to 19% of all posts published during the three-month period. Eighteen respondents (41%) report deleting zero posts during that period.

**Table 84. Frequency of post deletion in a three-month period by authorship**

Percentage of Posts	Single-Bloggers Frequency (%)	Co-Bloggers Frequency (%)	All Bloggers Frequency (%)
≥ 20%	-	-	-
1-19%	12 (55%)	14 (64%)	26 (59%)
0%	10 (45%)	8 (36%)	18 (41%)

The 44 respondents who had deleted posts were presented with a listing of ten reasons for post deletion. Respondents were asked to identify all of the reasons which contributed to their decisions to delete posts. Table 85 shows frequency of reasons for post deletion for all



respondents. The most cited reasons for deletion, reported by more than half of respondents, are post duplication (57%) and post regret (52%), the latter characterized as just not wanting the post out there anymore. While only 18% of those reporting editing posts have done so to remove sensitive or revealing information, 50% deleting posts report doing so for this reason. Another more frequently cited reason for post deletion is because the post was published prematurely (45%). Reasons cited by between 30 and 34% of respondents are: 1) in response to a request from someone (34%); 2) incorrect information (32%); and 3) because of the experimental or test nature of the post (30%). Reasons least cited for blog deletion are: loss of interest in subject of the post (5%); loss of belief in opinions expressed (11%); and issues related to readability (14%).

**Table 85. Frequency of reasons for intentional deleting of posts**

<b>Reason</b>	<b>Frequency</b>	<b>Percentage</b>
Because of readability (e.g., spelling, grammar, formatting errors; bad phrasing; bad links)	6	14%
Because information no longer believed correct	14	32%
Because of opinions no longer believed	5	11%
Because of loss of interest in subject of the post	2	5%
In response to a request from someone	15	34%
To remove sensitive or revealing content	22	50%
Because accidentally published before post was complete	20	45%
Because it was a test or “experimental” post	13	30%
Because it was a duplicate post	25	57%
Because of “post regret;” I just didn’t want it out there anymore	23	52%

#### 4.6.5. Micro-blogging

To explore how blog content might be republished in other places, respondents were asked questions concerning their use of Twitter. There are other ways in which blog content may be republished elsewhere. For example, one interviewee, in a discussion on her blog network and long-term blog management, commented on cross-posting, remarking that, “anything that goes up [on her co-blog] is also simultaneously posted to [her single-blog].”

Cross-posting and other related activities, however, were not explored further in either the questionnaire or interview portions of the study.

Respondents were asked they used Twitter. For all bloggers responding (N=152), eighty-three (55%) do not use Twitter, while five (3%) read tweets but do not personally Twitter. The 64 (42%) who Tweet were asked if and, subsequently, how their respective blogs and Twitter accounts were linked. Respondents were provided with four response categories. Three of these were derived from White and Winn's (2009) survey of bloggers: 1) My blog and Twitter account are not linked; 2) My blog posts are automatically syndicated to the Twitter account and I don't ever post Tweets that are not associated with the blog.; and 3) My blog posts are automatically syndicated to my Twitter account but I also post Tweets that are not associated with the blog. Additionally, for those respondents' whose blogs and Twitter accounts are linked but not in the ways described in the response categories detailed immediately above, a fourth response category was provided: "My blog and Twitter account are linked but in some other way." Respondents selecting this category were provided with an open-text box to describe, in their own words, how their blogs and Twitter accounts are linked. Table 86 shows the frequency of the associations between respondents' blogs and Twitter accounts.

**Table 86. Frequency of association between respondents' blog and Twitter accounts**

Association	Frequency	Percentage
My blog and Twitter account are not linked	20	31%
My blog posts are automatically syndicated to my Twitter account and I don't ever post Tweets that are not associated with my blog.	13	20%
My blog posts are automatically syndicated to my Twitter account but I also post Tweets that are not associated with my blog	18	28%
My blog and Twitter account are linked but in some other way	13	20%

A majority (69%) link their blogs and their Twitter accounts in some way. Five other associations emerged from the thirteen respondents (20%) reporting that their respective blogs and Twitter accounts were linked in some other way. These include: 1) Manually

Tweeting all blog posts (n=2); 2) Manually tweeting some or select blog posts (n=3); 3) Providing feeds of all or select Tweets on their blog (n=5); 4) Providing links from their Twitter account to blog and/or vice versa (n=2), and; 5) Announcing blog posts via Twitter (n=2)

#### 4.6.5.1. Twitter and the Library of Congress

There was no explicit question in the interview schedule regarding the Library of Congress' recent announcement of their acquisition of all public tweets since Twitter's inception in March 2006 and plans to digitally archive these (Raymond, 2010). Five interviewees, however, did raise this development during the course of the interview. Additionally, in the course of the interview, the announcement was also told to several interviewees and s/he were invited to comment. Overall, interviewees were not particularly keen on the announcement, especially when considering Tweets within the broader landscape of digital content and other communications, as well as the technological requirements for effectively archiving tweets for access and use into the future.

Several qualified discussion of Twitter and the Library of Congress by commenting on the challenge in selecting digital information meriting long-term preservation and where Tweets should fall in a continuum of preservation priorities. A head librarian assigned to the history cluster prefaced his reaction to the Library of Congress' announcement in that way, remarking:

The whole [scholarly] discourse is kind of interesting and so it's hard for me to say that one set of these is less important than the others. But on the other hand it is hard for me to really quite grasp why the Library of Congress wants to archive Twitter.

An interviewee in the BioChemPhys cluster, when asked about his reaction to the announcement, replied, "Thinking about it, this is really cool, but is this a priority?" He did

not question the Library of Congress' capability for the endeavor, but did continue to question the importance of it:

I don't have a problem with it, as far as trusting them [the Library of Congress] that they will deal with things like privacy correctly. I just think the choice of where their resources are going to go, this might not be the wisest thing. I'd rather see some more important things preserved than Tweets.

Another interviewee, a tenured associate professor of law, thought the archive could be, "a treasure trove," though she cautioned that, "since only one genre and is not fully expressed as text," it could contribute to giving future users, "a very stilted view of our world." In addition to issues of context, she also points out that, since only a portion of the population Tweets, the archive will only be representative of a small population of people.

Interviewees suggested several other content types s/he thought potentially more important to preserve for a variety of reasons, including informational value and implications for future use. These included, "one or more of the blog engines that have had quite substantial content in them;" "all the scientific papers which are going to ... rot in libraries in a thousand years;" and all of the other, "stuff that is digital." A tenure-track professor in economics, when told about the announcement during the interview and asked to share his reaction, framed his response with reference to an economic model: "Wow, really? Are you kidding me? ... unless the storage operates at nearly zero costs, I just can't imagine it passing any sort of cost-benefit test."

In returning to the issues raised by interviewees in regard to what or what not to keep, a tenured associate law professor concedes that the Library of Congress, "must think that stuff is important," but he does raise concerns in regard to the technological infrastructure required to digitally archive tweets:

It raises the question of what's important. And maybe it's a function of how much storage capacity we have. Maybe we have lots of storage capacity. But that for me

raises a question of how do you read all that stuff? With technology changing so quickly, do we have ... the technology ... the software to read it all?

Another interviewee, a tenured humanities professor, shared similar sentiments in his reaction to the announcement. He commented on technological requirements and the challenges of appraising today for use tomorrow. Also, as in an earlier example from another interviewee, he addressed issues related to context:

People in their own time are not great, necessarily, at identifying those things that will, at a later date ... define the evolution of a media, define the evolution of their thought, define the evolution of their culture. So in one way, a total archiving initiative of a particular media form is a great idea. On the other hand, it's of no good to anybody unless somehow someone at some point figures out a front-end that does something more than a standard keyword search ... that gives you some fuller insight into the surrounds of the text, the nature of the communication ... Preservation only makes sense [if done in a way] so that at some point when people don't even know what the form is that you're talking about, they can understand it.

During another interview, a Ph.D. in economics had not heard the Library of Congress and Twitter announcement but did share comments on how his blog and Twitter activities are linked and his preservation preference in regard to the latter:

I use it [Twitter] to say things that are inappropriate for the blog .. the last things I'd want to preserve are my tweets ... There are some people who could feel the same way about their blog posts. They're just kind of a chance to blow off a little steam and say something that's entirely not academic in tone, versus other people [including himself] ... who see it [blogs] as an academic communication.

#### 4.7. Blog Preservation Behaviors

##### 4.7.1. Questionnaire Respondents

Questionnaire respondents were presented with a series of questions concerning their experience in deliberately saving all or parts of their blogs. Collaborative authorship raises particular concerns. As such, single- and co-blogger respondents were presented with slightly different questions, and granular results are described separately in this sub-section. This is

followed by a high-level summary of both single- and co-blogger approaches to saving all or part of their respective blogs.

#### 4.7.1.1. Single-Blogger Respondents

Single-blogger respondents were asked if s/he purposefully save a copy of their entire blog, and if so, if s/he do so via an archiving service and/or independent of an archiving service. “Entire blog” was characterized in the questionnaire to include, “content and user element features, like commenting systems, blog rolls and archive features.” Additionally, examples of these two types of approaches – an archiving service or independent of an archiving service – were also provided. For the former, this included reference to the Internet Archive, an institutional repository, and a digital archive. For the latter, this included reference to exporting a database to a server.

Careful consideration was paid to the wording of these questions, and in particular, consideration of the two approaches. Respondents were first asked whether s/he save the blog via an archiving service. Three response categories were provided: yes; no; and maybe. The complete response for the latter option is, “Maybe I do, but I’m not sure if the service qualifies as an archiving service.” Table 87 shows the frequency of single-blogger responses across these three response categories.

**Table 87. Frequency of single-bloggers saving blog via an archiving service**

Status	Frequency	Percentage
Yes	5	7%
No	53	79%
Maybe	9	13%

Respondents were then asked whether s/he save the entire blog independent of an archiving service. Table 88 shows the frequency of responses. The two approaches – via or independent of an archiving system – are not exclusive. For the twenty-seven respondents

indicating s/he save their entire blog independent of an archiving service, two also report saving their blog via an archiving service and three selected, “maybe.”

**Table 88. Frequency of single-bloggers saving blog independent of an archiving service**

Status	Frequency	Percentage
Yes	27	40%
No	40	60%

Overall, twenty-seven respondents, or 40% of all single-bloggers in the sample, reported yes or maybe to saving their entire blog either via an archiving service, independent of an archiving service, or through both approaches. Those replying yes or maybe to saving their respective blogs via an archiving service were asked to provide the name of the service, or services, in an open-text response. Additionally, respondents reporting saving an entire blog independent of an archiving service were asked to identify their approach. These respondents were presented with a list of four strategies from which to choose: an export tool from the blog publishing and service provider; via a syndication service, such as RSS feeds; mirroring to another Website; and printing to paper. Respondents were also presented the option to describe other strategies in an open-text box. Further, since respondents may employ two or more of these strategies, they were instructed to select all that apply.

Twenty-two unique respondents, or 81% of all applicable respondents, replied. From a review of their responses, six categories of services or strategies were derived from researcher-supplied categories and respondents’ open-text responses: the Internet archive; export tools; syndication services; mirroring to another site; personal back-ups; and institutional back-ups. These approaches may be further characterized as either programmatic or personal. Personal denotes actions taken by the respective blogger. This may or may not include free- or fee-based services, such as subscription to a cloud hosting service. Programmatic denotes action by an organization, such as the Internet Archive or an

institution's back-up utility, rather than deliberate, personal action by the respective blogger.

Table 89 lists each strategy or service and the frequency of adoption by these 22 respondents.

**Table 89. Frequency of single-blogger approaches for saving entire blog**

Strategy or Service	Frequency	Percentage
Internet Archive	2	9%
Export tool from blog publishing and service provider	20	91%
Via a syndication service (e.g., RSS feed)	5	23%
Mirrored to another Website (e.g., to another blog)	1	5%
Personal back-up (e.g., database dump; manual backup to disk, server, workstation, or other location)	13	59%
Institutional back-up	3	14%

Programmatic approaches are few: two bloggers use the Internet Archive, and three report back-up of their blogs via their respective institution's established back-up services. Twenty respondents (91%) use an export tool provided by their respective blog's publishing and service provider, with some respondents specifically referencing export tools from WordPress, Blogger, and Moveable Type. Five respondents (23%) referenced saving their blogs via syndications services, like RSS feeds, while one reported their blog as mirrored to another Website. Over half of respondents (59%) reported a variety of personal back-up strategies. These included manual database dumps, with several noting specific locations of the back-ups, including Dropbox, local disks and servers. Several also referenced specific hosting and back-up services and other applications, including Blogger Backup (<http://bloggerbackup.codeplex.com>), Joyent Public Cloud Hosting (<http://www.joyent.com>), Website Source, and an unspecified Web crawler. One respondent described their strategy as "self-archiving," but did not specify what action(s) this entailed.

The 32 respondents not reporting saving their entire blogs were asked if s/he ever purposefully save specific components of their blog for their personal use. These specific components included: posts with comments; posts and comments, but separately; posts only; and comments only. Respondents were also able to select, "No, I save none of these



components.” For the 30 single-bloggers responding to this question, only seven (14%) report saving any of these components. Table 90 shows the frequency of respondents’ approximation of components saved by component type.

**Table 90. Frequency of blog components purposefully saved by single-bloggers**

Percentage of Components	Posts Only Frequency	Posts and Comments Frequency	All Components Frequency (%)
100%	1	1	2 (7%)
80-99%	-	-	-
60-79%	2	-	2 (7%)
40-59%	1	-	1 (3%)
20-39%	-	-	
1-19%	1	1	2 (7%)
0%	-	-	23 (77%)

These seven respondents were asked to identify or describe the strategies that s/he take in saving these components. A listing of five categories of strategies was provided: an export tool from the blog publishing and service provider; via a syndication service, such as RSS feeds; mirroring to another website; printing to paper; and deposit to an institutional repository or digital archive. Further, a fifth category was provided for respondents to enter any other strategy(ies) via an open-text box. Additionally, since respondents may employ two or more of these strategies, s/he were instructed to select all that apply.

Five categories of approaches, both in consideration of researcher-supplied categories and respondents’ open-text responses, were identified. These are: 1) Export tool (n=1); 2) Via syndication services (n=2); 3) Printed to paper (n=2); 4) Saved as a MSWord or other document file type (n=5), and; 5) Saved email notifications (n=1).

#### 4.7.1.2. Co-Blogger Respondents

As described in reference to single-bloggers in the preceding sub-section, co-bloggers were also asked if s/he purposefully save a copy of their entire blog, and if so, if s/he do so via an archiving service and/or independent of an archiving service.

Respondents were first asked whether they or a co-author save the blog via an archiving service. Due to respondents' blogs being co-authored, six response categories were provided: 1) Yes, I do; 2) Yes, but not me; a co-author does; 3) Maybe I do, but I'm not sure if the service qualifies as an 'archiving service;' 4) Maybe my co-author does, but I'm not sure if it qualifies as an 'archiving service;' 5) No, not me, but I'm not sure if a co-author does; and 6) No, neither my co-author nor I do.

For the 84 co-bloggers responding, Table 91 shows the frequency of co-bloggers' responses across these six response categories. Only 11 (13%) report known archiving, either by themselves or a co-author, via an archiving service. Additionally, 22 (26%) selected "maybe," unsure if the service qualifies as an archiving service.

**Table 91. Frequency of co-bloggers saving blog via an archiving service**

Status	Frequency	Percentage
Yes, I do	5	6%
Yes, but not me; a co-author does	6	7%
Maybe I do, but I am not sure if the service qualifies as an "archiving service"	9	11%
Maybe my co-author does, but I'm not sure if it qualifies as an "archiving service"	13	15%
No, not me, but I'm not sure if a co-author does	29	35%
No, neither my co-author nor I do	22	26%

Additionally, the 33 co-bloggers responding either yes or maybe were presented with two follow-up questions concerning permissions. First, respondents were asked how important they felt it is to get permission from their co-authors to save the entire blog via an archiving service. For the thirteen responding, only three (23%) found it important, characterized as either "somewhat" or "very important." The other ten respondents (77%) found it unimportant, with three selecting "somewhat unimportant" and seven selecting "not at all important."

Second, respondents were asked if they have the permission of their co-authors to save the entire blog via an archiving service. For the twelve responding, six (50%) replied in

the affirmative (yes), and five (42%) replied no. The remaining respondent replied that issues of permission are handled by the foundation hosting the blog.

Next, respondents were asked whether s/he or a co-author save the entire blog independent of an archiving service. Due to respondents' blogs being collaboratively authored, five categories were provided; these categories are shown in Table 92. One category, "No, only an archiving service is used," was applied to facilitate skip-logic and navigation through the remaining questions in this particular section of the survey. Those respondents reporting saving an entire blog, regardless of approach, were not asked to reply to later questions on saving blogs by component (e.g., posts, comments). This is detailed in a later part of this sub-section. Ultimately, the inclusion of this response category proved unsuccessful as nine of the 14 respondents selecting this option had not identified as personally using an archiving service or having knowledge of a co-author's use of an archiving service, as presented in a preceding question. These respondents' replies were not included in analysis of responses to this particular question. Removing for these respondents, overall, 73 respondents, or 85% of all co-bloggers in the completed sample, responded.

Table 92 shows the frequency of co-blogger responses across the five response categories. Only twelve bloggers (16%) report that their entire blog is saved independent of an archiving service, either directly by the respondent (8%) or indirectly by a co-author (8%). Further, three of these respondents also report personally saving their respective blogs via an archiving service. In consideration of both approaches, eight unique respondents, or 9% of the completed sample, report personally and directly saving their entire blog.

**Table 92. Frequency of co-bloggers saving blog independent of an archiving service**

Status	Frequency	Percentage
Yes, I do	6	8%
Yes, but not me; a co-author does	6	8%
No, not me but I'm not sure if a co-author does	38	52%
No, neither my co-author nor I do	18	25%
No, only an archiving service is used	5	7%

The six co-bloggers reporting that they directly save their entire blog independent of an archiving service were asked how important they felt it is to get permission from their co-authors to save the entire blog in this fashion. Only one felt it “somewhat important.” The remaining four (67%) felt it unimportant, with one finding it “somewhat unimportant” and four feeling it “not at all important.” Next, these respondents were asked if they had the permission of their co-authors to save the entire blog independent of an archiving service. Four (67%) replied yes. For the other two respondents, one replied that issues of permission are handled by the foundation hosting the blog and the other wrote, “I think it's assumed I'll back-up the blog.”

As with the questions posed to single-bloggers, co-bloggers responding “yes” or “maybe” concerning use of an archiving service were asked to provide the name of the service, or services, in an open-text response. Additionally, those co-bloggers reporting personally saving an entire blog independent of an archiving service were asked to identify the strategy employed from four listed strategies (export tool from the blog publishing and service provider; via a syndication service, such as RSS feeds; mirroring to another website; and printing to paper), and/or describe the strategy(ies) via an open-text box. Additionally since respondents may employ two or more strategies, they were instructed to select all that apply. Only eight respondents, overall, identified and/or described their strategies for saving blogs either through an archiving service or independent of an archiving service.

From a review of these eight responses, nine categories of strategies, both in consideration of the researcher-supplied categories and respondents' open-text responses, were derived. As with the strategies reported by single-bloggers, these may be further characterized as either programmatic or personal. Again, personal denotes actions taken by the respective blogger. This may or may not include free- or fee-based services, such as a subscription, third-party back-up service. Programmatic denotes action by an organization, such as the Library of Congress. Table 93 lists each strategy or service and the frequency of adoption by these eight respondents.

**Table 93. Frequency of co-blogger approaches for saving entire blog**

Strategy or Service	Frequency
Export tool from blog publishing and service provider	2
Via a syndication service (e.g., RSS feed)	1
Printed to paper	1
Personal back-up (e.g., database dump; manual backup to disk, server, workstation, or other location)	4
Library of Congress Legal Blawgs Web Archive	1
Lexis Nexis	2
Email	1
Don't know	2
Other	1

Programmatic approaches are less frequent: three bloggers (38%) report blog archiving by either Lexis Nexis or the Library of Congress. Fewer co-bloggers (n=2) than single-bloggers (n=20) use an export tool provided by their respective blog's publishing and service provider. Four (50%) employ personal back-up strategies, with one referencing use of a specific back-up service, Backupify (<http://www.backupify.com/>). One respondent described his/her approach as "recursive wget;" since unsure what this actually entails, this response was categorized as "Other" in Table 93. Two respondents reported not knowing what strategy was applied in saving their respective blogs.

Those co-bloggers who did not respond as directly responsible when asked if they saved their entire blog via or independent of an archiving service were asked if they ever

purposefully saved specific components of their blog for their personal use. These specific components included: their own posts with comments; their own posts and comments, but separately; their posts only; and comments to their posts only. Respondents were also able to select, “No, I save none of these components.” Sixty-nine co-bloggers were eligible to respond. For the 62 co-bloggers responding, Table 94 shows the frequency of component types, if any, saved, and Table 95 shows the frequency of respondents’ approximation of components saved by component type. A majority of respondents (64%) save none of these components. For the 18 respondents (29%) saving components, the most cited component type saved, reported by 16% of respondents, are “my posts only.” No one reported exclusively saving comments only.

**Table 94. Frequency of component types purposefully saved by co-bloggers**

Component	Frequency	Percentage
My posts only	11	16%
My posts with comments	6	9%
My posts and comments, but separately	1	1%
No components	44	64%

*Note.* Percentage is based on the number of co-bloggers (N=69) eligible to respond and presented with the question. Eight co-bloggers were not presented with this question as they had reported directly saving their blogs in their entirety. Further, nine co-bloggers eligible to respond, since s/he did not report directly saving of the blog in its entirety, were not presented with this question, as described in section 4.7.2.1. due to selection of a response category which triggered a skip-logic effect.

**Table 95. Frequency of extent of component types purposefully saved by co-bloggers**

Percentage of Components	Posts Only Frequency	Posts with Comments Frequency	Posts and Comments Frequency	All Components Frequency (%)
100%	5	3	-	8 (13%)
80-99%	2	1	-	3 (5%)
60-79%	2	1	-	3 (5%)
40-59%	1	-	-	1 (2%)
20-39%	1	1	1	3 (5%)
1-19%	-	-	-	-

As with applicable single-bloggers, the 18 co-bloggers (29%) reporting saving of blog components were asked to identify the strategy employed from five researcher-supplied strategies and/or describe the strategy(ies) via an open-text box. Additionally, since

respondents may employ two or more of these strategies, they were instructed to select all that apply.

Seven categories of approaches, both in consideration of researcher-supplied categories and respondents' open-text responses, were derived. These are presented in Table 96 by frequency. The most cited strategy, selected by eight respondents, is to save content as an MSWord or other document or text file type. One respondent described this as, "ctrl+c, ctrl+v." An "Other" category was also devised to reflect strategies not represented by the six approaches listed in Table 96. The strategies assigned to the "Other" category, in respondents' own words, are: "bookmarked pages;" "print to file on computer;" and "automatic permalink."

**Table 96. Frequency of co-blogger approaches for saving blog components**

Strategy or Service	Frequency
Export tool from blog publishing and service provider	2
Via a syndication service (e.g., RSS feed)	1
Mirrored to another Website	2
Printed to paper	1
Personal back-up (e.g., to hard drive)	2
Saved as MSWord or other document or text file type	8
Other	3

#### 4.7.1.3. All Respondents

Across both single- and co-blogger respondents, 70 unique respondents (46%) report affirmatively that s/he directly and purposefully save their entire blog via an archiving service or independent of an archiving service. Additionally, 23 unique respondents (16%) report saving parts of their respective blogs. Ultimately, 93 respondents (61%) take some action to save all or parts of their blogs. Table 97 shows the frequency of the extent of blog saving.

**Table 97. Frequency of all bloggers purposefully saving blog**

Extent	Frequency	Percentage
Entire Blog	70	46%
Blog component (e.g., posts only, posts with comments; post and comments)	25	16%

Overall, 54 single- and co-blogger respondents identified or described a variety of strategies for saving his/her respective blogs, in whole or in part. Table 98 presents combined frequency of all strategies reported. The most widely cited strategy, reported by 25 respondents (46%), is use of an export tool from respondents' blog publishing and hosting service. The second most cited is personal back-up strategies (35%). As mentioned in earlier sub-sections in regard to strategies, the listing in Table 98 is not mutually exhaustive. Respondents may report use of two or more of these approaches. Further, strategies described or identified may reveal both a tool for saving (i.e., blog export tool) as well as storage location (i.e., disk, server, and work station for personal back-ups).

**Table 98. Frequency of all bloggers' strategies for saving blog in whole or in part**

Approach	Frequency	Percentage
Internet Archive	2	4%
Library of Congress Legal Blawgs Web Archive	1	2%
Lexis Nexis	2	4%
Export tool from blog publishing and service provider	25	46%
Via a syndication service (e.g., RSS feed)	9	17%
Mirrored to another Website (e.g., to another blog)	3	6%
Personal back-up (e.g., database dump; manual backup to disk, server, workstation, or other location)	19	35%
Institutional back-up	3	6%
Printed to paper	4	7%
Saved as MSWord or other document or text file type	13	24%
Email	2	4%
Other	4	7%
Don't know	2	4%

#### 4.7.2. Interviewees

To gather additional perceptions on digital preservation, interviewees were asked two specific questions. First, putting aside blogs, interviewees were asked if they think about issues of digital preservation in their everyday research and writing activities. Many interviewees do think about it and demonstrate a sophisticated understanding of several key, underlying issues, even if not by the phrase, "digital preservation." As remarked on by a blogger in the BioChemPhys cluster: "I might not have the term, digital preservation, as



those words in my head [but] I am constantly thinking about the durability of what is posted online, be it scientific journals or blogging or anything else.”

While nearly all interviewees responded that they think about digital preservation, there were variations to the degree it is considered in their everyday research and writing activities, as well as variation as to what aspects are considered. About half of the interviewees remarked on the importance of digital preservation through referencing particular preservation approaches as well as challenges. The latter included issues related to link rot, forward portability, access, context, cost, and authenticity, among others. As illustration of the latter, a Ph.D. in economics remarked that, “digital archives are absolutely, probably, one of the most critical components of academic life.” In reference to the former, an interviewee assigned to the humanities cluster, and the only study participant working in a library, described digital presentation as, “a great concern,” and continued:

... particularly as more and more materials and more and more vaguely eccentric materials, you might say, or one-off kinds of things, become more and more important as scholars’ resources without any preservation schema in place.

A tenured associate professor of law commented that issues of digital preservation are particularly important to her field:

We are particularly concerned about preservation of the evidence – the thing – and we are concerned about what happens if the thing – the evidence – gets transformed. What happens if you transform the original and it disappears, or is transformed into something else? What happens if the digital thing for some reason is corrupted? This is particularly a problem, for example, if email vanishes. This is one of the issues we are confronted with.

A tenured humanities professor also felt digital preservation important, both personally as well as among members of his field:

I think a lot about preservation in general. I think historians have to because we are not only conscious of our reliance on archives ... most of us are pretty conscious of the series of strange accidents that tend to create archives.

One tenured professor in the BioChemPhys cluster responded that she is aware of the problem but does not actually do much in response:

[The] problem has been brought to my awareness because of occasional conversations with librarians or reading blog posts from science librarians who are concerned with this ... I recognize that it is all too obvious that the ... electronic media in which things have been stored have generally turned out to be really, really bad in terms of preservation ... whereas paper has held up really well ... Does that influence my action? No ... it doesn't change what I do. I don't make an effort to make sure that my papers are in an ... institutional repository. I have no idea whether the journals that I publish in have good policies to ensure the articles will continue to be available even if the journal goes bust.

When responding, several interviewees described their own digital preservation approaches or preferences. One tenured law professor provided a detailed account:

I have a rigid back-up system of everything. I keep one set at home, one set at the office in case one place burns down or one hard drive crashes. I think digital preservation is extremely important ... every other week I back up everything on the other place's drive ... When I am working on something currently, I maintain a separate email account and a separate provider that I send drafts to just to sit there, never to be retrieved ... I don't ever want to lose anything.

He also supplemented his description by commenting on one of his reasons for his strong sense of personal responsibility and his extensive preservation approach. When it comes to back-ups and other computing issues, he noted, "Personally, I trust no one."

One tenured law professor described his own experience for maintaining and preserving his teaching materials:

I am not trained in any of this, and I do not insist on library quality copies and I don't have the 'gold-standard' of what archiving is and I store it on my regular servers, but am very interested in keeping this stuff around.

His remarks reflect a good-enough approach to preservation since his actions may not reflect, in his words, the gold-standard. For another interviewee, back-ups are only a step, not a solution: "It's possible to do all the back-up in the world and if someone comes along and changes the URL scheme of their site, you're dead." He cautioned that people will get the

error message and conclude, “Hey it’s gone,” and not go any further to attempt to access the information.

When discussing digital preservation in her own academic life, a tenured professor in chemistry commented on the need to be proactive, not just in back-ups but also in regard to formal migration:

Every time formats look like they’re going to change, I’m sort of moving things and now I sort of have things floating in the cloud in the hopes that I can pull them back if I have them in four or five places floating out there.

Several interviewees commented on how the Web facilitates a sort of “lots of copies keeps stuff safe” approach, as well as a “preservation by use” approach. One tenured economics professor remarked: “I am more optimistic about exposure via the Internet than to be moldering in some library.”

A tenured professor in history and American studies also spoke of the assumed persistence of content on the Web. She extended this consideration to concerns over control of content published and made available online via proprietary services, identifying this as an outstanding area in need of investigation:

Of course, there’s this whole thing about how nothing ever disappears on the Internet ... don’t put anything up on Facebook that you don’t want everybody to have forever. On the other hand, how you make that kind of data searchable is an entirely different question. And how you make it searchable if it actually belongs to a private company ... My guess is that my blog belongs to me and it belongs to Google because I blog on Blogger, and that there’s a joint ownership of it. Now if Google wanted to take my blog and publish it as a book, I think I’d have a great intellectual property lawsuit, and I’d probably win. But, if they said they were archiving the blog, I’m not sure I would have a leg to stand on if I said, ‘Well, I don’t want my blog to be archived with you, I want it to go to [a specific library]’ ... but the question is if they don’t want to archive it but they still own it, how can I then call the [library] and say, ‘Hey, I want you guys to be archiving my blog’ ... I think that’s the kind of legal territory that no one’s explored yet.

Also, in regard to digital preservation in general, interviewees extended this “preservation by use” notion to “preservation by publication.” Several pointed out that it is safe to assume that content formally published via books, journals, and other formal channels will be available into the future. As one tenured history professor phrased it, “when something is already published, the threat of losing your work declines.”

Next, interviewees were asked if they had considered issues of digital preservation in regard to their respective blogs prior to taking the questionnaire. Over half responded that they either had not or had only cursorily considered it. As shared by a tenure-track associate professor of law, one of the reasons he had not thought about preservation of his blog, in part, was:

... because it is electronic and it is everywhere. I take what I write and cut and paste from my blog what I write and save it as a Word document but more for easy access if I want to look for something I wrote about; also as a quantitative measure, to demonstrate extent of writing.

Another interviewee, a Ph.D. in chemistry, commented that he has not thought so much about digital preservation of his blog, though he does occasionally “download and archive” it. Beyond that, in consideration of both his blog and all things “digital:”

I sort of feel like the future can take care of itself up to a point because we’re no longer so dependent on any particular physical medium and I would think that as long as you stick with some kind of format that is low-level ... not too wild a format, I would presume the future would find a way to read it.”

Another interviewee, a Ph.D. in economics, raised the point, however, that proprietary providers may throw a proverbial wrench into the plan if services are discontinued or providers’ close up shop.

I know that there is a kind of informal mechanism for maintaining content on the Web ... it’s gonna be something that whatever Google decides to keep around or Blogger ... if that company goes out of business, obviously a whole bunch of blogs might disappear.”

For another interviewee, a tenure-track assistant professor of business and economics, while he may not think deliberately about preservation, he does think about the future when blogging: “When I write the blog, I do kind of have an eye to the future.” He commented that he’s noticed things he has written in the past appear somewhere else and that, “Google knows everything and Google will know everything forever.” Another interviewee, who views his blog as ephemera, acknowledged that things online tend to “last,” and remarked that, “the only consideration I take to the future is the consideration I take to everything that goes online and that is, if you wouldn’t want your grandmother to read it on the front pages, don’t put it online.”

For the other interviewees who report thinking about blog preservation, it has not necessarily led to preservation action. Several alluded to reliance on service providers. One interviewee commented that, while he thinks a little about digital preservation of his blog, he does not do much about it: “I put a lot of effort into it. I sort of hope that the people maintaining the site are doing a professional back-up job.” He also described his reasons for selecting his particular provider:

Once upon a time I wrote my own blog software and maintained my own blog on my own Website ... the big reason I switched to where I switched over ... is that they would take over the maintenance of the software and that would include preserving of the blog. The whole reasons I switched to them is so I wouldn’t have to think about it.

An interviewee assigned to the humanities cluster, while commenting that he very much has considered digital preservation of his blog, reports not having done much about it:

I’m ashamed to say I don’t have a preservation solution for my own blogs ... because I am using free software and not paying as much attention as I should ... All my hard drives are backed up on a day-to-day [basis], occasionally completely, institutionally. But that doesn’t mean the stuff I am blogging, which is out somewhere in the cloud, is retrievable through that back-up and that concerns me.

Another interviewee referenced her own current scholarship on blogging as contributing to her thinking “a lot” about it since it has raised many concerns: “How those blogs will be preserved, under what conditions, who does the collections, what are the problems associated with having too much evidence?”

One interviewee expressed concern for the preservation of his blog but, due to his affiliation with a network, he feels there is not much that he can currently do about it:

That’s a problem [blog preservation]. A big problem because I don’t control it; it’s part of a network. I am scared to death. I don’t know what they are doing about that. That is a huge concern of mine because if it crashed I don’t have any copies.

In addition to thinking about it, several interviewees also described taking actions to preserve their respective blogs. One strategy is through regular back-ups of their blogs, reflective of results from the questionnaire portion of this study. One tenured biology professor described a suite of blog preservation back-up strategies as well as the occasional “print-to-paper” approach:

I don’t want to lose all that writing so I regularly back up the blog using the software that the blogging people provide and then from there, again, it’s the many copies approach. It’s sitting there in my Google Docs account, it’s on a computer at home, it’s on a computer at school. It sits there. And periodically I will simply just print out a run of text. I mean, I have printed out everything I wrote on the blog from the time I began it until this summer ....and just stuff it in a file as a hard copy.

In describing the print-to-paper approach, she commented on the stability of the carrier: “At least to me, preservation has to be a way that you can access it as well that is convenient, so paper has that advantage too.”

For a couple of other interviewees, back-ups are more irregular. This is illustrated by one interviewee, a tenured professor in the BioChemPhys cluster:

When I first started blogging I occasionally downloaded the whole content of the blog and pasted it into another file and then I got lazy and assumed that [service provider]

would look after it and I have not done anything since then. So I'm sort am aware of the possibility that the blog may not live forever.

A lack of skills and know-how was referenced by three interviewees as an impediment to personal blog preservation. A tenured associate professor of law remarked that she is not very "tech savvy," a sentiment also elaborated on by another interviewee in the law cluster: "I'm not technologically adept enough to think about the constraints on it being there forever. I suppose it has something to do with [service provider's] capacity ..."

Another blogger commented that he thinks about digital preservation of his blog, but is not sure how to best go about it, particularly when considering the dynamic and volatile nature of blogs as they move from one hosted location and application to another.

I don't really know the procedure. How would one go about it? I would need to learn more about it. Especially my blog has moved [references three different host locations] and it will move again within a month or two to another platform ... every move scrambles the formatting, some videos get lost. So, it deteriorates every time it moves from, you know, BlogSpot to Moveable Type to Wordpress to who knows what is next. Something gets messed up ... If there is a way to preserve it in a form and formatting that I like ... I would love that but I don't know exactly how to go about it. Just trust [service provider], that they're going to keep it forever? I don't know.

For two interviewees, they do not need to think about it as someone else is already doing it for them. Both participate in a programmatic preservation program. One, a tenured law professor, commented that the, "Library of Congress was good enough to tell me they were going to do it for me. That took that off the table. And to be honest, I had not worried about it much before that."

#### 4.7.2.1. Networks

Interviewees aligned with a network of academic blogs were asked if any plans about preservation and/or long-term management of their blog had ever been provided to them by the network's managers or sponsors. For the nine interviewees asked this question, both

current and former network bloggers, eight had no recollection of such information being provided. For one interviewee, this left her “a little dismayed.”

There’s no communication, it’s not two-way at all. It’s all one-way, top-down. They sort of give us the decisions they come to but they never shared any contracts that they’ve entered into; for example, with the advertisers. And that dismays me very much. I mean, I think if I had to do it over again, I would not enter into the network.

One commented that he hadn’t received any information, but that he also hadn’t asked for it either. Another interviewee anticipated what he would be told if a question about long-term management and preservation was posed to his network’s manager:

What I need to do ... I need to talk with them [network management] and find out what’s being done and I think he will just say it is up to the provider [blog application provider]. Right now it is all there. I can go back to my first entry and find it.

A former network blogger recalled that there was a little bit of chatter between bloggers and lower management but limited action and resolution:

... [laughing] it pretty much ended, like, you know, as long as the [publishing group] exists and can pay for the servers, the blogs will be there, including old stuff ... even my archives [that are] there will survive there as long as the [publishing group] exists, which may be another year [more laughter] ...but no serious discussion was there.

He also shared a specific concern about network affiliation and intellectual property rights, and sees this as an outstanding area of investigation:

It is very interesting in the blogging world with people who join networks and sign contracts. They always have questions: ‘who really owns my text?’ [As these new networks are forming] ... that will be an issue to definitely watch because something is going to happen where a blogger and the network are going to get in a row about it and will have a legal case and a precedent ... it’s going to happen. I don’t know who, where, when ... it’s very, very tricky legal language that those contracts have in them for bloggers and networks or bloggers writing for a media company.

#### 4.7.2.2. Co-Bloggers

Additionally, ten interviewees who published to co-blogs were asked if any plans about preservation and/or long-term management of their blog had ever been discussed



among the respective blogs' authors. Seven interviewees reported that there was either no discussion at all or, if there was any discussion, they had not been privy to it.

Three interviewees reported some discussion of long-term blog management with their respective co-bloggers, though these discussions had not been very extensive. For example, one commented that he and his co-bloggers agreed they do not, "live for the blog," but that they do have a "sense of obligation" to keep it going for now.

An interviewee assigned to the humanities cluster was involved in discussions with his co-bloggers, though these discussions have not led to any action:

We are aware there are cloud-based methods for doing this and other, perhaps more secure ways to do it, locally, but we haven't done it. I think it is quite common among [his particular blogging community] that I know that people are ignoring this.

4.8. Blog Preservation Preferences

Reported in this section are results from a series of questions presented to questionnaire respondents concerning their perceptions on preservation for their respective blogs. Also reported here are findings from questions posed to interviewees to gather more information in regard to perceptions on responsibility and capability for preservation, as well as questions to gather opinions on prioritizing blogs for preservation when considered within a broader landscape of scholarly communication units.

#### 4.8.1. Extent of Preservation for Private and Public Access and Use

In the questionnaire sections on blog preservation preferences, single- and co-bloggers were provided with a definition of digital preservation prior to the first question.

The definition, adapted from the OAIS Reference Model (2002), was provided as follows:

The act of maintaining information in a correct and independently understandable form for a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community. This period extends into the indefinite future.

Questionnaire respondents were asked to what extent they agreed or disagreed with a series of statements concerning their blog and preference for preservation. The statements were designed to capture perceptions in consideration of both public and private access as well as indefinite and short-term periods of time. In the questionnaire, “short-term” was characterized in the applicable responses categories by example dates; that is: “i.e., 2020, 2035.” Single-bloggers were provided with four statements and co-bloggers with eight. The reason co-bloggers were presented with additional statements was to capture perceptions in consideration of respondents’ respective content posted to their blog, as well as all content – both their own and their co-authors – posted to the blog.

Table 99 shows frequency of perceptions of agreement or disagreement for single-blogger respondents across four statements. Over eight out of ten single-bloggers agree, either strongly or somewhat, that their respective blogs should be preserved for personal access and use (86%) and public access and use (83%) into the indefinite future. They were less inclined to agree that their blogs should be preserved only into the short term future. Only 11 (18%) strongly or somewhat agreed their blogs should be preserved for their personal access and use into the short-term future, and 13 (21%) agreed in regard to public access and use. A majority disagreed, either strongly or somewhat, to preservation of their respective blog into the short-term future for both personal access and use (62%) and public access and use (57%).

**Table 99. Frequency of single-blogger agreement or disagreement for blog preservation**

Statement: “My blog should be ...”	Strongly Agree Freq (%)	Somewhat Agree Freq (%)	Neither Agree nor Disagree Freq (%)	Somewhat Disagree Freq (%)	Strongly Disagree Freq (%)
Preserved for my <b>personal</b> access and use into the <b>indefinite</b> future	37 (59%)	17 (27%)	6 (10%)	2 (3%)	1 (2%)
Preserved for <b>public</b> access and use into the <b>indefinite</b> future	39 (58%)	17 (25%)	7 (10%)	4 (6%)	-
Preserved for my <b>personal</b> access and use only into the <b>short-term</b> future	5 (8%)	6 (10%)	12 (20%)	16 (26%)	22 (36%)
Preserved for <b>public</b> access and use only into the <b>short-term</b> future	5 (8%)	8 (13%)	13 (21%)	13 (21%)	22 (36%)

Co-blogger respondents were asked to comment on their preferences for preservation in consideration of access level (public and private) and time (indefinite and short-term future) for two levels of their respective blogs’ content: respondents own content, worded as “my content” in the questionnaire, and the blog as a whole. Table 100 shows frequency of co-bloggers’ perceptions of agreement or disagreement.

**Table 100. Frequency of co-blogger agreement or disagreement for blog preservation**

Statement	Strongly Agree Freq (%)	Somewhat Agree Freq (%)	Neither Agree nor Disagree Freq (%)	Somewhat Disagree Freq (%)	Strongly Disagree Freq (%)
<b>My content</b> on the blog should be preserved for my <b>personal</b> access and use into the <b>indefinite</b> future	43 (51%)	24 (28%)	14 (16%)	1 (1%)	3 (4%)
<b>My content</b> on the blog should be preserved for <b>public</b> access and use into the <b>indefinite</b> future	36 (43%)	30 (36%)	11 (13%)	5 (6%)	2 (2%)
<b>My blog</b> , as a whole, should be preserved for myself and my co-author(s)’ <b>personal</b> access and use into the <b>indefinite</b> future	36 (42%)	24 (28%)	19 (22%)	3 (4%)	3 (4%)
<b>My blog</b> , as a whole, should be preserved for <b>public</b> access and use only into the <b>indefinite</b> future	35 (41%)	30 (35%)	13 (15%)	6 (7%)	1 (1%)
<b>My content</b> on the blog should be preserved for my <b>personal</b> access and use into the <b>short-term</b> future	4 (5%)	11 (13%)	30 (36%)	18 (22%)	20 (24%)
<b>My content</b> on the blog should be preserved for <b>public</b> access and use into the <b>short-term</b> future	4 (5%)	14 (17%)	30 (36%)	18 (22%)	17 (20%)
<b>My blog</b> , as a whole, should be preserved for myself and my co-author(s)’ <b>personal</b> access and use into the <b>short-term</b> future	2 (2%)	10 (12%)	29 (35%)	21 (26%)	20 (24%)
<b>My blog</b> , as a whole, should be preserved for <b>public</b> access and use only into the <b>short-term</b> future	3 (4%)	11 (14%)	29 (36%)	20 (25%)	18 (22%)

Nearly eight out of ten co-bloggers (79%) agree, either strongly or somewhat, that their respective content on the blog should be preserved for both personal and public access and use into the indefinite future. Further, seven out of ten (70%) agree their blogs, as a whole, should be preserved for personal access and use for themselves and their co-authors into the indefinite future, and nearly eight out of ten (79%) agree their blogs, as a whole, should be preserved for public access and use into the indefinite future.

Co-blogger respondents, as seen among single-blogger respondents, are less inclined to agree that their personal blog content and their blogs as a whole should be preserved only into the short term future. Only 15 respondents (18%) strongly or somewhat agreed their content published to the blog should be preserved for their personal access and use into the short-term future, and 23% agreed in regard to public access and use. Further, only 12 (14%) strongly or somewhat agreed that their blog, as whole, should be preserved for personal access and use for themselves and their respective co-authors into the short-term future, and 14 (18%) agreed in regard to public access and use. About half of co-bloggers disagreed, either strongly or somewhat, to preservation of their blogs into the short-term future for both personal access and use (50%) and public access and use (47%).

Agreement for preservation is also presented by cluster. Table 101 shows the frequency of agreement, either strongly or somewhat, for all respondents. For co-bloggers, agreement as presented in Table 101 represents agreement concerning the blog as whole; further, reference to personal access refers to access for both the co-blogger and their respective co-author(s).

Respondents assigned to the BioChemPhys cluster were slightly more inclined to agree their blogs should be preserved across all access and time levels than respondents in

other clusters. However, agreement for preservation for both personal and public access into the indefinite future is high across all clusters, representing the preference of a clear majority of respondents – between 73 and 83% – assigned to each respective cluster.

**Table 101. Frequency by all bloggers’ agreement for blog preservation by cluster**

Statement “My blog should be ...”	History Freq (%)	Economics Freq (%)	Law Freq (%)	BioChemPhys Freq (%)
Preserved for <b>personal</b> access and use into the <b>indefinite</b> future	22 (76%)	28 (74%)	37 (73%)	25 (83%)
Preserved for <b>public</b> access and use into the <b>indefinite</b> future	26 (84%)	28 (72%)	41 (79%)	26 (87%)
Preserved for <b>personal</b> access and use only into the <b>short-term</b> future	2 (7%)	7 (18%)	5 (10%)	9 (31%)
Preserved for <b>public</b> access and use only into the <b>short-term</b> future	6 (21%)	7 (19%)	5 (10%)	9 (31%)

Agreement for preservation is also presented for all respondents regardless of cluster, as shown in Table 102. It is noted that the same conditions concerning co-bloggers, as described for Table 101 immediately above, are applied in Table 102. For all respondents, regardless of authorship, most agree, either strongly or somewhat, that their respective blogs should be preserved into the indefinite future, with slightly more preferring their blogs be preserved for public access and use (80%) than for private access and use (76%).

**Table 102. Frequency of all bloggers’ agreement for blog preservation**

Statement “My blog should be ...”	Frequency	Percentage
Preserved for <b>personal</b> access and use into the <b>indefinite</b> future	112	76%
Preserved for <b>public</b> access and use into the <b>indefinite</b> future	121	80%
Preserved for <b>personal</b> access and use only into the <b>short-term</b> future	23	16%
Preserved for <b>public</b> access and use only into the <b>short-term</b> future	27	19%

#### 4.8.1.1. Blogs and Preservation Prioritization

While the majority of questionnaire respondents support preservation of their respective blogs, this finding falls short of how respondents might prioritize preservation of their blogs when considered among the variety of communications they produce and consume in their everyday scholarly lives. Interviewees commented extensively on the

challenges of selection and appraisal in regard to digital preservation. This was also implied in interviewees' remarks on the Library of Congress' plan to create an archive of public Tweets, as described in Section 4.6.5. Further, while respondents support blog preservation, what would it mean to them if their blogs, by chance, were "lost?" Two questions were asked of interviewees during the interview portion of this study to gather perceptions related to these issues.

Interviewees were presented with a "doomsday" scenario, asked how they would feel if their blog was unexpectedly and permanently deleted and what, if anything, they would do in response? Interviewees described a variety of reactions. One referenced the stages of grief when someone dies, commenting, "I would probably go through all those stages quickly." In terms of interviewees' emotive responses, five primary "sentiments" emerged. These are bulleted below, accompanied by comments from interviewees to illustrate the reactions described:

- 1) Sadness: "Pretty bad;" "Very bad;" "Sad;" "Pretty sad;" "Panicked;" "Devastated, both emotionally and professionally."
- 2) Anger: "Mad as hell;" "Pretty peeved;" "Pretty angry;" "Angry and upset;" "Frustrated;" "I'd do something drastic [in response] (i.e., legal action)."
- 3) Relief: "I don't have to do it anymore;" "I get half an hour of my life back."
- 4) C'est la Vie: "Pour another cup of coffee and get back to work;" "Probably have a drink and forget about it;" "Not welcomed but not tragic ... I'd get over it;" "Drop out of the blogosphere until something else comes along."
- 5) Doubt: "How would that happen?;" "It would take an extreme catastrophe;" "Hard to believe lost and unrecoverable."

Reactions were not mutually exclusive. Interviewees' expressed, for the most part, two or more of these sentiments. As described above, several doubted that such a loss could occur. Others made clear that they would be able to go on with the blog, specifically mentioning access to personal back-ups to recover "lost" content. A few made mention of resorting to such an approach in the past. In other activities to recover the blog, several commented that they would contact their respective service provider, host or network. One described that he would be distressed if his blog was lost, but this distress would not so much be related to the "loss" of the blog, but related to his feelings toward the service provider and their breach of the terms of service agreements. Another took responsibility for the blog's loss, commenting:

But I guess part of that would be my fault [if the blog was lost] because when it all started, you know, there was nothing to preserve and I had no idea that it would turn into the 21st most popular blog out of 4,000-some. I had no idea that it was going to be successful. I had no idea of the benefits it was going to provide. If I was doing it now [starting the blog], preservation would be one of the top things on the list.

Several interviewees, in discussing their emotive response, situated their blogs within their overall scholarly production. In the words of one interviewee, his blog, "has a short-shelf life," and in the words of another, "I'd go 'oh shit,' but it's not a big disaster ... it could be worse." Some provided specific descriptions of what would be worse in their reactions to this doomsday scenario: e.g., hard-drive failure, DropBox failure; loss of lab specimens. As another illustration of the "it could be worse" perspective, a Ph.D. in economics, when describing how he might react, commented:

The other thing that actually doesn't make me so upset about ... losing a blog compared to losing access to the American Economic Review is that the blog is where you do all the ideas and work with ideas, and is not necessarily the final output."

As further illustration the “it could be worse” viewpoint, interviewees were asked to consider their overall scholarly communication activities, both formal and informal. They were then asked, in consideration of a climate of finite resources, to describe their preservation preferences for these communications, placing them along a continuum of high to low priority. Additionally, interviewees were asked to specifically place their respective blogs within this continuum. Interviewees took two approaches when responding. Some took a ranking approach in their response, ranking between three to five sources as first, or most important to preserve, second, or next most important, and so on. Others ranked their preferences from high to low. Responses reflected a range of prioritization preferences.

Figure 1 shows all interviewees’ responses, in their own words, mapped horizontally along a continuum from high to low. These include both units and channels of communications. In Figure 1, units and channels interviewees’ identified as the highest priority to preserve are grouped on the left-hand side of the continuum (high). Units and channels considered a lower priority to preserve are grouped on the right-hand side of the figure (lower). Those units and channels considered a preservation priority, but neither the highest nor the lowest, are presented in the middle of the continuum. Each occurrence of channel and unit type identified and ranked by interviewees is mapped to the figure. Vertical placement (from top to bottom) is not important; all units and channels identified were simply plotted to the figure to facilitate grouping of similar responses among interviewees in regard to placement horizontally, along the continuum from left to right.



[illegible]

Most interviewees described their reasons for their placement of units and channels on the continuum of priorities to preserve. For example, one interviewee described placing his blog lower since he felt his blog is more representative of service rather than scholarship. Many commented on the ephemeral nature of blogs. A tenured law professor commented that he is, “All for preserving ephemera where it makes sense.” Issues of scalability were also raised in reference to the potentially ephemeral nature of blogs. Several interviewees commented that there is no need to save everything nor should there be attempts to save

everything, particularly when considered within the aggregate digital landscape. This recognition of scale contributes to concerns in regard to selection, as illustrated by remarks from a tenure-track associate professor of law:

I think there is real value in trying to preserve the blogs. I think at some point you have to recognize you can't retain everything but I think we need to be careful ... if we kept only the New York Times and the Wall Street Journal in terms of newspaper archives, we'd miss an awful lot, and I think that we run the same risk when it comes to things in blogging and other resources.

A few considered the question by reference to "economic models." This is illustrated by comments from a Ph.D. in economics: "If you wanted to put an economic model on it, you'd say the more work that went in per word, the more important it is to preserve it." He continued by using emails as an example: "When I go through [emails] what I delete are funny cat video forwards; I don't delete long letters."

A tenured professor of economics also commented on issues of cost associated with his scholarly output, but approached it from the angle that, ultimately, while blogs may have less informational value, they may be cheaper to preserve:

I teach, and that's a big component of the value that I provide as a professor. Digitizing that is something that may happen in the future and is happening increasingly, although very small proportions of the lectures are taped and put online. And that's a cost problem [it's expensive to do; i.e., recording, digitizing, putting it online] ... in terms of the spectrum, it's the net value of preservation ... When it comes to the blog, the blog is cheap to do. In terms of my scholarly value, it's a very, very small part of my scholarly value that I create. Single-digits, in percentage terms. It's cheap to preserve it, very cheap to preserve it, and so, from that point of view, it makes a lot of sense to do so.

Just as a few interviewees had found the doomsday scenario dubious, a few also questioned the need to prioritize communications for preservation. In terms of cost and value, others raised concerns about expending resources. A couple of interviewees raised the point that some organizations are already actively preserving blogs, such as the Internet Archive,

so blogs are less of a preservation priority for them. One tenured law professor's reaction to the question was: "Is that right? Let me question your premise. I mean, Archive.org is trying to archive the whole Web and they have a lot of resources. And I have found [identified two of his blogs] in there. So they're clearly doing that." Additionally, a Ph.D. in chemistry commented that, with low storage costs, preservation selection decisions are not, "a resource-limited problem," illustrating his viewpoint through reference to a ten dollar USB stick, and that he can store, "everything I have ever put on my blog since 2002 ... all the papers and patents I have ever published in my entire career ... so I have never even thought about prioritizing." This leads to the question: because we can keep it all, should we? Both these comments touch on aspects of scalability, value, and use.

Returning to perceptions of value, a tenure-track associate professor of economics remarked: "I value my blog quite a lot but I am not so arrogant enough to think it has value to the rest of the world." He expressed little public value in blog preservation. He responded by rephrasing the question:

What I think you might be leading up to, the big punch, is would this be a most valuable use of these scarce resources, if this stuff doesn't save itself for free? My answer would be, generally speaking, for most blogs, no ... and so many [blog] entries are so finely tuned to the specific issues of the day that if we just save these things, how much extra work will people have to go into to figure out exactly who are the players, what was the issue?

In further illustration of his viewpoint he addresses issues of appraisal and the potential value of blogs for reuse or even as artifacts:

So many people put playful, snarky thing in blogs ... some blogs have real serious, intellectual merit and others are just, 'oh, I found a great new restaurant, so if you're in Washington DC ....' We're going to save all of that? Is anyone ever going to go back and read it?

A couple of interviewees provided some reasons why blogs, as well as other, less formal publications, merit some preservation prioritization. A tenured professor in biology described her preservation priorities as a “parabola,” and her preference to preserve both very formal as well as very informal communications:

I almost view it as a kind of parabola. I’d want to keep the most informal sets of thoughts that I’ve got, and preserve those, and then the sort of very finished peer-reviewed projects ... and I’d also include the essays I’ve written for formal publications as well.

A tenured associate professor of law described her preservation preferences for two different types of blogs:

There are some blogs that are clearly intended to be contributions because of their content and then there are some blogs that are clearly intended to be ephemeral because the content is dissemination of news -- quick news, updates -- and the content, even the writer’s intent of the content, is simply to be quickly absorbed and then who cares.

While Figure 1 represents responses for all those who ranked unit types or channels along a continuum, it does not reflect all interviewees’ opinions. Several interviewees, when asked where to place blogs on the continuum, responded with, “it depends.” These responses are not shown in Figure 1. Instead, some interviewee comments are summarized here to further illustrate issues related to selection and appraisal. As one tenured economics professor remarked, “I don’t know. As you know, all these blogs are different and it’s back to the apples and oranges things. Mine are musings, and I don’t mind calling them that.”

One tenure-track associate professor of law commented that, initially:

... pretty clear conference proceedings and more traditional scholarship and articles would be first in line, but although that said, I think long-term, in some ways, blogs may be more beneficial in that if ... someone wants to go back and look at the record ... to ‘connect the dots’ after the fact.

This sentiment was echoed by another interviewee assigned to the history cluster:

But personally, although I value the blog immensely, I guess I'm less concerned about it being archived – I don't know why, maybe I should be more concerned about it – but I'm less concerned about that in relation to my other scholarly production ... the articles I've written and the ones that got written so far are obviously in libraries all over, so in digital and hard copy, it is being archived, and the blog isn't, so I don't know. This interview is making me more concerned about it [laughter].

Several interviewees described the challenge of selection among not just blogs, in general, but the content contained within their specific blogs. This contributes to consideration of preservation at the post- rather than the blog-level. Comments from two interviewees are provided below for illustration:

{Ph.D. in economics} If you put a gun to my head and said please tell me your top 100 posts out of 2700 ... I could filter out pretty quickly, but after awhile I'd get down to the agonizing decision of the top 100, and my criteria for saving them would be what I want to share with somebody else or what was valuable for developing my thought. So those blog posts are the ones worth keeping, but all the other ones are kind of ephemera ... It's just like conversation in a bar or a coffee shop with your friend. You don't need to maintain every word.

{Blogger in BioChemPhys cluster} As a blogger, I always think, 'okay, what's going to happen to my blog 100 years from now' ... I have 10,000 posts. I wouldn't cry if 9,000 of those disappear but there are posts that I hope would stay for a while, posts that I put a lot of energy and research and time into, that are longer, that are well-thought of and well-written, I hope. That serve as resources ... things that can be useful to others.

#### 4.8.2. Responsibility and Capability for Preservation

Questionnaire respondents were presented with two questions to assess their opinions on who should be most responsible for preservation of their respective blog, as well as who they perceive as most capable. Concerning responsibility, respondents were first asked to rate how much responsibility they feel select actors have for the preservation of their respective blogs. A researcher-derived list of "actors" was provided. Single-bloggers were presented with nine categories of actors. Co-bloggers were also presented with these categories, as well

as a tenth, additional category, “My blog’s co-author(s),” due to the co-authored nature of their blog.

These actor categories can be organized to represent three general approaches to preservation. These approaches, with accompanying actor roles provided in parentheses, are: 1) Personal (e.g., myself; my blog’s co-author/s); 2) Affiliated (e.g., my blog’s host, including service provider, hosting company or network; my institution’s library; my institution’s archive; my institution’s IT department); and 3) Unaffiliated (e.g., search engine; public trust; my national library; my national archive). Personal denotes direct action by blog authors. Affiliated denotes action by entities for which the blogger is directly affiliated. For example, as an employer, customer, or subscriber. And unaffiliated denotes action by organizations and other entities for which the blogger is not directly affiliated, necessarily, as an employer, customer, or subscriber.

Respondents were asked to rate each actor group across four levels of responsibility: much, some, a little or none. Additionally, respondents were also provided the option to select “Group does not apply” for each of the actor roles. This was done, in part, due to variety in respondents occupational roles and perceptions of affiliation to certain actor groups.

Table 103 shows the frequency of all respondents’ ranking of actor groups in regard to responsibility for blog preservation. Percentages reflect the frequency of those responding for each actor type. For example, the percentage for the frequency of those responding to “My blog’s co-author(s)” represents only the 84 co-bloggers responding; single-bloggers were not presented with this actor type.

**Table 103. Frequency by all bloggers' ranking of responsibility for preservation**

<b>Actor Group</b>	<b>Much Freq (%)</b>	<b>Some Freq (%)</b>	<b>A Little Freq (%)</b>	<b>None Freq (%)</b>	<b>Group Does not Apply Freq (%)</b>
Myself	87 (58%)	41 (27%)	11 (7%)	10 (7%)	1 (1%)
My blog's co-author(s)	20 (24%)	38 (45%)	12 (14%)	10 (12%)	4 (5%)
My blog's host (e.g., service provider, hosting company, or network)	62 (42%)	50 (34%)	11 (7%)	22 (15%)	4 (3%)
Search engine (e.g., Google)	18 (12%)	34 (23%)	27 (18%)	64 (43%)	5 (3%)
Public trust (e.g., Internet Archive)	14 (9%)	27 (18%)	43 (29%)	57 (39%)	7 (5%)
My institution's library	3 (2%)	7 (5%)	18 (12%)	103 (69%)	18 (12%)
My national library	6 (4%)	7 (5%)	21 (14%)	98 (65%)	18 (12%)
My institution's archive	-	9 (6%)	14 (9%)	104 (70%)	22 (15%)
My national archive	6 (4%)	6 (4%)	15 (10%)	102 (68%)	20 (13%)
My institution's IT department	2 (1%)	2 (1%)	11 (7%)	112 (75%)	22 (15%)

The majority of respondents (58%) identify themselves as having the most responsibility for the preservation of their respective blogs. Their blog's host, which may denote a service provider, hosting company or network, was identified as having the second most responsibility, selected by 42% of respondents. Those identified as having no responsibility represent both affiliated actor groups – library (69%), archive (70%) and IT department (75%) – and unaffiliated groups – national library (65%) and national archive (68%).

Concerning capability, respondents were next asked to rate how much capability they feel select actors have for the preservation of their respective blogs from the same listing provided in the question concerning responsibility. Table 104 shows the frequency of all respondents' ranking of capability level among actor groups in regard to blog preservation. Percentages reflect the frequency of those responding for each actor type. For example, the percentage for the frequency of those responding to "My blog's co-author(s)" only represents the 77 co-bloggers responding; single-bloggers were not presented with this actor type.

**Table 104. Frequency by all bloggers' ranking of capability for preservation**

<b>Actor Group</b>	<b>Much Freq (%)</b>	<b>Some Freq (%)</b>	<b>A Little Freq (%)</b>	<b>None Freq (%)</b>	<b>Group Does not Apply Freq (%)</b>
Myself	49 (35%)	43 (30%)	30 (21%)	19 (13%)	1 (1%)
My blog's co-author(s)	18 (23%)	30 (39%)	16 (21%)	11 (14%)	2 (3%)
My blog's host (e.g., service provider, hosting company, or network)	91 (64%)	37 (26%)	7 (5%)	4 (3%)	3 (2%)
Search engine (e.g., Google)	62 (44%)	46 (33%)	18 (13%)	10 (7%)	4 (3%)
Public trust (e.g., Internet Archive)	46 (33%)	47 (33%)	22 (16%)	16 (11%)	10 (7%)
My institution's library	13 (9%)	29 (20%)	31 (22%)	48 (34%)	21 (15%)
My national library	24 (17%)	21 (15%)	32 (23%)	46 (32%)	19 (13%)
My institution's archive	13 (9%)	24 (17%)	32 (23%)	48 (34%)	25 (18%)
My national archive	22 (16%)	22 (16%)	31 (22%)	46 (33%)	20 (14%)
My institution's IT department	16 (11%)	16 (11%)	30 (21%)	54 (38%)	25 (18%)

While a majority of bloggers (58%) identify themselves as having the most responsibility for the preservation of their respective blog, fewer (35%) identify themselves as having the most capability. Their blog's host, which may denote a service provider, hosting company or network, was identified as having the most capability for preservation, selected by 64% of respondents, followed by search engines (44%). As with preservation responsibility, the actor groups identified as having no capability represent both affiliated groups— library (34%), archive (34%) and IT department (38%) – and unaffiliated groups – national library (32%) and national archive (34%).

Perceptions on actors perceived as most responsible and most capable are also presented side-by-side in Table 105. In this table, “most” is characterized as those actors ranked as having much or some responsibility and capability. This information is also presented in Table 106 by rank order, with one being the highest ranked group, two the second highest, and so on. Group names have been modified slightly in this table.



**Table 105. Frequency of actors perceived as most responsible and most capable for blog preservation**

<b>Actor Group</b>	<b>Most Responsible Freq (%)</b>	<b>Most Capable Freq (%)</b>
Myself	128 (85%)	92 (65%)
My blog's co-author(s)	78 (69%)	48 (62%)
My blog's host (e.g., service provider, hosting company, or network)	112 (75%)	128 (90%)
Search engine (e.g., Google)	52 (35%)	108 (77%)
Public trust (e.g., Internet Archive)	41 (28%)	93 (66%)
My institution's library	10 (7%)	42 (30%)
My national library	13 (9%)	45 (32%)
My institution's archive	9 (6%)	37 (26%)
My national archive	12 (8%)	44 (31%)
My institution's IT department	4 (3%)	32 (23%)

**Table 106. Ranking of actors perceived as most responsible and most capable for blog preservation**

<b>Rank</b>	<b>Responsible</b>	<b>Capable</b>
<b>One</b>	Blogger	Blog Host
<b>Two</b>	Blog Host	Search Engine
<b>Three</b>	Co-author(s)	Public Trust
<b>Four</b>	Search Engine	Blogger
<b>Five</b>	Public Trust	Co-Author(s)
<b>Six</b>	National Library	National Library
<b>Seven</b>	National Archive	National Archive
<b>Eight</b>	Institutional Library	Institutional Library
<b>Nine</b>	Institutional Archive	Institutional Archive
<b>Ten</b>	Institutional IT Dept.	Institutional IT Dept.

Half of the actor groups converge in terms of perceptions on responsibility and capability. Respondents perceive libraries, archives, and institutional IT departments as both least responsible and least capable for preservation of respondents' respective blogs. Another group, the blog host, converges somewhat. This group, comprising blog service providers, hosting companies, and networks, is perceived as the most capable and the second most responsible among all actor groups. While bloggers perceive themselves and, if applicable, their co-authors, as the most responsible actor group for preservation of their respective blogs, they perceive themselves as the fourth most capable and their co-authors, if applicable, as fifth. Additionally, search engines and public trusts, while found to be ranked as the fourth and fifth most responsible actor groups, are ranked second and third, respectively, in terms of capability.

#### 4.8.2.1. Additional Perspectives on Responsibility and Capability

Perceptions on blog preservation responsibility and capability were also gathered during the interview. As with questionnaire respondents, the majority of interviewees also identified as being personally responsible for preservation of their own blogs. For several interviewees, responsibility and “ownership” aligned. One tenured economics professor remarked: “The blog is self-published, as it were, so I would regard the publisher as the one who’s primarily responsible for that.” Another interviewee commented that while he is, “happy that Archive.org and Google or whatever have cached copies of it [his blog] and everything,” that does not mean he should, “just slack off and let them do it.” He concludes: “It just seems kind of philosophically alien to me to expect other people to do that for me.” Another expressed their perception succinctly, emphasizing not primary responsibility, but ultimate responsibility: “The final responsibility rests with the individual who is writing.”

A tenure-track assistant professor of economics does not see much value in blog preservation and, as such, does not “expect other people to pay the costs to record something that I feel is not particularly significant.” He goes on to make his case against blog preservation:

I have conversations with friends, colleagues, students, and former students in restaurants and in lines at Kroger that I view to be as meaningful as my blog posts and no one’s ever going to suggest following me around with a microphone just to catch the occasional nuggets of wisdom I might let fall ...”

As shown in the questionnaire results, while the majority of respondents may see themselves, or their co-author(s), as most responsible, they rate themselves as having less capability than service providers, search engines, and public trusts. One network-affiliated tenured law professor commented on the need for more tools to facilitate personal capability: “What I would like, personally, would be to have the ability myself to download the entire

thing and keep it ... like an export of everything so I'd have it sitting right here on my computer." This call for tools to facilitate personal preservation was echoed by another interviewee:

One of the things that I would like to see is that the service providers ... provide guidance or assistance towards a back-up; on the whole, they don't because they don't think of it as their responsibility. They may in the future.

Some tools for personal archiving are already available, as demonstrated in the questionnaire results as reported in Section 4.7. However, these comments point to the need to raise awareness of some of these tools among communities of bloggers. One interviewee talked with some familiarity of available third-party services, at cost, commenting: "I think that individuals actually have a lot of power to preserve their own [blogs]. I know they can set up their own preservation ... through a third party provider if they have the money to do that."

Some interviewees also commented on their expectations for service providers, hosts, and networks to share in responsibility. While interviewees may perceive providers as responsible and capable, a few mentioned being unaware if and, subsequently, how providers address archiving and preservation issues in their terms of service agreements. As one tenured law professor remarked: "It doesn't make any sense to me that a third party would have any responsibility to archive it. I think it would be [provider's] responsibility, though I have never read our terms of service." He goes on to present an alternative:

If for some reason I went and read the terms of service, and they totally disclaim all responsibility for permanence or archiving, it would then be our responsibility to figure out how to archive our posts. The worst case scenario would be that we should save every post before we post it or if there any third party services that will agree to archive your blog for you for a fee. I don't remember any of those services ever contacting us, if they exist.

A tenured professor of law raised several questions in regard to service providers: “How long is [provider] going to keep my blog? At any point, will it fall off? How big is the [provider’s] service?” She continues, commenting that she, “Never worried about the privacy aspect of it, but now I sort of worry about the permanence aspect of it. How permanent is it? If [provider] went bankrupt, would we just have to start over on a different server?”

Another interviewee, a tenured humanities professor, commented on the complicated, co-produced nature of blogs, in terms of content creators and service providers, and implications for control of and access to content:

I don’t want to be beholden to a [service provider] ... I’m always nervous about what that means for control over content that is created. We know only too well if that service goes away, if they unilaterally change their terms of service – which, you know, has become a standard if wretched practice for a software provider – where are you going to be ... We are all hostage to the fact that everything you think about the rights that you have in one of those relationships can change. They can change without you consenting to it. The only power you have is to withdraw from that relationship at the time of the change.

Further, most interviewees who publish to a blog network feel the network also shares in some preservation responsibility and capability, as well as have an interest in it. One interviewee described a recent trend, remarking that he feels that more bloggers are, “blogging on behalf of institutions, like research institutions and [magazine] publishers.” He sees these established organizations as, “much more able to preserve this information and to make sure it gets shared on Facebook, Twitter, and what have you than I am. And they also have, presumably, a much greater interest in making sure it’s preserved.”

Other reasons emerged for this viewpoint of network obligation. One is that since the blogs are hosted and published by the network, the bloggers have given up some control of the content and container; hence, by having control, networks, by default, also have some responsibility for preservation. One interviewee commented that his blog posts, “always start

as a Word document on [his] computer so [he] save those things,” and that, while that might mean the blog can be recovered, “that does not mean it’s available,” since his network is “hosting it [the blog].” As a result, the network should be responsible for, “keeping it on the Internet.”

Several mentioned appreciation for organizations currently capturing and archiving blogs, including the Library of Congress and the Internet Archive. Interviewees, for the most part, were supportive of organizational involvement, though, preferably, not their own institutions. Some see organizations as having more capability due to their access to resources. Several interviewees specifically mentioned Google. One tenure-track associate professor of law remarked: “Places like Google and others have capability because of their business model ... they have a capability to record and maintain things in a way that perhaps some other areas don’t.” A tenured professor of economics referenced Google, his service provider, and commented, ““I’m okay to be optimistic that Google will be around for a while ... I’ve not really worried about that.”

Another interviewee, a tenured professor in history and American studies, however, expressed trepidation at the potential reach and role of Google: “It makes me very nervous, frankly, that companies like Google might, in the end, be responsible for preserving everything.” She cited the work of Siva Vaidhyanathan, and his upcoming book, *The Googlization of Everything*,<sup>65</sup> commenting on the, “real dangers of one company growing so big and being in charge of preserving so much of our cultural heritage,” and “that the idea that a corporation will undertake this just seems natural and normal. And I think it is the job of nations to preserve national heritage, not corporations.”

---

<sup>65</sup> Vaidhyanathan, S. (2009, September 9). The Googlization of Everything to be published in early 2011. [Web log message]. Retrieved from: [http://www.googlizationofeverything.com/2010/09/the\\_googlization\\_of\\_everything\\_1.php](http://www.googlizationofeverything.com/2010/09/the_googlization_of_everything_1.php)

The Internet Archive was also specifically referenced by several interviewees, and seen to have ample capability for preservation. In expressing appreciation for the Internet Archive, a Ph.D. in economics described it as, “an amazing public good that’s being produced by someone who really had their head on straight when they started in ’96.”

Several reasons emerged for resistance to intuitional-based preservation. One is fiscal. This was raised by several interviewees. As one tenured associate law professor noted: “I have to say, being at an academic institution that is publicly funded and knowing that publicly funded institutions have a lot of demand on their finances, I don’t know how much we can expect publicly funded institutions to do that [blog preservation], frankly.” Another interviewee cautioned that institutions have to be careful with their resources: “... they [his institution] would have more than a little capability of tracking the blogs of all of our professors, perhaps, but ... we certainly don’t have the resources to make ... an additional archive available for us on the scale of the blogosphere.”

Another reasons that emerged is that, while the bloggers may be representatives of their respective institutions, it does not mean their blog is representative. As one tenured law professor commented: “I don’t consider my institution responsible for these things because none of them are done through the institution except the classroom ones [blogs] ... does the world really need to know when page 75 was due?” Another remarked: “I don’t feel like my blog is their [her institution’s] problem.”

And, while bloggers may be affiliated with a particular institution today, it does not mean they will continue to be affiliated at a later date. As pointed out by a couple of interviewees, institutional responsibility for preservation is complicated since institutional affiliations may change, as academics move among universities and colleges. They

eventually also retire. This last scenario was proposed by one interviewee who wondered if her institution would maintain her personal archive after she leaves. Also, as pointed out by a Ph.D. in economics, academics may not be aligned with an institution. Consider the questionnaire results; as presented in section 4.2.4, most, though not all, questionnaire respondents are employed by post-secondary institutions.

Another theme that emerged was that the risk of loss to the blogger is much greater than the risk of loss to their respective institution. Further, the investment is that of the blogger, rather than the institution. This is illustrated by comments from a tenured biology professor:

I think for me the stakes are high so it's important for me to keep track ... the stakes are not as high for the university to keep track of it so that's why I feel like it's my responsibility to do that ... I am the one that has the investment and I am the one who will lose or win depending on my ability to retrieve the information.

A few perceived their institutions as having less capability for preservation based on past experiences. As one interviewee commented, "the track record at [her institution] has not been great ... I think my institution isn't well set up to think about these things." She referenced a past experience in which they had a set of back-up tapes of her research data and misplaced them. She fortunately had back-ups, but it could have been a substantial loss without her personal back-ups. In regard to her institution and capability, she commented: "The capability is not there. And if the capability is not there, I don't want to give them the responsibility."

Though questionnaire respondents see their institutions as having little responsibility or capability for blog preservation, this does not necessarily reflect a devaluing of their roles in the preservation of scholarship. Actually, as pointed out by a couple of interviewees, that is exactly the role for which libraries and archives are valued in regard to more traditional

forms of scholarship. A tenured associate professor of law, in talking about digital preservation in her everyday scholarly life, referenced working with librarians on a regular basis. Another tenured law professor commented: “I trust my library to preserve books and articles and things. If I cite to them, I don’t worry that people will have trouble finding them in the future ... But for other more ephemeral materials, it’s in my interest to keep a copy so that if it is no longer online when they are looking for it and if Archive.org doesn’t have it, I have one.”

A tenured humanities professor commented on the divergence between the technical know-how of digital preservation and the “philosophical” know-how, and identified librarians as able to bridge this gap:

The people who on one hand are most technically capable of doing it tend to be people identified with information technology in some fashion [e.g., ITS on campus] or people that are tied to the current economy of publication and dissemination through digital means. ... If you turned to your average group of academics and you say ‘preserve your electronic communications,’ up and down my own hall that would inspire either panic or a blank stare. People would have no idea where to begin that effort in technical terms. But they might very well have a better handle on ... why they might want to do it or why it’s conceptually important and that’s where I think a lot of the people technically capable ... tend to fall down.

He summarized these complications succinctly: “We tend to just say ‘well, when we outsource the technical questions, we outsource the philosophical questions’ ... the only people who are in the in-between right now are librarians.”

While personal responsibility was emphasized by a majority of interviewees, some did comment on how libraries and archives might position themselves to take on responsibility for some blog preservation:

I don’t think I would ever think it was anyone’s responsibility. I could see, however, a university library creating an archive ... that they might think that was in their mission just as it is to accumulate volumes and subscribe to different services. I don’t



think I have an opinion that it's in the public interest for a state or local or federal library to do that.

Her comments are also reflective of another theme that arose: A couple of interviewees expressed not having strong views, one way or another, on who should be responsible or capable for preservation of their respective blogs. This viewpoint is shared by a tenured professor in economics in his comments on the shifting roles of traditional organizations and the challenge in predicting what will come next:

As we keep saying, the ground is shifting so rapidly that what we once thought of as a strictly library activity, is now such a vaster thing ... ten years ago there was no Blogger. So, the roles have shifted so drastically. And as far as I can tell, the libraries aren't even sure where they're going to end up, and just like traditional publishers don't know where they're going to end up, traditional journals don't know where they're going to end up. Traditional journals, with their clunker two-year turnaround times, aren't long for this world, in my view. And so, I'm sure they're all scrambling and trying to figure out what the hell to do with themselves and I have no idea how they will come out ... These questions are vague, and not sure I have strong views .

Another interviewee also makes mention of this uncertainty in his response in supporting his viewpoint that bloggers are primarily responsible:

Because they [library, archive, and IT department] have the expertise doesn't mean they have the ability to do it, necessarily. I think they will have the ability to do it in the future and perhaps the capacity to do it, but I don't know that we have a clear understanding of the capacities required because I don't think we have a clear understating of what the members' of the community, who might think of that archive/library as their repository, are actually doing. But I still think the primary responsibility, at least at this time, rests on the authors, so to speak.

Other interviewees offered suggestions as to how institutions could participate in blog preservation. A Ph.D. in economics proposed institutions adopt an "opt-in model" for blog preservation, akin to a self-deposit model for an institutional repository:

Assume that [there is] some mechanism for moving archives or even centralizing them. What I would want as an academic is an email from the archive department saying: We have this service; if you'd like to use it, input your blog name here and press 'please archive.' So ... the archiving people don't go running around trying to find everything that academics are doing because some academics do not want to

have their stuff archived and they don't have to worry about this whole searching, archiving everything. And also it gives academics an easy way of getting their work archived without worrying about it, especially the ones who don't even think about it, so they can have it shared out there. And of course they will be able to put granular controls on who gets access to that archive.

Another interviewee, a tenured humanities professor, proposed a role for institutional involvement, but acting collaboratively rather than in isolation, and to do so not just in pursuit of the goal – preservation – but also in consideration of the *why*:

Ideally, I would like to see our own institutions acting in concert in large networks to engage in preservation because we most clearly ought to understand why we want to preserve as well as why we're underwriting knowledge production and dissemination in the first place.

While some interviewees did not see responsibility for blog preservation as belonging to their respective institutions' libraries, archives, or IT departments, three interviewees did see a role for national cultural heritage institutions. A tenured professor in history and American studies shared her perspective on the role for federal archives and museums:

Ideally, preservation would be undertaken by places like the Smithsonian or the National Archives. And I think particularly the Smithsonian has the kind of intellectual scope and has access to the scientists to really be able to think about this. A lot of what the Smithsonian does is scientific and cultural history. So, I think it falls very much in their purview. But I also think, as someone who spends a lot of time at the National Archives and at the National Archives online, it's really a natural for them ... it seems to me, given the fact that the Internet is, in fact, a government thing and the airwaves on which all of our digital messages are traveling now are also a government thing, it seems to me that the preservation of electronic and digital materials sort of naturally fall in the purview of the federal government.

Lastly, while interviewees, as well as questionnaire respondents, shared a strong preference for personal responsibility, that does not mean sole responsibility. In consideration of personal or programmatic approaches to preservation, a Ph.D. in chemistry shared his own "all-in" perspective, reflecting that it is not a *me or you or us or them* scenario:

I say let a hundred flowers bloom. I think people should make their own archives and I think institutions should have their own. The more places things are stored together, the chance they have of being saved.

## CHAPTER 5

### RESULTS: BLOG ANALYSIS

This chapter presents quantitative results from the blog analysis data collection phase of the study, describing and quantifying blog structural elements and content elements as they relate to issues of identity, authority, composition, use, and dependencies, with the latter in reference to software publishers and other co-producers. Additionally, select results from the interview portion of the study are provided to illustrate and provide context for several of the findings reported here.

#### 5.1. Completed Samples: Blogs and Interviews

The blog sample, originally shown in Table 6 in Section 3.1.2., is reproduced here as Table 107. It shows the count of blogs, by authorship and cluster, comprising the completed sample for the blog analysis portion of the study. While the sample represented 50% of all blogs in the sampling frame, ultimately one co-blog assigned to the economics cluster and selected for the sample was excluded during data collection. Due to issues of accessibility, with the blog continuously timing out, only partial data points could be collected. Table 107 reflects this revision to the completed sample for the blog analysis portion of the study.

**Table 107. Blog sample**

Clusters	Single-Blogs Count	Co-Blogs Count	Total Blogs Count
History	16	7	23
Economics	17	8	25
Law	11	13	24
BioChemPhys	17	4	21
All Clusters	61	32	93

Additionally, in regard to the interview portion of the study and as presented in the preceding two chapters, the completed interview sample is comprised of 13 single- and 11 co-bloggers representing all four clusters. For the completed interview sample by authorship and cluster, see Table 21 in Section 4.1.2. It is not reproduced here.

## 5.2. Identity and Authorship

This section reports on three characteristics of identity and authorship: access points for identifying personal information on blog authors, including name, occupation, and/or education; descriptive measures of co-blogs' primary co-authors; and single-bloggers' use of real names and pseudonyms.

### 5.2.1. Access Points

All single- and co-blogs comprising the sample are authored by bloggers meeting the parameters for scholars, as presented in Section 3.1.1.1. As originally presented there, a single-blogger is deemed a "scholar" if one of the following four sub-criteria is met:

- a) One or more of the following keywords or phrases are used to describe the blogger:  
Ph.D., Dr., Professor, Reader, Lecturer, Doctoral Student, or Doctoral Candidate;
- b) One or more of the following keywords or phrases are used to describe the blogger:  
Scholar, Academic, Researcher, Research Director, Fellow, and/or Other Role Identifier (i.e., biologist); **and** institutional affiliation is provided, either specifically by name or through a general description denoting a post-secondary education or research setting;
- c) A link to blogger's curriculum vitae (CV), either full or partial, or other listing, either full or partial, of blogger's publications is accessible, **with** at least one citation to a journal article published or co-published by the blogger, or;

- d) The blogger is described as a graduate student **and** explicit reference to area of study or pursuant degree is provided.

For co-blogs, only those co-bloggers publishing to their respective blogs within one-month from the date of assessment for sampling were considered for inclusion based on scholar parameters a, b, c, or d.

For all blogs in the completed sample, the access point from which to locate information to assess these parameters was recorded. These access points were assessed in the following order, based on immediacy of visibility to blog visitors: 1) indentifying information on the first page; 2) identifying information via an internal blog link; 3) identifying information via an external link; and 4) a combination of both internal and external links, specifically for co-blogs. While bloggers' information may be available through two or more of these access points, only one access point was recorded, with the exception of co-bloggers in regard to internal and external links. This information was collected to inform how a casual reader of the blog may be exposed to information on bloggers' identity, whether shown directly on first page, through an internal link, through an external link, or, for co-blogs, a combination of both internal and external links.

Table 108 shows the frequency of access points for identifying information on blog authors. Sixty-three blogs in the sample (67%) have identifying information either posted to the first page of the blog or available via an internal blog link.

**Table 108. Frequency of access points for information on bloggers' identity**

Degree	Single-Blog Frequency (%)	Co-Blog Frequency (%)	All Blogs Frequency (%)
<b>First page</b>	23 (3%)	7 (22%)	<b>30 (32%)</b>
<b>Internal link</b>	22 (3%)	11 (34%)	<b>33 (35%)</b>
<b>External links</b>	16 (26%)	7 (22%)	<b>23 (25%)</b>
<b>Internal and External Links</b>	-	7 (22%)	<b>7 (8%)</b>

### 5.2.2. Co-Authors: Extent

For the 32 co-blogs in the blog sample, the number of all primary co-bloggers assigned to each blog was collected. As described in Section 3.1.1.1., primary denotes that the co-blogger is a regular, current contributor to the blog. Co-bloggers may be described in author-identification fields as blogger, author, editor, co-editor or contributor. Excluded are bloggers identified as guest editors, emeriti bloggers, or other categories that denote either a former or temporary relationship to the blog. For the 32 co-blogs in the sample, 334 co-bloggers were identified. Descriptive measures on the number of primary authors for co-blogs were calculated as follows: mean 10; median 6.5; mode 9; and range 33. Additionally, descriptive measures by cluster are presented in Table 109.

**Table 109. Descriptive measures of number of primary authors for co-blogs by cluster**

Measure	History	Economics	Law	BioChemPhys
Mean	8	13	11	8
Median	3	8.5	6	8
Mode	3	2, 13	5	7, 9
Range	21	34	32	2

From these 334 identified co-blog authors, only 124 (37%) were considered active, characterized as posting to their respective co-blog within the previous 31 days from date of assessment for sampling. Only active co-bloggers were considered for further assessment, in line with the scholar parameters described in the previous sub-section. For these 124 co-bloggers, 114 (92%) met these scholar parameters.

### 5.2.3. Single-Authors: Name and Location

For the 61 single-blogs in the blog sample, their respective authors' names were identified and then coded to distinguish between real and pseudonymous names. Five name categories were used: 1) pseudonym; 2) first name or nickname; 3) first name and initial; 4) initial and last name; and 5) full name, characterized, at minimum, as first and last name.

Additionally, since bloggers' may identify themselves by more than one name type on their respective blog, each applicable name type was coded. For six single-blogs (10%), the blogs' respective authors used two name types. See Table 110 for frequency of name types for the 61 single-blogs in the sample.

**Table 110. Frequency of name types**

<b>Name Type</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Pseudonym</b>	13	21%
<b>First name or nickname</b>	1	2%
<b>First name and initial</b>	-	-
<b>Initial and last name</b>	2	3%
<b>Full name</b>	51	84%

Overall, 51 blogs (84%) feature authors' full names. Eight of the blogs (13%) were authored by bloggers using pseudonyms exclusively; five (8%) use pseudonyms in addition to another name type. The cluster with the highest representation of authors using a pseudonym, exclusively or in combination with another name, is history (n=7). Exclusively pseudonymously authored blogs, by cluster, are: history (n=5), economics (n=1) and BioChemPhys (n=2). This finding lines up with interview comments from a tenured professor in history and American studies. She remarked that, "historians get fewer opportunities than, say, political scientists or economists or sociologists to bridge the world of the academic and public." She goes on to connect this to implications for pseudonymous authorship:

I think people who are in those fields are a lot less suspicious of blogging than people who, say, are in history or in English. And I think how this plays itself out is that history bloggers and English lit. bloggers have a much stronger tendency to be anonymous or pseudonymous.

The access point, or points, from which bloggers' names, whether real and/or pseudonymous, were located on the blog was also recorded. These access points were: 1) blog header; 2) blog title tag; 3) blog URL; 4) blog sidebar; and 5) post author field. If name



was not available at any of these access points, then it was coded as such. Additionally, since bloggers' names may appear at two or more access points, each applicable location was coded. For forty-five single-blogs (74%), bloggers' names were located at two or more of these access points. Table 111 lists frequency of access points for the 61 single-blogs in the sample.

**Table 111. Frequency of access points for single-bloggers' name**

<b>Access Point</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Header</b>	14	23%
<b>Title tag</b>	8	13%
<b>URL</b>	25	41%
<b>Sidebar</b>	39	64%
<b>Post author field</b>	44	72%
<b>None of these</b>	3	5%

A majority of blogs have author name accessible in the post author field (72%) and sidebar (64%). Additionally, author name, whether in whole or in part, is included in 41% of blogs' URLs. Author name was not available at any of the five access points for three blogs (5%).

### 5.3. Blog Elements and Features

This section reports on frequency of select blog characteristics for all single- and co-blogs in the sample. These are characteristics found at the blog-level rather than the post-level. Select post-level features, such as permalinks, categories, and commenting system, are presented in Section 5.4.

#### 5.3.1. Select Feature Set

Blogs were coded as having or not having seven features: 1) blog title, as shown in blog header field; 2) additional descriptive information in title tag field; 3) archiving system; 4) categories; 5) blogroll; 6) search tool; and 7) commercial advertisements. For clarification on some of these features, categories are labels generally found in the blog's sidebar and may also be identified by other headings, including keywords, tags, tag cloud, or

topics. Blogrolls, generally found in a blog's sidebar, may also be identified by other headings, including friends, subscriptions, favorites, or other label, and feature links to other blogs. Commercial advertisements were identified as badges appearing in the sidebar. These may include individual advertisements or advertisements affiliated with an advertisement network, such as GoogleAds, BlogAds, or Law Blog Ad Network. Additionally, the archiving category was not coded during the blog analysis stage. Rather, it was coded when deriving the blogger sample. All 93 blogs in the completed sample have an archive system, as this was a requirement for eligibility in the overall study. These blogs were not re-coded to confirm presence of an archiving feature as the time lapse between sampling frame construction and blog analysis coding was only 30 to 45 days.

Table 112 shows the frequency of blogs with these features from all blogs in the sample. All blogs are identifiable by blog name on the first page of the blog, as all feature a header providing the blog title. Further, most (74%) include supplemental title information in a tag field, including personal identifiers like blogger name, title, or institutional affiliation, and scope-related information. A majority feature blogrolls (74%) and categories (66%), which may aid in access and use as it provides additional contextual information.

**Table 112. Frequency of blog user interface elements and other features**

Element	Frequency	Percentage
<b>Title: Header</b>	93	100%
<b>Title: Tag</b>	69	74%
<b>Archiving system</b>	93	100%
<b>Categories</b>	61	66%
<b>Blogroll</b>	69	74%
<b>Search tool</b>	66	71%
<b>Commercial advertisements</b>	28	30%

*Note.* Concerning commercial advertisements, included is one blog that featured ads at the post-level instead of the sidebar, where all other ads were located. While this particular instance represents a post-level feature, it is included among these blog-level results because it was the only instance of advertisements at the post-level. No category concerning commercial advertisement is provided in Section 5.4. (Post Elements and Features).

In consideration of blogrolls, the number of links in each respective blog's blogroll was tallied. Table 113 shows frequency of blogs listed to blog rolls across five count intervals for the sixty-nine blogs in the sample coded as having a blog roll.

**Table 113. Frequency of blogroll links**

<b>Links</b>	<b>Frequency</b>	<b>Percentage</b>
<b>≤ 9</b>	8	12%
<b>10-24</b>	19	28%
<b>25-49</b>	19	28%
<b>50-99</b>	14	20%
<b>≥ 100</b>	9	13%

### 5.3.2. Application Software

Information on blog application was collected for each blog in the sample. This information was gathered from a review of the first page of the blog and/or the blog URL. Table 114 shows the frequency of applications for all blogs in the sample. Two application categories in the table merit clarification. For blogs coded as using a WordPress application, these represent both WordPress.com blogs and WordPress.org blogs.<sup>66</sup> Additionally, the category, "WordPress theme," represents those blogs that feature a WordPress theme from a third-party provider.

**Table 114. Frequency of blog application**

<b>Blog Application</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Blogger</b>	35	38%
<b>WordPress</b>	23	24%
<b>WordPress theme and other application/service</b>	5	4%
<b>TypePad</b>	9	10%
<b>Movable Type</b>	7	8%
<b>LiveJournal</b>	1	1%
<b>blog.co.uk</b>	1	1%
<b>Unknown</b>	14	15%

Blogger and WordPress are the most popular applications, representing 62% of all blogs in the sample. For 14 blogs (15%), application was not identified from a review of the blog's first page or by a reference in the URL.

---

<sup>66</sup> To aid in distinction between these two WordPress designations, simply WordPress.org is a blog publishing application, and WordPress.com is a blog publishing and hosting service.

### 5.3.3. Network Affiliation

Information on membership in a network was collected for each blog in the sample.

Networks are distinguished as aggregates of blogs with a shared host location, typically organized by topical treatment or institutional affiliation, as well as commonly sponsored by an organization, such as a publisher or a university. Eighteen blogs in the sample (19%) were coded as belonging to a network. Eleven unique networks were identified. Table 115 shows the frequency of network affiliation for these eighteen blogs. ScienceBlogs and the Law Professor Blogs Network are the most represented, each with five blogs in the sample.

**Table 115. Frequency of blogs network affiliations**

<b>Network</b>	<b>Frequency</b>
<b>ScienceBlogs</b>	5
<b>Law Professor Blogs Network</b>	5
<b>Kellogg Faculty Blogs</b>	2
<b>Discover Blogs</b>	1
<b>George Mason University's History News Network</b>	1
<b>Nature Network Blogs</b>	1
<b>blogs.worldbank.org</b>	1
<b>ScienceForums.Net</b>	1
<b>Weblogs at Harvard Law School</b>	1

Additionally, network affiliation is shown by cluster in Table 116. More blogs in the BioChemPhys and law clusters are network-affiliated than blogs in the economics and history clusters.

**Table 116. Frequency of network affiliation for blogs by cluster**

<b>Cluster</b>	<b>Frequency</b>
<b>History</b>	1
<b>Economics</b>	3
<b>Law</b>	6
<b>BioChemPhys</b>	8

### 5.4. Post Elements and Features

This section reports on frequency of select post elements and features for all single- and co-blogs in the sample. Posts were coded as having or not having five features and

elements: 1) post title field; 2) commenting system; 3) categories; 4) permalink; and 5) syndication feeds.

For clarification, “categories” are labels assigned to posts, generally found in the post header or footer, and may also be identified by other headings, including keywords, labels, or tags. “Permalink” also includes posts labeled as “permanent links.” Links identified as “trackbacks” were not included in this category. Syndication includes RSS, atom, or other syndication feeds. Syndication indicators (i.e., an icon or “subscribe” widget) may be found in the post header or footer and/or the blog footer or sidebar. Further, syndication may allow subscription to feeds for posts, posts and comments, or comments only. The type or level of syndication was not captured in this analysis. All blogs were coded as having syndication if a syndication icon, text statement, or subscription widget, whether at the post- or blog-level, was present.

Table 117 shows the frequency of blogs with these post features for all blogs in the sample. All , or nearly all , have categories (100%), syndication services (98%), and commenting systems (90%). One in four blogs make available permalinks or permanent links at the post-level.

**Table 117. Frequency of blogs post elements and features**

<b>Element</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Title</b>	93	100%
<b>Commenting system</b>	84	90%
<b>Categories</b>	71	100%
<b>Permalink</b>	23	25%
<b>Syndication</b>	91	98%

### 5.5. Posting Activity

While information on posting activity at time of sampling was captured, information on date of most recent post was also captured when coding for the blog analysis portion of the study. Date of most recent post was subtracted from the date the blog was coded for

analysis. This was done to provide an indicator for posting activity at time of coding, as well as to provide some insight into accuracy of blogs as current as determined at sampling.

Concerning date of most recent post published to the blog at the time of coding, seven intervals of time were derived: 1) 0 days ago, meaning that blog was posted to on same date as it was coded; 2) one to three days ago; 3) four to six days ago; 4) seven to thirteen days ago; 5) two to four weeks ago; 6) one to two months ago; and 7) more than two months ago.

Table 118 shows frequency of recent posts across these seven intervals. This is also presented by cluster in Table 119. At the time of coding, 86% of blogs had been published to within the previous two weeks, and 78% within the previous week. This is nearly the same percent as reported in the questionnaire portion of the study: 77% of questionnaire respondents reported updating their blogs one or more times a week.

**Table 118. Frequency of recent posting**

<b>Time Lapse</b>	<b>Frequency</b>	<b>Percentage</b>
<b>0 days ago</b>	23	25%
<b>1 to 3 days ago</b>	36	39%
<b>4 to 6 days ago</b>	13	14%
<b>7 to 13 days ago</b>	7	8%
<b>2 to 4 weeks ago</b>	9	10%
<b>1 to 2 months ago</b>	4	4%
<b>More than 2 months ago</b>	1	1%

*Note.* While all blogs were screened for language, with publication in English a requirement for participation, one blog in the competed sample was found to have published the most recent post in French at time of coding. This blog was still retained in the sample since, from a thorough inspection during coding, a majority of blog content was in English. Recent posting for this blog was calculated based on most recent post published in English. A more recent post had been published in French, but since content of post could not be analyzed in-line with other post-level coding categories reported, this post was not analyzed.

**Table 119. Frequency of recent posting by cluster**

<b>Time Lapse</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>0 days</b>	3 (13%)	9 (36%)	8 (33%)	3 (15%)
<b>1 to 3 days ago</b>	8 (35%)	9 (36%)	11 (46%)	8 (40%)
<b>4 to 6 days ago</b>	3 (13%)	4 (16%)	1 (4%)	5 (25%)
<b>7 to 13 days ago</b>	3 (13%)	1 (4%)	2 (8%)	
<b>2 to 4 weeks ago</b>	3 (13%)	2 (8%)	1 (4%)	3 (15%)
<b>1 to 2 months ago</b>	2 (9%)	-	1 (4%)	1 (5%)
<b>More than 2 months ago</b>	1 (4%)	-	-	-

*Note.* See note posted at Table 119.

Additionally, information on the total number of posts in a complete one-month period preceding the date of coding was also collected. Across all blogs, 2,726 posts were published between August 1-31, 2010. Descriptive measures on these posts were calculated as follows: mean 29; median 14; mode 0; and range 324. Additionally, descriptive measures by cluster are presented in Table 120. Based on the mean, about thirty posts are published to blogs in the sample each month, though in recognition of a range between 0 to 324 posts.

**Table 120. Descriptive measures of posts published to all blogs between August 1-31, 2010 by cluster**

Measure	History	Economics	Law	BioChemPhys
Mean	8	32	53	22
Median	6	20	31	14
Mode	0	2	5	4
Range	24	175	321	110

## 5.6. Post Content

The most recent post published to each blog at the time of coding was evaluated in regard to content and commenting. Concerning content, posts were coded for word count, embedded or imported content, links, and comments.

### 5.6.1. Text

All words in the main entry body of the most recent blog post were counted. Any text affiliated with the post, but not present in the main entry body, such as title, author, date, and other footer or header fields, were excluded from this count. Text in the entry body was coded at three levels: total words, quoted words, and original words. This was achieved through cutting and pasting the text of the entry body into a MS Word document to identity total word count. From a visual inspection, text in the entry body was considered to be quoted from another source if the text was distinguished by quote marks, italics, indentation, or highlighting, or if text was an excerpt accompanied by a link(s) to original source(s). This text was isolated and word count was derived. While the text is considered quoted and

originating from another source that does not necessarily mean from another author. As will be described later in this section when describing link characteristics, some posts may include internal links to other entries posted to the same blog. Hence, the quoted content may originate from another source but does not necessarily, in all cases, originate from another author.

Text assumed to be original, in that it was not identified as quoted text, was calculated by subtracting the count of quoted words from all words in the main entry body.

For the 93 posts coded, 50,565 total words were counted, with 7,361 of these identified as quoted. Fifty-seven posts (60%) were identified as having no quoted words. Based on total word count and quoted word count, 43,204 words were assumed to be original to the 93 posts. Descriptive measures on the three levels of word count are presented in Table 121 below. Also, Table 122 shows frequency of word count for these 93 posts by ten count intervals, ranging 0 to 3000 or more words. On average, the most recent posted entry contained 465 words.

**Table 121. Descriptive measures of total, quoted, and original word count for most recent blog posts**

Measure	Total Words	Quoted Words	Original Words
Mean	544	79	465
Median	254	0	254
Mode	26	0	26
Range	4,400	860	3,540

**Table 122. Frequency of most recent posts' total, quoted, and original word count**

Word Count	Post Total Words Frequency (%)	Post Quoted Words Frequency (%)	Post Original Words Frequency (%)
0	1 (1%)	56 (60%)	1 (1%)
1-49	9 (10%)	5 (5%)	9 (10%)
50-99	9 (10%)	6 (6%)	9 (10%)
100-199	19 (20%)	13 (14%)	19 (20%)
200-299	14 (15%)	8 (9%)	14 (15%)
300-499	11 (12%)	2 (2%)	19 (20%)
500-999	15 (16%)	3 (3%)	9 (10%)
1000-1999	11 (12%)	-	9 (10%)
2000-2999	1 (1%)	-	2 (2%)
≥ 3000	3 (3%)	-	2 (2%)



Additionally, word count means at the three levels is also presented by cluster, as shown in Table 123. Posts in the history cluster tend to have more total and original words than posts in the other clusters. Posts in the law cluster tend to have the fewest words in regard to total and original word counts, but tend to have more quoted words than posts in the other clusters.

**Table 123. Means of total, quoted, and original words for most recent blog posts by cluster**

<b>Word Count</b>	<b>History Mean</b>	<b>Economics Mean</b>	<b>Law Mean</b>	<b>BioChemPhys Mean</b>
<b>Total</b>	700	525	458	493
<b>Quoted</b>	90	59	120	44
<b>Original</b>	610	466	338	449

#### 5.6.2. Other Content Types

In addition to examination of text as a content type, seven other content types within the main entry body for the most recent post were coded: 1) photo elements; 2) static graphic elements; 3) interactive graphic elements; 4) other image elements (e.g., clip art, drawings, screen shots, etc.); 5) video elements; 6) audio (music) elements; and 7) audio (spoken) elements. This content may be embedded or imported and may originate from the blog author or somewhere else. No action was taken to identify where content originated; posts were only coded to reflect if they contained any of these elements and, if so, how many.

For the most recent posts in the 93 sampled blogs, 62 (67%) contained none of these elements; the only content type present for these particular posts was text. Thirty-one (33%) posts contained between 1 and 18 of these elements. Table 124 shows the frequency of all content types, including text, contained within blogs' most recent posts. All but one post contain text; the post with no text contains a photo. Photo elements (15%) and other image elements (16%) are the most represented content types, after text (99%).

**Table 124. Frequency of content types in most recent blog post**

<b>Content Type</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Text</b>	92	99%
<b>Photo elements</b>	15	16%
<b>Static graphic elements</b>	4	4%
<b>Interactive graphic elements</b>	-	-
<b>Other image elements (e.g., clip art, drawings, screen shots)</b>	16	17%
<b>Video elements</b>	2	2%
<b>Audio (music) elements</b>	-	-
<b>Audio (spoken) elements</b>	1	1%

### 5.6.3. Links

Next, text-based hyperlinks within the main entry body were counted. Across the most recent post for the 93 blogs in the sample, 76 (82%) contained at least one text-based hyperlink. Links were further characterized and coded as either internal to the blog or external to the blog. Table 125 shows frequency of posts with hyperlinks. Additionally, some links were “bad,” characterized as leading to a 404 or similar error message, and did not allow for determination of whether the link was internal or external. Those links were also coded and are presented as “bad links” in Table 125. Ultimately, internal or external linkages could be determined for 73 posts (78%) in the sample. Links for embedded or imported content elements, as described in Section 5.6.2., are not included in these link counts.

**Table 125. Frequency of text-based hyperlinks in most recent blog post**

<b>Text-Based Links</b>	<b>Frequency</b>	<b>Percentage</b>
<b>Internal links</b>	19	20%
<b>External links</b>	70	75%
<b>Bad links</b>	3	3%
<b>No links</b>	17	18%

For the 73 posts containing links, excluding those with bad links, only three (4%) contained internal links exclusively. Links in 54 posts (74%) were external only. Sixteen posts (22%) contained both internal and external links.

Additionally, descriptive measures for all 73 posts containing internal and external text-based links are provided in Table 126. Based on the mean, these posts contain about five internal and external links on average, though in recognition of a range between 1 to 42 links.

**Table 126. Descriptive measures of internal and external links in most recent post**

Measure	Internal Links	External Links	Internal & External Links
Mean	4	4	5
Median	2	2	2
Mode	1	1	1
Range	23	38	41

#### 5.6.4. Comments

For the 84 blogs in the sample with commenting systems, as reported in Section 5.3.1., comment activity for the most recent post was captured. Table 127 shows frequency of recent posts with comments. A majority of posts (57%) had no comments. Fourteen posts (17%) had been commented on six or more times. These comment counts reflect all comments affiliated with the respective posts. No further action was taken to distinguish unique commentators among all comments identified.

**Table 127. Frequency of comments to most recent post**

Comments	Frequency	Percentage
0 comments	48	57%
1 comment	6	7%
2 to 5 comments	16	19%
6 to 10 comments	8	10%
11 to 19 comments	4	5%
≥ 20 comments	2	2%

#### 5.7. Rights, Policies, and Disclaimers

All blogs in the sample were examined in line with issues related to copyright as well as policies or disclaimer statements in regard to blog content, commenting, and post editing.

##### 5.7.1. Rights and Use Statements

From a review of blogs, two broad types of rights and use statements were identified: Creative Commons licenses, and text-based rights statements, both with and without the copyright formality (e.g., “© Copyright All Rights Reserved. Contact post author for permission.”). Location information was also collected for these statements (i.e., sidebar, footer, “about” page).

Table 128 shows frequency of rights statements by broad type for all blogs in the sample. Since eight blogs feature two or more statements, frequency, as reported in Table 128, reflects this. For 47 blogs (51%) rights statements were not identified at either the blog or post level. Additionally, Table 129 shows frequency of rights statement by cluster. Blogs in the BioChemPhys (71%) and law (58%) clusters had a higher percentage of rights statements attached to their blogs than blogs in the history (35%) and economics (36%) clusters.

**Table 128. Frequency of rights and use statements by type for all blogs**

Type	Frequency	Percentage
None	47	51%
Creative Commons	13	14%
Text statement	34	37%

**Table 129. Frequency of rights and use statements by cluster**

Rights and Use Statement	History Freq (%)	Economics Freq (%)	Law Freq (%)	BioChemPhys Freq (%)
0	15 (65%)	16 (64%)	10 (42%)	6 (29%)
1 or more	8 (35%)	9 (36%)	14 (58%)	15 (71%)

For the 46 blogs (49%) in the sample with some type of rights statement identified, Table 130 shows rights statement by both type and location. The most occurring rights statement type and location is a text-based statement in the footer (50%), followed by a text statement in the sidebar (22%). In regard to just type, 14 Creative Commons licenses were identified (30%), found across four locations.

**Table 130. Frequency of rights statement type and location for all blogs**

Type and Location	Frequency	Percentage
Creative Commons license: Sidebar	8	17%
Creative Commons license: Footer	4	9%
Creative Commons license: "Policy" or similar page	1	2%
Creative Commons license: "About" or similar page	1	2%
Text statement: Sidebar	10	22%
Text statement: Footer	23	50%
Text statement: "About" or similar page	5	11%
Text statement: "Policy" or similar page	1	2%

Extent of coverage of these explicit rights and use statements was also coded. This was done to determine if rights statements posted to blogs distinguished between different levels of content (e.g., blog, post, and comments). Three categories of extent were derived: 1) all right reserved; 2) some rights reserved, and 3) no specification of extent.

To clarify, statements directly stating, “all rights reserved,” were coded as such. Additionally, statements identifying blog (i.e., © Blog Name), name of blogger(s), “this work,” and/or reference to content (i.e., “all content,” “text on this blog,”) were also coded for extent as “all rights reserved.” Also included are blogs with Creative Commons licenses where all rights might not necessarily be reserved, based on license type, but the blog in its entirety is subject to the terms of the respective license. For blogs with statements that identify specific content on the blog as reserved, these were coded as “some rights reserved.” Those with no extent information were coded as such.

Nearly all statements (97%) reflected all right reserved or, in consideration of Creative Commons licenses, all content subject to license. Only two blogs’ statements were not coded as such; one was coded as “some rights reserved,” and the other had no information on extent.

#### 5.7.2. Disclaimers

A review of a blog’s first page, “about” or similar page, and/or “policy,” “use,” or similar page was done to identify either explicit or implicit disclaimer statements or policies. Explicit is characterized as a specific, identified statement, policy, or page concerning responsibility for content. Implicit is characterized as a statement on responsibility not explicitly identified, but nonetheless present. For example, a paragraph on

responsibility included in a broad description of a blog’s overall scope and intent, as found at an “about” page, would be characterized as implicit.

Table 131 shows frequency of disclaimer statements, either explicit or implicit, for all blogs in the sample. Implicit or explicit disclaimer statements were not identified for 69 blogs (74%). Eleven (12%) had explicit disclaimers and 13 (14%) were found to have implicit disclaimers. Additionally, Table 132 shows frequency of disclaimer statements by cluster. Blogs in the BioChemPhys (38%) and economics (36%) clusters had a higher percentage of disclaimer statements attached to their blogs than blogs in the law (21%) and history (9%) clusters.

**Table 131. Frequency of disclaimer statements for all blogs**

<b>Disclaimer Statement</b>	<b>Frequency</b>	<b>Percentage</b>
<b>None</b>	69	74%
<b>Explicit</b>	11	12%
<b>Implicit</b>	13	14%

**Table 132. Frequency of disclaimer statements by cluster**

<b>Disclaimer Statement</b>	<b>History Freq (%)</b>	<b>Economics Freq (%)</b>	<b>Law Freq (%)</b>	<b>BioChemPhys Freq (%)</b>
<b>None</b>	21 (91%)	16 (64%)	19 (79%)	13 (62%)
<b>Explicit or implicit</b>	2 (9%)	9 (36%)	5 (21%)	8 (38%)

Information on location of explicit disclaimer statements was also captured. Location and frequency for the 11 blogs with explicit statements are: disclaimer and/or terms of use page (n=4); sidebar (n=5); and footer (n=2).

The 24 explicit and implicit disclaimer statements were reviewed to assess descriptive content of statements. Eighteen (75%) made clear that opinions expressed were blogger(s)’ own and did not represent those of others. Others specifically mentioned include bloggers’ employers, universities, funders, blog network or other sponsoring organization, or, as found on one blog, even the blogger’s own previously held opinions. Six statements (25%) made clear that blog content did not constitute advice, whether medical, legal, investment, or

otherwise. Four statements (17%) addressed responsibility for content, whether on blog or content of linked resources, and stated that blogger(s) were not responsible for content, do not guarantee it is correct, and/or do not necessarily agree with it. One statement also addressed issues of confidentiality, stating that there was no guarantee communications made with blogger would remain confidential. Additionally, another statement addressed issues of privacy, stating that person(s) described on the blog presented composites of people, unless person(s) specifically identified.

### 5.7.3. Commenting Policy

Next, a review of a blog's first page, "about" or similar page, and/or "policy," "use," or similar page was done to identify either explicit or implicit commenting policies. Explicit is characterized as a specific, identified statement, policy, or page concerning blog commenting. Implicit is characterized as a statement on commenting not explicitly identified, but nonetheless present. For example, a paragraph on commenting guidelines included in a broad description of a blog's overall scope and intent, as found at an "about" page, would be characterized as implicit. Statements that only encourage comments (i.e., "feel free to submit comments") are not considered commenting policies or statements, and were not coded as such.

Table 133 shows frequency of disclaimer statements, either explicit or implicit, for all blogs in the sample. Seventy-six blogs (82%) do not have either an implicit or explicit commenting policy associated with their blog.

**Table 133. Frequency of commenting policies for all blogs**

<b>Commenting Policy</b>	<b>Frequency</b>	<b>Percentage</b>
<b>None</b>	76	82%
<b>Explicit</b>	8	9%
<b>Implicit</b>	9	10%

Information on location of explicit commenting policies was also captured. Location and frequency for the eight blogs with explicit statements are: commenting policy or related policy page (n=5) and sidebar (n=3).

The 17 explicit and implicit commenting policies were reviewed to assess descriptive content of statements. While themes emerged relating to several issues, including moderation of comments (i.e., reviewing, editing, deleting) and acceptable tone and language in support of a civil and respectful discourse (i.e., no cussing, no name-calling, no spam), one specific theme is addressed here concerning comments and intellectual property rights.

No explicit rights and use statements, as described in Section 5.7.1, specifically identified comments in regard to rights though, from references to “this work” or “all rights reserved,” it might be assumed rights’ statements extend to contributions from commentators (or ... it might also be assumed that they do not). From a review of all commenting policies, both implicit and explicit, three (18%) addressed comments and intellectual property rights. Two are very brief, with one noting that comments, “go to the public domain,” and the other that comments remain, “the property of their respective authors.” The third is much more extensive. Per the detailed commenting policy, all comments are public and commentators, through submitting comments, agree to a, “perpetual, royalty-free license to use, distribute, reproduce, edit, and publish this information in other formats.” Further, commentators agree that the content does not infringe on any, “third party’s copyright or other proprietary rights.”

#### 5.7.4. Blog Editing Policy

Next, a review of a blog’s first page, “about” or similar page, and/or “policy,” “use,” or similar page was done to identify either explicit or implicit post editing policies or statements. For the 93 blogs in the sample, only two included such a statement. For one of



these blogs, the statement is included in a dedicated “policies” page. It described conditions under which posts would be edited with no acknowledgement of change (i.e., spelling errors or other style-related edits). If post is determined to be “too misleading to leave posted,” or, “if we change our minds about something, we will fess up or just move on, but we won't be sneaky about it.” Otherwise, posts will remain posted, even if, “we later regret it.”

For the other blog, the policy is listed in the blog sidebar. It informs readers that any change to a published post is listed at the bottom of the respective post. Two exceptions apply: if post was very recently published (i.e., if need for edit discovered immediately after publishing post to blog); and if change is made simply to revise labels assigned to the respective post.

During the interview portion of the study, one tenure track assistant professor of business and economics, while not commenting directly on editing blog posts, did mention the benefit of ease of editing in the digital medium, illustrating his comments by reference to an online magazine to which he contributes. His comments are in-line with one of aspect of the blog post editing policy described immediately above:

One of the things I like about the digital medium is it allows for very easy correction of mistakes. If something is in print, it's in print. If you make a mistake in an online piece, you can correct it, add a note at the bottom ... people then can not only have a record of what you actually said, but if they can get ... notes on mistakes and comments and what not, they actually have a record of the process by which the ideas developed.

#### 5.8. Age, Access, and Authority

This section reports results concerning several aspects of sampled blogs intended to inform considerations of authority, audience, and availability.

### 5.8.1. Blog Age

All blogs in the sample, per a condition for inclusion in the study as detailed in Section 3.1.2.1, are at least one year old. Additional information was collected at the blog analysis stage, including date of oldest blog post. This date was identified from the earliest post accessed via a blog's archive. Blog age, reported by year, is calculated as the number of years between date of oldest blog post and date at time of coding. For example, if a blog's oldest post is dated July 30, 2002 and the blog was coded for analysis on September 16, 2010, then blog age is calculated as eight years. Table 134 shows frequency of blog age for all blogs in the sample, and Table 135 shows blog age by cluster.

**Table 134. Frequency of blog age by authorship**

<b>Years</b>	<b>Single-Bloggers Frequency (%)</b>	<b>Co-Bloggers Frequency (%)</b>	<b>All Bloggers Frequency (%)</b>
<b>1</b>	2 (3%)	3 (9%)	<b>5 (5%)</b>
<b>2</b>	3 (5%)	2 (6%)	<b>5 (5%)</b>
<b>3</b>	11 (18%)	2 (6%)	<b>13 (14%)</b>
<b>4</b>	11 (18%)	6 (19%)	<b>17 (18%)</b>
<b>5</b>	17 (28%)	12 (38%)	<b>29 (31%)</b>
<b>6</b>	10 (16%)	4 (13%)	<b>14 (15%)</b>
<b>7</b>	5 (8%)	3 (9%)	<b>8 (9%)</b>
<b>8</b>	2 (3%)	-	<b>2 (2%)</b>

**Table 135. Frequency of blog age by cluster**

<b>Years</b>	<b>History Frequency (%)</b>	<b>Economics Frequency (%)</b>	<b>Law Frequency (%)</b>	<b>BioChemPhys Frequency (%)</b>
<b>1</b>	1 (4%)	2 (8%)	-	2 (10%)
<b>2-3</b>	7 (30%)	6 (24%)	2 (8%)	3 (14%)
<b>4-5</b>	7 (30%)	11 (44%)	15 (63%)	13 (62%)
<b>≥ 6</b>	8 (35%)	6 (24%)	7 (29%)	3 (14%)

The oldest blog age is eight years. The average age for all blogs in the sample is 4.55 years. Blogs in the law cluster are older than those in other clusters, with 92% ages four to eight years old. Overall, blogs in the sample can be characterized as established. Only ten (10%) are between one and two years old. A majority (57%) are between five and eight years old.

### 5.8.2. Blog and Page Views

The 93 blogs in the sample were reviewed to identify if a specific blog statistics widget, Site Meter (<http://www.sitemeter.com>), was available on the first page of the blog. While other statistical widgets are available (e.g., PostRank, Quantcast, StatCounter, etc.), only Site Meter was considered for this analysis. This was done intentionally for homogeneity in blog statistics reported. Additionally, from a preliminary review of blog instances done to inform population source selection at the very early stages of this research's design, Site Meter was observed to be a popular choice among those instances. Forty-two blogs in the sample (45%) have blog statistics available via Site Meter. Reported here, as shown in Table 136, are descriptive results from four measures as identified via Site Meter. These are provided to illustrate diversity in visits and page views among blogs in the sample, as is clear from looking at the range. This data is taken from another source, and while it may present blog visits and page views, that does not characterize how and who – or, in regard to robots and scripts, what – visited the blog or viewed the page.

**Table 136. Descriptive measures of blog visits and page views from Site Meter**

Measure	Total Visits	Avg. Visits Per Day	Total Page Views	Avg. Page Views Per Day
Mean	2,551,778	1,866	4,058,312	2,793
Median	434,164	243	731,061	372
Mode	-	41	-	372
Range	32,814,982	26,730	56,273,228	40,501

### 5.8.3. Blog Access and Indexing

While all blogs in the sample are publicly available, in line with the conditions established at sampling, four actions were taken to determine if blogs are “findable,” both in consideration of today and into the future. This was done through four off-site measures using two indexing and search services, Google’s Advanced Blog Search and Technorati’s blog search service, and two digital archive services, the Internet Archives Wayback

Machine and the Library of Congress' Legal Blawgs Web Archive. Reported here are results from searching these four sources.

#### 5.8.3.1. Google Blog Search

All blogs in the sample were searched using Google's Advanced Blog Search feature ([http://blogsearch.google.com/blogsearch/advanced\\_blog\\_search](http://blogsearch.google.com/blogsearch/advanced_blog_search)). Simply, the URL for each blog was entered in the section, "In Blog," and in the field, "at this URL." Ninety-one blogs in the sample (98%) were retrieved using this technique. Two blogs (2%) were not. For those two blogs, the blog title, as found in the respective blog's header field, was entered in the search field, "with these words in the blog title," and searched. They were still not retrieved, and no other search actions were taken. Hence, 91 blogs, or 98% of the completed sample, are considered to be indexed by Google.

#### 5.8.3.2. Technorati

Next, all blogs in the sample were searched using Technorati's blog search tool (<http://technorati.com/>). The blog URL was entered in the search box and the icon "blog," as opposed to post, was selected. If the blog was not retrieved, then the search was redone but using the blog's title instead of the URL. Ultimately, 72 blogs (77%) were retrieved. The two blogs not retrieved using Google's Advanced Blog Search were also not indexed at Technorati.

#### 5.8.3.3. Internet Archive Wayback Machine

Next, all blogs in the sample were searched in the Internet Archive Wayback Machine (<http://www.archive.org/web/web.php>). The blog URL was entered in the search box for all media types. Seventy-four blogs (80%) were retrieved. For the 19 blogs not found (20%), no

other actions were taken. Additionally, no other information was collected regarding these archived instances (i.e., dates of archived results).

Table 137 shows frequency of all blogs archived to the Internet Archive Wayback machine by cluster. All law blogs in the cluster are archived there, as are a majority of blogs in the other three clusters.

**Table 137. Frequency of blogs archived at the Internet Archive Wayback Machine by cluster**

Status	History Frequency (%)	Economics Frequency (%)	Law Frequency (%)	BioChemPhys Frequency (%)
Archived	20 (87%)	15 (60%)	24 (100%)	15 (71%)

#### 5.8.3.4. Library of Congress Legal Blawgs Web Archive

Next, all blogs in the sample assigned to the law cluster (N=24) were searched at the Library of Congress' Legal Blawgs Web Archive (<http://www.archive.org/web/web.php>). These were retrieved by using the "browse by title" feature. If the blog was not identified using this approach, then the archive was browsed by name(s) of the blog's respective author(s). No other actions were taken beyond this step. Ultimately, 12 law blogs, representing half of all blogs in the cluster, were identified. No other information was collected regarding these archived instances (i.e., date captured).

Table 138 shows the frequency of all blogs in the sample indexed by Google and Technorati, as well as those captured and indexed by the Internet Archive Wayback Machine and the Library of Congress's Legal Blawgs Web Archive.

**Table 138. Frequency of blogs indexed to Google, Technorati, Internet Archive and Library of Congress**

Indexing Service or Archive	Frequency	Percentage
Google	91	98%
Technorati	72	77%
Internet Archive Wayback Machine	74	80%
Library of Congress Legal Blawgs Web Archive	12	50%
None of these	2	2%

*Note.* The percentage for the Legal Blawgs archive is based on the completed sample for the law cluster (N=24); all other percentages are based on the completed sample across all clusters (N=93).

#### 5.8.4. Blog Authority

Several actions were taken during coding to assess blog authority and recognition. This entailed one on-site measure (awards listed to blog), and two off-site measures (Technorati Authority number and Technorati Top 100 ranking). While there are blog ranking sources available,<sup>67</sup> only rankings from Technorati were coded. This was due to extent of Technorati's index of blogs and explicit information on calculating a blog's authority.

##### 5.8.4.1. Technorati Authority Number and Top 100 Ranking

For all blogs in the sample indexed at Technorati (N=72), the blog's "site details" page at Technorati was reviewed. For each listed blog, Technorati Authority number was recorded. Per the Technorati Authority FAQ, authority, "measures a site's standing and influence in the blogosphere," and is, "calculated based on a site's linking behavior, categorization and other associated data over a short, finite period of time."<sup>68</sup> Authority numbers are scalar, ranging from 0 to 1000, with 0 being the lowest authority score and 1000 the highest. Per Technorati's authority ranking system, the higher a blog's authority number, the more authority.

For blogs in the sample indexed at Technorati, authority scores ranged from 1 to 757. Table 139 shows frequency of Technorati Authority scores. Slightly more than half (52%) have low authority scores, characterized as below 200. A little over three out of ten (32%) have higher authority scores, characterized as between 500 and 757.

---

<sup>67</sup> For example, see Wikio's Top Blogs, available at <http://www.wikio.com/blogs/top>

<sup>68</sup> Technorati Authority FAQ: <http://technorati.com/what-is-technorati-authority/>

**Table 139. Frequency of Technorati Authority scores for blogs indexed to Technorati**

<b>Score</b>	<b>Frequency</b>	<b>Percentage</b>
<b>0-99</b>	19	26%
<b>100-199</b>	19	26%
<b>200-299</b>	-	-
<b>300-399</b>	-	-
<b>400-499</b>	18	25%
<b>500-599</b>	11	15%
<b>600-699</b>	3	14%
<b>700-799</b>	2	3%

Information about each blog on their respective “site details” page was also reviewed to identify if any instances were included in the Technorati Top 100 Rank and/or Technorati Top 100 Topical Rank among the 1.2 million blogs indexed to Technorati. These rankings are based on Technorati Authority scores. None of the sample’s seventy-two blogs indexed at Technorati were listed to the Technorati Top 100 Rank. Six blogs (8%) were listed to the Technorati Top 100 Topical Rank, with two blogs listed to two Top 100 Topical groups. The topical areas for these six blogs, with representation by count presented in parentheses, are: Top 100 Business (n=3); Top 100 Science (n=2); Top 100 Health (n=1); Top 100 Finance (n=1); and Top 100 Green (n=1).

#### 5.8.4.2. Blog Awards

For on-site measures of blog authority, all blogs in the sample were reviewed and coded for any awards, characterized as formal recognition from a specific organization serving as the awarding group. Awards typically appear as badges in a blog’s sidebar. Also, reference to awards is also found as text statements in blogs’ sidebars as well as badges or text statements in other locations, including “about,” “profile,” “contact,” or similar pages. Excluded are badges which point to a blog’s “top” ranking from a blog indexing service, such as Technorati or Wikio. However, top rankings from non-indexing related services, like journals, are considered “awards” in this analysis.

Eighty blogs in the sample (86%) had no reference to any award. For the 13 blogs that did, the number of awards per blog ranged from one to eleven, though ten blogs listed just one award. Examples of awarding organizations and/or awards listed include: the American Bar Association (ABA) Journal “Blawg 100;” the Research Blogging Awards 2010, and the History News Networks’ Cliopatria Awards for History Blogging.



## **CHAPTER 6**

### **DISCUSSION**

This chapter discusses key findings and implications from the questionnaire, interview, and blog content analyses in response to the study's research questions and objectives. Presented first are considerations for the development of a profile of scholars who blog. Next, each research question, as presented in the Methods chapter, is addressed. This section concludes with discussions on the study's design and implementation, including recognition of its limitations.

#### **6.1. Scholars and their Blogs?**

This study relied on a big assumption going in: that is, the population under investigation were indeed scholars. Extensive consideration was paid at the sampling stage to confirm scholar status among participants, characterized by occupation, education, research area and other combinations of attributes. Findings from the questionnaire and interview portions of the study confirmed this. Respondents have, at minimum, a graduate-level degree, while a majority hold a doctorate. A majority are tenured, post-secondary faculty employed at the associate professor rank or higher. All have published and provided service. A majority have published to more than seven or more "unit-types" of scholarly communications, including peer reviewed journal articles, book chapters, and papers in conference proceedings. Most have served as a referee for a scholarly journal and evaluated a book or manuscript for a publisher. Additionally, most blog under their full names, characterized, at minimum, as first and last name. The average blog age is 4.5 years old.

For faculty respondents, average professional age is fifteen years, while average real age is forty-five. While these are averages, the range shows representation at all age levels, both professional age – ranging from 0 to 39 years – and real age – ranging from age 25 to 75 years old.

So then, who is a scholar blogger? Based on demographic characteristics of respondents to this survey, with representation across all levels of professional age, real age, rank, tenure status, publication, and service history, he may be any one of your colleagues. And “he” here is intentional because a prevalent demographic characteristic is gender: a clear majority are men (which, depending on your discipline and institution, may also be prevalent among your immediate colleagues).

Table 140 presents dominant characteristics of sampled bloggers, as described above, as well as characteristics of sampled blogs, by cluster.

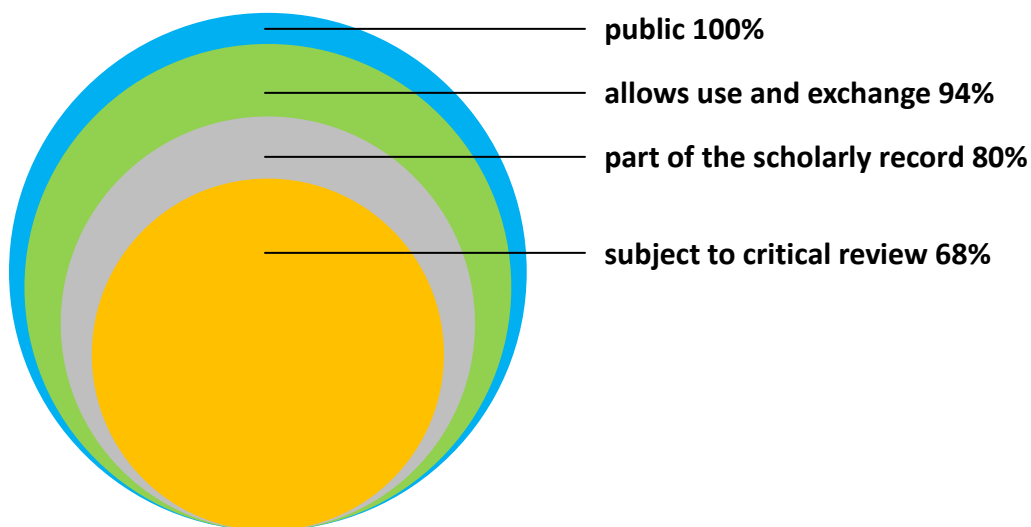
**Table 140. Characteristics of majority of sampled blogs and bloggers by cluster**

Characteristic	History	Economics	Law	BioChemPhys
<b>Avg. Professional Age</b>	10.5 years	18 years	16 years	12.5 years
<b>Avg. Real Age</b>	48 years	48 Years	44 years	41 years
<b>Highest Degree Earned</b>	Ph.D.	Ph.D.	J.D.	Ph.D.
<b>Gender</b>	♂71%   ♀29%	♂95%   ♀5%	♂70%   ♀30%	♂78%   ♀22%
<b>Occupation</b>	Faculty	Faculty	Faculty	Faculty
<b>Rank</b>	Tenured	Tenured	Tenured	Tenured
<b>Blog Age</b>	4.70 years	4.24 years	5.13 years	4.10 years
<b>Blog Application</b>	Blogger or WordPress	Blogger or WordPress	Blogger or TypePad	Blogger or Network
<b>Blogging Identity</b>	Full Name	Full Name	Full Name	Full Name
<b>Avg. Monthly Posts</b>	8	32	53	22
<b>Avg. Original Word Count/Post</b>	610	466	338	449

## 6.2. Blogs and the ‘Scholarly Record’

The first research question asked: how do scholars who blog perceive their blog in relation to their cumulative scholarly record? A majority agree their blogs are a part of their scholarly record. They also feel their blog satisfies the criteria for scholarship for unpublished scholarly outcomes and publications, as presented by Braxton et al. (2002), citing Shulman & Hastings (1999): the blog is public; the blog is subject to critical review; and the blog is in a form that allows for use and exchange by other members of the scholarly community. Agreement is summarized in Figure 2.

**Figure 2. Perceptions of blogs in relation to the scholarly record and scholarship criteria**



Further, in response to a series of questions in both the questionnaires and interviews intended to gauge perceptions of blogs in relation to the scholarly objectives for research, teaching, and service, a majority see their blogging as benefiting eight of ten aspects of their scholarly life, contributing to a sense of improvement in their teaching, writing, research, and communications with peers. Blogging was also perceived to contribute to invitations to publish to ten different communication unit-types, with two – scholarly paper for a

conference and non-fiction essay, article or op-ed piece – selected by a majority of questionnaire respondents. Respondents also perceive their blogs as having led to invitations to present, provide service, and collaborate.

Blogging, thus, can be seen to facilitate some traditional processes within the system of scholarly communication when viewed from the perspective of functions for registration, certification, and awareness, as presented by Roosendaal et al. (2001), and legitimization and dissemination, as summarized by Borgman (2007). In regard to audience, scholars blog with their community in mind and identify professional peers as their primary audience. Blogging can also be seen to contribute to perceptions of internal as well as external rewards (Murray and Moore, 2006), and institutionalized or elementary recognition (Hagstrom, 1965), whether directly or, in the case of invitations to publish and the like, indirectly.

#### 6.2.1. Characterizing the ‘Scholarly Blog’

The nature of this research leads to an obvious question: how can we move toward a genre-classification of scholar blogs for facilitating identification both within the blogosphere as well as in the academe? My response, as supported by remarks from some interviewees: Do we have to? In consideration of blog “typing,” researcher assessments of blogs by type are challenging. For example, Herring et al. (2005), in their blog analysis study, specified classification based on a “clear majority” and coded for “mixed” and “other” if type not clear (p. 147). As White and Winn (2009) found in their survey of bloggers in general, categories of type or blog “style” are not mutually exclusive. Several interviewee respondents presented their own classification schemes of scholarly or scientific blogs. Presented below are excerpts from three interviewees from the sciences, the humanities and the social sciences. (And, interestingly enough, each took a three-level approach to their classifications):

{Tenured professor in chemistry}: With science blogs ... I think there are more or less three kinds. There's sort of the 'reports on current research with some of my thoughts about it' blogs. There are blogs that are written by scientists but in fact are really not about science but about the doing of science in a particular time or place, about being a professor or assistant professor at a research I ... and don't really say much about the science they do. One can read it for some time and not even know if they are a biologist or a chemist. And then I think there are the people who are blogging about the science they are doing or the science their students are doing [references several of these bloggers]. So I think these are the three sort of types and I think they run as relatively independent threads. I think very few people kind of cross-read them and each of them think of themselves as science bloggers.

{A tenured history professor}: Some blogs seem to me to be pretty close to that [scholarly] in their character. Some academic blogs seem to me to be evolving toward being something else altogether. Something that engages, let's say, entirely new forms of creating authority or expertise, that are engaging some sort of crowd-sourcing network. And some blogs seem to me to be not at all attempting to be any form of expertise or scholarship even though they are written by academics and happen to concern, at times, academia.

{Tenured economics professor}: Some [have] short, almost bite-sized points to make, which is kind of what I put in my mine. There are others which are sort of medium length, weekly reports ... and there are still others which are quite lengthy and quite involved which are almost at the level of, I don't want to say a journal paper, but a journal communication ... shorter published pieces. So the range is quite large, and I think I take them each at face-level. In other words, each of them have their place as far as I'm concerned ... because otherwise it's apples and oranges.

Another way to consider types of blogs is to do it from the perspective of the benefits of blogging in consideration of other communication channels. Several interviewees stressed that a benefit of publishing to their blog is that it removes the restrictions and conventions of formal scholarly publishing. One tenured professor in history and American studies commented:

The academy is obsessed with rules. And part of what makes scholars different from other writers is the fact that we have all these rules [makes reference to publication conventions, like peer-review and gate-keeping]... one of the things about blogging that makes academics really nervous: how will their writing be perceived if it doesn't sort of pass through all of these gate-keepers, and how do you understand someone's writing if it hasn't passed through the gate-keepers?

A clear majority of questionnaire respondents reported that their blogs improved their creativity in examining research in new ways, their ability to share prepublication materials with colleagues, and their overall enjoyment of their work as a scholar. So then, an exercise in trying to define or standardize the form may actually contribute to restricting the form, potentially proving consequential for both blog authors and consumers.

One aspect of scholarly life – opportunity for promotion at one’s institution – was reported by questionnaire respondents as the least improved by blogging. An interviewee described the need for further understanding of how blogs can be seen to have some impact in the system of scholarly communication, whether as a, “contribution or as an academic output ... either formally or informally.” However, he also provided a caveat in reflection of this need:

If blogging was made a part of an academic’s output, or it was a measurable contribution, or even worse, it was mandatory for advancement, then what we would see is a huge rush to create a huge number of very bad blogs.

So, the take-away on what is a scholarly blog ... well, it depends. This leads to a reconsideration of categorizing scholars who blog. It has been intentional – and hopefully consistent – throughout this chapter and preceding chapters to refer to “scholars who blog” as opposed to “scholar bloggers.” Scholars have diverse reasons why they elect to blog and they have a diverse range of perceived benefits from blogging. What is a scholarly blog, and who is a scholar blogger, depends on how the respective blogger defines these for themselves and elects to fit it within their own professional and academic life. While findings from this study make clear that blogs are a unit and channel for communication within the system of scholarly communication, that does not necessarily mean that all see their blogs as scholarly. Interviewees made this clear: Some identify themselves as a scholar blogger or science

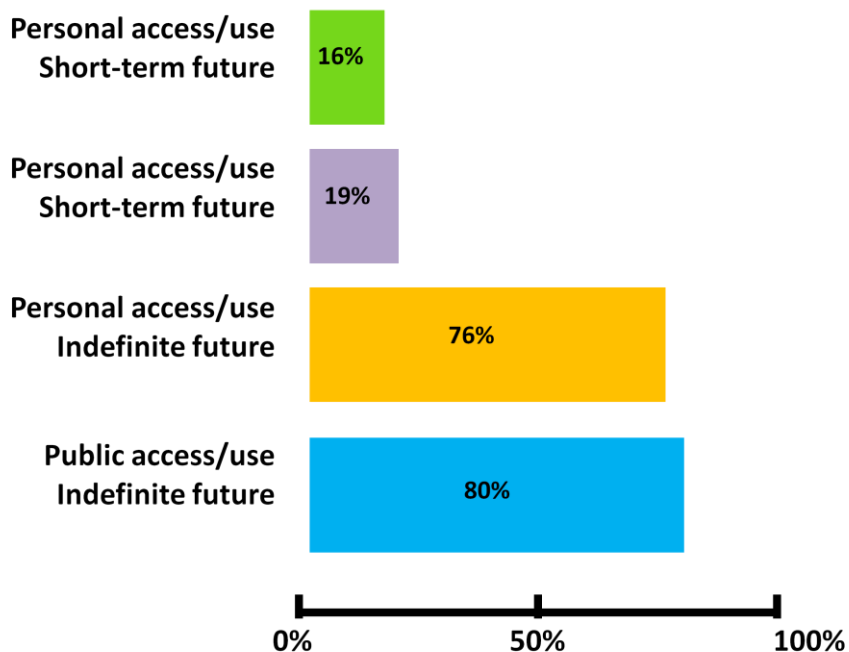
blogger, while others identify as a scholar who blogs. As one tenured history professor worded it:

Everybody has to make a kind of ad-hoc determination – everyone who’s blogging as a scholar – about what it is that they think they are doing and how it is that they want what they’re doing to be regarded because every single one of these things is different and there is no agreed upon standard and there is no institutional norm.

### 6.3. Perceptions and Preferences for Preservation

The second research questions asked: how do scholars who blog perceive their blog in relation to long-term stewardship and, subsequently, who, if anyone, do scholars perceive as responsible as well as capable for blog preservation? Concerning the former, over three out of four respondents to the questionnaire feel their blogs should be preserved, both for personal access and use as well as public access and use. They are also interested in doing so for the indefinite, rather than short-term, future. Figure 3 shows respondents’ perceptions across four levels of preservation.

**Figure 3. Agreement for blog preservation**

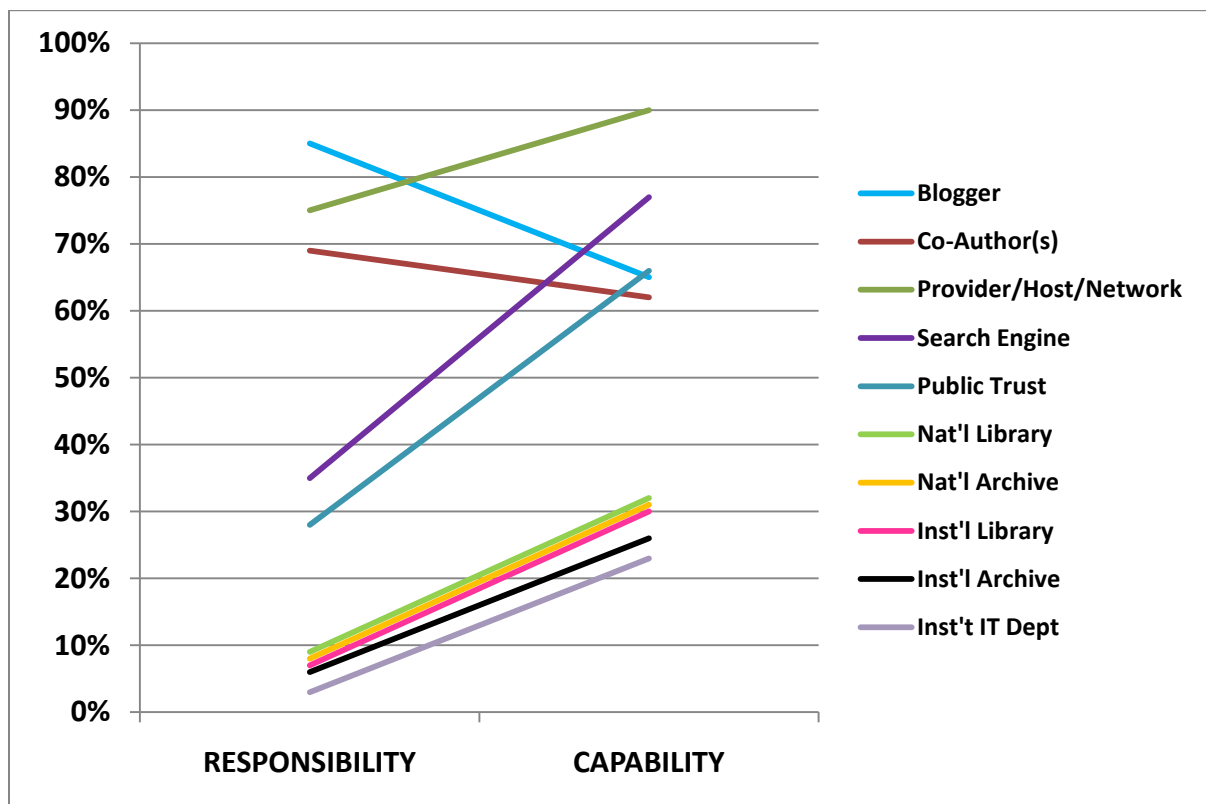


Concerning the latter, a clear majority of respondents feel themselves first and foremost as the most responsible for preservation of their blogs. In terms of capability, respondents feel the blog's service provider, hosting company, and/or network is most capable.

### 6.3.1. Personal Responsibility and Provider Capability

A fundamental principle of digital preservation is that content creators are the first line of defense (Garrett & Waters, 1995). In consideration of respondents' perceptions toward responsibility for blog preservation, it can also be seen that content creators are the last line of defense, as they feel a personal responsibility for preservation. Capability however, does not necessarily align. Figure 4 shows questionnaire respondents' perceptions of responsibility and capability across the ten actor groups presented in the questionnaires.

**Figure 4. Perceptions on responsibility and capability for blog preservation**





In regard to responsibility, the second and third groups perceived as most responsible are the blog's service providers, hosting companies, and networks, and the blog's co-authors. For capability, bloggers rank fourth, while co-authors, if any, rank fifth. Libraries, archives, and institutional IT departments are perceived as the least responsible and least capable. Interviewees were also invited to comment on the responses s/he indicated in the questionnaire. Findings from both of these data collection activities contribute to several implications for these various actor groups. These implications are described in Section 6.5.

#### 6.4. Characteristics and Behaviors Impacting Preservation

The third research question asked: How do blog characteristics and blogger publishing behaviors and preferences impact preservation? As reported in other studies on blog publishing (e.g., Lenhart, Fallows, & Horrigan, 2004; Lenhart & Fox, 2006; Rainie, 2005; Hank et al, 2007; Sheble et al, 2007; White & Winn, 2009), findings from this study confirmed the dynamic, co-produced, strongly contextualized nature of blogs. While scholars feel their blogs should be preserved into the indefinite future for private and public access and use, many factors may inhibit such activity when considered in-line with the goals for preservation (Caplan, 2008). However, findings also indicate that scholars are aware of issues impacting long-term management of their blog, and some take specific actions in response.

##### 6.4.1. Publishing Behaviors and Preferences

All blogs in the sample, as a condition of inclusion, are publicly available. As shown in the results, they are also findable since indexed to Google and Technorati. The blogs are regularly published to, with most scholars' blogging either more or about the same as when they first began to blog. Scholars not only publish to the blog identified at sampling, but for

more than four out of ten of questionnaire respondents, they also publish to other singularly- and collaboratively-authored blogs.

Scholars add new posts to their blogs, at minimum, several times a week. They also modify existing content, editing or deleting previously published posts, as well as changing the structure, or “skin” their blog. In consideration of the ease in which bloggers may modify content published to the blog, with posts a particular focus, only two blogs in the blog sample were found to have a versioning policy posted to their blog. Some blogs do feature explicit or implicit disclaimer-type policies or statements that address information posted to the blog, with most making clear the opinions and ideas published are the scholars’ own and not that of his/her employer, funder, network, or other affiliation-type. Some also make clear that information is not guaranteed to be correct or accurate, nor is it intended to serve as advice.

Blogs are highly contextualized, as seen at both the blog-level, via blogrolls and other linked content, and at the post-level, as demonstrated through imported content and links. While rich contextualization is valuable for immediate and short-term use and interpretation of the content, it may not prove viable into the future due to issues of link-rot and other factors which may obstruct access to or interpretation of contextual content or markers.

Content in the post’s main entry body, as found during the blog analysis portion of the study, is predominantly text-based. From a review of the most recent posts published to the blog, posts contained an average of 544 words, with 80 of these representing text originating from somewhere else. Scholars also import or embed photos, images and other elements. In terms of seeking permission to republish content to their blogs, scholars are pretty much split down the middle on how important or unimportant they feel it is to verify permission.

Scholars are also fairly evenly split in regard to how important they feel it is for others to seek permission to re-publish content posted to the scholars' respective blogs. About half of the blogs in the sample publish copyright right statements to their blogs, though rights and use licenses, such as a Creative Commons licenses, are less common. Nearly all blog under their real names. They also provide contact mechanisms, such as email. These two characteristics help to facilitate permission-seeking in regard to dissemination and re-use.

Additionally, scholars publish other personal identifiers to their blogs, such as occupation, employer, and research area. These overt identifiers facilitate some assessment of a blog's credibility and authenticity, as do other on-blog indicators, including award badges and blog statistical widgets posted to the sidebar. Further, using an off-blog indicator, for those blogs in the sample indexed to Technorati, a majority were found to have Technorati Authority Rank scores between 400 to 757.

#### 6.4.2. Co-Producer Dependencies

A clear majority of blogs in the sample are published via a blog publishing application, with Blogger and WordPress the most represented applications. While bloggers may be seen as the primary contributors of original content to blogs, in the form of text, images, audio, and other graphic elements, these blog publishing applications are also producers through the provision of the "container." Choice in blog application, as well as host location or membership in a blog network, has implications for preservation in regard to control of and access to content as well as context, with the latter manifested through user element features.

To illustrate, consider service providers. Service providers may stipulate in terms of service agreements that, while bloggers retain control of contributed content (i.e., posts), the service provider holds the copyright to the web publishing software (i.e., presentation structure).<sup>69</sup> As another example, a violation of a term of the service agreement might result in discontinuation of the service. For the most part, bloggers, as reported by Viegas (2005), assume the persistency of their blogs unless deliberate action is taken by a blogger to remove it. Even if no longer maintained, it is assumed the blog will still be available via search engine caches. This assumes no other influences impacting service.

Further, consider networks. Almost one out of five blogs were identified as members of a network in the blog analysis portion of the study. During the interview portion of the study, scholars' blogging as part of a network were asked to comment on whether any plans or policies for the long-term management of their respective blogs had been provided by the network administrators or editors. A couple interviewees commented on policies in terms of expectations for active publishing, but no extensive or explicit policies or procedures for the long-term management of their blogs had ever been disclosed.

Another co-producer dependency concerns blog commenting systems, a standard feature of blogs in the sample. Commentators can be seen as co-producers since they contribute content to the blog in the form of commentary published via a blog's commenting system. Few blogs feature either explicit or implicit policies or statements in regard to commenting. Further, only a few address content originating and published to the blog by commentators and other contributors in their posted rights and use statements. These concerns, however, should also be viewed in consideration of actual commenting activity rather than "potential" commenting. While commenting is supported, it does not mean

---

<sup>69</sup> See, for example, Blogger's Terms of Service at <http://www.blogger.com/terms.g>

commenting occurs. For almost six out of ten blogs in the sample, the most recently published post, at the time of coding, had no comments associated with it.

#### 6.4.3. Back-Up and Archiving Behaviors and Strategies

A majority of scholars responding to the questionnaires take action to save either their entire blog or components of the blog (e.g., posts; posts with comments; post and comments, but separately). They also reported a range of strategies for doing so, including service provider export tools, database dumps or manual back-ups, as well as a variety of storage locations, including institutional back-up servers, personal servers or workstations.

Concerning saving of the blog in its entirety, co-blogs introduce particular concerns since the content represents the work of two or more primary authors. Co-bloggers did not feel it important to get permission from their co-authors to save the blog, though a majority reported that s/he did have permission to do so. As one respondent wrote, “I think it’s assumed I’ll back up the blog.” Co-bloggers participating in interviews were asked to comment on any discussions s/he had with their co-authors in regard to long-term management of the blog. For those reporting discussions with their co-bloggers, most characterized these as cursory, with no resulting action. Others weren’t aware if any such conversations had ever taken place.

##### 6.4.3.1. Known and Unknown Archiving

For questionnaire respondents reporting that s/he save their blog in it’s entirety, few report doing so via an archiving service. However, as found in the blog analysis portion of the study, half of the law blogs in the sample were archived to the Library of Congress’ Legal Blawgs Web Archive. Further, while questionnaire respondents may not have reported known and personal use of an archiving service, eight out of ten blogs in the blog sample

were found to be archived to the Internet Archive Wayback Machine. While the extent of how recently and frequently the blog is crawled and indexed to the Internet Archive varies across the sample – and further, was not coded to that granularity of detail – it does show that most blogs in the sample have been digitally archived to some extent by a third-party, public trust. So, while questionnaire respondents reported some examples of known archiving by a digital archiving service, blog analysis showed a much greater extent of potentially “unknown” archiving as well.

### 6.5. Implications

The attributes of the blogs of scholars and scholars’ blog publishing practices, perceptions and preferences provide rich, descriptive information from which to consider blog preservation activities and the role of blogging within the academe. Implications from the overall findings are presented in this section, along with opportunities for different actor-groups in the preservation of scholars’ blogs.

#### 6.5.1. Back-up or Preservation? Or, Back-up is Preservation?

Scholars report strong preference for preservation of their blogs, both for private and public access and use, and employ several strategies in their predominantly personal approaches to saving their blogs, in whole or in part. But this leads to the question: Are these strategies preservation? When considering the definitions and goals of preservation as used in the information and library science literature (E.g., Caplan, 2008; Moore, 2008; WGDDP, 2007), no, not really. The strategies can better be described as back-ups. This, however, leads to another important question: Do we need to qualify any preservation action, particularly for personal preservation approaches, in line with the definitions and goals used by researchers in the information and library science fields? Consider the issue raised by Harvey (2007)

that, when talking about the need for digital preservation, stories of data recovery are used in support of these calls, but not stories of actual data loss. Also, as Lynch (2005) points out, we lack the tools to test what approaches or strategies will work in terms of access and rendering of digital content into the future:

... without the digital analogs of physical accelerated ageing test beds, most research is either about tools, about identifying approaches that don't work, or is highly speculative in nature – how do you prove that your approach in fact will preserve data for a thousand years without having to wait that long? (You can prove it will fail in much less time, of course). (para. 14).

As was evidenced when constructing the sampling frames, blogs which had not been updated in several years, and assumed to be abandoned, were still available and, even more importantly, still able to perform. That is, from a cursory examination in assessing these blogs for inclusion, the first page could be viewed.

It is impossible to write of the topic of digital preservation without acknowledgement of the concern: is it hype, to some degree, or is it reality? Two camps arose from the interview portion of the study in regard to perceptions on preservation. One camp could be categorized as “moderate doomsayers,” characterized as cautionary and concerned. These interviewees alluded to the implications of blog availability into the future and the potential for disruption due to a variety of issues, including forward migration, link rot, and availability of continued support from blog publishing and hosting services and blog networks and sponsors, and other hosting companies. The other camp could be categorized as “satisficing optimists.” Their remarks were characterized by a recognition that, through publication to the Web and a sort of “preservation by proliferation” approach in terms of multiple copies available via crawling and caching, content will continue to live on ... even without more direct or purposeful intervention. Several interviewees alluded to the Web’s inability to forget. It is reflective of the Web as society’s attic, where stuff will be stored

away when no longer an immediate need.<sup>70</sup> The case can be made for both of these camps. Further, interviewees were not necessarily seen as being of one or the other camp but, rather, their comments were reflective of both viewpoints.

Further discussion regarding what is and what is not “true” preservation is beyond the scope of this chapter and, ultimately, the scope of this research. It is raised here in recognition of varying perceptions on digital preservation among “laypersons,” used to distinguish between those concerned about preservation, and others actively researching and working on issues of digital preservation and curation. Even these digital preservation researchers and practitioners have a hard time coming to consensus on the terminology, as noted by the DPE (2007). However, regardless of extent or degree of preservation intended, there are many compelling issues in regard to managing blog content into the future, whether for personal or public access and use. These are described throughout the remainder of this section.

#### 6.5.2. Preserve What? Issues of Representation, Selection, and Permissions

Discussion on blog preservation leads to the question, “preserve what?” This pertains to the format (i.e., blog-level, post-level, etc.), as well as extent (i.e., entire blog since inception, select posts, etc.), and issues of performance (i.e., linking, commenting, etc.). Interviewees raised some of these considerations during their interviews. The “what” of preservation also brings in issues of selection. Many interviewees remarked on the ephemeral nature of blogs and their opinions on those which, either by particular instance or type, do and do not merit preservation. Interviewees also addressed the challenges of making decisions in regard to future use. Their responses were reflective of not only the amount of

---

<sup>70</sup> And, as commented on by several interviewees, the Web can also be seen as society’s haunted attic, where stiff long forgotten can come back and disturb or unsettle you later.



material it is feasible to preserve, but the amount that it is feasible for future users to be able to negotiate and make sense of.

Two strong themes emerged from these remarks. One, blogs are context-rich, as demonstrated through linkages and associations with other content published to the Web. The loss of such context in the long-term may diminish their value for future use as well as obfuscate the original message as intended when initially published to the Web. Two, blogs are only one type of communication unit in a diverse, complicated, and overloaded infosphere. Too much information can also lead to consequences of future use. Just because we can save things doesn't necessarily mean we should. This was very much reflected in interviewees' reactions to the Library of Congress' announcement about archiving all public Tweets since Twitter's inception in 2007. To summarize in one word the sentiment of the majority who commented: "Why?"

In consideration of the diverse units of communication and other products and outputs resulting from the manufacture of scholarship, interviewees were asked to comment on a continuum of priorities for digital preservation and, subsequently, to place their respective blog within the continuum. While responses varied, with a few placing their blogs higher in the continuum, others placed blogs dead last in terms of prioritization. Further, most agreed that traditional units of communication, such as journal articles and monographs, are more important to preserve than their respective blogs.

Regardless of the resolution of these "whats" of preservation, the nature of the format and the publishing activities of bloggers contributes to many concerns and, possibly, impediments to responsible preservation. "Responsible" is used here intentionally to represent the chaotic legal and regulatory framework in which blogging takes place.

Consider the “what” of performance. Since the interactions supported by blogs, such as the ability to comment, tag, and link, are principle contextual characteristics of the medium, terms of service agreements may allow content of posts to be preserved external to the blog publishing application – such as in a text document, a strategy reported by questionnaire respondents – but not within the context of how the post was published to the Web. In particular, consider commenting systems. It is assumed that comments are, generally speaking, copyrightable expressions and that, absent of any direct policy statement to the contrary published to the blog, permission would need to be sought to preserve such comments. Any preservation action begins with the simple act of copying, and copying is an exclusive right of a copyright holder. Thus, clearing permissions to preserve comments could be a herculean task sans explicit policy statements published to the blog.

These legal issues are provided here at a very high level. The full range of implications is beyond my purview. It is interesting to note, however, that for scholars assigned to the law cluster, a majority report never or seldom checking permission to use imported content in their blogs though a majority find it important to verify permissions for using content originating elsewhere.<sup>71</sup> For content originating at scholars’ own blogs, respondents in the law cluster were fairly split down the middle in terms of how important or unimportant they feel it is for others’ to verify permission. Additionally, a majority of law cluster blogs in the blog sample have one or more rights and use statements posted to the blog, though blogs in the BioChemPhys cluster have the highest percentage of rights and use statements among all clusters in the sample. While not all bloggers in the law cluster are assumed to have expertise in intellectual property rights, they might be assumed to have a

---

<sup>71</sup> These responses were based on a specific time period: the one month period preceding respondents’ completion of the questionnaire.

little more than bloggers assigned to the other clusters as well as this researcher. Bearing in mind this assumption, these findings are presented here because it may imply that, quite possibly, issues of copyright may not be as considerable as assumed for responsible, personal preservation. Personal is added here in reflection of the strategies scholars' report using in saving all or parts of their blog, and their identification as being primarily responsible for preservation. This would be a very interesting area of further investigation, particularly when considering a copyright holders' exclusive rights as well as exceptions to those rights, such as Fair Use.

#### 6.5.3. Perceptions of Actor Roles in Blog Preservation

Questionnaire findings show that scholars are less likely to identify libraries, archives, and institutional IT departments as either responsible or capable for blog preservation. They do, however, identify themselves, their co-authors, and service providers, including blog publishing and hosting services, networks, and other hosting companies, as being most responsible. Further, scholars' feel service providers, search engines, and public trusts, such as the Internet Archive, are the most capable. These findings were explored further in the interview portion of the study to understand the "why" behind these responses.

Interviewees see themselves as most responsible because, well, it's their content. They identify responsibility as resting primarily with authorship. They also attribute responsibility as well as capability to their service providers. Scholars assume both the capacity and responsibility for these providers to maintain content for at least as long as interviewees maintain their status as subscribers, but also, potentially, longer since these providers are the ones who, ultimately, control the content. The former is in consideration of their respective terms of service agreements, with interviewees commenting on their

expectations as consumers that these providers will hold up their end of the bargain; that is, the expectation that their blogs will be available and maintained as long as providers' terms of use are not violated. It is interesting to point out, however, that several interviewees remarked that they are not familiar with these terms.

It is interesting to reference the finding from Marshall et al. (2007) that people rely heavily on service providers and often take no extra precautions to ensure the longevity of their personal digital assets. Findings from this study do reflect expectations for service providers, but not the lack of "extra precautions" as found by Marshall et al (2007). Scholars responding to this study take action to save all or parts of their blogs. While many may use an export tool provided by their respective service provider, s/he are the ones making use of the tool, while the service providers are simply make it available.

In consideration of the other actors groups identified as most capable for blog preservation, interviewees commented on both the technical capacity as well as financial resources of search engines, specifically identifying Google. They also commented on the positive reputation and expertise of the Internet Archive. A few also commented on a role for involvement by national cultural heritage institutions, with a couple specifically mentioning their own involvement, and satisfaction, with the Library of Congress Legal Blawgs Web Archive program.

But what about institutional-based libraries, archives, and IT departments? Interviewees did not dismiss the value of these organizations. If anything, for some, it is exactly this value that contributes to perceptions of libraries and archives' low responsibility and capability. That is, libraries and archives have more important things to do.

That was just one theme that emerged from interviews. Another is that scholars see their blogs as their own products, born of their initiatives. Their blogs are not representative of the work done on behalf of their respective institution and, hence, should not be seen as representative of the institution, making them inappropriate for institutional-based archiving. A related, sub-theme is that scholars feel their blogs are outside the scope of library and archival collections, at both the institutional and national-level.

This sense of scholars' personal sense of their blog as their own contributes to another theme. Since the blogs are scholars' products, born of their initiatives, scholars cannot expect these unaffiliated actor groups to be responsible for their preservation. A final theme, mentioned by the fewest interviewees, was that scholars were skeptical of their respective libraries, archives, or IT departments' capacity to preserve their blog, and provided examples of personal experience with these organizations that informed this viewpoint.

Regardless of the degree of perceptions (most or least) concerning capability and responsibility, overall findings indicate opportunities for these various actor groups to contribute to blog preservation, whether directly or indirectly. This is detailed in section 7.1. in the next chapter.

## 6.6. Limitations

Discussion of research design leads to addressing the proverbial “elephant in the room,” that elephant being the blogosphere. The blogosphere is considerable, with millions of blogs co-produced by millions of bloggers, readers, and scores of service providers. The sheer size of the blogosphere does not lead to simply a “needle in a haystack” scenario. Considering the variety in blogs and bloggers and the low barrier of entry for publication, it is more akin to a chameleon in a haystack. Blogs are added, deleted, and forgotten on a daily

basis. Lifespan varies, from blogs that are active – added to, modified and maintained – for mere days or weeks to blogs remaining active for years.

No assumption on the representativeness of the samples in this study are made. The sampling frames were derived from purposive sampling. The findings are not representative for all scholar bloggers; it is only representative of those scholars and blogs listed to the Academic Blog Portal in the area of history, economics, law, biology, physics and chemistry. While sampling techniques were implemented to exclude non-eligible participants and increase homogeneity, respondents to the questionnaire and interview portions of the study were still diverse. Further, blogs selected in the sample, while also subject to assessment for eligibility and enhanced homogeneity, were also diverse. Since there is no way of knowing how many bloggers in the blogosphere identify as scholar bloggers or, subsequently, their blogs as scholarly blogs, it is inappropriate to attempt to assess the degree of representativeness of the samples. However, this is also in recognition that the population of scholar bloggers and scholar blogs is unknowable. While the samples are small and unrepresentative, the methods employed, including sampling and data collection, were considerate.

As described in Chapter 3, the research was designed in recognition of this complicated landscape, and in particular, specific characteristics of the population under study. From an earlier exploratory study of general bloggers' perceptions on digital preservation, a finding made very clear was that blogger to blog is not a simple 1:1 ratio (Hank et al, 2007; Sheble et al., 2007). Bloggers not only publish to two or more blogs; blogs are also collaboratively authored. This led to two specific considerations in the design of this study for assessing scholars' blog publishing practices and preferences and perceptions for

digital preservation. First, for bloggers publishing to two or more blogs, perceptions, preferences and practices may vary across instances. Secondly, for co-blogs, perceptions, preferences and practices may vary across the blogs' authors. Findings from this study confirmed those of the earlier study: 42% of questionnaire respondents publish to two or more blogs, and 56% of the questionnaires' completed sample was comprised of co-bloggers.

The design was also complicated by two elusive questions – who is a scholar, and what is a scholarly blog – as well as the challenge of designing the instrumentation in consideration of different professional practices, whether these be disciplinary, occupational, or geographical. In regard to the latter, while blogs published in English were a requirement for inclusion in the study, there was no requirement in terms of geographic location.

In response to the above considerations, six particular actions were taken. Additional actions were also taken, as described in Chapter 3, but these six deserve particular attention here. First, explicit inclusion criteria and a protocol for assessment were developed and used in selecting both blogs and bloggers for the samples. Second, two questionnaires were developed and administered, one for single-bloggers and the other for co-bloggers. Third, questionnaire respondents were asked to respond based on a particular blog instance, identified by both blog name (header title field) and URL, with this instruction provided in all recruitment materials and the questionnaires. Fourth, co-bloggers were given further instruction to reply based on their own preferences rather than how s/he think her/his co-authors might also respond. Fifth, for questions relating to publication and service history, earlier surveys of scholars across disciplines were referenced to ensure comprehensive responses categories were provided. And sixth, since tenure practices vary across countries,

this was considered in the design of the question stem and response categories for a tenure-related question.

While these specific actions were taken to counter the considerations described earlier, they also introduced some limitations.

#### 6.6.1. Potential Exclusion of Eligible Respondents

While the inclusion criteria and protocol for assessment, as applied at sampling and described in Section 3.1.1.1., are deemed thorough by this researcher, this may have contributed to members of the sampling frame being excluded who may otherwise qualify as scholars under other parameters. Additionally, scholars may meet the inclusion criteria, but were still excluded because the information to assess eligibility was incomplete or unavailable. No additional effort was made to verify credentials or contact information beyond the specific procedures detailed in the Sample Coding System (See Appendix A). Further, for co-blogs, only those co-bloggers publishing within one month of the time of sampling were considered. Hence, some scholars contributing to co-blogs were excluded even if there was evidence that s/he met all other criteria for inclusion.

While not a limitation, it does merit pointing out that the preparation of the sampling frames and resulting samples was time-intensive. Each of the 644 blogs identified at the source and, subsequently, bloggers publishing to these blogs, were manually assessed in line with the eligibility criteria.

#### 6.6.2. Participation

In the questionnaire stage of the study, besides an initial invite email and two subsequent reminder emails, no other actions were taken to solicit participation from non-



respondents. While follow-up interviews with non-respondents could be informative, it was beyond the time frame and resources of this study.

No inducements to participate were provided at the time of recruitment, beyond contributing to knowledge, to participants in either the questionnaire or interview phases of the study. While tangible inducements have been shown to contribute to higher response rates (E.g., Dillman, 2007; Punch, 2003), it is assumed the potential of a free iPod, Amazon gift certificate, or other tangible inducement would not be particularly alluring (or needed) for the population of professionals under study. An inducement was provided to respondents completing the questionnaire. At the end of the questionnaire, respondents could indicate if s/he would be interested in receiving a final report from the study. Only those completing the questionnaire through to the end were told of the final report and provided opportunity to request one.

#### 6.6.2.1. Ineligible Sample Members

In their invitation to participate in the questionnaire phase of the study, sample members were asked to complete the questionnaire based on a particular blog instance. The recruitment email also provided context for the study, including study title, brief description, and goals. As described in Section 4.1.1., four invitees responded that s/he did not consider the blog scholarly, themselves as scholars, or themselves as active bloggers, and, ultimately self-reported as ineligible to participate. While these bloggers, who had qualified as eligible based on the inclusion parameters, only represented one percent of all sample members invited to participate, it is demonstrative of the challenges in distinguishing between blogs published to by scholars and “scholarly blogs.” Additionally, one blog selected in the sample for the blog analysis portion of the study was excluded after data collection commenced. The

decision was made not to replace the blog since it only represented a 0.5 point percentage change in the sampling ratio (from 50% to 49.5%).

#### 6.6.2.2. Interview Recruitment

Only questionnaire respondents were eligible to participate in the interview portion of the study. Additionally, not all respondents who identified themselves as interested were selected to do so, nor did they have an equal chance of being selected to participate. The recruitment and ultimate selection of interview subjects was motivated by three factors: 1) time of questionnaire completion, 2) respondents' discipline and author-status (single or collaborative); and 3) availability during the time interviews were being scheduled (August 18 to September 3, 2010). Hence, those respondents expressing interest who returned questionnaires at a later point in the collection period were excluded, as were those sharing characteristics with bloggers already invited to participate and those only available after September 3, 2010. In future investigations, more time, as allowed, will be built in to the data collection period so that a random sample could be drawn of all interested participants, clustered and stratified by authorship, to ensure both representation across clusters of single- and co-bloggers in the sample as well as an equal chance of being invited to participate.

#### 6.6.3. Instrument Design and Analysis

While the questionnaires went through extensive revision and testing, there are two specific issues that, in hindsight, merit consideration, both for future studies on this population as well as for future use of the instruments, as provided in Appendices C and D, by this author or by others.

First, a PIN was used to facilitate administration of the questionnaire. One reason among several for electing to use a PIN was to track questionnaire access to inform reminder

invites. This was done to be conscientious and not to remind someone to take a questionnaire s/he had either already completed or, at the least, accessed. Unfortunately, due to an oversight in the questionnaire design stage, four members of the sample who accessed the welcome screen and then selected “No Thanks” to participating were then taken to an exit screen rather than the screen to enter her/his PIN. Thus, these respondents may have received unnecessary reminder emails as their access to the questionnaire could not be tracked.

Second, publishing genres and conventions vary across disciplines. Questions on unit-types of scholarly communication to assess publication history were carefully considered. Attempts were made to comprehensively list published units of communication commonly published by scholars for questions on unit-types of scholarly communication to assess publication history. However, in consideration of the diverse fields of sample members, a specific publication was left out at the design stage. That is, the law review article, which does not qualify, per se, as a peer reviewed article since, generally, law reviews are reviewed and edited by law school students. This exclusion was intentional so as to make the questionnaires applicable to all sample members. In hindsight, I would include this particular unit-type. It also merits noting that a few respondents went out of their way to send an email after completing the questionnaire to point out the absence of this option. I am appreciative of their feedback, particularly as they have already invested time in completing a long questionnaire comprised of between 41 and 62 questions.

Another question concerned tenure status. This question could be seen, for the most part, to be applicable to North American academics, but not necessarily to academics in other parts of the world. Based on feedback at the questionnaire pre-test stage, the response categories were slightly revised. A future attempt with a geographically diverse sample of

academics would refine this question further, as to avoid confusion, as well as to offer clear demonstration of the researcher's understanding of the diversity in tenure conventions.

Third, concerning questionnaire analysis, only completed questionnaires, characterized as 80% or more of all applicable questions completed, were accepted for analysis. Nearly all returned questionnaires were completed at a 95% or higher completion rate. While the four partial questionnaires returned may have yielded useful information, these were excluded from analysis.

Fourth, in regard to the blog analysis portion of the study, only two blog indexing services were checked to investigate blog availability, Google and Technorati. Other blog indexing services are readily available, with some also representing an extensive index of blogs. Checking if blogs were indexed to any of these other services was not done for two basic reasons: one, the popularity and scale of Technorati and Google's indexes; and two, simply time constraints for data collection. Google and Technorati were considered sufficient for assessing blogs availability within the blogosphere, and checking across other services was considered redundant.

Fifth, in regard to analysis overall, the findings reported here are high-level descriptive statistics to describe the population under study in response to the specific research questions. While all results presented here are final, the exploration of the data is not. Specific areas of further analysis of the study's aggregate data (questionnaires, interviews, and blog analysis) are planned. Further analysis will focus on specific respondent attributes across nine variables, potentially leading to statistically significant differences in attitudes, behaviors, and characteristics of scholars in regard to blogs within the system of scholarly communication and issues of digital preservation. These variables include: 1) Basic

demographics (e.g., gender, age, and professional age); 2) Cluster (history, economics, law and physics, biology and chemistry); 3) Employing organization (e.g., Carnegie Classification of Institutions I-V); 4) Occupation type: (e.g., post-secondary faculty, post-secondary non-faculty, non-post-secondary, self-employed); 5) Identity (e.g., real name, pseudonym); 6) Authorship (single or collaborative); and 7) Network membership.

#### 6.6.4. Researcher Bias

I could not approach an investigation of the preservation of scholars' blogs without first broaching the issue of whether, and how, these blogs are perceived as scholarly or not. The study was designed, from the beginning, in recognition of the need to limit my own bias, which I will divulge here. I do believe blogs are representative of scholarly communication and that they meet the parameters of scholarship for unpublished communications. I also believe some blogs merit preservation action, whether at a personal or organizational level. But I just as strongly believe that some don't. Further, I feel the preference of content creators is paramount to any preservation action.

These views were not shared with respondents, either in email correspondence or during the interviews. I think it also merits noting that, while I read blogs, I do not actually maintain a blog. So my own biases are not from the viewpoint of a blog author, but as a blog consumer. Just as interviewees' motivations for blogging were described in Section 4.2.5., I would like to describe my motivations not to blog here: one, simply, is that during these dissertating years, my time has tended to manage me rather than me my time; and secondly, and tied into my awareness of my own biases, I did not want to blog on the topic of my research in that it may leave a trail and expose my biases and tendencies to potential study participants.

## **CHAPTER 7**

### **CONCLUSIONS AND FUTURE DIRECTIONS**

Scholars feel blogging contributes to their enjoyment of their work as scholars and positively impacts aspects of their research, teaching, and/or service. While blogging was not seen to contribute to opportunities for promotion at one's respective institutions, blogging did contribute to scholars' personal sense of intrinsic rewards, including greater visibility as a scholar, as well as more formal recognition through invitations to publish, present, provide service, and collaborate.

Considering the functions of scholarly communication, as summarized by Borgman (2007) – legitimization, dissemination, and access, preservation and curation – blogs, as channels for communication, facilitate the dissemination function by providing a mechanism for scholars' to make known their "ideas, knowledge, and works." Blogging can also be seen, in part, to contribute to the legitimization function by serving as a sort of "proving ground" in the development and communication of their scholarship and, as referenced above, leading to invitations to publish and present. In regard to the function for access, preservation and curation, scholars feel their blogs should be preserved for personal and public access and use into the indefinite future. And, further, that they are the ones most responsible for doing so.

Scholars take some action in regard to the long-term stewardship of their blogs, identifying a number of personal strategies for saving all or parts of their respective blogs. While this reflects some capability on their part, scholars identify the blogs' service providers, hosting companies or sponsoring networks as most capable. Due to the co-

producer dependencies between bloggers and service providers, this intersection between most responsible and most capable contributes to opportunities for personal preservation action as supported by third-party services.

“Personal” is qualified in that preceding sentence because, while scholars see a role for service providers in the preservation of their respective blogs, as well as other organizations like Google and the Internet Archive, few scholars feel libraries, archives, or institutional IT departments are either responsible or capable for the preservation of scholars’ blogs. This may very well be due to the medium itself. While scholars feel their blogs should be preserved, on a continuum of priorities, other types of communications, such as journal articles, monographs, and, in the words of one interviewee, “things where my audience is clearly my peers,” take priority. This does not mean scholars don’t value the traditions and services of libraries and archives; it is just that, well, they might think their time and resources be better spent elsewhere. One tenure-track associate professor of economics, addressing libraries in particular, appealed:

Don’t spend your resources. Libraries are too squeezed and fighting too hard to overcome just the idiocratization of culture where a Google search on keywords substitutes for a scholarly search among students. Putting money into archiving blogs is only going to detract from what I view libraries ought to be doing. Don’t do it. Please don’t do it.

Additionally, interview participants strongly identified their blogs as their personal productions. Hence, even if affiliated with an institution as a faculty member or in another role, some do not see their contributions as aligned with their institutional responsibilities or as representative of products of their institution. The former, as found in the blog analysis portion of the study, was also evidenced by the publication of explicit or implicit disclaimer statements.

Garrett and Water's (1996) characterized content creators as the first line of defense. In regard to scholars and their blogs, I propose two extensions of this fundamental principle. First, content creators are a literal line of defense. As blogs are a co-produced medium, many players, besides the respective blog's authors, have a hand in both the direct preservation of the blog – e.g., providers' terms of service agreements and the availability of application-based export and archiving tools – and in the indirect preservation of the blog – e.g., issues of rights, use, and context in consideration of content published to the blog by commentators, made available on the blog through embedding, importing, and linking from other sources, and “made possible” on the blog by software providers and hosting services.

Second, scholars who blog, as primary content creators – characterized as those publishing original content to the blog and identifying as the blog's' sole author or, in the case of co-blogs, co-author – are also the last line of defense. A clear majority of questionnaire respondents and interviewees see themselves as most responsible for their blog's preservation, followed by service providers, host companies, and networks. So, in this way, bloggers can be seen to be both the first and last lines of defense.

In regard to the first line, this is seen through their preferences and behaviors in regard to composition and dissemination. To name just a few, this includes choice of blog application or service, host, or network provider, and the publishing of descriptive information to the blog that helps in assessing the informational value (i.e., authority, credibility, and trustworthiness of both authors and content) and rights, licenses, and permissions for use. And in regard to the latter, this is seen through their actions and initiatives in maintaining and planning for the long-term management of their respective blogs, demonstrated through a variety of personal archiving practices, including regular



back-ups, and through participation in web archiving programs, whether known, such as participation in the Library of Congress Legal Blogs Web Archive, or potentially “unknown,” such as making their blogs available for crawling to the Internet Archive Wayback Machine.

#### 7.1. Recommendations to Facilitate Blog Preservation

Findings provide support for several recommendations for involvement, either direct or indirect, by co-producers and other vested parties for the preservation of blogs. One would be the development and provision of exemplars of a range of policy and disclaimer statements for scholars to incorporate into their own blogs. These policies and statements should address explicit issues related to rights and use of content published to the blog. This could include commenting guidelines, including agreement to a non-exclusive deposit license to make all comments or other contributions available for preservation, and that the respective commentators are within their own rights to make such an agreement. Additionally, editing policies would enable improved interpretation by end-users of a blog’s authenticity and integrity, with such statements making clear how, if any, changes to previously published content are annotated directly to the affected post.

Service providers and networks could make these policy exemplars directly available to subscribers. Librarians and archivists can be involved in crafting such policies, based on their long-established practices and traditions in managing both analog and digital collections. Further, libraries and archives could promote use of such policies during other interactions with scholars focused on effective data and digital asset management.

Second, other documentation could be made available to scholars, describing different strategies for preservation and, if deemed appropriate, recommended third-party,

back-up and Web archiving service providers. One interviewee, in commenting on tools he would like to have available to him, described the equivalent of his providers' export tool. Explicit documentation would help to make scholars aware that such options may be available to them.

Third, blog application providers or developers could include additional fields and tools at the point of composition, or at a later date after publication, to allow authors' to indicate what posts merit keeping and which do not. For example, this could be through the inclusion of an "archive it" feature in the blog's composition editor, which would allow bloggers to indicate those posts s/he would like preserved (i.e., longer pieces on a recent legal case), as well as to indicate those posts that should not (i.e., posts about the view from their conference hotel room). Further, programs using web crawlers could recognize these particular "archive" posts, which would facilitate capture of content intended to last, instead of all content published to the blog.

Further, service providers can consider providing their own archiving services or contracting such services to a third-party, and making these available to their subscribers. From the exploratory study on general bloggers' perceptions on digital preservation, respondents expressed some interest in subscribing to fee-based preservation services (Hank et al., 2007; Sheble et al, 2007). While such services currently do exist, such as Backupify, few scholars in either the questionnaire or interview stages of the study reported use of these services. If archiving and back-up services were identified directly at the blog application and hosting service providers' site, then scholars may be more aware of such services and, hence, more inclined to participate.

Fourth, service providers should consider the language and terms as explained in their terms of service agreements, in reflection of interviewees' expectations of service. Further, subscribers may not be able to assess changes to their terms of service agreements if (1) they are not notified of changes, and (2) if terms of service agreements are not dated. Those providers that do not date their agreements or inform subscribers of changes, say through an email alert or alert posted to all blogs' composition editors, should probably consider doing so.

Fifth, while the findings show low perceptions in regard to libraries and archives' responsibility and capability for blog preservation, that does not necessarily imply that there should be no involvement. Much appreciation for the Library of Congress' Legal Blogs Web Archive was mentioned by legal scholars participating in the interview portion of the study. Additionally, another interviewee, in reflecting on the role of libraries and archives, remarked that he would be a willing participant in an easy-to-use, opt-in blog preservation program if one were offered at his institution. Rather, libraries and archives, as is their practice, are encouraged to consider the needs of their respective institution's scholars before engaging in developing or expanding an institutional blog preservation program.

## 7.2. Recommendations for Further Research

Discussion on the place and role of blogs and blogging in the system of scholarly communication are found in the blogosphere, as well as in other reports in the literature, as are discussions on blogs and digital preservation. Discussions of both however – scholar blogs and digital preservation – are limited. Further, empirical research on scholars, blogs, and digital preservation is even more limited. A few studies have looked at some of these issues, but in isolation rather than in combination. This study was done to fill that gap.

Findings present many opportunities to continue investigation in these areas. Several are presented here. It also merits noting that interviewees provided their own ideas for extending this research. A grateful acknowledgement for sharing these ideas, several of which are incorporated in the suggestions for future research that follow.

#### 7.2.1. Further Analysis

As mentioned in Section 6.6., the wealth of data collected provides much opportunity for extending analysis of the descriptive results presented in this dissertation. In addition, a particular investigation is planned but for blogs not included in the sample. While this study reported findings on active blogs and bloggers, in line with the research questions, 456 blogs listed at the sample source were excluded from participation following assessment of inclusion criteria. A more thorough treatment of these instances will be completed. Of particular interest are those blogs that, while still accessible, are not being currently published to, with some not showing recent posts in at least the past three years. Some excluded blogs went “dark” with no mention (or warning) posted in the most recently published post, while others specifically described reasons the blog will no longer be published to and updated, as well as why the decision was made to keep the blog in the blogosphere, rather than to delete it. Future investigation may contribute to discussions on permanence of the medium and implications for scholarly merit and appraisal (i.e., if the blogger lets a blog go “dark,” then why shouldn’t the collective “we?”). This will be complemented from findings from the questionnaire in regard to blog deletion. It is curious to consider what, if any, differences there are in motivations to delete a blog – an experience reported by 18% of questionnaire respondents – compared to motivations to stop blogging, as found during the blog sampling frame assessment.

One interviewee, in describing his thoughts on his blog and digital preservation, remarked, “It is what it is for as long as it is and then when it isn’t ... it can go off to the great blog heaven in the sky.” It will be interesting to go back and look at these inactive blogs that linger in the blogosphere. Maybe, it’s more of a “blog purgatory” than a blog heaven.

#### 7.2.2. Future Research

This particular class of blogs and bloggers presents many opportunities for further investigation to better understand the phenomenon of blogging by scholars and, likewise, implications for blog preservation. The attitudes and preferences of non-blogging scholars was beyond the scope of this study. An investigation of this population would be informative on perceptions of blogs within the system of scholarly communication and preferences for preservation. It could be assumed that non-blogging scholars would be less inclined to view blogging as a product and process of scholarship, as well as less inclined to support preservation. It would potentially contribute to interesting comparisons between scholars who do not blog and their counterparts who do. Further, this research could inform additional analysis begun here in regard to how scholars feel others in their field perceive blogs as opposed to how they feel they should be perceived, a theme explored during the interview phase of the study. Additionally, when considering the reward function of the system, this research could contribute to enhanced understanding of how blogs are perceived by administrators, peers, and others involved in decisions regarding tenure and other opportunities for promotion.

In regard to these opportunities for promotion and other types of formal recognition, the research reported here presented brief findings. For example, most respondents to the

questionnaire did not see their blogs as improving opportunities for promotion at their institutions. A few interviewees mentioned blogs in consideration of tenure, with one relating his own recent experience when going up for tenure (as well as a less successful experience of a peer). In a study looking at academic librarians' perceptions of blogs as scholarship, Hendricks (2010) asked interview questions regarding blogs and tenure. It would be interesting to extend future research to an examination of post-secondary faculty in other disciplines who blog and how s/he have integrated, or elected not to integrate, their respective blogs in their tenure packages. Further, it would be interesting to examine the attitudes of how those evaluating these tenure packages considered inclusion of these blogs.

While a majority of scholars responding to the questionnaire feel their respective blogs meet the three criteria for scholarship, as proposed by Braxton et al. (2002), in citing Shulman & Hastings (1999), the criterion for "critical review" was selected less frequently than the criteria for public availability, use, and exchange. No definition of what was meant by critical review was provided in the questionnaire. However, considering respondents' well-established traditional publishing history, it is assumed – and was expected – respondents' would reply based on their personal interpretation of critical review from their multiple actor-roles: authors, informal and formal reviewers, gatekeepers, and consumers. Future investigation of blogs in consideration of critical review may serve to inform practices related to extrinsic rewards and formal recognition. One interviewee, a PhD. In economics, suggested a design for such a study, getting to the question of: Is it the content or the container? He proposed a sort of "bait and switch" design. First, have a set of select blog posts and articles evaluated by reviewers. Then, reformat these so the blog posts appear as articles and the articles appear as blog posts, and have these evaluated as well. The intent

would be to investigate if and how reviewers' evaluations may have changed based on form rather than content.

The blog sample included both, "blogs in the wild," characterized as blogs not directly affiliated with a sponsor or established network, as well as blogs affiliated with a network (e.g., ScienceBlogs). Future examinations looking specifically at scholars aligned with these networks, and their respective blogs, would be compelling. As demonstrated in the findings to this study, such affiliation impacts bloggers' perceptions and preferences in regard to digital preservation since networks are seen as having some capability and responsibility for preservation. Further, it also implies some curatorial activities on the part of the networks' sponsors, editors, or managers in regard to selection and assessment of informational value. This can be seen, for example, by how network's recruit and support bloggers in the network. It would also be interesting, from a network perspective, to examine bloggers' experiences in both joining and, when applicable, leaving networks. In regard to the former, one interviewee mentioned this as a particular interest area, commenting:

I think it would be interesting to see, as you work on this ... the degree to which people in different fields are moving toward institutional blogging, like ... blogs for Forbes, the New York Times and what have you and various research institutes ... Moving toward that and away from individual Blogger or Wordpress sites.

It would also be interesting to examine how networks' communicate long-term service availability and blog management expectations with bloggers or, as commented on by interviewees, don't. Further, a thorough examination of practices and procedures – both technical, in terms of content models and preservation approaches, as well as organizational, such as collection policies – of institutions with extant blog preservation programs would be informative. For example, a case study approach to the Library of Congress' Legal Blogs Web Archive, with representation from library personnel, the authors of the archived

“blawgs,” and end-users, would yield potentially useful findings for other organizations and institutions interested in adopting or retooling their own blog preservation programs. It could also, as well, inform approaches, tools, and strategies that could be implemented in personal preservation initiatives.

Blog preservation raises a number of regulatory and legal issues. I am not a copyright scholar, hence some of these issues are beyond my expertise to examine and expertly comment on. However, future examinations of these issues would be welcomed by parties with such expertise. Also, many interviewees referenced terms of service agreements when speaking about their perceptions on the role of blog publishing and hosting service providers in blog preservation. An examination of terms of service agreements, and bloggers’ interpretations of the terms, could provide interesting findings in regard to the divide between what providers promise and what consumers expect.

### 7.3. Let the Conversation Continue

Borgman (2007) signed off with the invitation, “Let the conversation begin,” in the preface to her book on scholarship in a digital world (p. xix). This dissertation is the voices of 153 scholars in the humanities, social sciences, sciences, and law contributing to the conversation. And scholars blogging are very interested in the dialogue around issues related to blogs, scholarly communications, and/or digital preservation. This is indicated by the moderate-to-high response rate (RR1 52%) to the questionnaire portion of the study, particularly considering the length of the instruments – between 41 and 62 questions – as well as respondents’ interest in receiving a final report from the study, indicated by 88%, and participating in a follow-up telephone interview, as indicated by 47%. Several interviewees concluded their respective interviews by remarking on their interest in the topics under



consideration. As one tenured professor commented: “We all have questions about all this stuff as it’s evolving before our eyes.” The blogs of scholars provide an informative lens from which to consider the current state of scholarly communication, whether described as an evolution, transformation, collision, or other descriptor. The conclusions and recommendations presented within will hopefully serve to continue the conversation.

## APPENDIX A: SAMPLE CODING SYSTEM

### SAMPLING FRAME 1: BLOGS

---

#### Identification information

---

**A. Blog Title:** Blog title as it appears in source listing (Academic Blog Portal)

Enter text

**B. Blog ID:** Unique 3-digit number assigned when captured for sampling frame

Enter in digits

**C. Acquisition Date:** Date blog assessed for sampling frame

mm\_dd\_yyyy

**D. URL:** URL of blog homepage as it appears in source listing

Enter text

**E. Source Note | Scope:** Is there an internally linked “about page” at source listing?

(0) No; (1) Yes

**F. Source Note | Author:** Is there an author(s) listed next to blog title as listed at source?

(0) No; Yes [Text: Enter name(s)]

**G. Source Note | Additional Info:** Is there another note next to blog title as listed at source?

(0) No; Yes [Text: Enter note]

---

#### Criteria 1: Blog is publicly available

---

**H. Availability:** Blog is publicly accessible

(0) No; (1) Yes; (2) Yes but URL change, alerted by note in most recent post

**I. If H-2, provide new URL**

Enter text

**J. If H-2, reason given for change:**

(0) Migration to new platform; (1) No reason given; Other [Text: Briefly describe]

**K. If H-0, reason blog not accessible:**

(0) Password protected; (1) Dud URL; Other [Text: Briefly describe]

**L. If K-1, describe:**

(0) Domain sale; (1) 404 Error; (2) Server Error/Server Not Found, or Page Not Found, but not identified as 404 error; (3) 410 Error; (4) Site Temporarily Unavailable message; (5) Spinning URL; nothing happens; (6) Wrong link, i.e., unknown Web site content; (7) Other

**M. If L-6 or L-7, briefly describe:**

Enter text

\*\*\* If H-0, stop coding. Blog is not eligible for study. \*\*\*

**N. For H-1 or H-2, is blog header different from blog title listed at source?**

(0) No; Yes [Text: Enter title]

---

**Criteria 2: Blog is published in English**

---

**O. Language | English:** Blog is published in English

No (0); Yes (1)

\*\*\* If O-0, stop coding. Blog is not eligible for study. \*\*\*

---

**Criteria 3: Blog is a knowledge or personal blog**

---

**P. Blog Style:** Blog is a journal style blog characterized as knowledge blog or personal blog

(0) No; (1) Yes; (2) Cannot determine

**Q. If P-0, the style of blog is:**

(0) Class blog; (1) News Service/Aggregator; (2) Website using blog software; (3) Twitter aggregator; (4) Filter Blog; (5) Other [Text: Briefly describe]

\*\*\* If P-0 or P-2, stop coding. Blog is not eligible for study. \*\*\*

---

**Criteria 4: Blog posts are time-stamped**

---

**R. Time-Stamping:** Posts accessible on first page of blog are explicitly time-stamped, denoting date of publication

(0) No; (1) Yes

\*\*\* If R-0, stop coding. Blog is not eligible for study. \*\*\*

---

**Criteria 5: Blog is active, characterized as most recent post dated within previous two weeks**

---

**S. Recency:** Most recent post published within previous three months from acquisition date

(0) No; (1) Yes

**T. If S-0, date of most recent post:**

mm\_dd\_yyyy

**U. If S-0, is there an official statement on most recent post implying that blog is no longer being published to:**

(0) None; (1) Yes; (2) No, but statement about intent to re-launch, but currently in need of authors; (3) No, but statement about intent to take break and come back to blog; (4) No, but statement acknowledging that blog has fallen silent, but no mention of stopping, etc.; Other [Text: Briefly describe]

\*\*\* If S-0, stop coding. Blog is not eligible for study. \*\*\*

**V. If S-1, was the most recent post published:**

(0) More than 1 month ago; (1) in the last month; (2) in the last two weeks

**W. If V-0 or V-1, date for most recent post:**

mm\_dd\_yyyy

\*\*\* If V-0 or V-1, stop coding. Blog is not eligible for study. \*\*\*

---

**Criteria 6: Blog is at least 1 year old.**

---

**X. Archive:** Blog has a working (i.e., clickable) archive feature enabled, reflecting “blog-life span” and allowing user to access published posts

(0) No; (1) Yes

**Y. If X-0, can posts older than 1 year be identified through another navigation feature to take user to last page of blog:**

(0) No; Yes [Text: Briefly describe]

\*\*\* If X-0, stop coding. Blog is not eligible for study. \*\*\*

**Z. Age:** Blog at least 1 year old, calculated from acquisition date

(0) No; (1) Yes

\*\*\* If Z-0, stop coding. Blog is not eligible for study. \*\*\*

---

**Criteria 7: Identifying information on authorship.**

---

**AA. Authorship:** Identifying information on authorship is available from first page of blog, including: about information posted to header, title tag, or sidebar; internal link to “about” or similar page; or external link to “about” page, personal website, or similar page:

(0) No; (1) Identifying info on first page; (2) Internal link; (3) External link; (4) Both internal and external Links

\*\*\* If AA-0, stop coding. Blog is not eligible for study. \*\*\*

**AB. Singular or collaborative authorship:** Special instructions for authorship: Only count author(s) identified as primary, regular author. May be described in author-related fields as: blogger, author, editor, co-editor or contributor. Do not include guest editors, emeriti

bloggers, or other categories of bloggers that denote a former or temporary relationship to blog:

(0) Single; (1) Collaborative

**AC. If AB-1 (collaboratively-authored), how many authors:**

Enter total in digits

---

**Criteria 8: Blogger qualifies as scholar**

---

**AD. If AB-1 (collaboratively-authored), how many authors have published posts within the previous 1 month (approximated as 31 days) from acquisition date:**

Enter total in digits; if cannot be determined, enter “Unknown”

**AE. Scholar Status:** For all single-authored bloggers and any collaborative bloggers identified in AD, are s/he a scholar, evidenced by: A) One or more of the following keywords or phrases: *Ph.D.; Dr.; Professor; Reader; Lecturer; Doctoral Student; Doctoral Candidate*; B) One or more of the following keywords or phrases: *Scholar, Academic, Researcher, Research Director, and/or Fellow* AND identified Institutional Affiliation, either by name or description, denoting a higher education or research setting; C) Link to CV (full or partial) or other listing (full or partial) of publications published by identified blogger with one or more journal articles authored by blogger; or D) Blogger described as a graduate student AND reference to scope/area of study is provided:

(0) No/none; (1) Some; (2) Yes – all; (3) Cannot Determine based on criteria A-D; (4) Both 1 and 3 (4)

**AF. If AE-3 or AE-4, how many total bloggers per instance cannot be determined:**

Enter total in digits

**AG. If AE-1 (some) or AE-4, how many meet researcher-defined parameters as “scholar:”**

Enter total in digits.

\*\*\* If AE-0 or AE-3, stop coding. Blog is not eligible for study. \*\*\*

**AH. Does blog, and subsequently the blog’s authors, qualify for the sampling at this point?**

(0) No; (1) Yes; (3) Duplicate listing

\*\*\* If AH-1, proceed to “Sampling Frame 2: Bloggers” to continue assessment of blogger eligibility.

**SAMPLING FRAME 2: BLOGGERS**

---

**Criteria 9: Blogger contact information**

---

**A. Blog Title** | Title as it appears in Sampling Frame 1 (Columns A or N):

Enter Title

**B. Blog ID** | Unique 3-digit number assigned (see Sampling Frame 1, Column B):

Enter as: ###

**C. Blogger ID** | Unique 5-character ID assigned to blogger:

Enter as: XX-###

**D. Acquisition Date** | Date blogger contact information assessed for sampling frame:

mm\_dd\_yyyy

**E. URL** | Blog URL as it appears in Sampling Frame 1 (Columns D or I):

Enter URL

**F. Authorship** | For sampled instance as it appears in Sampling Frame 1 (Column AB):

(0) Single; (1) Collaborative

**G. Blogger name | Type:**

(0) Pseudonym; (1) First name, including transparently derived nickname; (2) First name + Initial; (3) Initial + last name; (4) Full name, characterized, at minimum, first and last names; Other [Text: Briefly describe]

**H. Blogger name:** Record name to fullest extent possible:

Enter name

**I: Contact Email:**

(0) No; (1) Yes, found on first page of blog; (2) Yes, found via an internal link; (3) Yes, found via an external link (i.e., to CV, personal website, Blogger profile); (4) Yes, through simple Google search when blogger's name and institutional affiliation made known in "about" information posted to blog

**J. If I-1, 2, 3, or 4, record email address:**

Enter email address

**K. If I-0, is there a contact form on the blog where a personalized, private message may be sent:**

(0) No; (1) Yes

\*\*\* If I-0, stop coding. Blogger is not eligible to participate \*\*\*

**L. Does blogger qualify for sample?**

(0) No; (1) Yes; (3) Duplicate

**M. PIN** | Unique 4-digit number assigned for questionnaire administration:

Enter: ####

**N. Salutation** | Address to be used in personalized email correspondence:

(0) Pseudonym; (1) First Name; (2) Professor Surname; (3) Dr. Surname; (4) Mr. Surname; (5) Ms. Surname; Other [Text: Briefly describe]

## APPENDIX B: DISQUALIFIED BLOGS

The 644 blogs listed at the sample source were assessed for eligibility between July 30 and August 4, 2010. Presented below are the counts and percentages of blogs (n=456), by cluster, that did not meet the eligibility criteria. Results are presented in the order by which the criteria were assessed.

Criterion	Ineligible History Blogs (N=143)	Ineligible Economics Blogs (N=141)	Ineligible Law Blogs (N=72)	Ineligible BioChemPhys Blogs (N=100)
<b>One: Blog is <u>not</u> publicly available</b>	21 (15%)	29 (21%)	6 (8%)	18 (18%)
<b>Two: Blog is <u>not</u> published in English</b>	9 (6%)	12 (9%)	2 (3%)	3 (3%)
<b>Three: Blog is a <u>not</u> a knowledge or personal blog</b>	13 (9%)	11 (8%)	18 (25%)	4 (4%)
<b>Four: Blog posts are <u>not</u> time-stamped</b>	1 (1%)	-	-	1 (1%)
<b>Five: Blog is <u>not</u> actively published to</b>	77 (54%)	57 (40%)	35 (49%)	56 (56%)
<b>Six: Blog is <u>less than</u> one year old</b>	10 (7%)	17 (12%)	5 (7%)	8 (8%)
<b>Seven: Blog <u>does not</u> contain personal identifiers in regard to authorship</b>	5 (4%)	7 (5%)	5 (7%)	6 (6%)
<b>Eight: Blog is <u>not</u> authored by one or more bloggers meeting scholar parameters</b>	7 (5%)	8 (6%)	1 (1%)	4 (4%)



## APPENDIX C: QUESTIONNAIRE I (SINGLE-BLOGGERS)

**Note on Appendix C:** Reproduced here is the questionnaire, as presented in Qualtrics. Some directional statements, distinguished by brackets, are included to show how questionnaire appeared to respondents. These directional statements were not included in the administered version of the questionnaire. Specifically: 1) page breaks are included to reflect the number of questions presented to respondents on each page of the questionnaire; 2) skip logic is noted to reflect respondents' movement through the questionnaire, based on their respective responses; and 3) questions are numbered in this appendix to support navigation through the instrument. Questions were unnumbered in the administered questionnaire due to skip logic. Instead, a progress bar was presented at the bottom of each page of the questionnaire, showing the percentage of questions completed. Additionally, since questions unnumbered, a note at the beginning of each section was included to inform respondents as to how many questions comprised each respective section of the questionnaire. Lastly, other elements present in the administered questionnaire were a header and a footer, presented on each page of the questionnaire. The header contained an abbreviated title of the study, Scholars, Blogs & Digital Presentation. Displayed in the footer was the institutional name (University of North Carolina at Chapel Hill), and the IRB study number.

---

### WELCOME

Please read through this statement to learn more about this study. Then, click YES at the bottom of this page to continue to the survey, or click NO THANKS to exit the survey.

You're being asked to participate because the blog identified in your emailed invitation was listed at the Academic Blog Portal. You are one of about 300 scholars invited to take the survey. Your participation is completely voluntary.

Completion of the survey should take between 20 to 40 minutes. Depending on your answers, you will be asked between 41 to 58 primarily closed-ended questions.

You are free to answer or not answer any question and are not obligated to complete answering questions once you begin.

Every effort will be taken to protect your identity. All data obtained will be reported as group data. No individual, institution, or blog will be identified when reporting findings.

There are no risks anticipated should you participate. This study investigates the publishing behaviors and motivations of scholars who blog, the characteristics of their scholarly blogs, and their preferences and perceptions regarding digital preservation. Through participating, you will be helping to further understanding of blogging by scholars and implications for digital preservation. There is no financial benefit for your participation. The cost to

participate is your time, and for that I am very appreciative. You may contact me, Carolyn Hank, with any questions at either 919-259-3191 or hcarolyn@email.unc.edu.

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB\_subjects@unc.edu. If you contact the IRB, please refer to study number 10-1254.

Electronically submitting your completed survey connotes your consent to be a participant in this study.

**Would you like to continue on to the first survey question?**

Yes *[Skip Logic: Go to Next Page]*

No Thanks (EXIT) *[Skip Logic: Go to Exit Screen]*

---

**[Page Break]**

[1] To continue to the first question of the survey, please enter the 4-Digit PIN provided in your emailed invitation to participate:

(Single-line Text Box)

---

**[Page Break]**

**BACKGROUND**

*In this section of the survey, depending on your responses, you will be presented between 4 and 5 questions.*

[2] **Question:** Which one of the following best describes your field of study?

History

Economics

Law

Biology

Chemistry

Physics

None of the above (please describe): (Single-line Text Box)

[3] **Question:** What is the highest academic degree you have earned (e.g., BA, BS, MS, MA, LBB, JD, JSD, PhD, EdD)?

(Single-line Text Box)

[4] **Question:** In what year did you earn this degree?

(Single-line Text Box)

---

**[Page Break]**

---

[5] **Question:** Besides the blog referenced in your emailed invitation to participate, do you publish any other blogs?

Yes [*Skip Logic: If Yes, go to 6*]

No [*Skip Logic: If No, go to 7*]

---

**[Page Break]**

---

[6] **Follow-Up Question:** Approximately, how many other blogs do you publish for which:

You are the only author? (Drop down menu: 0, 1, 2, 3, 4, 5, 6 or more)

You are one of 2 or more authors? (Drop down menu: 0, 1, 2, 3, 4, 5, 6 or more)

---

**[Page Break]**

---

**Please Read:** Respond to the remaining questions in this survey based on the blog identified in your emailed invitation to participate, identified throughout as “your blog,” “my blog,” or “this blog,” unless specifically instructed otherwise.

## **BLOGGING AND SCHOLARLY COMMUNICATION**

*This section of the survey contains 7 questions.*

[7] **Question:** To what extent do you agree or disagree that your blog satisfies these criteria?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
It is public				
It is subject to critical review				
It is in a form that allows use and exchange by other members of the scholarly community				

[8] **Question:** The Association of Research Libraries defines the scholarly record as the message communicated by scholars, which may be published or unpublished and hence either formally or informally communicated. These communications may be reflective of scholarly works, knowledge or ideas.

To what extent do you agree or disagree with this statement: “My blog is a part of my scholarly record.”

Strongly agree

Somewhat agree

Somewhat disagree

Strongly disagree

---

**[Page Break]**

---

[9] **Question:** Do you feel your blog has ever led to any of the following invitations to publish?

	Yes	No
Invitation to publish or co-publish a textbook		
Invitation to edit or co-edit a scholarly book or monograph		
Invitation to author or co-author a scholarly book or monograph		
Invitation to author or co-author a chapter in a scholarly book or monograph		
Invitation to author or co-author an article in a peer-reviewed journal		
Invitation to comment in a peer-reviewed journal		
Invitation to prepare a book review in a scholarly journal		
Invitation to prepare a book review for a newspaper or a general magazine		
Invitation to prepare a nonfiction essay, article, or op-ed piece in a newspaper or general magazine		
Invitation to prepare a scholarly paper for a conference		

**[Page Break]**

[10] **Question:** Do you feel your blog has ever led to any of the following invitations to provide service?

	Yes	No
Invitation to evaluate a scholarly book or manuscript for a publisher		
Invitation to serve as an editor or associate editor of a scholarly journal		
Invitation to serve your scholarly or professional society in some capacity (e.g., conference planning committee, working group, etc.)		

**[Page Break]**

[11] **Question:** Do you feel your blog has ever led to an invitation to serve as a speaker or panelist at a conference?

Yes  
No

[12] **Question:** Do you feel your blog has ever led to an invitation to collaborate on a research project?

Yes  
No

**[Page Break]**

[13] **Question:** To what extent do you feel that publishing your blog has improved or impaired the following aspects of your scholarly life? (Skip any statement that does not apply).

	Greatly improved	Somewhat improved	Neither improved nor impaired	Somewhat impaired	Strongly impaired
My research productivity					
My creativity in examining research in new ways					
My ability to share pre-publication material with colleagues					
The overall quality of my research					
The quality of my writing					
My writing efficiency					
The quality of my teaching					
My enjoyment of my work as a scholar					
My greater visibility as a scholar					
Opportunities for promotion at my institution					

**[Page Break]**

## **BLOGGING ACTIVITY, IDENTITY, AND AUDIENCE**

This section of the survey contains **6 questions**.

[14] **Question:** On average, how often do you update this blog? Updating includes adding new posts, modifying existing posts, responding to comments, modifying links, and other activities where you create new content or modify content published to your blog.

- 10 times a day or more
- 5 to 9 times a day
- 3 or 4 times a day
- 1 or 2 times a day
- A few times a week
- At least once a week
- Once every few weeks
- Once a month
- Less than once a month

---

**[Page Break]**

---

[15] **Question:** Which statement best describes how frequently you update your blog now compared to when you first began to publish your blog?

- I blog a lot more frequently
- I blog somewhat more
- I blog about the same amount
- I blog somewhat less
- I blog a lot less

---

**[Page Break]**

---

[16] **Question:** Where is your blog currently hosted? If hosted at more than one location, please select the location you identify as your blog's primary host.

- A free third party blog publishing and hosting service (e.g., Blogger, Wordpress.com)
- A paid third party blog hosting service (e.g., IPOWERWEB, Go Daddy)
- A managed blog network (e.g., ScienceBlogs, Law Professors Blogs)
- A personal server
- A work server
- A commercial server
- None of these (please describe): (Single-line Text Box)
- I do not know

[17] **Question:** Have you ever migrated your blog from one hosted location to another?

- Yes, just once
- Yes, more than once
- Yes and no; I didn't but my service provider or network did
- No

---

**[Page Break]**

---

[18] **Question:** Which group listed below do you feel best represents the primary audience for your blog?

- Colleagues and professional peers
- Family and friends
- Myself
- General public
- Students
- None of the above (please describe): (Single-line Text Box)

[19] **Question:** Which group listed below do you feel best represents the secondary audience for your blog?

I only blog for my primary audience  
Colleagues and professional peers  
Family and friends  
Myself  
General public  
Students  
None of the above (please describe): (Single-line Text Box)

---

**[Page Break]**

---

**BLOG PUBLISHING BEHAVIORS AND PREFERENCES**

*In this section, depending on your responses, you will be presented between 7 and 10 questions.*

[20] **Question:** Do you make all of the posts you publish to your blog publicly available?

Yes [Skip Logic: If Yes, go to 22]

No [Skip Logic: If No, go to 21]

---

**[Page Break]**

---

[21] **Follow-Up Question:** Approximately, how many posts on your blog would you say are not publicly available?

All (100%)  
Most (80-99%)  
A majority (60-79%)  
About half (40-59%)  
Some (20-39%)  
A minority (1-19%)

---

**[Page Break]**

---

[22] **Question:** For blog posts you have published during the past month, did you import content originating from someone other than yourself? Content includes text, images, video, and audio.

Yes [Skip Logic: If Yes, go to 23]

No [Skip Logic: If No, go to 24]

---

**[Page Break]**

---

[23] **Follow-Up Question:** For the imported content, about how often do you check whether or not there are any rights and use statements associated with the content before publishing it (i.e., Copyright statement, Creative Commons license)?

Always  
Usually

About half the time  
Seldom  
Never

---

**[Page Break]**

---

[24] **Question:** How important do you feel it is for you to verify permission to post content you believe may be protected by copyright before you publish this content to a post in your blog? Verifying permission can be through multiple means, including directly contacting authors or complying with copyright or use statements applied to the content at the source location.

Very important  
Somewhat important  
Somewhat unimportant  
Not at all important

[25] **Question:** How important do you feel it is for others to verify permission to share content originating in a post you published to your blog?

Very important  
Somewhat important  
Somewhat unimportant  
Not at all important

---

**[Page Break]**

---

[26] **Question:** Have you ever responded to comments posted to your blog?

Yes  
No  
I can't because my blog does not have a commenting system  
I can't for another reason (please describe): (Single-line Text Box)

[27] **Question:** On average, how often have you contributed comments to someone else's blog during the past year?

More than once a day  
About once a day  
More than once a week  
About once a week  
Two to three times a month  
About once a month  
A few times  
Not at all

---

**[Page Break]**

---



**[28] Question: Do you Twitter?**

No *[Skip Logic: If No, go to 30]*

No, I read tweets but I do not personally Twitter *[Skip Logic: If No, go to 30]*

Yes, I Twitter *[Skip Logic: If Yes, go to 29]*

---

**[Page Break]**

---

**[29] Follow-Up Question:** Is your Twitter account linked in any way to your blog?

Yes, my blog posts are automatically syndicated to my Twitter account and I don't ever post tweets that aren't associated with my blog.

Yes, my blog posts are automatically syndicated to my Twitter account, but I also post tweets that aren't associated with my blog.

Yes, but in some other way (please describe): (Single-line Text Box)

No, my Twitter account and my blog are not linked.

---

**[Page Break]**

---

**BLOG REVISIONS**

*In this section, depending on your responses, you will be presented between 3 and 8 questions.*

**[30] Question:** Have you ever intentionally edited a blog post after it was published to your blog?

Yes *[Skip Logic: If Yes, go to 31]*

No *[Skip Logic: If No, go to 33]*

---

**[Page Break]**

---

**[31] Follow-Up Question:** Approximately, for posts published to your blog over the past three months, how many posts do you estimate you have intentionally edited after initial publication?

All (100%)

Most (80-99%)

A majority (60-79%)

About half (40-59%)

Some (20-39%)

A minority (1-19%)

None (0%)

**[32] Follow-Up Question:** Have you ever intentionally edited blog posts after initial publication for any of the following reasons?

	Yes	No
To make corrections (e.g., spelling or grammatical errors, formatting errors)		

To made revisions (e.g., rework phrasing, add new information, update links)		
To revise information I no longer believe is correct		
To revise opinions I no longer believe		
In response to a request from someone		
To remove sensitive or revealing content		
Because accidentally published post before it was complete		

#### [Page Break]

[33] **Question:** Have you ever intentionally deleted a blog post after it was published to your blog?

Yes [*Skip Logic: If Yes, go to 34*]

No [*Skip Logic: If No, go to 36*]

#### [Page Break]

[34] **Follow-Up Question:** Approximately, for posts published to your blog over the past three months, how many posts do you estimate you have intentionally deleted after publication?

All (100%)

Most (80-99%)

A majority (60-79%)

About half (40-59%)

Some (20-39%)

A minority (1-19%)

None (0%)

[35] **Follow-Up Question:** Have you ever intentionally deleted blog posts after initial publication for any of the following reasons?

	Yes	No
Because of readability (e.g., spelling, grammar, formatting errors, bad phrasing, bad links)		
Because of information I no longer believe is correct		
Because of opinions I no longer believe		
Because of loss of interest in the subject of the post		
In response to a request from someone		
To remove sensitive or revealing content		
Because I accidentally published post before it was complete		
Because I considered it a test or “experimental” post		
Because it was a duplicate post		
Because of “post regret;” I just didn’t want it out there anymore		

#### [Page Break]

[36] **Question:** Have you ever changed the “skin” of your blog? Skin refers to the layout and design of your blog.

Yes [*Skip Logic: If Yes, go to 37*]

Yes and no; I didn’t but my service provider or network did [*Skip Logic: If Selected, go to 38*]

No [*Skip Logic: If No, go to 38*]

---

**[Page Break]**

[37] **Follow-Up Question:** Approximately, how often do you change your blog’s skin?

At least once a month

Every few months

A couple of times a year

About once a year

Less than once a year

---

**[Page Break]**

## **BLOG PRESERVATION BEHAVIORS**

*In this section, depending on your responses, you will be presented between 3 and 5 questions.*

[38] **Question:** Currently, do you purposefully save a copy of your entire blog, including content and user element features, like commenting systems, blog roll, and archive feature, via an archiving service (e.g., Internet Archive, an institutional repository or digital archive)?

Yes [*Skip Logic: If Yes, go to 39*]

No [*Skip Logic: If No, go to 40*]

Maybe, I am not sure if the service qualifies as an “archiving service” (*Skip Logic: If Selected, go to 39*)

[39] **Follow-Up Question:** Please identify the service or, if more than one, services you use:

(Single-line Text Box)

---

**[Page Break]**

[40] **Question:** Currently, do you purposefully save a copy of your entire blog, including content and user element features, like commenting systems, blog roll, and archive feature, independent of such a service (e.g., exporting database to a server)?

Yes [*Skip Logic: If Yes, go to 41*]

No [*Skip Logic: If No, go to 42*]

No, I only use an archiving service [*Skip Logic: If No, go to 48*]

---

**[Page Break]**

[41] **Follow-Up Question:** Which of the strategies below best describes how you have saved your blog independently of an archiving service? Select all that apply.

- Export tool from my blog publishing and service provider
- Via a syndication service (e.g., RSS feed)
- Mirrored to another Web site (e.g., to another blog)
- Printed to paper
- Some other strategy (please describe): (Single-line Text Box)

*[Skip Logic: After question displayed, go to 48]*

---

**[Page Break]**

[42] **Question:** Currently, do you purposefully save any of these components of your blog for personal use?

- Posts with comments *[Skip Logic: If selected, go to 43]*
- Posts and comments, but separately *[Skip Logic: If selected, go to 44]*
- Posts only *[Skip Logic: If selected, go to 45]*
- Comments only *[Skip Logic: If selected, go to 46]*
- No, I save none of these components *(Skip Logic: If No, go to 48)*

---

**[Page Break]**

[43] **Follow-Up Question:** Approximately how many posts with comments have you purposefully saved?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

*[Skip Logic: After question displayed, go to 47]*

---

**[Page Break]**

[44] **Follow-Up Question:** Approximately how many posts and comments have you purposefully saved?

	All (100%)	Most (80-99%)	A majority (60-79%)	About half (40-59%)	Some (20-39%)	A minority (1-19%)
POSTS						
COMMENTS						

*[Skip Logic: After question displayed, go to 47]*

---

**[Page Break]**

---

[45] **Follow-Up Question:** Approximately how many posts have you purposefully saved?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

*[Skip Logic: After question displayed, go to 47]*

---

**[Page Break]**

---

[46] **Follow-Up Question:** Approximately how many comments have you purposefully saved?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

---

**[Page Break]**

---

[47] **Follow-Up Question:** Which of the strategies below best describes how you have saved this content? Select all that apply.

- Export tool from my blog publishing and service provider
- Via a syndication service (e.g., RSS feed)
- Mirrored to another Web site (e.g., to another blog)
- Printed to paper
- Deposited in an institutional repository or digital archive
- Some other strategy (please describe): (Single-line Text Box)

---

**[Page Break]**

---

## **BLOG PRESERVATION PREFERENCES**

The following **3 questions** concern your perceptions on digital preservation. Digital preservation, as used in this section, is defined as: *“The act of maintaining information in a correct and independently understandable form for a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community. This period extends into the indefinite future.”*

[48] **Question:** To what extent do you agree or disagree with each of the following statements regarding your blog?

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
My blog should be preserved for my PERSONAL access and use into the INDEFINITE future.					
My blog should be preserved for PUBLIC access and use into the INDEFINITE future.					
My blog should be preserved for my PERSONAL access and use only into the SHORT-TERM future (i.e., 2020, 2035).					
My blog should be preserved for PUBLIC access and use only into the SHORT-TERM future (i.e., 2020, 2035).					

**[Page Break]**

[49] **Question:** Please rate each of the following groups with regard to how much responsibility you feel each has for the preservation of your blog:

	Much	Some	A Little	None	Group Does Not Apply
Myself					
My blog's host (e.g., service provider, hosting company, or network)					
Search engines (e.g., Google)					
Public trust (e.g., Internet Archive)					
My institution's library					
My national library					
My institution's archive					
My national Archive					
My institution's IT department					

[50] **Question:** Please rate each of the following groups with regard to how much capability you feel each has for the preservation of your blog:

	Much	Some	A Little	None	Group Does Not Apply
Myself					

My blog's host (e.g., service provider, hosting company, or network)					
Search engines (e.g., Google)					
Public trust (e.g., Internet Archive)					
My institution's library					
My national library					
My institution's archive					
My national Archive					
My institution's IT department					

### [Page Break]

## OTHER BLOGGING ACTIVITIES

*In this section, depending on your responses, you will be presented between **2 and 5** questions.*

[51] **Question:** Is this blog the first blog you have ever published?

Yes [*Skip Logic: If Yes, go to 53*]

No [*Skip Logic: If No, go to 52*]

### [Page Break]

[52] **Follow-Up Question:** In what year did you publish your first blog?

2009

2008

2007

2006

2005

2004

2003

2002

2001

2000

1999

1998

1997 or earlier

### [Page Break]

[53] **Question:** Across your overall blogging activities, apart from the blog identified in your emailed invitation to participate, have you ever deleted an entire blog?

Yes, on purpose [*Skip Logic: If Yes, go to 54*]

Yes, but on accident [*Skip Logic: If Yes, go to 55*]

Yes, both on purpose and accidentally [*Skip Logic: If Yes, go to 54*]

No [*Skip Logic: If No, go to 56*]

---

**[Page Break]**

---

[54] **Follow-Up Question:** Have you ever intentionally deleted an entire blog for any of the following reasons?

	Yes	No
Because too much time required to maintain blog		
Because too expensive to maintain blog		
Because blog was unpopular or under-read		
Because of dissatisfaction with blog publishing application		
Because of dissatisfaction with blog hosting service or location		
Because of loss of interest in the subject of the blog		
In response to a request from someone		
Because it was a project-based or time-limited blog		
Because it was a duplicate blog		
Because of “blog regret,” I just didn’t want it out there anymore		
Because blog received too much spam		

[55] **Follow-Up Question:** Approximately, how many blogs do you estimate you have deleted?

ON PURPOSE (Drop down menu: 0, 1, 2, 3, 4, 5 or more)

ON ACCIDENT (Drop down menu: 0, 1, 2, 3, 4, 5 or more)

---

**[Page Break]**

---

## DEMOGRAPHICS

The questionnaire is nearly complete. The remaining questions are designed to collect demographic information. Again, your answers are confidential. All data obtained from the study will be reported as group data. No individual or institution will be identified.

[56] **Question:** What is your gender?

Male

Female

[57] **Question:** What is your age? Enter in digits:

(Single-line Text Box)

---

**[Page Break]**

---

[58] **Question:** Have you (with or without co-authors) ever published any of the following items?



	Yes	No
Textbook		
Scholarly book or monograph (as editor or co-editor)		
Scholarly book or monograph (as author or co-author)		
Chapter in a scholarly book		
Article in a peer-reviewed journal		
Comment in a peer-reviewed journal		
Book review in a scholarly journal		
Book review for a newspaper or general magazine		
Scholarly paper in a conference proceedings		
Nonfiction essay, article, or op-ed piece in a newspaper or general magazine		

[59] **Question:** Have you ever:

	Yes	No
Served as a referee for a scholarly journal article?		
Evaluated a scholarly book manuscript for a publisher?		
Served as an editor or associate editor of a scholarly journal?		

#### [Page Break]

There are only a few questions left. These remaining questions ask information about your current occupation. Please respond for your **PRIMARY** occupation; for example, if you are a tenured Associate Professor but also work as a Consultant, please respond in reference to the former.

[60] **Question:** Do you most identify as self-employed in regard to your primary employment?

Yes [*Skip Logic: If Yes, go to 61*]

No [*Skip Logic: If Yes, go to 63*]

#### [Page Break]

[61] **Follow-Up Question:** Please provide a brief description of your primary job-related activities.

(Multi-line Text Box)

[62] **Follow-Up Question:** Where do you live? If you maintain two or more residences, places report for your primary residence:

State/Province: (Single-line Text Box)

County: (Single-line Text Box)

*[Skip Logic: After question displayed, go to 66]*

---

**[Page Break]**

[63] **Question:** What is your job title?

(Single-line Text Box)

[64] **Follow-Up Question:** What is your tenure status?

Not applicable; not in a tenure track position

Tenured

Non-tenured in tenure track position

Prefer not to answer

[65] **Follow-Up Question:** What is the name of the institution where you are currently employed?

(Multi-line Text Box)

---

**[Page Break]**

**Thank you for taking the time to complete this questionnaire. Your input is greatly appreciated.**

The next phase of this study involves phone interviews. An interview will take approximately 15 to 25 minutes to complete.

[66] Would you be interested in participating in a follow-up phone interview?

Yes *[Skip Logic: If Yes, go to 67]*

No *[If No, go to 68]*

---

**[Page Break]**

[67] Thank you for your interest. You may be contacted within the next week to schedule a phone interview. You are free to decline participation at that time. Please provide your contact information to facilitate an interview request. Again, this does not obligate you to continue your participation.

Name: (Single-line Text Box)

Email: (Single-line Text Box)

Telephone (optional): (Single-line Text Box)

---

**[Page Break]**

[68] A final report from this study, including findings from the survey you just completed, will be available in January 2011.

Would you like to be contacted by email when the report is available?

Yes

No Thanks

---

**[Page Break]**

---

Thank you for your time spent taking this survey. Your response has been recorded

*[End of questionnaire]*

## APPENDIX D: QUESTIONNAIRE II (CO-BLOGGERS)

**Note on Appendix D:** Reproduced here is the questionnaire, as presented in Qualtrics. Some directional statements, distinguished by brackets, are included to show how questionnaire appeared to respondents. These directional statements were not included in the administered version of the questionnaire. Specifically: 1) page breaks are included to reflect the number of questions presented to respondents on each page of the questionnaire; 2) skip logic is noted to reflect respondents' movement through the questionnaire, based on their respective responses; and 3) questions are numbered in this appendix to support navigation through the instrument. Questions were unnumbered in the administered questionnaire due to skip logic. Instead, a progress bar was presented at the bottom of each page of the questionnaire, showing the percentage of questions completed. Additionally, since questions unnumbered, a note at the beginning of each section was included to inform respondents as to how many questions comprised each respective section of the questionnaire. Lastly, other elements present in the administered questionnaire were a header and a footer, presented on each page of the questionnaire. The header contained an abbreviated title of the study, Scholars, Blogs & Digital Presentation. Displayed in the footer was the institutional name (University of North Carolina at Chapel Hill), and the IRB study number.

---

### WELCOME

Please read through this statement to learn more about this study. Then, click YES at the bottom of this page to continue to the survey, or click NO THANKS to exit the survey.

You're being asked to participate because the blog identified in your emailed invitation was listed at the Academic Blog Portal. You are one of about 300 scholars invited to take the survey. Your participation is completely voluntary.

Completion of the survey should take between 20 to 40 minutes. Depending on your answers, you will be asked between 41 to 62 primarily closed-ended questions.

You are free to answer or not answer any question and are not obligated to complete answering questions once you begin.

Every effort will be taken to protect your identity. All data obtained will be reported as group data. No individual, institution, or blog will be identified when reporting findings.

There are no risks anticipated should you participate. This study investigates the publishing behaviors and motivations of scholars who blog, the characteristics of their scholarly blogs, and their preferences and perceptions regarding digital preservation. Through participating, you will be helping to further understanding of blogging by scholars and implications for digital preservation. There is no financial benefit for your participation. The cost to

participate is your time, and for that I am very appreciative. You may contact me, Carolyn Hank, with any questions at either 919-259-3191 or hcarolyn@email.unc.edu.

All research on human volunteers is reviewed by a committee that works to protect your rights and welfare. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the Institutional Review Board at 919-966-3113 or by email to IRB\_subjects@unc.edu. If you contact the IRB, please refer to study number 10-1254.

Electronically submitting your completed survey connotes your consent to be a participant in this study.

**Would you like to continue on to the first survey question?**

Yes [*Skip Logic: Go to Next Page*]

No Thanks (EXIT) [*Skip Logic: Go to Exit Screen*]

---

**[Page Break]**

[1] To continue to the first question of the survey, please enter the 4-Digit PIN provided in your emailed invitation to participate:

(Single-line Text Box)

---

**[Page Break]**

**BACKGROUND**

*In this section of the survey, depending on your responses, you will be presented between 4 and 5 questions.*

[2] **Question:** Which one of the following best describes your field of study?

History

Economics

Law

Biology

Chemistry

Physics

None of the above (please describe): (Single-line Text Box)

[3] **Question:** What is the highest academic degree you have earned (e.g., BA, BS, MS, MA, LBB, JD, JSD, PhD, EdD)?

(Single-line Text Box)

[4] **Question:** In what year did you earn this degree?

(Single-line Text Box)

---

**[Page Break]**

---

[5] **Question:** Besides the blog referenced in your emailed invitation to participate, do you publish any other blogs?

Yes [*Skip Logic: If Yes, go to 6*]

No [*Skip Logic: If No, go to 7*]

---

**[Page Break]**

---

[6] **Follow-Up Question:** Approximately, how many other blogs do you publish for which:

You are the only author? (Drop down menu: 0, 1, 2, 3, 4, 5, 6 or more)

You are one of 2 or more authors? (Drop down menu: 0, 1, 2, 3, 4, 5, 6 or more)

---

**[Page Break]**

---

**Please Read:** Respond to the remaining questions in this survey based on the blog identified in your emailed invitation to participate, unless specifically instructed otherwise. This blog is identified throughout as “your blog,” “my blog,” or “this blog,” *even though it is known that you are one of two or more contributors, or authors, to this blog*. The use of “your blog” or “my blog” is done solely to facilitate data collection.

Your blog's co-author(s) may or may not have been invited to participate. Please respond to questions in this survey based on your experiences, perceptions, and preferences, rather than how you feel your co-author(s) might respond.

## **BLOGGING AND SCHOLARLY COMMUNICATION**

*This section of the survey contains 7 questions.*

[7] **Question:** To what extent do you agree or disagree that your blog satisfies these criteria?

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
It is public				
It is subject to critical review				
It is in a form that allows use and exchange by other members of the scholarly community				

[8] **Question:** The Association of Research Libraries defines the scholarly record as the message communicated by scholars, which may be published or unpublished and hence either formally or informally communicated. These communications may be reflective of scholarly works, knowledge or ideas.

To what extent do you agree or disagree with this statement: “My blog is a part of my scholarly record.”

Strongly agree  
Somewhat agree  
Somewhat disagree  
Strongly disagree

**[Page Break]**

[9] **Question:** Do you feel your blog has ever led to any of the following invitations to publish?

	Yes	No
Invitation to publish or co-publish a textbook		
Invitation to edit or co-edit a scholarly book or monograph		
Invitation to author or co-author a scholarly book or monograph		
Invitation to author or co-author a chapter in a scholarly book or monograph		
Invitation to author or co-author an article in a peer-reviewed journal		
Invitation to comment in a peer-reviewed journal		
Invitation to prepare a book review in a scholarly journal		
Invitation to prepare a book review for a newspaper or a general magazine		
Invitation to prepare a nonfiction essay, article, or op-ed piece in a newspaper or general magazine		
Invitation to prepare a scholarly paper for a conference		

**[Page Break]**

[10] **Question:** Do you feel your blog has ever led to any of the following invitations to provide service?

	Yes	No
Invitation to evaluate a scholarly book or manuscript for a publisher		
Invitation to serve as an editor or associate editor of a scholarly journal		
Invitation to serve your scholarly or professional society in some capacity (e.g., conference planning committee, working group, etc.)		

**[Page Break]**

[11] **Question:** Do you feel your blog has ever led to an invitation to serve as a speaker or panelist at a conference?

Yes  
No

[12] **Question:** Do you feel your blog has ever led to an invitation to collaborate on a research project?

Yes  
No

**[Page Break]**

[13] **Question:** To what extent do you feel that publishing your blog has improved or impaired the following aspects of your scholarly life? (Skip any statement that does not apply).

	Greatly improved	Somewhat improved	Neither improved nor impaired	Somewhat impaired	Strongly impaired
My research productivity					
My creativity in examining research in new ways					
My ability to share pre-publication material with colleagues					
The overall quality of my research					
The quality of my writing					
My writing efficiency					
The quality of my teaching					
My enjoyment of my work as a scholar					
My greater visibility as a scholar					
Opportunities for promotion at my institution					

**[Page Break]**

## **BLOGGING ACTIVITY, IDENTITY, AND AUDIENCE**

This section of the survey contains **6 questions**.

[14] **Question:** On average, how often do you update this blog? Updating includes adding new posts, modifying existing posts, responding to comments, modifying links, and other activities where you create new content or modify content published to your blog.

10 times a day or more  
5 to 9 times a day  
3 or 4 times a day  
1 or 2 times a day  
A few times a week



At least once a week  
Once every few weeks  
Once a month  
Less than once a month

---

**[Page Break]**

---

[15] **Question:** Which statement best describes how frequently you update your blog now compared to when you first began to publish this blog?

I blog a lot more frequently  
I blog somewhat more  
I blog about the same amount  
I blog somewhat less  
I blog a lot less

---

**[Page Break]**

---

[16] **Question:** Where is your blog currently hosted? If hosted at more than one location, please select the location you identify as your blog's primary host.

A free third party blog publishing and hosting service (e.g., Blogger, Wordpress.com)  
A paid third party blog hosting service (e.g., IPOWERWEB, Go Daddy)  
A managed blog network (e.g., ScienceBlogs, Law Professors Blogs)  
A personal server  
A work server  
A commercial server  
None of these (please describe): (Single-line Text Box)  
I do not know

[17] **Question:** Have you ever migrated your blog from one hosted location to another?

Yes, just once  
Yes, more than once  
Yes and no; I didn't but my service provider, network, or co-author did  
No

---

**[Page Break]**

---

[18] **Question:** Which group listed below do you feel best represents the primary audience for your blog?

Colleagues and professional peers  
Family and friends  
Myself  
General public  
Students  
None of the above (please describe): (Single-line Text Box)

[19] **Question:** Which group listed below do you feel best represents the secondary audience for your blog?

- I only blog for my primary audience
- Colleagues and professional peers
- Family and friends
- Myself
- General public
- Students
- None of the above (please describe): (Single-line Text Box)

---

**[Page Break]**

---

## **BLOG PUBLISHING BEHAVIORS AND PREFERENCES**

*In this section, depending on your responses, you will be presented between 7 and 10 questions.*

[20] **Question:** Do you make all of the posts you publish to your blog publicly available?

- Yes [*Skip Logic: If Yes, go to 22*]
- No [*Skip Logic: If No, go to 21*]

---

**[Page Break]**

---

[21] **Follow-Up Question:** Approximately, how many of the posts you publish to your blog would you say are not publicly available?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

---

**[Page Break]**

---

[22] **Question:** For blog posts you have published during the past month, did you import content originating from someone other than yourself? Content includes text, images, video, and audio.

- Yes [*Skip Logic: If Yes, go to 23*]
- No [*Skip Logic: If No, go to 24*]

---

**[Page Break]**

---

[23] **Follow-Up Question:** For the imported content, about how often do you check whether or not there are any rights and use statements associated with the content before publishing it (i.e., Copyright statement, Creative Commons license)?

- Always
- Usually
- About half the time
- Seldom
- Never

---

**[Page Break]**

---

[24] **Question:** How important do you feel it is for you to verify permission to post content you believe may be protected by copyright before you publish this content to a post in your blog? Verifying permission can be through multiple means, including directly contacting authors or complying with copyright or use statements applied to the content at the source location.

- Very important
- Somewhat important
- Somewhat unimportant
- Not at all important

[25] **Question:** How important do you feel it is for others to verify permission to share content originating in a post you published to your blog?

- Very important
- Somewhat important
- Somewhat unimportant
- Not at all important

---

**[Page Break]**

---

[26] **Question:** Have you ever responded to comments made to a post you published to your blog?

- Yes
- No
- I can't because my blog does not have a commenting system
- I can't for another reason (please describe): (Single-line Text Box)

[27] **Question:** On average, how often have you contributed comments to someone else's blog during the past year?

- More than once a day
- About once a day
- More than once a week
- About once a week

Two to three times a month  
About once a month  
A few times  
Not at all

---

**[Page Break]**

---

**[28] Question: Do you Twitter?**

No *[Skip Logic: If No, go to 30]*  
No, I read tweets but I do not personally Twitter *[Skip Logic: If No, go to 30]*  
Yes, I Twitter *[Skip Logic: If Yes, go to 29]*

---

**[Page Break]**

---

**[29] Follow-Up Question:** Is your Twitter account linked in any way to your blog?

Yes, my blog posts are automatically syndicated to my Twitter account and I don't ever post tweets that aren't associated with my blog.  
Yes, my blog posts are automatically syndicated to my Twitter account, but I also post tweets that aren't associated with my blog.  
Yes, but in some other way (please describe): (Single-line Text Box)  
No, my Twitter account and my blog are not linked.

---

**[Page Break]**

---

**BLOG REVISIONS**

*In this section, depending on your responses, you will be presented between 3 and 8 questions.*

**[30] Question:** Have you ever intentionally edited a blog post after you published it to your blog?

Yes *[Skip Logic: If Yes, go to 31]*  
No *[Skip Logic: If No, go to 33]*

---

**[Page Break]**

---

**[31] Follow-Up Question:** Approximately, for all of your posts published to this blog over the past three months, how many posts do you estimate you have intentionally edited after initial publication?

All (100%)  
Most (80-99%)  
A majority (60-79%)  
About half (40-59%)  
Some (20-39%)  
A minority (1-19%)

None (0%)

[32] **Follow-Up Question:** Have you ever intentionally edited blog posts after initial publication for any of the following reasons?

	Yes	No
To make corrections (e.g., spelling or grammatical errors, formatting errors)		
To made revisions (e.g., rework phrasing, add new information, update links)		
To revise information I no longer believe is correct		
To revise opinions I no longer believe		
In response to a request from someone		
To remove sensitive or revealing content		
Because accidentally published post before it was complete		

**[Page Break]**

[33] **Question:** Have you ever intentionally deleted a blog post after you published it to your blog?

Yes *[Skip Logic: If Yes, go to 34]*

No *[Skip Logic: If No, go to 36]*

**[Page Break]**

[34] **Follow-Up Question:** Approximately, for all of the posts published to your blog over the past three months, how many posts do you estimate you have intentionally deleted after publication?

All (100%)

Most (80-99%)

A majority (60-79%)

About half (40-59%)

Some (20-39%)

A minority (1-19%)

None (0%)

[35] **Follow-Up Question:** Have you ever intentionally deleted blog posts after initial publication for any of the following reasons?

	Yes	No
Because of readability (e.g., spelling, grammar, formatting errors, bad phrasing, bad links)		
Because of information I no longer believe is correct		
Because of opinions I no longer believe		
Because of loss of interest in the subject of the post		
In response to a request from someone		

To remove sensitive or revealing content		
Because I accidentally published post before it was complete		
Because I considered it a test or “experimental” post		
Because it was a duplicate post		
Because of “post regret;” I just didn’t want it out there anymore		

---

**[Page Break]**

---

[36] **Question:** Have you ever changed the “skin” of your blog? Skin refers to the layout and design of your blog.

Yes [*Skip Logic: If Yes, go to 37*]

Yes and no; I didn’t but my service provider, network, or co-author did [*Skip Logic: If Selected, go to 38*]

No [*Skip Logic: If No, go to 38*]

---

**[Page Break]**

---

[37] **Follow-Up Question:** Approximately, how often do you change your blog’s skin?

At least once a month

Every few months

A couple of times a year

About once a year

Less than once a year

---

**[Page Break]**

---

## **BLOG PRESERVATION BEHAVIORS**

*In this section, depending on your responses, you will be presented between 3 and 10 questions.*

[38] **Question:** Currently, do you or a co-author purposefully save a copy of your entire blog, including content and user element features, like commenting systems, blog roll, and archive feature, via an archiving service (e.g., Internet Archive, an institutional repository or digital archive)?

*There are several responses to choose from since your blog is co-produced. From the responses below, please select only the response you feel best answers the question:*

Yes, I do [*Skip Logic: If Selected, go to 39*]

Yes, but not me; a co-author does [*Skip Logic: If Selected, go to 42*]

Maybe I do, but I am not sure if the service qualifies as an “archiving service” [*Skip Logic: If Selected, go to 39*]

Maybe my co-author does, but I am not sure if the service qualifies as an “archiving service” [*Skip Logic: If Selected, go to 42*]

No, not me, but I am not sure if a co-author does [*Skip Logic: If Selected, go to 42*]

No, neither my co-author nor I do [*Skip Logic: If Selected, go to 42*]

---

**[Page Break]**

---

[39] **Follow-Up Question:** How important do you feel it is to get permission from your co-author(s) to save the entire blog via an archiving service?

- Very important
- Somewhat important
- Somewhat unimportant
- Not at all important

[40] **Follow-Up Question:** Do you have permission from your co-author(s) to save the entire blog via an archiving service?

- Yes
- No
- Sort of (please explain): (Single-line Text Box)

[41] **Follow-Up Question:** Please identify the service or, if more than one, services you use:

(Single-line Text Box)

---

**[Page Break]**

---

[42] **Question:** Currently, do you or a co-author purposefully save a copy of your entire blog, including content and user element features, like commenting systems, blog roll, and archive feature, independent of such a service (e.g., exporting database to a server)?

*Again, there are several responses from which to choose. Please select the response you feel best answers the question:*

- Yes, I do [*Skip Logic: If Selected, go to 43*]
- Yes, but not me; a co-author does [*Skip Logic: If Selected, go to 46*]
- No, not me, but I am not sure if a co-author does [*Skip Logic: If Selected, go to 46*]
- No, neither my co-author nor I do [*Skip Logic: If Selected, go to 46*]
- No, only an archiving service is used [*Skip Logic: If Selected, go to 52*]

---

**[Page Break]**

---

[43] **Follow-Up Question:** How important do you feel it is to get permission from your co-author(s) to save the entire blog independently of an archiving service?

- Very important
- Somewhat important
- Somewhat unimportant
- Not at all important

[44] **Follow-Up Question:** Do you have permission from your co-author(s) to save the entire blog independently of an archiving service?

Yes

No

Sort of (please explain): (Single-line Text Box)

[45] **Follow-Up Question:** Which of the strategies below best describes how you have saved your blog independently of an archiving service? Select all that apply.

Export tool from my blog publishing and service provider

Via a syndication service (e.g., RSS feed)

Mirrored to another Web site (e.g., to another blog)

Printed to paper

Some other strategy (please describe): (Single-line Text Box)

*[Skip Logic: After question displayed, go to 52]*

---

**[Page Break]**

---

[46] **Question:** Currently, do you purposefully save any of these components of your blog for personal use?

My posts with comments *[Skip Logic: If selected, go to 47]*

My posts and comments, but separately *[Skip Logic: If selected, go to 48]*

My posts only *[Skip Logic: If selected, go to 49]*

Comments to my posts only *[Skip Logic: If selected, go to 50]*

No, I save none of these components *(Skip Logic: If No, go to 52)*

---

**[Page Break]**

---

[47] **Follow-Up Question:** Approximately how many of your posts with comments have you purposefully saved?

All (100%)

Most (80-99%)

A majority (60-79%)

About half (40-59%)

Some (20-39%)

A minority (1-19%)

*[Skip Logic: After question displayed, go to 51]*

---

**[Page Break]**

---

[48] **Follow-Up Question:** Approximately how many of your posts and comments have you purposefully saved?



	All (100%)	Most (80-99%)	A majority (60-79%)	About half (40-59%)	Some (20-39%)	A minority (1-19%)
POSTS						
COMMENTS						

*[Skip Logic: After question displayed, go to 51]*

---

**[Page Break]**

---

[49] **Follow-Up Question:** Approximately how many of your posts have you purposefully saved?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

*[Skip Logic: After question displayed, go to 51]*

---

**[Page Break]**

---

[50] **Follow-Up Question:** Approximately how many comments to your posts have you purposefully saved?

- All (100%)
- Most (80-99%)
- A majority (60-79%)
- About half (40-59%)
- Some (20-39%)
- A minority (1-19%)

---

**[Page Break]**

---

[51] **Follow-Up Question:** Which of the strategies below best describes how you have saved this content? Select all that apply.

- Export tool from my blog publishing and service provider
- Via a syndication service (e.g., RSS feed)
- Mirrored to another Web site (e.g., to another blog)
- Printed to paper
- Deposited in an institutional repository or digital archive
- Some other strategy (please describe): (Single-line Text Box)

---

**[Page Break]**

---

## BLOG PRESERVATION PREFERENCES

The following **3 questions** concern your perceptions on digital preservation. Digital preservation, as used in this section, is defined as: *“The act of maintaining information in a correct and independently understandable form for a period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community. This period extends into the indefinite future.”*

[52] **Question:** To what extent do you agree or disagree with each of the following statements regarding your blog?

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree
My CONTENT on the blog should be preserved for my PERSONAL access and use into the INDEFINITE future.					
My CONTENT on the blog should be preserved for PUBLIC access and use into the INDEFINITE future.					
My BLOG, as a whole, should be preserved for myself and my co-author(s)' PERSONAL access and use into the INDEFINITE future.					
My BLOG, as a whole, should be preserved for PUBLIC access and use into the INDEFINITE future.					
My CONTENT on the blog should be preserved for my PERSONAL access and use only into the SHORT-TERM future (i.e., 2020, 2035).					
My CONTENT on the blog should be preserved for PUBLIC access and use only into the SHORT-TERM future (i.e., 2020, 2035).					
My BLOG, as a whole, should be preserved for myself and my co-author(s)' PERSONAL					

access and use only into the SHORT-TERM future (i.e., 2020, 2035).					
My BLOG, as a whole, should be preserved for PUBLIC access and use only into the SHORT-TERM future (i.e., 2020, 2035).					

**[Page Break]**

[53] **Question:** Please rate each of the following groups with regard to how much responsibility you feel each has for the preservation of your blog:

	Much	Some	A Little	None	Group Does Not Apply
Myself					
My blog's co-author(s)					
My blog's host (e.g., service provider, hosting company, or network)					
Search engines (e.g., Google)					
Public trust (e.g., Internet Archive)					
My institution's library					
My national library					
My institution's archive					
My national Archive					
My institution's IT department					

[54] **Question:** Please rate each of the following groups with regard to how much capability you feel each has for the preservation of your blog:

	Much	Some	A Little	None	Group Does Not Apply
Myself					
My blog's co-author(s)					
My blog's host (e.g., service provider, hosting company, or network)					
Search engines (e.g., Google)					
Public trust (e.g., Internet Archive)					
My institution's library					
My national library					
My institution's archive					
My national Archive					
My institution's IT department					

---

**[Page Break]**

---

**OTHER BLOGGING ACTIVITIES**

*In this section, depending on your responses, you will be presented between 2 and 5 questions.*

[55] **Question:** Is this blog the first blog you have ever published?

Yes [*Skip Logic: If Yes, go to 57*]

No [*Skip Logic: If No, go to 56*]

---

**[Page Break]**

---

[56] **Follow-Up Question:** In what year did you publish your first blog?

2009

2008

2007

2006

2005

2004

2003

2002

2001

2000

1999

1998

1997 or earlier

---

**[Page Break]**

---

[57] **Question:** Across your overall blogging activities, apart from the blog identified in your emailed invitation to participate, have you ever deleted an entire blog?

Yes, on purpose [*Skip Logic: If Yes, go to 58*]

Yes, but on accident [*Skip Logic: If Yes, go to 59*]

Yes, both on purpose and accidentally [*Skip Logic: If Yes, go to 58*]

No [*Skip Logic: If No, go to 60*]

---

**[Page Break]**

---

[58] **Follow-Up Question:** Have you ever intentionally deleted an entire blog for any of the following reasons?

	Yes	No
Because too much time required to maintain blog		
Because too expensive to maintain blog		
Because blog was unpopular or under-read		

Because of dissatisfaction with blog publishing application		
Because of dissatisfaction with blog hosting service or location		
Because of loss of interest in the subject of the blog		
In response to a request from someone		
Because it was a project-based or time-limited blog		
Because it was a duplicate blog		
Because of "blog regret;" I just didn't want it out there anymore		
Because blog received too much spam		

[59] **Follow-Up Question:** Approximately, how many blogs do you estimate you have deleted?

ON PURPOSE (Drop down menu: 0, 1, 2, 3, 4, 5 or more)

ON ACCIDENT (Drop down menu: 0, 1, 2, 3, 4, 5 or more)

### [Page Break]

## DEMOGRAPHICS

The questionnaire is nearly complete. The remaining questions are designed to collect demographic information. Again, your answers are confidential. All data obtained from the study will be reported as group data. No individual or institution will be identified.

[60] **Question:** What is your gender?

Male

Female

[61] **Question:** What is your age? Enter in digits:

(Single-line Text Box)

### [Page Break]

[62] **Question:** Have you (with or without co-authors) ever published any of the following items? Co-authors, as used here, refers to co-authors across your scholarly publishing activities rather than specifically to your blog's co-author(s):

	Yes	No
Textbook		
Scholarly book or monograph (as editor or co-editor)		
Scholarly book or monograph (as author or co-author)		
Chapter in a scholarly book		
Article in a peer-reviewed journal		
Comment in a peer-reviewed journal		
Book review in a scholarly journal		
Book review for a newspaper or general magazine		

Scholarly paper in a conference proceedings		
Nonfiction essay, article, or op-ed piece in a newspaper or general magazine		

[63] **Question:** Have you ever:

	Yes	No
Served as a referee for a scholarly journal article?		
Evaluated a scholarly book manuscript for a publisher?		
Served as an editor or associate editor of a scholarly journal?		

#### [Page Break]

There are only a few questions left. These remaining questions ask information about your current occupation. Please respond for your **PRIMARY** occupation; for example, if you are a tenured Associate Professor but also work as a Consultant, please respond in reference to the former.

[64] **Question:** Do you most identify as self-employed in regard to your primary employment?

Yes *[Skip Logic: If Yes, go to 65]*

No *[Skip Logic: If Yes, go to 67]*

#### [Page Break]

[65] **Follow-Up Question:** Please provide a brief description of your primary job-related activities.

(Multi-line Text Box)

[66] **Follow-Up Question:** Where do you live? If you maintain two or more residences, places report for your primary residence:

State/Province: (Single-line Text Box)

County: (Single-line Text Box)

*[Skip Logic: After question displayed, go to 70]*

#### [Page Break]

[67] **Question:** What is your job title?

(Single-line Text Box)

[68] **Follow-Up Question:** What is your tenure status?

Not applicable; not in a tenure track position  
Tenured  
Non-tenured in tenure track position  
Prefer not to answer

[69] **Follow-Up Question:** What is the name of the institution where you are currently employed?

(Multi-line Text Box)

---

**[Page Break]**

---

**Thank you for taking the time to complete this questionnaire. Your input is greatly appreciated.**

The next phase of this study involves phone interviews. An interview will take approximately 15 to 25 minutes to complete.

[70] Would you be interested in participating in a follow-up phone interview?

Yes [*Skip Logic: If Yes, go to 71*]  
No [*If No, go to 72*]

---

**[Page Break]**

---

[71] Thank you for your interest. You may be contacted within the next week to schedule a phone interview. You are free to decline participation at that time. Please provide your contact information to facilitate an interview request. Again, this does not obligate you to continue your participation.

Name: (Single-line Text Box)  
Email: (Single-line Text Box)  
Telephone (optional): (Single-line Text Box)

---

**[Page Break]**

---

[72] A final report from this study, including findings from the survey you just completed, will be available in January 2011.

Would you like to be contacted by email when the report is available?

Yes  
No Thanks

---

**[Page Break]**

---

Thank you for your time spent taking this survey. Your response has been recorded

*[End of questionnaire]*

## APPENDIX E: QUESTIONNAIRE PRE-TEST CORRESPONDENCE

---

### QUESTIONNAIRE I (SINGLE-BLOGGER) TESTERS

---

**Subject Line:** Help in testing my Scholars, their Blogs and Digital Preservation survey

Hello [Name],

I am preparing to release my questionnaires in support of my dissertation research study, Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation. As someone who both blogs and researches issues related to digital preservation, I would appreciate your help in testing one of the instruments. Below you will find the personalized recruitment letter I plan to email invited participants. If you are able to help with the pre-testing, I ask that you read through the letter and then, following the link provided in the letter, complete the questionnaire. Beyond taking the questionnaire, you are welcome to provide feedback in a follow-up email to me. I am particularly interested in hearing about:

1. Time to complete: In testing so far, time to complete has ranged between 25 to 45 minutes. Was that your experience?
2. Clarity: Were the questions and answer choices clear? Was there any confusing language, or were there any questions/answer choices you felt would benefit from the provision of "definitions?"
3. Recall and Comprehension: Keeping in mind that the questionnaire is intended for scholars who have maintained active blogs -- characterized as updates, at minimum, once every two weeks -- for at least one year in the areas of history, law, economics, biology, chemistry and physics, do you feel that the questions were easy to answer. That is, did any demand an excessive amount of time (for recollection/comprehension) on the part of the respondent?
4. Any other suggestions you feel would be helpful would be much appreciated.

If you are able to participate in this pre-test, the survey (link provided below) will be open for you until August 3<sup>rd</sup>.

Thank you in advance for your help. Any feedback you can provide is much appreciated.

Best,  
Carolyn

-----



## RECRUITMENT EMAIL

Subject Line: Study: Scholars and Their Blogs

Dear [Name]:

The neologisms, bloggership and blogademia, have emerged in recent years to describe scholars' adoption of blogs as units of communication. But, are blogs scholarship? Where do they fit in relation to one's cumulative scholarly record? Consider preservation, a primary function of the system of scholar communication. Due to the speed for which digital communications change – as well as our own publishing behaviors and preferences – will the scholar blogs of today be available into the future?

I hope you will consider sharing your opinions on some of these outstanding questions by participating in a survey in support of my research study, "Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation." You will be asked questions about your own blog publishing behaviors, your perceptions of your blog in relation to your scholarly activities, and your thoughts on preservation. Scholars who blog in the areas of biology, chemistry, economics, history, law and physics have been invited to participate. The survey should take between 25 to 45 minutes to complete.

Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [INSERT BLOG NAME], available at: [INSERT URL]. When accessing the survey, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

The survey is NOW OPEN, and will remain open until AUGUST 3, 2010 at the URL provided below:

SURVEY URL: [SURVEY URL]

YOUR PIN: [XXXX]

If you have any questions about participating, feel free to contact me. This study has been approved by the UNC-CH Institutional Review Board (IRB). If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the IRB at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu). If you contact the IRB, please refer to study number 10-1254.

My research, in part, is supported by a Eugene Garfield Doctoral Dissertation Fellowship, awarded by Beta Phi Mu. The findings from this stage of the research will be complemented with data from analysis of over 100 scholar blogs and interviews with about 20 scholar bloggers.

Thank you in advance for considering participation in my study.

Sincerely,

Carolyn Hank

Carolyn Hank  
Doctoral Candidate  
University of North Carolina at Chapel Hill  
School of Information and Library Science  
tel 919.259.3191  
fax 919.962.8071  
<http://ils.unc.edu/~hcarolyn/>

---

## QUESTIONNAIRE II (CO-BLOGGER) TESTERS

---

**Subject: Help in testing my Scholars, their Blogs and Digital Preservation survey**

Hello [Name],

I am preparing to release my questionnaires in support of my dissertation research study, Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation. As someone who both blogs and researches issues related to digital preservation, I would appreciate your help in testing one of the instruments. Below you will find the personalized recruitment letter I plan to email invited participants. If you are able to help with the pre-testing, I ask that you read through the letter and then, following the link provided in the letter, complete the questionnaire. Beyond taking the questionnaire, you are welcome to provide feedback in a follow-up email to me. I am particularly interested in hearing about:

1. Time to complete: In testing so far, time to complete has ranged between 25 to 45 minutes. Was that your experience?
2. Clarity: Were the questions and answer choices clear? Was there any confusing language, or were there any questions/answer choices you felt would benefit from the provision of "definitions?"
3. Recall and Comprehension: Keeping in mind that the questionnaire is intended for scholars who have maintained active blogs -- characterized as updates, at minimum, once every two weeks -- for at least one year in the areas of history, law, economics, biology, chemistry and physics, do you feel that the questions were easy to answer. That is, did any demand an excessive amount of time (for recollection/comprehension) on the part of the respondent?
4. Any other suggestions you feel would be helpful would be much appreciated.

If you are able to participate in this pre-test, the survey (link provided below) will be open for you until August 3<sup>rd</sup>.

Thank you in advance for your help. Any feedback you can provide is much appreciated.

Best,  
Carolyn

-----

#### EMAIL RECRUITMENT EMAIL

Subject Line: Study: Scholars and Their Blogs

Dear [Name],

The neologisms, bloggership and blogademia, have emerged in recent years to describe scholars' adoption of blogs as units of communication. But, are blogs scholarship? Where do they fit in relation to one's cumulative scholarly record? Consider preservation, a primary function of the system of scholar communication. Due to the speed for which digital communications change – as well as our own publishing behaviors and preferences – will the scholar blogs of today be available into the future?

I hope you will consider sharing your opinions on some of these outstanding questions by participating in a survey in support of my research study, "Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation." You will be asked questions about your own blog publishing behaviors, your perceptions of your blog in relation to your scholarly activities, and your thoughts on preservation. Scholars who blog in the areas of biology, chemistry, economics, history, law and physics have been invited to participate. The survey should take between 25 to 45 minutes to complete.

Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [BLOG TITLE], available at: [BLOG URL]. When accessing the survey, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

Additionally, I realize you are just one of the scholars publishing to this blog. Your co-authors may or may not have been selected to participate. When completing the questionnaire, please respond based on your own experiences, perceptions, and preferences, rather than how you feel your co-authors might respond.

The survey is now OPEN, and will remain open until AUGUST 3, 2010 at the URL provided below:

SURVEY URL: [SURVEY URL]

YOUR PIN: [XXXX]

If you have any questions about participating, feel free to contact me. This study has been approved by the UNC-CH Institutional Review Board (IRB). If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the IRB at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu). If you contact the IRB, please refer to study number 10-1254.

My research, in part, is supported by a Eugene Garfield Doctoral Dissertation Fellowship, awarded by Beta Phi Mu. The findings from this stage of the research will be will be complemented with data from analysis of over 100 scholar blogs and interviews with about 20 scholar bloggers.

Thank you in advance for considering participation in my study.

Sincerely,

Carolyn Hank

Carolyn Hank  
Doctoral Candidate  
University of North Carolina at Chapel Hill  
School of Information and Library Science  
tel 919.259.3191  
fax 919.962.8071  
<http://ils.unc.edu/~hcarolyn/>

## APPENDIX F: QUESTIONNAIRE RECRUITMENT CORRESPONDENCE

---

### INITIAL INVITATION: QUESTIONNAIRE I (SINGLE-BLOGGERS)

---

**Subject Line:** Study: Scholars and their Blogs

Dear [Title and Name],

The phrases “bloggership” and “blogademia” have emerged in recent years to describe scholars’ adoption of blogs as units of communication. But, are blogs scholarship? Where do they fit in relation to one’s cumulative scholarly record? Consider preservation, a primary function of the system of scholar communication. Due to the speed for which digital communications change – as well as our own publishing behaviors and preferences – will the scholar blogs of today be available into the future?

I hope you will consider sharing your opinions on some of these outstanding issues by participating in a survey in support of my research study, “Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation.” You will be asked questions about your publishing behaviors, your perceptions of your blog in relation to your scholarly activities, and your thoughts on preservation. Scholars who blog in the areas of biology, chemistry, economics, history, law and physics have been invited to participate.

Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [BLOG TITLE, available at: [URL] When accessing the survey, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

The survey is now open. It will remain open until midnight (EDT) 27 August 2010 at the URL provided immediately below. The survey should take between 20 to 40 minutes to complete.

SURVEY URL: [QUESTIONNAIRE I URL]

YOUR PIN: [XXXX]

This study has been approved by the Institutional Review Board (IRB) at the University of North Carolina at Chapel Hill. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the IRB at 919-966-3113 or by email to IRB\_subjects@unc.edu. If you contact the IRB, please refer to study number 10-1254.

My research is supported in part by a Eugene Garfield Doctoral Dissertation Fellowship, awarded by Beta Phi Mu. If you have any questions about participating, feel free to contact me.

Many thanks,

Carolyn Hank

Carolyn Hank  
Doctoral Candidate  
University of North Carolina at Chapel Hill  
School of Information and Library Science  
tel 919.259.3191  
fax 919.962.8071  
<http://ils.unc.edu/~hcarolyn/>

---

INITIAL INVITATION: QUESTIONNAIRE II (CO-BLOGGERS)

---

**Subject Line:** Study: Scholars and their Blogs

Dear [Title and Name],

The phrases “bloggership” and “blogademia” have emerged in recent years to describe scholars’ adoption of blogs as units of communication. But, are blogs scholarship? Where do they fit in relation to one’s cumulative scholarly record? Consider preservation, a primary function of the system of scholar communication. Due to the speed for which digital communications change – as well as our own publishing behaviors and preferences – will the scholar blogs of today be available into the future?

I hope you will consider sharing your opinions on some of these outstanding issues by participating in a survey in support of my research study, “Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation.” You will be asked questions about your publishing behaviors, your perceptions of your blog in relation to your scholarly activities, and your thoughts on preservation. Scholars who blog in the areas of biology, chemistry, economics, history, law and physics have been invited to participate.

Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [BLOG TITLE] available at: [BLOG URL] Additionally, I realize you are just one of the scholars publishing to this blog. Your co-author(s) may or may not have been selected to participate. When accessing the survey, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

The survey is now open. It will remain open until midnight (EDT) 27 August 2010 at the URL provided immediately below. The survey should take between 20 to 40 minutes to complete.

SURVEY URL: [QUESTIONNAIRE II URL]

YOUR PIN: [XXXX]

This study has been approved by the Institutional Review Board (IRB) at the University of North Carolina at Chapel Hill. If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the IRB at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu). If you contact the IRB, please refer to study number 10-1254.

My research is supported in part by a Eugene Garfield Doctoral Dissertation Fellowship, awarded by Beta Phi Mu. If you have any questions about participating, feel free to contact me.

Many thanks,

Carolyn Hank  
[Signature Line]

---

FIRST REMINDER: QUESTIONNAIRE I & II

---

**Subject Line:** Scholars and Their Blogs: What Do You Think?

Dear [Title and Name],

Do you consider your blog scholarship? Should your blog be preserved? I invite you to share your opinions on these and other outstanding questions by taking part in a 20-to-40 minute survey.

Last week I emailed you and told you about the survey I'm administering in support of my research study, "Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation." I am surveying scholars who blog in the areas of biology, chemistry, economics, history, law, and physics.

If you choose to participate, you will be asked questions about your publishing behaviors, your perceptions of your blog in relation to your scholarly activities, and your thoughts on preservation.

Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [BLOG TITLE], available at: [BLOG URL] When accessing the survey, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

The survey is now open and will remain open until midnight (EDT) 27 August 2010 at the URL provided immediately below:

SURVEY URL: [QUESTIONNAIRE I OR II URL]

YOUR PIN: [XXXX]

If you have any questions about participating, feel free to contact me.

Additionally, this study has been approved by the University of North Carolina at Chapel Hill's Institutional Review Board (IRB). If you have questions or concerns about your rights as a research subject you may contact the IRB at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu). If you contact the IRB, please refer to study number 10-1254.

Sincerely,

Carolyn  
[Signature Line]

---

**FINAL REMINDER: QUESTIONNAIRE I & II**

---

**Subject Line:** Scholars and Their Blogs Study: Five Days Left

Dear [Title and Name],

I am emailing you as a final reminder about the survey I am administering in support of my research study, Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation.

If you already accessed the survey, my apologies for sending an unnecessary reminder and much thanks for your participation.

If you have not accessed the survey, but would like to participate, the survey will remain open until midnight (EDT) 27 August 2010. Because I realize many bloggers publish to more than one blog, please respond based on your specific blog, [BLOG TITLE], available at: [BLOG URL] When accessing the questionnaire, you will be prompted for a 4-digit PIN. Your PIN is [XXXX].

More information about the study is provided on the first page of the questionnaire, available at: URL: [QUESTIONNAIRE I OR II URL]

If you have any questions about participating, feel free to contact me. This study has been approved by the University of North Carolina at Chapel Hill's Institutional Review Board (study number 10-1254).

And lastly, thank you for your consideration.

Best,  
Carolyn  
[Signature Line]



## APPENDIX G: INTERVIEW CONSENT SCRIPT AND SCHEDULE

---

### CONSENT SCRIPT

---

Hello [NAME], it's Carolyn Hank, calling for our scheduled phone interview. How are you today? **(Wait for reply).**

Good to hear. Before we begin, I first wanted to thank you for completing the questionnaire and agreeing to participate in an interview. Also, I have a prepared consent script as I will need to get your consent to go forward with the interview. May I read you the consent script now? **(Wait for reply).**

Thank you. As you already know, I am conducting research on scholars, their blogs, and implications for digital preservation. I am a Doctoral Candidate at the School of Information and Library Science at the University of North Carolina at Chapel Hill.

Your participation in this interview is completely voluntary. This means that you do not have to participate in this interview unless you want to. You don't have to answer any question that you choose not to answer. We will just skip that question and go on to the next one.

I will not identify you, your blog, or your institution, or use any information you share during the interview that would make it possible for anyone to identify you in any presentation or written report about this study. I will not even record your name with your responses. If it is okay with you, I might want to use direct quotes from you, but these would only be quoted as coming from a person of a certain label or title, like "one tenured professor at a Carnegie 1 research institution said."

There are no other expected risks to you for helping me with this study. There is no financial benefit for your participation. The cost to participate is your time, and for that I am very appreciative.

I would like to record this interview. The digital recording will be maintained by me, with only access by me, in a secure, password protected location. I will permanently delete the recording after completion of my analysis. Further, if a transcript is made, in no way will your name be associated with the transcript. Any identifying personal information received in the course of this interview will be de-identified. If a transcript of our interview is produced, I will be the only person with access to the transcript. Further, it will be destroyed within 1 year.

Lastly, as I shared in the email inviting you to participate in the interview, you may contact my Institutional Review Board, anonymously if you wish, with any questions or concerns,

Thank you. At this point, do you have any questions for me? [Respond to any inquiries].

Do I have your permission to begin recording and asking you questions? (**If yes, continue**)

**If No:** We can still complete the interview even if I cannot record you. Do you give me permission to take notes during our interview? These notes may include direct quotes from you. Again, in no way will these notes be associated with you in any way when reporting findings from this study. (**If yes, continue**)

**If No to Note-Taking:** Thank you. I very much appreciate the time you took to talk with me today and for your past participation in the study. This ends the interview. At this point, can I answer any questions for you? [Respond to any inquiries]. Again, thank you for your time. [End call].

---

## INTERVIEW SCHEDULE

---

- 1) First, how do you define the system of scholarly communication in your own professional life?
- 2) Second, how do you feel your blog fits within the system of scholarly communication?
- 3) Next, in your opinion, how do you think others actually perceive blogs within the system of scholarly communication?
- 4) Now, in your opinion, how do you think others should perceive blogs within the system?
- 5) I would like to now ask you a few questions in regard to digital preservation. Just for some context, digital preservation has been described as a “communication with the future.” Putting aside blogs for the moment, do you think about issues of digital preservation in your everyday research and writing activities?
- 6) Prior to taking the questionnaire, had you ever considered issues of digital preservation in regard to your blog?
- 7) (Co-bloggers only): Since your blog is collaboratively authored, have you and your co-bloggers ever discussed issues related to digital preservation or other issues about the long-term management of your blog?
- 8) (Network-affiliated bloggers only): Since your blog is part of the [NETWORK NAME], were any plans about preservation and/or long-term management of your blog ever provided to you by the network’s editors or managers?
- 9) (Bloggers publishing to 2 or more blogs only): Since you publish to [Count of Blogs] different blogs, how do your opinions on digital preservation vary for your multiple instances, if at all?

- 10) In a climate of finite resources, decisions on what to digitally preserve and what not to preserve are challenging. Consider your own scholarly publishing, both published and unpublished, and formally and informally communicated. In your opinion, what types of communications, besides your blog, would be a high priority to preserve, and what types would be a low priority?
- 11) In your opinion, where do blogs fit in a continuum of priorities to preserve, from high to low, across the different types of communications you produce?
- 12) In the questionnaire, you were asked who should be most responsible for preservation of your blog, as well as who would be most capable. Would you like to comment on why you responded in the way you did. If you'd like a refresher on how you responded, I'd be happy to provide it.
- 13) One last question: How would you feel if your blog was unexpectedly and permanently deleted right now? What, if anything, would you do in response?
- 14) Do you have any other comments you would like to share with me in regard to blogging, scholarly communication, and digital preservation?

Thank you. That concludes our interview session. Thank you for taking the time and allowing me to interview you today. Your valuable input has been very much appreciated.

## APPENDIX H: INTERVIEW DEBRIEFING SHEET

---

### Identification

---

Date of Interview:

Start Time:

End Time:

Blogger ID (Unique 5-character ID assigned to blogger):

---

### Recording

---

Was there a problem with the tape recording? Yes | No

If yes, describe:

What, if any, measures will be taken to correct problem:

Is recording still viable? Yes | No

Was recording saved digitally in secure location and labeled accordingly (mmddyy\_interview ID number assigned by researcher\_rec): Yes | No

---

### Note-taking

---

Were brief notes taken during interview session? Yes | No

Were brief notes saved digitally in secure location and labeled accordingly (mmddyy\_interview ID number assigned by researcher\_prel\_notes): Yes | No

Were extensive notes taken immediately following interview session? Yes | No

Were extensive notes saved digitally in secure location and labeled accordingly (mmddyy\_interview ID number assigned by researcher\_notes): Yes | No

---

### Interview and Schedule

---

Were any unexpected topics introduced by subject: Yes | No

If yes, should these topic(s) be considered in future interviews? Yes | No

Were there questions that did not work? Yes | No

If yes, describe:

How many minutes did it take to complete the interview, from start to finish:

Was there sufficient time to complete the interview? Yes | No

If no, please estimate reason why:

---

**Additional Comments**

---

Please list/describe any additional information regarding this particular interview session, if not covered in previous questions:

## APPENDIX I: INTERVIEW RECRUITMENT CORRESPONDENCE

---

### INVITATION EMAIL

---

**Subject Line:** Scholars and Their Blogs: Interview Request

Dear [Title and Name],

Thank you for participating in my study, “Scholars and their Blogs: Characteristics, Preferences and Perceptions Impacting Digital Preservation.” Your completion of the questionnaire was much appreciated.

The intent of this email is to invite you to participate in the next phase of my research study, a 15-to-25 minute phone interview. I am very interested to hear more about your perceptions on blogging, scholarly communication, and digital preservation. About 20 participants are being interviewed at this stage. I do plan to record the interviews. I will be the only person with access to the recordings. If you choose to participate, I will go over more of my procedures and your rights as a participant at the beginning of the interview.

I will be conducting these interviews between Wednesday, August 18th and Friday, September 3rd. If you are interested in participating, please respond to this email with the following information:

- 1) Your preferred contact information, whether by telephone or VoIP (e.g., Skype)
- 2) Your location (e.g., state/province and country)
- 3) Preferred call times, based on your geographic location
- 4) Preferred days and/or dates

Once I receive your reply, I will send a confirmation email within a day or two that will include a date and time for the interview. If the date and time are no longer open for you, I will gladly reschedule. I will also send a reminder email a day or two before we are scheduled to talk.

Thank you in advance for considering continued participation in my research study. If you have any questions about participating, feel free to contact me. As shared in an earlier email, this study has been approved by the UNC-CH Institutional Review Board (IRB). If you have questions or concerns about your rights as a research subject you may contact, anonymously if you wish, the IRB at 919-966-3113 or by email to [IRB\\_subjects@unc.edu](mailto:IRB_subjects@unc.edu). If you contact the IRB, please refer to study number 10-1254.

Best,  
Carolyn

Carolyn Hank  
Doctoral Candidate  
University of North Carolina at Chapel Hill  
School of Information and Library Science  
tel 919.259.3191  
fax 919.962.8071  
<http://ils.unc.edu/~hcarolyn/>

---

CONFIRMATION EMAIL

---

**Subject Line:** RE: (Subject Line of Respondents' Reply Email)

Dear [Title and/or Name],

Thank you for agreeing to continue participating in my study. Your willingness to be interviewed is much appreciated.

The intent of this email is to confirm the interview date, time, and contact number. Based on the preferences you provided, I have scheduled your phone interview as follows:

Date: [Day, Date, Month, Year]

Time: [xx:xx (EDT/CDT/MDT/PDT)]

Number: [xxx-xxx-xxxx]

If you are no longer able to participate on this date and/or at this time, or if the preferred contact number has changed, please let me know by email to [hcarolyn@email.unc.edu](mailto:hcarolyn@email.unc.edu).

Again, the interview will take approximately 15 to 25 minutes to complete. I will send a confirmation reminder email to you 1 to 2 days before our scheduled interview. You may choose to withdraw from this study at anytime.

Thank you again for your continued participation. If you have any questions about participating, feel free to contact me at any time.

Best,  
Carolyn  
[Signature Line]

---

REMINDER EMAIL

---

**Subject Line:** Reminder: Scholars and Their Blogs Interview

Dear [Title and/or Name],

The intent of this email is to remind you of our interview date, time, and contact number. I will be contacting you tomorrow, [DATE], at [TIME] at [TELEPHONE NUMBER]. The interview will take approximately 15 to 25 minutes to complete.

If you are no longer able to participate on this date and/or at this time, or if your preferred contact number has changed, please let me know by email to hcarolyn@email.unc.edu or phone (919-259-3191).

Also, just to let you know, while my contact number is 919-259-3191, I will be calling you from a different number (area code 440).

I look forward to speaking with you tomorrow.

Cheers,  
Carolyn  
[Signature Line]

---

DECLINE EMAIL

---

**Subject Line:** Follow-up: Scholars and their Blogs Study

Dear [Title and Name],

Thank you for completing my study's questionnaire, and for expressing interest to continue participating in my study. Your willingness to be interviewed is much appreciated.

At this time, however, I received more interest in interviewing than anticipated. As a result, I will not be contacting you further in regard to participating in this, the final phase, of my study.

Thank you again for your continued interest. If you have any questions, feel free to contact me at any time. Also, as you indicated when completing your questionnaire, I will be contacting you later on this year when the final report from the study is available.

Appreciatively yours,

Carolyn  
[Signature Line]



## APPENDIX J: SINGLE- AND CO-BLOG CODING SYSTEM

---

### Data Management

---

**A. Blog ID #:** Unique 3-digit number assigned at sampling

Enter in digits

**B. Sampling Date:** Date blog selected for sample

mm\_dd\_yyyy

**C. Code Date:** Date blog coded for analysis

mm\_dd\_yyyy

**D. Code Time:** Approximate start and finish

hh:mm\_hh:mm

**E. Code Time:** Total duration, in minutes

Enter in digits

---

### Authorship (Special Instruction: Code F – N for Single-Blogs Only)

---

**F. Blogger's Name:** From first page, “about” page, "profile" page and/or “contact” page only

(0) None; (1) Pseudonym; (2) First name, including transparently derived nickname;  
(3) First name + initial; (4) Initial + last name; (5) Full name, characterized, at  
minimum, first and last names; (6) 2 or more of these; (7) Cannot determine

**G. If F-6 (2 or more):** List each occurrence by corresponding number in F, separating each  
occurrence with a semicolon

Enter in digits

**H. If F-7 (Cannot determine):** Briefly explain

Enter Text

**I. Blogger's Name | Location:**

(0) in Header; (1) in Title tag; (2) in URL; (3) in Sidebar; (4) in Author Field in Post;  
(5) 2 or more of these locations; (6) None of these locations

**J. If I-5 (2 or more):** List each occurrence by corresponding number in I, separating each  
occurrence with a semicolon

Enter in digits

**K. Photographic Representation of Blogger:** From first page, “about” page, "profile" page  
and/or “contact” page only. *Special Instructions: Do not include logos or other graphics used  
as icons; only consider photographic elements.*

(0) None; (1) Yes – supports identification ; (2) Yes – does not support identification (i.e., back of head, foot)

**L. Occupation | Type:** From first page, “about” page, "profile" page and/or “contact” page only

(0) None; not provided; (1) Generic– e.g., Faculty, Researcher, Graduate Student, Scholar; (2) Specific – e.g., Associate Professor, Research Scientist, Chemist

**M: Institutional Affiliation | Type:** From first page, “about” page, "profile" page and/or “contact” page only

(0) None; not provided; (1) General – e.g., College, Research organization; (2) Specific – e.g., Antioch College, National Institute of Health (NIH); (3) Self-employed

**N. Highest Degree | Type:** From first page, “about” page, "profile" page and/or “contact” page only

Enter abbreviation (e.g., BA, BS, MS, MA, LBB, JD, JSD, PhD, EdD)

---

### Blog Elements and Features

---

**O. Title | Header:** From first page of blog. *Special Instruction: Header refers to title of blog.*  
(0) None; (1) Yes; (2) Yes, but blog software logo

**P. Title | Tag :** From first Page of blog  
(0) None; (1) Yes

**Q. Title | URL:** *Special Instruction: Exclude host domain name – e.g., BlogSpot.*  
(0) Personal Name, Complete; (1) Personal Name, Partial; (2) Pseudonym, Complete; (3) Pseudonym, Partial; (4) Header (Blog Title), Complete; (5) Header, Partial; (6) Blog Network, Partial; (7) Blog Network, Complete; (8) 2 or more of these; (9) None of these

**R. If Q-8 (2 or more):** List each occurrence by corresponding number in Q, separating each occurrence with a semicolon.  
Enter in digits

**S. Blog Software:** URL or first page, including side bar and footer only  
(0) Unknown; (1) Blogger; (2) Wordpress.com; (3) Wordpress.org; (4) Wordpress, not specified; (5) Spaces - Windows Live; (6) LiveJournal; (7) Xanga; (8) TypePad; (9) MovableType; (10) Vox; (11) Blog-City; (12) Diaryland; (13) Serendipity; (14) Typo; (15) Other

**T. If S-15 (Other):** Provide application name and/or describe  
Enter in text

**U. Commenting System | Posts:**

(0) No; (1) Yes

**V. Categories | Posts:** *Special Instruction: Categories also are identified as “keywords,” “labels,” or other name. Categories, or other label assigned, are descriptors applied to posts generally found in a post's footer or header.*

(0) No; Yes [Enter label(s) as text; if more than one label, separate each label by a semicolon]

**W. Categories | Blog:** *Special Instruction: Categories may also be identified as “keywords,” “tags,” “tag cloud,” or other name. Categories, or other label assigned, are descriptors applied to blog generally found in a blog’s sidebar.*

(0) No; Yes [Enter label(s) as text; if more than one label, separate each label by a semicolon]

**X. Permanent Links | Posts:**

(0) No; (1) Yes - "Permanent Link"; Yes - but by another Label, e.g., Permalink (Enter label(s) as text, separating each occurrence by a semicolon]

**Y. Blogroll:** First page only. *Special Instruction: May also be identified as “friends,” “subscriptions,” “links,” “webring,” “bookmarks,” or other name. Blogrolls, or other label assigned, are lists of links to other blogs generally found in a blog’s sidebar.*

(0) No; (1) Cannot determine; Yes [Enter label(s) as text; if more than one label, separate each label by a semicolon]

**Z. If Y-2 (Cannot determine):** Briefly describe.

Enter as Text

**AA. Syndication:** From first page header, sidebar or footer, or post footer only

(0) No; (1) Yes; (2) Cannot Determine

**AB. Search tool:** From first page only):

(0) No; (1) Yes

**AC. Commercial Advertisements:** From first page header, sidebar or footer, or post header or footer only

(0) No; (1) Yes

---

**Rights and Disclaimers**

---

**AD. Intellectual Property Rights (IPR) Statement:** From first page header, sidebar or footer, post header or footer, or “about” page, “profile” page and/or “contact” page only

(0) No; (1) Creative Commons badge - Side bar; (2) Creative Commons badge - Post; (3) Creative Commons Badge - Other location; (4) Other rights badge; (5) Text

statement - Side bar [e.g., ©]; (6) Text statement - Footer ; (7) Text statement - About or profile page; (8) Text statement - Other location; (9) 2 or more locations; Other [Text]

**AE. If AD-9 (2 or more):** List each occurrence by corresponding number in AD, separating each occurrence with a semicolon.

Enter in digits

**AF. IPR Statement | Extent Specification:** From first page header, sidebar or footer, post header or footer, or “about” page, “profile” page and/or “contact” page only

(0) N/A - No Rights Statement; (1) No specification of extent; (2) “Blog” e.g., “This blog,” This [blog name]”; (3) Posts only; (4) Comments only; (5) Posts and Comments; (6) “This work” – no further specification; (7) “All rights reserved”; Other [Text]

**AG. Policy Statement | Disclaimer:** From first page, about page, or “policy” page only

(0) No; Yes [Text: Identify location on blog and briefly describe]

**AH. Policy Statement | Content Policy:** From first page, about page, or “policy” page only

(0) No; Yes [Text: Identify location on blog and briefly describe (e.g., Content policy page: Does not represent content is accurate or timely]

**AI. Policy Statement | Editing and Revisions:** From first page, about page, or “policy” page only

(0) No; Yes [Text: Identify location on blog and briefly describe]

**AJ. Policy Statement | Comments:** From first page, about page, or “policy” page only

(0) No; Yes [Text: Identify location on blog and briefly describe]

---

### Authority and Audience

---

**AK. Blog Award | Badge:** From first page, “about” page, “profile” page and/or “contact” page only

(0) No; Yes [Enter total number of unique awards listed]

**AL. If AK-Yes:** List each award by name, separating each occurrence with a semicolon.

Enter in text

**AM. Network | Blog:** From first page, “about” page, “profile” page and/or “contact” page only. *Special Instructions: e.g., Law Professor Blogs; ScienceBlogs, etc. Do not include advertisement networks, such as Google Ads or AD Sense.*

(0) No; Yes [List each unique network by name; if more than one, separate each occurrence with a semi-colon]

**AN. Advertisement Network | Blog:** From first page, “about” page, "profile" page and/or “contact” page only

(0) No; Yes [List each unique ad network by name; if more than one, separate each occurrence with a semi-colon]

**AO. Certification | Blog:** From first page, “about” page, "profile" page and/or “contact” page only. E.g., HONcode Certification Badge

(0) No; Yes [Enter total number of certifications listed]

**AP. If AO-Yes:** List each certification by name, separating each occurrence with a semicolon.

Enter in text

**AQ. Number of Blogs in Blogroll:** From first page only. *Special Instruction: If no blogroll, leave blank.*

Enter in digits.

**AR. Blog Stats | Badges and Widgets:** From first page, “about” page, "profile" page and/or “contact” page only. *Special Instruction: Exclude Blogger “Profile” stats.*

(0) No; Yes [List; separate each listing with semi-colon, e.g., SiteMeter; StatCounter]

**AS. If AR-Yes (Blog Stats) | Unique visitors:** *Special Instructions: From statistics plug-in or from linked location; skip if AR- 0.*

(0) No stats provided; Yes [Enter total in digits, followed by source; e.g., 335,034 SiteMeter]

**AT. If AR-Yes (Blog Stats) | Total Visits:** *Special Instructions: From statistics plug-in or from linked location; skip if AR- 0.*

(0) No stats provided; Yes [Enter total in digits, followed by Source; e.g., 335,034 SiteMeter]

**AU. If AR-Yes (Blog Stats) | Avg. Visits Per Day:** *Special Instructions: From statistics plug-in or from linked location; skip if AR- 0.*

(0) No stats provided; Yes [Enter total in digits, followed by Source; e.g., 335,034 SiteMeter]

**AV. If AR-Yes (Blog Stats) | Total Page Views:** *Special Instructions: From statistics plug-in or from linked location); skip if AR- 0.*

(0) No stats provided; Yes [Enter total in digits, followed by Source; e.g., 335,034 SiteMeter]

**AW. If AR-Yes (Blog Stats) | Avg. Page Views per Day:** *Special Instructions: From statistics plug-in or from linked location); skip if AR- 0.*

(0) No stats provided; Yes [Enter total in digits, followed by Source; e.g., 335,034 SiteMeter]

**AX. Indexed on Blog Search Engine | Google:** *Special Instructions: Search at [http://blogsearch.google.com/blogsearch/advanced\\_blog\\_search?hl=en](http://blogsearch.google.com/blogsearch/advanced_blog_search?hl=en); select “advanced blog search;” then paste blog URL at “In Blogs” in field, “at this URL,” and then select “posts written in past month” from drop down menu.*

(0) No; (1) Yes

**AY. Indexed on Blog Search Engine | Technorati:** *Special Instructions: Search at <http://technorati.com>; paste blog URL into search box, and then select “Blogs.”*

(0) No; (1) Yes

**AZ. If AY-1 (Yes) | Technorati Authority Number**

Enter in digits

**BA. If AY-1 (Yes) | Technorati Top 100 Rank**

(0) None; Yes [Enter in digits]

**BB. If AY-1 (Yes) | Technorati Top 100 | Topical**

(0) None; Yes [Enter “Topic,” separating each occurrence with a semicolon; e.g., Top 100 Business]

---

### **Publishing Activity**

---

**BC. Date of Most Recent Post:** *Special Instruction: At time of analysis, not at time of sample frame construction or sampling.*

mm\_dd\_yyyy

**BD. Number | Total Posts in August 2010:** *Special Instruction: Coded for a complete, 1-month reporting period.*

Enter in digits

**BE. Number | Total Posts with Comments in August 2010:** *Special Instruction: Coded for a complete, 1-month reporting period.*

Enter in digits

**BF. Blog | Date of Oldest Post:**

mm\_dd\_yyyy

---

### **Post Features**

---

*Special Instruction for BG – BT: Report data for the most recent post in entry body.*

**BG. Most Recent Post | Title:**

(0) No; (1) Yes

**BH. Most Recent Post | Total Words in Entry Body:** *Special Instructions: Exclude title of entry and other supplemental content (i.e., tags) provided in post's header or footer. To obtain word count, copy and paste text from entry body into a .docx document.*

Enter in digits

**BI. Most Recent Post | Total Words in Entry Body Quoted from Another Source:** *Special Instructions: Count part(s) of entry body assumed to be quoted from another source(s), as distinguished by quotation marks, indentation, italics and/or link to originating source.*

Enter in digits

**BJ. Most Recent Post | Total Links in Text of Entry Body:**

Enter in digits

**BK. Most Recent Post | Total Internal (in-Blog) Links in Text of Entry Body:**

Enter in digits

**BL. Most Recent Post | Total External (Off-Blog) Links in Text of Entry Body:**

Enter in digits

**BM. Most Recent Post | Total Photo Elements Embedded or Imported:**

Enter in digits

**BN. Most Recent Post | Total Static Graph Elements Embedded or Imported:**

Enter in digits

**BO. Most Recent Post | Total Interactive Graph Elements Embedded or Imported:**

Enter in digits

**BP. Most Recent Post | Total Other Image Elements Embedded or Imported (e.g., including clip art, drawings, etc.):**

Enter in digits

**BQ. Most Recent Post | Total Video Elements Embedded or Imported:**

Enter in digits

**BR. Most Recent Post | Total Audio – Music Elements Embedded or Imported:**

Enter in digits

**BS. Most Recent Post | Total Audio – Spoken Elements Embedded or Imported:**

Enter in digits

**BT. Most Recent Post | Total Comments:**

(0) None; 1 or more [Enter total in digits]

**BU. For Post Published Approximately 1 Year Earlier | Total Comments: Special**

*Instructions: Approximation defined as post published in 2009 within 2 weeks of most recent post's month/day; if more than one post published on that date, refer to post that was posted earliest in the day.*

(N/A) No posts published within 2 week time period for the previous year; (0) None - No Comments for post from 1-year ago; 1 or more comments [Enter total in digits]

**BV. FOR BU-0 (None) or BU-1: Date of Post Analyzed:**

mm-dd-yyyy

---

**Archiving**

---

**BW. Law Cluster Blogs | Indexed at Law Library of Congress Blog Archive: Special**

*Instruction: Only code for sampled blogs in the Law Cluster. To determine, browse by blog title at: [http://lcweb2.loc.gov/diglib/lcwa/browsetitles?query="+collection:mrva0015&start=A](http://lcweb2.loc.gov/diglib/lcwa/browsetitles?query=).*

(0) No; (1) Yes

**BX. All Blogs | Indexed in the Internet Archive: Special Instruction: To determine, search by UR under "all media types" at: <http://www.archive.org/web/web.php>.**

(0) No; (1) Yes



## REFERENCES

- Academic Blog Portal. (2006, October 5). Frequently asked questions [Wiki page]. Retrieved from [http://www.academicblogs.org/wiki/index.php/Frequently\\_Asked\\_Questions](http://www.academicblogs.org/wiki/index.php/Frequently_Asked_Questions)
- Althouse, A. (2006). *Why a narrowly defined legal scholarship blog is not what I want: An argument in pseudo-blog form*. (Legal Studies Research Paper Series No. 1021). Madison, WI: University of Wisconsin Legal Studies. Retrieved from <http://ssrn.com/abstract=898171>
- American Association for Public Opinion Research. (2009). *Standard definitions: Final dispositions of case codes and outcome rates for surveys* (6th ed.). Deerfield, IL: Author. Retrieved from [http://www.aapor.org/AM/Template.cfm?Section=Standard\\_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814](http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions1&Template=/CM/ContentDisplay.cfm&ContentID=1814)
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Association of Research Libraries. (1986). *The changing system of scholarly communication*. Washington, DC: Author.
- Association of Research Libraries. (2009). *The research library's role in digital repository services: Final report of the ARL Digital Repository Issues Taskforce*. Washington, DC: Author. Retrieved from <http://www.arl.org/bm~doc/repository-services-report.pdf>
- Babbie, E. (1990). *Survey research methods* (2nd ed.). Belmont, CA: Wadsworth Publishing.
- Bailey, C.W., Coombs, K., Emery, J., Mitchell, A., Morris, C., Simons, S., & Wright, R. (2006). *SPEC kit 292: Institutional repositories*. Washington, DC: Association of Research Libraries.
- Barton, M.R., & Waters, M.M. (2004-2005). *Creating an institutional repository: LEADIRS workbook*. Cambridge, MA: MIT Libraries. Retrieved from <http://www.dspace.org/implement/leadirs.pdf>
- Bauer, M.W. (2000). Classical content analysis: A review. In M.W. Bauer and G. Gaskell (Eds.), *Qualitative researching with text, image and sound: A practical handbook* (pp. 131-151). London, UK: Sage Publications.
- Bazerman, C. (1994). *Constructing experience*. Carbondale, IL: Southern Illinois University Press.

- Bazerman, C. (2008). *Handbook of research on writing: History, society, school, individual, text*. New York, NY: L. Erlbaum Associates.
- Bearman, D. (2005, August 1). How might we improve blogs as an academic medium [Web log message]. Retrieved from [http://prawfsblawg.blogs.com/prawfsblawg/2005/08/how\\_might\\_we\\_im.html](http://prawfsblawg.blogs.com/prawfsblawg/2005/08/how_might_we_im.html)
- Beattie, H.R. (2007). *"The texture of the everyday: Appraising the values of women's diaries and weblogs"* (Master's thesis). Retrieved from ProQuest Dissertation and Theses. (AAT MR36295).
- Berg, B.L. (2004). *Qualitative research methods for the social sciences* (5th ed.). Boston, MA: Pearson Education.
- Blanchard, A. (2004). Blogs as virtual communities: Identifying a sense of community in the Julie/Julia Project. In L. J. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, and J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. St. Paul, MN: University of Minnesota. Retrieved from [http://blog.lib.umn.edu/blogosphere/blogs\\_as\\_virtual.html](http://blog.lib.umn.edu/blogosphere/blogs_as_virtual.html)
- Blouin, F. (1999). Archivists, mediation, and constructs of social memory. *Archival Issues*, 24(2), 101-112.
- Borgman, C.L. (1990). Editor's introduction. In C.L. Borgman (Ed.), *Scholarly communication and bibliometrics* (pp. 7-27). Newbury Park, CA: Sage Publications.
- Borgman, C.L. (2000). Digital libraries and the continuum of scholarly communication. *Journal of Documentation*, 56(4), 412-430.
- Borgman, C.L. (2007). *Scholarship in the digital age: Information, infrastructure and the Internet*. Cambridge, MA: MIT Press.
- Braxton, J.M., Luckey, W., & Helland, P. (2002). *Institutionalizing a broader view of scholarship through Boyer's four domains: ASHE-ERIC higher education report*. San Francisco, CA: Jossey-Bass.
- Brown, J.S., & Duguid, P. (2000). The Social life of information: Chapter one: Limits to information. *First Monday*, 5(4). Retrieved from [http://firstmonday.org/issues/issue5\\_4/brown\\_chapter1.html](http://firstmonday.org/issues/issue5_4/brown_chapter1.html)
- Bruns, A. (2006). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. New York, NY: Peter Lang.
- Caplan, P. (2008). The preservation of digital materials. *Library Technology Reports*, 44(2), 5-38.
- Carnaby, P., & Sutherland, S. (2009, August). *Citizen-created content, digital equity and the preservation of community memory*. Paper presented at the World Library and

- Information Congress: 75th IFLA General Conference and Council, Milan, Italy.  
Paper retrieved from <http://www.ifla.org/files/hq/papers/ifla75/146-carnaby-en.pdf>
- Caron, P.L. (2006). Are scholars better bloggers? Bloggership: How blogs are transforming legal scholarship. *Washington University Law Review*, 84(5), 1025-1042.
- Caron, P.L., & Hodnicki, J.A. (n.d.). Welcome [Web blog message]. Retrieved from <http://www.lawprofessorblogs.com/>
- Carr, L., & Harnad, S. (2005). *Keystroke economy: A study of the time and effort involved in self-archiving*. (Technical Report). Southampton, UK: School of Electronics and Computer Science, University of Southampton. Retrieved from <http://eprints.ecs.soton.ac.uk/10688/>
- Ceccarelli, L. (2001). *Shaping science with rhetoric: The cases of Dobzhansky, Schrodinger, and Wilson*. Chicago, IL: University of Chicago Press.
- Center for Research Libraries & Online Computer Library Center. (2007). *Trustworthy repositories audit and certification: Criteria and checklist*. Chicago, IL: CRL and Dublin, OH: OCLC. Retrieved from [http://www.crl.edu/sites/default/files/attachments/pages/trac\\_0.pdf](http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)
- Chan, L. (2004). Supporting and enhancing scholarship in the digital age: The role of open-access institutional repositories. *Canadian Journal of Communication*, 29(3). Retrieved from <http://www.cjc-online.ca/index.php/journal/article/viewArticle/1455/1579>
- Cline. (2006, July 28). Rhetorica: Professors who blog [Web page]. Retrieved from [http://rhetorica.net/professors\\_who\\_blog.htm](http://rhetorica.net/professors_who_blog.htm)
- Consultative Committee for Space Data Systems. (2002). *Reference model for an open archival information system (OAIS)*. (CCSDS 650.0-B-1: Blue Book, Issue 1). Washington, DC: National Aeronautics and Space Administration. Retrieved from <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- Council on Library and Information Resources. (2001). *The evidence in hand: Report of the Task Force on the Artifact in Library Collections*. (CLIR Publication No. 103). Washington, DC: Author. Retrieved from <http://www.clir.org/pubs/reports/pub103/contents.html>
- Courant, P.N. (2008). Scholarship: The wave of the digital future. In R.N. Katz (Ed.), *The tower and the cloud* (pp. 202-211). Boulder, CO: Educause.
- Cronin, B. (2005). *The hand of science: Academic writing and its rewards*. Lanham, MD: Scarecrow Press.

- Crow, R. (2002). *The case for institutional repositories: A SPARC position paper*. Washington, DC: Scholarly Publishing and Academic Resources Coalition. Retrieved from [http://www.arl.org/sparc/bm~doc/IR\\_Guide\\_&\\_Checklist\\_v1.pdf](http://www.arl.org/sparc/bm~doc/IR_Guide_&_Checklist_v1.pdf)
- Czaja, R., & Blair, J. (2005). *Designing surveys: A guide to decisions and procedures*. Thousand Oaks, CA: Pine Forge Press.
- Dattalo, P. (2008). *Determining sample size: Balancing power, precision and practicality*. New York, NY: Oxford University Press.
- Day, M. (2003). *Prospects for institutional e-print repositories in the United Kingdom*. (ePrints UK Supporting Study, No. 1). Bath, UK: UKOLN. Retrieved from <http://eprints-uk.rdn.ac.uk/project/docs/studies/impact/>
- Dayal, G. (2005a, April 5). PH.Dotcom: What if professors could lecture 24-7? Blog culture invades academia. *Village Voice*. Retrieved from <http://www.villagevoice.com/2005-04-05/art/ph-dotcom/>
- Dayal, G. (2005b, April 5). A brief guide to blogademia. *Village Voice*. Retrieved from <http://www.villagevoice.com/2005-04-05/art/a-brief-guide-to-blogademia/>
- Dempsey, L. (2007, October 27). Quotes of the day (and other days?): Persistent academic discourse [Web log message]. Retrieved from <http://orweblog.oclc.org/archives/001467.html>
- DigitalPreservationEurope. (2007). *DigitalPreservationEurope research roadmap*. (DPE Project no. 034761). Glasgow, UK: Humanities Advanced Technology and Information Institute. Retrieved from [http://www.digitalpreservationeurope.eu/publications/reports/dpe\\_research\\_roadmap\\_D72.pdf](http://www.digitalpreservationeurope.eu/publications/reports/dpe_research_roadmap_D72.pdf)
- Digital Curation Centre & DigitalPreservationEurope. (2007). *DCC and DPE digital repository audit method based on risk assessment, v. 1.0*. Retrieved from <http://www.repositoryaudit.eu/>
- Dillman, D.A. (2007). *Mail and Internet surveys: The tailored design method* (2nd ed.). Hoboken, NJ: Wiley and Sons.
- Dillman, D.A., Smythe, J.D., & Christian, L.M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: Wiley and Sons.
- Entlich, R. (2004). Blog today, gone tomorrow? Preservation of Weblogs. *RLG DigiNews*, 8(4). Retrieved from <http://www.worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file3518.html#article3>
- Farrell, H., & Drezner, D.W. (2008). The power and politics of weblogs. *Public Choice*, 134(1/2), 15-30. doi: 10.1007/s11127-007-9198-1

- Farrell, S., Ashley, K., Davis, R., Guy, M., Kelly, B., & Pinsent, E. (2010). *A guide to web preservation: Practical advice for web and records managers based on best practices from the JISC-funded PoWR project*. London, UK: University of London Computer Center and Bath, UK: UKOLN.
- Ewins, R. (2005). Who are you? Weblogs and academic identity. *E-Learning*, 2(4), 368-376. doi: 10.2304/elea.2005.2.4.368
- Fitzgerald, B., & O'Brien, D. (2006). Bloggers and the law. In A. Bruns & J. Jacobs (Eds.), *Uses of blogs* (pp. 223-234). New York, NY: Peter Lang.
- Fowler, F.J. (2002). *Survey research methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Garrett, J., & Waters, D. (1996). *Preserving digital information: Report of the Task Force on Archiving of Digital Information*. (CLIR Publication No. 63). Washington, DC: The Commission on Preservation and Access. Retrieved from <http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>
- Garvey, W. D., & Griffith, B. C. (1971). Scientific communication: Its role in the conduct of research and creation of knowledge. *American Psychologist*, 26(4), 350-362.
- Gillmor, D. (2006). *We the media*. Sebastopol, CA: O'Reilly Media.
- Ginsparg, P. (2001). Creating a global knowledge network. *Proceedings of the Second ICSU Press-UNESCO Conference on Electronic Publishing in Science*. doi: 10.1244/02.eps2/005.1.00.
- Ginsparg, P. (2006) As we may read. *Journal of Neuroscience*, 26(38), 9606-9608. doi: 10.1523/JNEUROSCI.3161-06.2006.
- Glenn, D. (2003). Scholars who blog. The soapbox of the digital age draws a crowd of academics. *Chronicle of Higher Education*, 49(39).
- Gouge, M.K. (2004). Blogs as a means of preservation selection for the World Wide Web. *Proceedings of the American Society for Information Science and Technology*, 41(1). Silver Spring, MD: American Society for Information Science and Technology. doi: 10.1002/meet.1450420139
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380.
- Gurak, L., Antonijevic, S., Johnson, L., Ratliff, C., & Reyman, J. (Eds.). (2004). *Into the blogosphere: Rhetoric, community, and culture of weblogs*. St. Paul, MN: University of Minnesota. Retrieved from <http://blog.lib.umn.edu/blogosphere/>
- Hagstrom, W.O. (1965). *The scientific community*. New York, NY: Basic Books.

- Halavais, A. (2006). Scholarly blogging: Moving toward the visible college. In A. Bruns & J. Jacobs (Eds.), *Uses of blogs* (pp. 117-126). New York, NY: Peter Lang.
- Halford, M. (2009, June 10). Blogademia. *The New Yorker*. Retrieved from <http://www.newyorker.com/online/blogs/books/2009/06/blogademia.html>
- Hank, C. (2009). *Science and scholarship in the blogosphere: Blog characteristics, blogger behaviours and implications for digital curation*. Poster presented at the Fifth International Digital Curation Conference, London, UK. Retrieved from [http://ils.unc.edu/~hcarolyn/idcc09poster\\_hank.pdf](http://ils.unc.edu/~hcarolyn/idcc09poster_hank.pdf)
- Hank, C., Choemprayong, S., & Sheble, L. (2007). Blogger perceptions on digital preservation. In *Proceedings of the 7th ACM/IEEE Joint Conference on Digital Libraries: Building and Sustaining the Digital Environment* (p. 77). New York, NY: Association for Computing Machinery.
- Hank, C., Sheble, L., & Choemprayong, S. (2007). *Informing blog appraisal through bloggers' perspectives on selection and preservation*. Paper presented at the Conference on Appraisal in the Digital World, Rome, Italy. Retrieved from [http://ils.unc.edu/~hcarolyn/delos\\_blog\\_nov2007.pdf](http://ils.unc.edu/~hcarolyn/delos_blog_nov2007.pdf)
- Hartley, J. (2008). *Academic writing and publishing: A practical handbook*. New York, NY: Routledge.
- Harvard University. (2009). *Report of the Task Force on University Libraries*. Retrieved from [http://www.provost.harvard.edu/reports/Library\\_Task\\_Force\\_Report.pdf](http://www.provost.harvard.edu/reports/Library_Task_Force_Report.pdf)
- Harvey, R. (2005). *Preserving digital materials*. Munich, Germany: K.G. Sauer.
- Harvey, R. (2008). *So where's the black hole in our collective memory: A provocative position paper*. (DigitalPreservationEurope Position Paper). Glasgow, UK: DigitalPreservationEurope and the Humanities Advanced Technology and Information Institute. Retrieved from [http://www.digitalpreservationeurope.eu/publications/position/Ross\\_Harvey\\_black\\_hole\\_PPP.pdf](http://www.digitalpreservationeurope.eu/publications/position/Ross_Harvey_black_hole_PPP.pdf)
- Hedstrom, M. (1991). Understanding electronic incunabula: A framework for research on electronic records. *American Archivist*, 54(3), 334-354.
- Hedstrom, M. (1998). Digital preservation: A time bomb for digital libraries. *Computers and the Humanities*, 31(3). Retrieved from <http://deepblue.lib.umich.edu/handle/2027.42/42573>
- Hedstrom, M., Ross, S., Ashley, K., Christensen-Dalsgaard, B., Duff, W., Gladney, H., ... Neuhold, E. (2003). *Invest to save: Report and recommendations of the NSF-DELOS Working Group on Digital Archiving and Preservation*. Retrieved from [http://eprints.erpanet.org/94/01/NSF\\_Delos\\_WG\\_Pres\\_final.pdf](http://eprints.erpanet.org/94/01/NSF_Delos_WG_Pres_final.pdf)

- Heery, R., & Anderson, S. (2005). *Digital repositories review*. Bath, UK: UKOLN and the Arts and Humanities Data Services. Retrieved from <http://ahds.ac.uk/preservation/digital-repositories-review-2005.pdf>
- Hendricks, A. (2010). Bloggership, or is publishing a blog scholarship? A survey of academic librarians. *Library Hi Tech*, 28(3), 470-477. doi: 10.1108/07378831011076701
- Herring, S.C., Kouper, I., Paolillo, J.C., Scheidt, L.A., Tyworth, M., Welsch, P., Wright, E., & Yu, N. (2005). Conversations in the blogosphere: An analysis “from the bottom up.” In *Proceedings of the 38th Hawaii International Conference on System Science*. Los Alamitos, CA: IEEE Press.
- Herring, S.C., Scheidt, L.A., Wright, E., & Bonus, S. (2005). Weblogs as a bridging genre. *Information Technology & People*, 18(2), 142-171. doi: 10.1108/09593840510601513
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 1(3). Retrieved from <http://www.ijdc.net/index.php/ijdc/article/viewFile/69/48>
- Huang, C.Y., Shen, Y.Z., Lin, H.X., & Chang, S.S. (2007). Bloggers' motivations and behaviors: A model. *Journal of Advertising Research*, 47(4), 472-484. doi: 10.2501/S0021849907070493
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow, UK: Longman.
- Jensen, M. (2007, June 15). The new metrics of scholarly authority. *The Chronicle Review*, 53(41), B6.
- Johns, A. (1998). *The nature of the book: Print and knowledge in the making*. Chicago, IL: University of Chicago Press.
- Jones, R., Andrews, T., & MacColl, J. (2006). *The institutional repository*. Oxford, UK: Chandos Publishing.
- Jyh, P.WU.H., & Leng, T.Y. (2005). *Webblog archives: Achieving the recordness of web archiving*. Paper presented at the International Cultural Heritage Informatics Meeting, Paris, France. Retrieved from <http://www.archimuse.com/publishing/ichim05/Wu.pdf>
- Keen, A. (2007). *The cult of the amateur*. New York, NY: Doubleday.
- Kjellberg, S. (2009). Scholarly blogging practice as situated genre: An analytical framework based on genre theory. *Information Research*, 14(3). Retrieved from <http://InformationR.net/ir/14-3/paper410.html>

- Kling, R., & McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. *Journal of the American Society for Information Science*, 50(10), 890-906.
- Knorr-Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Kramer, A.D.I., & Rodden, K. (2007). Applying a user-centered metric to identify inactive blogs. In B. Begoles, S. Payne, E. Churchill, R. St. Amant, D. Gilmore, & M.B. Rosson (Eds.), *Conference on Human Factors in Computing Systems: CHI '07 extended abstracts* (pp. 2525–2530). New York, NY: Association for Computing Machinery. doi: 10.1145/1240866.1241035
- Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Cambridge, MA: Harvard University Press.
- Lampa, G. (2004). Imagining the blogosphere: An introduction to the imagined community of instant publishing. In L. J. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. St. Paul, MN: University of Minnesota. Retrieved from [http://blog.lib.umn.edu/blogosphere/imagining\\_the\\_blogosphere.html](http://blog.lib.umn.edu/blogosphere/imagining_the_blogosphere.html)
- LaPorte, R.E., Marler, E., Akazawa, S., Sauer, F., Gamboa, C., Shenton, C., ... Maclure, M. (1995). The death of biomedical journals. *BMJ*, 310(6991), 1387-1390.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
- Lavoie, B.F., & Dempsey, L. (2004). Thirteen ways of looking at digital preservation. *D-Lib Magazine*, 10(7/8). doi: 10.1045/july2004-lavoie
- Lemann, N. (2006, August 7). Amateur hour: Journalism without journalists. *The New Yorker*. Retrieved from [http://www.newyorker.com/archive/2006/08/07/060807fa\\_fact1](http://www.newyorker.com/archive/2006/08/07/060807fa_fact1)
- Lenhart, A., Fallows, D., & Horrigan, J. (2004). *Content creation online*. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://www.pewinternet.org/Reports/2004/Content-Creation-Online.aspx>
- Lenhart, A., & Fox, S. (2006). *Bloggers: A portrait of the Internet's new story tellers*. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://www.pewinternet.org/Reports/2006/Bloggers.aspx>
- Lievrouw, L.A. (1990). Reconciling structure and process in the study of scholarly communication. In C.L. Borgman (Ed.), *Scholarly communication and bibliometrics* (pp. 59-69). Newbury Park, CA: Sage Publications.



- Liu, S.H., Liao, H.L., & Zeng, Y.T. (2007). Why people blog: An expectancy theory analysis. *Issues in Information Systems*, 8(2). Retrieved from [http://jicis-online.org/iis/2007\\_iis/PDFs/Liu\\_liao\\_Zeng.pdf](http://jicis-online.org/iis/2007_iis/PDFs/Liu_liao_Zeng.pdf)
- Lynch, C.A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL* (226), 1-7. Retrieved from <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>
- Lynch, C.A. (2004). Editor's interview with Clifford A. Lynch. *RLG DigiNews*, 8(4). Retrieved from <http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070513/viewer/file1789.html#>
- Lynch, C.A. (2005). Where do we go from here: The next decade for digital libraries. *D-Lib Magazine*, 11(7/8). doi:10.1045/july2005-lynch.
- Lynch, C.A., & Lippincott, J.K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine*, 11(9). doi: 10.1045/september2005-lynch
- Markey, K., Rieh S.Y., St. Jean, B., Kim, J., & Yakel, E. (2007). *Census of institutional repositories in the United States: MIRACLE project research findings*. (CLIR Publication No. 140). Washington, DC: Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub140/pub140.pdf>
- Marlow, C. (2004). *Audience, structure and authority in the Weblog community*. Paper presented at the International Communication Association Conference, New Orleans, LA. Retrieved from <http://alumni.media.mit.edu/~cameron/cv/pubs/04-01.pdf>
- Maron, N.L., & Smith, K.K. (2008). *Current models of scholarly communication: Results of an investigation conducted by Ithaka for the Association of Research Libraries*. Washington, DC: Association of Research Libraries.
- Marshall, C.C., Bly, S., & Brun-Cottan, F. (2006). The long term fate of our digital belongings: Toward a service model for personal archives. *Archiving 2006 Final Program and Proceedings*, 3 (pp. 25-30). Springfield, VA: Society for Imaging Sciences and Technology.
- Marshall, C.C., McCown, F., & Nelson, M.L. (2007). Evaluating personal archiving strategies for Internet-based information. *Archiving 2007 Final Program and Proceedings*, 4 (pp. 151-161). Springfield, VA: Society for Imaging Sciences and Technology. Retrieved from <http://arxiv.org/ftp/arxiv/papers/0704/0704.3647.pdf>
- Mayer-Schonberger, V. (2009). *Delete: The virtue of forgetting in the digital age*. Princeton, NJ: Princeton University Press.
- McCullagh, K. (2008). Blogging: Self-presentation and privacy. *Information and Communication Technology Law*, 17(1), 3-23.

- McDonough, M., & Randag, S. (2009, December 2). Cover story: Third Annual ABA Journal Blawg 100 [Web log message]. Retrieved from [http://www.abajournal.com/magazine/article/third\\_annual\\_aba\\_journal\\_blawg\\_100](http://www.abajournal.com/magazine/article/third_annual_aba_journal_blawg_100)
- Meyer, W. (2005, October 26). Archiving Coyote Blog in print form [Web log message]. Retrieved from [http://www.coyoteblog.com/coyote\\_blog/2005/10/archiving\\_coyot.html](http://www.coyoteblog.com/coyote_blog/2005/10/archiving_coyot.html)
- Miller, C.A., & Shepherd, D. (2004). Blogging as a social action: A genre analysis of the weblog. In L. J. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the blogosphere: Rhetoric, community, and culture of weblogs*. St. Paul, MN: University of Minnesota. Retrieved from [http://blog.lib.umn.edu/blogosphere/blogging\\_as\\_social\\_action\\_a\\_genre\\_analysis\\_of\\_the\\_weblog.html](http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html)
- Miura, A., & Yamashita, K. (2004). Psychological and social influences on blog writing: An online survey of blog authors in Japan. *Journal of Computer-Mediated Communication*, 12(4). Retrieved from <http://jcmc.indiana.edu/vol12/issue4/miura.html>
- Mooney, C., & Kirshenbaum, S. (2009). *Unscientific America: How scientific illiteracy threatens our future* [Paperback edition]. New York, NY: Basic Books.
- Moore, R. (2008). Towards a theory of digital preservation. *International Journal of Digital Curation*, 1(3). Retrieved from <http://www.ijdc.net/index.php/ijdc/article/viewFile/63/42>
- Moore, S. (2003). Writers' retreats for academics. Exploring and increasing the motivation to write. *Journal of Further and Higher Education*, 27(3), 333-343. doi: 10.1080/0309877032000098734
- Mortensen, T., & Walker, J. (2002) Blogging thoughts: Personal publications as an online research tool. In A. Morrison (Ed.) *Researching ICTs in context: InterMedia report 3/2002* (pp. 249-278). Oslo, Norway: InterMedia.
- Morton, H.C., & Price, A.J. (1989). *The ACLS survey of scholars: Final report of views on publications, computers, and libraries*. Washington DC: Office of Scholarly Communication and Technology, American Council of Learned Societies.
- Murray, R., & Moore, S. (2006). *The handbook of academic writing: A fresh approach*. Maidenhead, UK: Open University Press.
- Nardi, P.M. (2003). *Doing survey research: A guide to quantitative methods*. Boston, MA: Allyn and Bacon.
- Nardi, B.A., Schiano, D.J., & Gumbrecht, M. (2004). Blogging as a social activity, or, would you let 900 million people read your diary? *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, 6(3) (pp. 222-231). New

- York, NY: Association for Computing Machinery. Retrieved from <http://home.comcast.net/%7Ediane.schiano/CSCW04.Blog.pdf>
- Nardi, B.A., Schiano, D.J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, 47(12), 41-46.
- National Enquiry into Scholarly Communication. (1979). *Scholarly communication: The report of the National Enquiry*. Baltimore, MD: Johns Hopkins University Press.
- National Science Foundation & Library of Congress. (2003). It's about time: Research challenges in digital archiving and long-term preservation. Retrieved from [http://www.digitalpreservation.gov/library/resources/pubs/docs/about\\_time2003.pdf](http://www.digitalpreservation.gov/library/resources/pubs/docs/about_time2003.pdf)
- Neuman, W.L. (2006). *Social research methods: Qualitative and quantitative approaches* (6th Ed.). Boston, MA: Pearson Education.
- Odlyzko, A. (2002). The rapid evolution of scholarly publication. *Learned Publishing*, 15(1), 7-19.
- Oi, I. (2006). Blogging the legal commons. In A. Bruns & J. Jacobs (Eds.), *Uses of blogs* (pp. 81-88). New York, NY: Peter Lang.
- Okerson, A. (1992). The missing model: A 'circle of gifts.' *Serials Review*, 18(1/2), 92-96.
- Olsen, D., Berlin, E., Olsen, E., McLean, J., & Sussman, M. (2009). *State of the blogosphere: 2009*. Technorati. Retrieved from <http://technorati.com/blogging/feature/state-of-the-blogosphere-2009/>
- O'Sullivan, C. (2005). Diaries, on-line diaries, and the future loss to archives: Or, blogs and the blogging bloggers who blog them. *American Archivist*, 68(1), 53-73.
- O'Toole, J.M. (1989). On the idea of permanence. *American Archivist*, 52(Winter 1989), 10-25.
- Parks, R.P. (2002). The Faustian grip of academic publishing. *Journal of Economic Methodology*, 9(3), 317-335.
- Paulus, M.J. (2006). Blogging for the record: A study of blogging from an archival point of view. *Journal of Archival Organization*, 4(3/4), 31-41.
- Payette, S., & Lagoze, C. (1998). Flexible and extensible digital object and repository architecture (FEDORA). In C. Nikolaou & C. Stephanidis (Eds.), *Lecture Notes in Computer Science: Research and Advanced Technology for Digital Libraries*, 1513 (pp. 41-59). Berlin, Germany: Springer-Verlag. doi: 10.1007/3-540-49653-X\_4
- Pearce-Moses, R. (2005). *A glossary of archival and records terminology*. Chicago, IL: Society of American Archivists. Retrieved from [http://www.archivists.org/glossary/term\\_details.asp?DefinitionKey=1681](http://www.archivists.org/glossary/term_details.asp?DefinitionKey=1681)

- Pennock, M. (2007). Digital curation: A life-cycle approach to managing and preserving usable digital information. *Library & Archives Journal*. Advance online publication. Retrieved from [http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch\\_curation.pdf](http://www.ukoln.ac.uk/ukoln/staff/m.pennock/publications/docs/lib-arch_curation.pdf)
- Peters, T.A. (2002). Digital repositories: Individual, discipline-based, institutional, consortial or national? *Journal of Academic Librarianship*, 28(6), 414-417.
- Peterson, R.A. (2000). *Constructing effective questionnaires*. Thousand Oaks, CA: Sage Publications.
- Podgor, E.S. (2006). Blogs and the promotion and tenure letter. *Washington University Law Review*, 84(5), 1109-1111.
- Powell, R.R., & Connaway, L.S. (2004). *Basic research methods for librarians* (4th ed.). Westport, CT: Libraries Unlimited.
- Price, D.J.D. (1986). *Little science, big science ... and beyond*. New York: Columbia University Press.
- Primary Research Group. (2009). *The survey of higher education faculty: Evaluation of library efforts to index, preserve and catalog blogs, websites, email, archives, and other cyber resources*. New York, NY: Author.
- Punch, K.F. (2003). *Survey research: The basics*. London: Sage Publications.
- Qian, H., & Scott, C. R. (2007). Anonymity and self-disclosure on weblogs. *Journal of Computer-Mediated Communication*, 12(4). Retrieved from <http://jcmc.indiana.edu/vol12/issue4/qian.html>
- Quiggin, J. (2006). Economic blogs and blog economics. In A. Bruns & J. Jacobs (Eds.), *Uses of blogs* (pp. 69-79). New York, NY: Peter Lang.
- Rainie, L. (2005). *The state of blogging*. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://www.pewinternet.org/Reports/2005/The-State-of-Blogging.aspx>
- Raymond, M. (2010, April 14). How Tweet it is!: Library acquires entire Twitter archive [Web log message]. Retrieved from <http://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>
- Rea, L.M., & Parker, R.A. (1997). *Designing and conducting survey research: A comprehensive guide* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Rice, J. (2009). Networked exchanges, identity, writing. *Journal of Business and Technical Communication*. Advanced online publication. doi: 10.1177/1050651909333178.

- Roloff, M E. (1981). *Interpersonal communication: The social exchange approach*. Beverly Hills, CA: Sage Publications.
- Roosendaal, H.E., Guerts, P.A. Th.M., & van der Vet, P.E. (2001). Developments in scientific communication: Considerations on the value chain. *Information Services & Use*, 21(1), 13-32.
- Rosenberg, S. (2009). *Say everything: How blogging began, what it's becoming and why it matters*. New York, NY: Three Rivers Press.
- Ross, S. (2000). *Changing trains at Wigan: Digital preservation and the future of scholarship*. (NSO Preservation Guidance Occasional Paper). London, UK: National Preservation Office, British Library. Retrieved from <http://www.bl.uk/npo/pdf/wigan.pdf>
- Rothenberg, J. (1995). Ensuring the longevity of digital documents. *Scientific American*, 272(1), 42-47.
- Saper, C. (2006). Blogademia. *Reconstruction*, 6(4). Retrieved from <http://reconstruction.eserver.org/064/contents.shtml>
- Schiffer, A. J. (2007, August). *Between pajamas and Pulitzers: Distributed gatekeeping and the potential of blogs as news media*. Paper presented at the Annual Meeting of the American Political Science Association, Chicago, IL. Retrieved from [http://www.allacademic.com/meta/p208737\\_index.html](http://www.allacademic.com/meta/p208737_index.html)
- Schmidt, J. (2007). Blogging practices: An analytical framework. *Journal of Computer-Mediated Communication*, 12(4). Retrieved from <http://jcmc.indiana.edu/vol12/issue4/schmidt.html>
- Serfaty, V. (2005). *The mirror and the veil: An overview of American online diaries and blogs*. Amsterdam and New-York: Rodopi.
- Shearer, K. (2006). The CARL institutional repositories project: A collaborative approach to addressing the challenges of IRs in Canada. *Library Hi Tech*, 24(2), 165-172.
- Sheble, L., Choemprayong, S., & Hank. C. (2007). *Preservation in context: Survey of blogging behaviors*. Paper presented at the Third International Digital Curation Conference, Washington DC.
- Sifry, D. (2004, October 10). State of the blogosphere, October 2004 [Web log message]. Retrieved from <http://www.sifry.com/alerts/archives/000245.html>
- Sifry, D. (2007, April 2). State of the blogosphere/State of the live Web [Web log message]. Retrieved from <http://www.sifry.com/stateoftheliveweb/>
- Smith, G.D. (2006). A case study in bloggership. *Washington University Law Review*, 84(5), 1135-1144.

- Solomon, D.J. (2002). Talking past each other: Making sense of the debate over electronic publication. *First Monday*, 7(8). Retrieved from [http://www.firstmonday.org/issues/issue7\\_8/solomon/index.html](http://www.firstmonday.org/issues/issue7_8/solomon/index.html)
- Solum, L. (2006a). Blogging and the transformation of legal scholarship. *Washington University Law Review*, 84(5), 1071-1088.
- Solum, L. (2006b). Download it while its hot: Open access and legal scholarship. *Lewis and Clark Law Review*, 10(4), 841-867.
- Sue, V.M., & Ritter, L.A. (2007). *Conducting online surveys*. Thousand Oaks, CA: Sage Publications.
- Thomas, C., & McDonald, R.H. (2007). Measuring and comparing participation patterns in digital repositories: Repositories by the numbers, part 1. *D-Lib Magazine*, 13(9/10). doi: 10.1045/september2007-mcdonald
- Tibbo, H.R. (2003). On the nature and importance of archiving in the digital age. In M. Zelkowitz (Ed.), *Advances in Computers: Information Repositories*, 57 (pp. 1-67). San Diego, CA: Elsevier.
- Treloar, A., & Groenewegen, D. (2007). ARROW, DART and ARCHER: A quiver full of research repository and related projects. *Ariadne*, (51). Retrieved from <http://www.ariadne.ac.uk/issue51/treloar-groenewegen/>
- Van de Sompel, H. (2000, December). *Closing keynote address*. Paper presented at the Coalition for Networked Information Fall Meeting, San Antonio, TX. Retrieved from <http://www.cni.org/tfms/2000b.fall/handout/HVDS-CNI-2000Ftf.pdf>.
- Van de Sompel, H., Payette, S., Erickson, J., Logoze, C., & Warner, S. (2004). Rethinking scholarly communication: Building the system that scholars deserve. *D-Lib Magazine*, 10(9). doi: 10.1045/september2004-vandesompel
- Viegas, F.B. (2007). Bloggers' expectations of privacy and accountability: An initial survey. *Journal of Computer-Mediated Communication*, 10(3). Retrieved from <http://jcmc.indiana.edu/vol10/issue3/viegas.html>
- Walker, J. (2006). Blogging from inside the ivory tower. In A. Bruns & J. Jacobs (Eds.), *Uses of blogs* (pp. 127-138). New York, NY: Peter Lang.
- Ware, M. (2004). *Pathfinder research on web-based repositories*. Bristol, UK: Publisher and Library/Learning Solutions.
- Wegner, K. (2005, August 2). Topical versus generalist blogging [Web log message]. Retrieved from [http://prawfsblawg.blogs.com/prawfsblawg/2005/08/ontopic\\_bloggin.html](http://prawfsblawg.blogs.com/prawfsblawg/2005/08/ontopic_bloggin.html)

- White, D., & Winn, P. (2009). *State of the blogosphere: 2008*. Technorati. Retrieved from <http://technorati.com/blogging/feature/state-of-the-blogosphere-2008/>
- Wildemuth, B.M. (Ed.). (2009). *Applications of social research methods to questions in information and library science*. Westport, CT: Libraries Unlimited.
- Willinsky, J. (2000). Proposing a knowledge exchange model for scholarly publishing. *Current Issues in Education*, 3(6). Retrieved from <http://cie.asu.edu/volume3/number6/>
- Working Group on Defining Digital Preservation. (2007). *Definitions of digital preservation*. Chicago, IL: Preservation and Reformatting Section, Association for Library Collections and Technical Services, American Library Association. Retrieved from <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf>
- Worthen, S. (2009). Medievalist weblogs (sorted alphabetically) [Web page]. Retrieved from <http://fishpond.owlfish.com/medievallogs.html>