INFORMATION SEEKING BEHAVIOR OF GEOLOGISTS WHEN SEARCHING FOR PHYSICAL SAMPLES

Sarah Ramdeen

A dissertation submitted to the faculty at 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
2017

Approved by:
Claudia Gollop
Betty Adrian
Jane Greenberg
Barbara Moran
Barbara Wildemuth
ABSTRACT

Sarah Ramdeen: Information Seeking Behavior of Geologists when Searching for Physical Samples
(Under the Direction of Claudia Gollop)

Information seeking is “a conscious effort to acquire information in response to a need or gap” in your knowledge (Case, 2007, p. 5). In the geosciences, physical samples such as cores, cuttings, fossils, and rocks are primary sources of information; they represent “the foundation of basic and applied geoscience research and education, and underpin industry programs to discover and develop domestic natural resources” (National Research Council, 2002, p. 8). This dissertation investigates the information seeking behavior of geologists when searching for physical samples. It takes a unique approach by looking at physical objects as information sources, as opposed to past studies which focused on print literature (Bichteler & Ward, 1989; Joseph, 2001).

Data collection was twofold. First I administered two questionnaires to state geological surveys. Thirty-five state geologists and 28 repository managers responded. The results capture an overview of these science data centers and their handling of collections of physical samples. State geological surveys were selected as they are a distinct type of facility which are similar to libraries. Like libraries, these institutions’ missions dictate maintaining a collection as well as providing access to their diverse data holdings.

In the second stage, I interviewed 15 geologists, primarily users of state geological survey collections. Responses highlighted various search behaviors which were used to develop
a model of their information seeking behavior. Some behaviors were dependent on one’s role within an organization, suggesting a division of labor in the research team. Many behaviors related to a researcher’s knowledge of the domain, e.g., knowing where to look, who to talk to, and how to determine the quality of the information found. The most frequently used search process by interview participants was relying on their social network to recommend or locate samples.

The results of this study suggest a number of recommendations and research opportunities for science data centers, including: 1) developing infrastructure which supports discovery and access, 2) further exploring the nature of task and role in searching, 3) developing training for searchers and curators, 4) developing standards for metadata creation related to physical samples, and 5) developing tools to aid in the search process.
This work is dedicated to my parents.
ACKNOWLEDGEMENTS

There have been so many people to support me though this work, both personally, and professionally. First, I would like to thank my family. Mom – you have always supported my adventurous ideas, a PhD included. I learned to think about things differently from you, and that has been vital in being a researcher. Dad – You might tease me about being in school for the rest of my life, but you more than anyone understood and respected this process. Thank you, Carrie, Zeb, Kim for everything. I have such a big family, and they have always been there for me. In no particular order, Grandma, Grandpa, Barbara, Ellen, Kathy, Kurt, Kenny, Richie, Taylor, Robin, and Cathy.

I am eternally grateful and very appreciative of my committee: Claudia Gollop, Betty Adrian, Jane Greenberg, Barbara Moran, and Barbara Wildemuth. Thank you for your time, your guidance, and your encouragement. This has been a long process and I truly appreciate your continual support.

Thank you to everyone in the Earth Science community. Thank you to the folks at the Florida Geological Survey. Jon Arthur gave me my first job as a geologist, Ken Campbell supported my efforts to pull in LIS ideas into my work, and Jackie Lloyd, you had such faith in me and my abilities and you have been a wonderful example and mentor.

Thank you to everyone in the Earth Science Information Partners, the Earth and Space Science Informatics Focus Group of the American Geophysical Union, the Research Data Alliance, and EarthCube. Thank you for allowing me to share my ideas, and for the research and
funding opportunities. I have been privileged to work with so many amazing women in the geosciences. In particular Ruth Duerr and Denise Hills. I don’t know what I would do without you two, thanks for keeping me sane!

Thank you to the alumni of the Archival Education and Research Initiative. AERI provides a safe space for students to explore their research ideas and really develop a community. Devan Donaldson and Edward Benoit, you two are amazing.

Thank you to all of my friends. The lines blur sometimes between my research communities and my friends and that is a good thing. But that means I wasn’t sure where to list some of you! Thank you to Elizabeth Parker, Diana Greenslade, and Jessie Nelson. You three are more than my best friends, you are sisters. Thank you to all of my classmates over the years at UNC: Amber, Rachael, Julia, Laura, Emily, Alex, Amanda, John, Sumayya, Angela, Tessa, Kjersti, Jewel, and Simon. I needed all of you. Brenda, Jeff, Josiah, and Eric you should be honorary members of that list. Thank you to the Solorooers, my family of hippies. Ross, Kathy, Shea, Lucas, Ryan, Glenn, Dan, Tom, and so many more. You all have known me since before I even thought about getting a PhD. You kept me grounded and calm. Thank you to all my friends from various parts of my life: Jenn, Kim, Candice, Gaines, Brent, Elise, Tyler, Eric, Melody, Daniel, Rachael, Nancy, and Ona. Thank you to my friend Kenji Butler for letting me ramble. And thank you to Michael Hysell for the encouragement and support.

Last but not at all least, thank you to all of the teachers, librarians, and mentors I have had over my many, many years of education. Steven Valdespino used to tease me about being a doctor. Looks like you were right. Bill Reaney, you shaped my life in so many ways. I learned to be part of a community, to work hard, and to always be caring and inclusive. Thank you to Bill Parker, you were my advisor as an undergrad and now my friend.
PREFACE

The introduction chapter of this dissertation was previously published in a special issue of GeoResJ. The theme of the special issue is rescuing legacy data for the future. This is particularly relevant for physical samples, which are often overlooked when we consider data rescue.
# TABLE OF CONTENTS

LIST OF TABLES...................................................................................................................... xvi

LIST OF FIGURES ...................................................................................................................... xvii

LIST OF ABBREVIATIONS........................................................................................................ xviii

CHAPTER 1: Introduction .......................................................................................................... 1

Geological collections at state geological surveys.................................................................. 3

Geological data.......................................................................................................................... 3

Evolution from museums to libraries......................................................................................... 8

Users of geological data............................................................................................................ 9

Geological surveys..................................................................................................................... 11

Stewardship challenges............................................................................................................. 13

Historical view on preservation................................................................................................. 13

Current stewardship efforts...................................................................................................... 13

Needs related to stewardship...................................................................................................... 17
CHAPTER 2: Literature review

Information in the earth sciences

Physical objects as information sources in the earth sciences.

Geological data.

Representation of knowledge.

Information organization.

Why do we keep scientific data collections?

Scientific collections of information.

Information behavior.

Ellis’s model of information seeking.

Criticism and limitations of Ellis’s model.

CHAPTER 3: Methods

Research question.

Phase one: Questionnaires.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Descriptive survey’ – understanding existing conditions</td>
<td>52</td>
</tr>
<tr>
<td>Questionnaire distribution</td>
<td>55</td>
</tr>
<tr>
<td>Data analysis</td>
<td>57</td>
</tr>
<tr>
<td>Phase two: Interviews</td>
<td>58</td>
</tr>
<tr>
<td>Question development</td>
<td>58</td>
</tr>
<tr>
<td>Sample</td>
<td>60</td>
</tr>
<tr>
<td>Interview procedures</td>
<td>61</td>
</tr>
<tr>
<td>Data analysis</td>
<td>62</td>
</tr>
<tr>
<td>Trustworthiness of the findings</td>
<td>66</td>
</tr>
<tr>
<td>CHAPTER 4: Results</td>
<td>69</td>
</tr>
<tr>
<td>Phase one: Questionnaires</td>
<td>69</td>
</tr>
<tr>
<td>Results</td>
<td>69</td>
</tr>
<tr>
<td>Summary</td>
<td>83</td>
</tr>
<tr>
<td>Phase two: Interviews</td>
<td>83</td>
</tr>
<tr>
<td>Participants</td>
<td>84</td>
</tr>
</tbody>
</table>
Results................................................................................................................................. 86

Summary ........................................................................................................................................ 104

Comparisons across research stages .................................................................................. 104

Use of collections .................................................................................................................. 105

Regions represented ............................................................................................................. 106

Analysis of other materials .................................................................................................. 106

Databases ................................................................................................................................. 107

State survey Data at risk evaluation ....................................................................................... 108

Community discussions ........................................................................................................ 109

Services ....................................................................................................................................... 110

CHAPTER 5: Discussion ............................................................................................................. 111

Information seeking ................................................................................................................. 111

Model ......................................................................................................................................... 114

Seeking ....................................................................................................................................... 117

Social networks and personal contact .................................................................................... 119

Secondary data .......................................................................................................................... 120
Resources by region................................................................. 121

Education in searching.......................................................... 121

Connection to science data centers...................................... 122

Nature of the data.................................................................. 126

Limitations of this study...................................................... 127

CHAPTER 6: Conclusion.......................................................... 129

Future research.................................................................... 131

APPENDIX A: GLOSSARY.......................................................... 133

APPENDIX B: PROTOCOLS FOR ARRANGING A VISIT TO STATE COLLECTIONS... 136

Alaska. ................................................................................. 136

Illinois. .................................................................................. 136

Indiana................................................................................. 136

APPENDIX C: INITIAL EMAIL AND QUESTION REQUESTS................................. 138

Email to state geologists...................................................... 138

Email to repository managers/contacts listed in the first questionnaire. .......... 139
APPENDIX F: RECRUITMENT SCRIPT ................................................................. 156

APPENDIX G: LIST SERVE RECRUITMENT ...................................................... 158

APPENDIX H: INTERVIEW GUIDE .................................................................... 159
  Obtain consent ................................................................................................. 159
  Interview guide ............................................................................................... 159

APPENDIX I: INFORMATION AND CONSENT ............................................... 162

APPENDIX J: EXAMPLE OF SEARCH BEHAVIOR ......................................... 165
  Example one .................................................................................................... 165
  Example two ................................................................................................... 166

REFERENCES .................................................................................................... 167
LIST OF TABLES

Table 1. Examples of materials in Geoscience Collections ........................................29

Table 2. Three types of knowledge ........................................................................32

Table 3. Categories for Ellis’s Models ..................................................................44

Table 4. Requirements, policies and plans related to collections...............................73

Table 5. Facility needs ..........................................................................................74

Table 6. Discovery and access (N=32) ..................................................................76

Table 7. Physical samples at state surveys .............................................................77

Table 8. Primary geographic region .....................................................................77

Table 9. How does your organization provide outside access to this collection? (Select all that apply) .................................................................81

Table 10. Description of participants ....................................................................85

Table 11. Definition of behavior categories ..........................................................108
LIST OF FIGURES

Figure 1. Examples of materials in a geological collection ........................................4

Figure 2. Examples of materials in deterioration ..............................................................7

Figure 3. Sample-subsample relationship .......................................................................30

Figure 4. Subtypes of science data collections .................................................................39

Figure 5. Information Behavior ......................................................................................41

Figure 6. Wilson (1999a)’s proposed diagram of Ellis’s model ........................................45

Figure 7. Utah Core Research Center Usage ....................................................................57

Figure 8. GSA Regions .................................................................................................71

Figure 9. Percentage of their collection that is documented ...........................................79

Figure 10. Percentage of the collection that is accessible digitally .................................79

Figure 11. Primary Use ..................................................................................................82

Figure 12. Yearly visitors ...............................................................................................83

Figure 13. Information seeking behavior model ............................................................115
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AASG</td>
<td>Association of American State Geologists</td>
</tr>
<tr>
<td>AEC</td>
<td>Atomic Energy Commission</td>
</tr>
<tr>
<td>AGI</td>
<td>American Geosciences Institute</td>
</tr>
<tr>
<td>CDA</td>
<td>Common Data Access Limited</td>
</tr>
<tr>
<td>CODATA</td>
<td>Committee on Data for Science and Technology</td>
</tr>
<tr>
<td>CSDCO</td>
<td>Continental Scientific Drilling Coordination Office</td>
</tr>
<tr>
<td>DataONE</td>
<td>Data Observation Network for Earth</td>
</tr>
<tr>
<td>ESIP</td>
<td>Earth Science Information Partners</td>
</tr>
<tr>
<td>GCMS</td>
<td>Geologic Collections Management System</td>
</tr>
<tr>
<td>GMC</td>
<td>Alaska Geologic Materials Center</td>
</tr>
<tr>
<td>GMRWG</td>
<td>Geologic Materials Repository Working Group</td>
</tr>
<tr>
<td>GSA</td>
<td>Geological Society of America</td>
</tr>
<tr>
<td>IGSN</td>
<td>International Geo Sample Numbers</td>
</tr>
<tr>
<td>IODP</td>
<td>Integrated Ocean Drilling Program</td>
</tr>
<tr>
<td>LacCore</td>
<td>National Lacustrine Core Facility</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>LIS</td>
<td>Library and Information Science</td>
</tr>
<tr>
<td>MGRRE</td>
<td>Michigan Geological Repository for Research and Education</td>
</tr>
<tr>
<td>NGGDPP</td>
<td>National Geological and Geophysical Data Preservation Program</td>
</tr>
<tr>
<td>NGDS</td>
<td>National Geothermal Data System</td>
</tr>
<tr>
<td>NRC</td>
<td>National Research Council</td>
</tr>
<tr>
<td>NSTC</td>
<td>National Science and Technology Council</td>
</tr>
<tr>
<td>OAIS</td>
<td>Open Archival Information System</td>
</tr>
<tr>
<td>RDA</td>
<td>Research Data Alliance</td>
</tr>
<tr>
<td>SDC</td>
<td>Scientific Data Collections</td>
</tr>
<tr>
<td>SESAR</td>
<td>System for Earth Sample Registration</td>
</tr>
<tr>
<td>USGS</td>
<td>United States Geological Survey</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

The first state geological survey was founded in 1823, and by 1840 there were at least 15 states with geological surveys (Association of American State Geologists (AASG), 2007). The function of a state geological survey varies from state to state, but the underlying purposes of the state surveys are consistent: to collect and maintain information about their state’s geology and to share these resources with the public. “In all states, a major purpose was to locate, describe, and publicize such natural resources as salt and mineral springs, building stones, shales, clays, slates, coal, and ores” (Hendrickson, 1961, p. 361). At the time of their founding in the mid 1800’s, state geological surveys represented “government-supported science” and were often “the first contact between the public and science” (Buchanan, 1992, p. 62).

“The early state surveys were intended by the legislatures to be only short-term undertakings” (Boscoe, 2003, p. 293). Initially, geologists (usually academics) were hired to conduct specific tasks, but as states evolved and the industry changed, roles changed as well (Buchanan, 1992). States’ collections often began as personal, individual collections, which later became the base of the state geological repositories. Given that they are personal, they contain qualities that make them unusual and perhaps more difficult to manage, similar to small science collections.

In 2008, Western Michigan University's Michigan Geological Repository for Research and Education (MGRRE) and the state’s geological repository, acquired a large amount of rock

---

1A version of this chapter previously appeared as an article in GeoResJ. The original citation is as follows: Ramdeen, S. (2015). Preservation challenges for geological data at state geological surveys. GeoResJ, 6, 213-220.
core from the Mosaic company. The company owned a potash mine from which the cores were drilled. They no longer wanted to store them and offered the cores as a donation to the university (Zipp, 2013). Two administrators at the MGRRE, realizing the research value of these materials, drove their own vehicle to pick up the boxes of cores. “It took four pick-up loads to bring all the material down to Kalamazoo” (Zipp, 2013). These samples were later used to verify the quality of amount of potash (a mineral used in fertilizers) in a rediscovered mineral deposit in West Michigan (Zipp, 2013). This discovery is estimated to be valued at $65 billion dollars and has the potential to have a major impact on the local economy. It might lower the costs of farming in the Midwest where farmers must pay to import potash from Mexico, Canada, and Russia. A new mine could create construction jobs as well as full time jobs at the site (Zipp, 2013). These ‘unwanted’ samples have become a major resource for the state of Michigan, and it was fortuitous that MGRRE saw the value in them as legacy data and had the opportunity (and resources) to preserve them.

On January 17th, 2001, a natural gas explosion occurred in downtown Hutchinson, Kansas. Two local businesses burned down as a result. Two days later, another leak occurred under a mobile home, and two people were killed. As a safety precaution, the city was evacuated. Ultimately, residents were not able to return until March (National Research Council (NRC), 2002). During the intervening months, KGas, the local gas company, collaborated with the Kansas Geological Survey to investigate the leaks. “Everyone involved in the crisis came to quickly value the geologic data and samples the Kansas Geological Survey had collected and archived for decades” (Allison, 2001, p. 14). Among the materials used were a collection of cores drilled in the 1960s by the U.S. Atomic Energy Commission (AEC). The AEC was investigating the possibility of nuclear storage in Kansas (Cutler & Maples, 2002). The Kansas
Geological Survey had maintained these legacy data as part of their repository. This reuse, use beyond their original purpose, helped the investigators better understand how the natural gas was leaking from a nearby underground storage facility (Cutler & Maples, 2002, p.16). As the NRC (2002) summarized, “having immediate access to critical geoscience data and information played a crucial role in facilitating rapid response to a local crisis” (NRC, 2002, p. 1).

The examples above demonstrate the importance of geological collections, their continued maintenance, and their potential for reuse. In the U.S. Geological Survey Geologic Collections Management System (GCMS), reuse is referred to as re-sampling or secondary sampling and is “the act of retrieving an archived sample for the purpose of additional scientific inquiry” (Geologic Materials Repository Working Group (GMRWG), 2015, p. 33). This study investigates information seeking behavior in relation to the reuse of geologic samples. In particular, objects which have been gathered with the intention of representing scientific geological information. Participants were recruited from the patrons of the core and cutting repositories within state geological surveys. Results of this qualitative inquiry were used to develop a model.

**Geological collections at state geological surveys**

*Geological data.*

In some subdomains of geology, physical samples or specimens are key to research. Scientists gather data from these items, analyze these data and produce scientific outcomes. These physical objects become data once they have been used in research, along with their associated metadata and descriptions. This metadata and documentation is also used to enable discovery and access for reuse as well as to capture geological information. There is a transition from a rock being just a rock, to it now representing scientific knowledge with this connection to
the documentation. If this connection is lost, the value as data becomes lost and the physical item just becomes a rock again.

Such physical geological data include items such as rock core and cuttings, thin sections and fossils, as illustrated in Figure 1. Most physical geological materials, when properly maintained, can be stored for future access without the risk of major sample degradation. For example, a properly stored and curated core sample from a well drilled in 1907 can produce new knowledge today and in the future. The data that these samples hold can be reused, reanalyzed, potentially using previously unavailable technology, and contribute to studies beyond the scope of the original project. The materials in these collections may be 1) examples of earlier observations or results, 2) standards, kept for the base of future comparisons, 3) resources for research into geological issues, 4) collections of rare or valuable items, 5) resources used for education and training future geologists, and 6) proactively collected materials for future use (National Science and Technology Council (NSTC), 2009).

Figure 1. Examples of materials in a geological collection

Cores
Cuttings
Thin Sections
Fossils

Note Author’s own photographs (2014).
In a recent White House memorandum, Holdren (2014) states “scientific collections provide an essential base for developing scientific evidence and are an important resource for scientific research, education, and resource management. Scientific collections represent records of our past and investment in our future”. It is important to maintain collections of scientific data not just for new research but to confirm previous work. As geologist and historian Jackson (1999) explains, “a fundamental tenet in science is the need for viable checking and reproducibility of results. Re-analyses may not be undertaken for some time after the original research, but require preservation of the original material worked on in order to be of any value” (p. 423). Raw data, which may include physical samples, may be used to conduct reliability and validity checks on the work being produced. Heidorn (2008) stresses the idea that science is based on theories and theories are created based on replicable data. If the data are inaccessible and the theory cannot be replicated, scientific results would be unsubstantiated. “The availability of the data behind experiments helps to insure scientific integrity by keeping the process open to external evaluation” (Heidorn, 2008, p. 286).

There are many ways to categorize data, some of which may not be mutually exclusive, e.g. big, small, dark, legacy, etc. Legacy data are part of what Heidorn (2008) termed the ‘long tail of science data’. Heidorn (2008) suggests the long tail of science data represents smaller individual collections, which never get inventoried and live in drawers or closets. These may also be categorized as dark data collections. Wallis, Rolando and Borgman (2013) suggest that these types of collections are similar to those covered by the term small science. Small science includes specialized datasets collected by individual and small teams of scientists rather than large groups. These larger groups collect “big data from big science [which] are intended for sharing among big teams” (Wallis, Rolando, & Borgman, 2013, p. 3).
The examples in the introduction demonstrate the value of geological collections. However, long term management and storage has not always been factored in to the data collection process. Differences in management might depend on the intended use, the focus of metadata, and other institutional variances. This may lead to valuable collections being abandoned or left deteriorating (see Figure 2) at the end of a project. This is not due to neglect, or lack of care, but a lack of resources and focus. State geological surveys face a variety of preservation challenges in relation to their geological data collections. Many facilities would like to have full maintenance for their samples, however these organizations do not have the proper resources to do so or lack a standard procedure for curation (NRC, 2002); resources include staffing, funding, and space.

The NRC (2002) provides a number of examples of potential loss of geological data collections. For example, in 2002, cores collected by the Tennessee Valley Authority and the Department of Energy were being stored outside, in the elements. The cores are from such important locations as the Clinch River Breeder Reactor site and the Oak Ridge Reservation. Exposure to air and humidity can cause boxes to decay, hand written labels to be lost, and for minerals to decay (see Figure 2, bottom right for an example of pyrite to oxidizing). When minerals decay, they no longer represent what the rocks and minerals represented in situ. When metadata on boxes becomes unreadable, or when samples change, their scientific value may be lost or diminished.
Long tail data are important as they are “a breeding ground for new ideas and never before attempted science” (Heidorn, 2008, p. 282). When they are inaccessible, these sets of data may be lost to the public beyond the finished publication. In his 2014 testimony, Gooding explains that state geological surveys get many of the items in their collections from donations (H.R. 5066, 2014b). These donations come from a wide range of individuals including scientists from “coal, oil and gas, mining, highway construction, and environmental investigations; construction projects; quarry operators; university research; and federal and state projects” (H.R. 5066, 2014b, p.3). Each has their own method of documentation, data collection, and curation. This can lead to complicated hybrid collections at the state survey level that, owing to their complicated curation schemes and lack of standardization, and may become lost.

Concerns for physical items also includes concern for their digital surrogates. In order to discover and access these geological collections, adequate metadata, records and other text based
materials are needed. These may be found in paper records, but are increasingly being digitized or digital born. Without this documentation, various aspects of scientific information contained in physical geological materials may be lost.

**Evolution from museums to libraries.**
The origins of museums and geological collections are closely linked. Geology uses analytical or comparative ways of ‘knowing’; research in geology involves deconstructing strata “into elements, in order to make classifications, or to better understand (and regulate) technical processes” (Pickstone, 1994, p. 113). Pickstone (1994) calls it a ‘museological science’ because “geology and mineralogy [are] also, in part, sciences of collections” (p. 117). When geology was still developing as a scientific field in the 1800’s, the role of the curator became very important in managing geological collections (Knell, 2000; Taylor, 1988). These managers were often expert geologists (Knell, 2000). Museums recognized the need to have someone manage and curate a collection to not only prevent it from falling into disrepair but also because of the value added to the collection through arrangement, identification, and cataloging of the rocks, fossils and other materials in the collection. For example, ‘amateur’ scientist William Smith’s skills were much sought after as a curator in order to get the full value from geological collections (Knell, 2000). Smith is now known for pioneering the field of stratigraphy. Museums keep collections for a number of reasons, including making them available for research, responding to public requests, educational activities, publishing scholarly works, and general interest publishing (Orna & Pettitt, 1980).

During their development, state geological surveys held similar principles as museums. Merrill notes in his 1920 history that 22 states, in describing or authorizing their state surveys, also include requirements for a museum, collection or other cabinet to store the materials gathered during research. This included records, documents and maps as well as collections.
Many states specifically included in their legislation the requirement that the museum or collection should be curated by a geologist (Hendrickson, 1961). In Kentucky, this was tasked to the inspector of mines, who “in addition to his duties as such inspector, shall be curator of the cabinet and other property of the geological survey or department” (Merrill, 1920, p.123).

While many states still maintain geological collections, the terminology has changed as state geological museums have evolved into active, living collections. These changes developed as the state surveys moved from projects led by a single person or small team, to permanent institutions which had distinctly different needs and interacted with a larger number of stakeholders. These once-museums are now called repositories\(^2\) or sample libraries. Specifically, state surveys have what the United States Geological Survey’s (USGS) GMRWG call active repositories – “a permanent facility that assumes responsibility for the long-term storage and maintenance of a collection (or collections) of related materials” (2015, p.11). In line with this move away from museums, in 1975 the AASG formed a ‘core and sample library’ committee, which still exists today (Cobb, 2008). As of 2015, of the twenty states which reference their collections in the annual State Geologist Journal. Twelve states refer to their collections as core or sample libraries, four states refer to these collections as repositories, and only one state refers to their collection as a museum (AASG, 2015). The term sample library provokes the concept Cragin and Shankar (2006) call ‘reference collections’ which are large-scale, and typically serve science and education, among others.

**Users of geological data.**
According to Gooding, manager of the Kentucky Geological Survey’s well sample and core library, state geological data collections “are used by scientists from the U.S. government,

\(^2\)“Historically, the term ‘repository’ is indicative of a distinct physical location where samples or documents are housed and curated” (GMRWG, 2013, p.7).
geological surveys, educators from academia, exploration, development and industry geologists, consultants, operators, students and the general public” (H.R. 5066, 2014b, p.1). Hawaii’s state survey reports “requests for information, technical data, and rock samples are handled on a routine basis” (Socolow, 1988, p.90). The makeup of the current user community and how they access the collections is unclear which supports the need for this research.

As part of the first granting phase of the National Geological and Geophysical Data Preservation Program (NGGDPP), the USGS sent out a questionnaire to the state geological surveys. The questionnaire included questions about frequency of collection use, demographic information about users, and other metrics related to the state surveys’ sample repositories. This is invaluable information, which was never published. There are currently no consistent reporting methods or published findings related to this type of information. Even the NRC’s 2002 report is missing this information. The exact variation in the size and scope of these collections is unclear. Troutman (2009) curates a list of these types of collections on his website, but it is incomplete and was last updated in 2009. The annual State Geologists Journal includes some of this information as volunteered by the state surveys but it is inconsistent. Some states report linear feet of core reviewed per year but not the number of sample sets reviewed or the number of visitors. This format of describing usage is also incomplete. A useful analogy to this situation is the documentation of the number of pages checked out from the library without including information on the number of patrons, the number of books, or perhaps the length of each book. The first phase of this research collects information on the variations and use of these collections.

“Information Seeking Behavior is the purposive seeking for information as a consequence of a need to satisfy some goal” (Wilson, 2000, p.49). It is used to describe how a
person searches for information about knowledge, which includes scientific data. Understanding the user community is a challenge with scientific data. Parsons and Duerr (2005) suggest data providers may have many assumptions about the users of their data. It may be difficult to precisely identify the user community beyond the organization that originally collected the materials. Organizations that provide access to their collections through online portals can collect usage statistics. Collections that require in-person visits often keep a visitor’s log, but that does not give a clear picture of potential users.

When managing data, there may be many assumptions about users. For example, one may assume all users will have similar knowledge of how the data should be used or that these users have similar education or disciplinary backgrounds. This assumption does not take into account lack of knowledge and customs of the specific organization that collected the data. The NRC (2002) states “using past observations in entirely new ways not envisioned when the data were initially collected” will help expand our understanding of the universe, this includes the unexpected user (p. 13).

In 1911, in response to feedback from the community of geologists and engineers, Hayes (1911) summarized, “information concerning state geological surveys is difficult or impossible to obtain” (p. 5). Is this still true today? Cragin and Shankar (2006) suggest a key problem with the small or dark science is that “[w]e do not fully understand the ways in which small (often local) data collections become more public or shared collections” (p.187). Addressing these questions will involve engaging the archival and information science fields (Ramdeen, 2013).

**Geological surveys.**
The USGS and the various individual state geological surveys are the largest public collectors of geological data in the United States (NRC, 2002). These data are not preserved and curated according to the same standards. Individual institutions have developed their own
methods of maintaining their collections and there is a need for consistency. Additionally, as noted above by Gooding (2014), state survey collections often house donated materials and donors standards vary. Concerns over sustainability and interoperability are high as scientists are concerned about accessibility for future researchers, particularly with regards to metadata for these physical objects.

The GMRWG (2015) states that the role of the USGS is to provide “the Nation with fundamental geochemical and geophysical data necessary to address major societal issues involving geologic hazards and disasters, climate variability and change, energy and mineral resources, ecosystems and human health, and ground-water quality and availability” (p. 22). In addition to the Federal survey, state geological surveys “play an important role in generating and disseminating information related to mineral resources” (NRC, 2003, p.25). The NRC (2003) stresses that the role of these organizations is to be an unbiased “source of science and information” (p. 25). These collections represent irreplaceable materials as “some of the sources of these collections are now reclaimed, flooded, or otherwise inaccessible” (NRC, 2002, p.53). In addition to being irreplaceable, these collections cost millions of dollars to acquire and would be expensive to recollect (H.R. 5066, 2014b). The report warns that sources of information from the private sector are often proprietary and not made available to the public. Private organizations have been known to discontinue their collections because of the significant costs, specifically when the costs are weighed in relation to the benefits to the individual organization. This is in contrast to Federal and state government surveys which serve a larger population which “from society’s perspective the aggregate benefits would justify the costs” (NRC, 2003, p. 25).
In industry, the value of these collections may only be realized once the materials have been de-acquisitioned or destroyed. In 2011 Common Data Access Limited (CDA) & Schlumberger issued a report addressing the “business value case for data management.” They suggest that the effort involved with recreating these types of legacy data are difficult or potentially impossible and yet the commercial value is great. This had led to a decision to re-find the original materials among retired workers rather than recreate them (CDA, 2011).

**Stewardship challenges**

**Historical view on preservation.**

Historically, state geological surveys have collections of geological data and act as both active and historical science data centers for their region (NRC & Leighton, 1932; Cragin & Shankar, 2006; & GMRWG, 2015). At the 1926 annual meeting of the AASG, representatives voted to support the NRC’s efforts to salvage well records, “this was perhaps the first attempt to preserve samples and records, a movement that is undergoing a resurgence in the 2000’s” (Cobb, 2008, p. 61). During the 1950’s, the AASG decided “that the archiving of drill core and data was a function primarily of the states and not of a national organization” (Cobb, 2008, p.103). In the 1970’s, the AASG passed a resolution to endorse the Bureau of Mines’ efforts “to develop an efficient and economical national system of repositories for drill cores and other samples of geologic materials” (Cobb, 2008, p.116). During this same time period, the state surveys voted against the formation of regional sample repositories, instead supporting a national core catalog, which is only now being developed.

**Current stewardship efforts.**

According to the USGS’s data lifecycle, providing access to materials is only one aspect of stewardship (Faundeen et. al., 2013). In a letter to the editor from 1817, the idea of public use and access to geological collections is highlighted as an important impact on scientific knowledge. A publicly available collection list enables researchers “to repeat the observations of
preceding inquirers and to extend their observations to other places” (Philosophical Magazine Series 1, 1817, p. 271). In the past decade, there has been growing attention to preserving and managing legacy collections of geoscience data, a trend first attempted by the AASG in the 1920’s and now seeing a return to importance starting in the early 2000’s (Cobb, 2008). This includes efforts by the USGS and the AASG to promote access to collections, as well as position statements from member societies such as the Mineralogical Society of America, American Geophysical Union, Geological Society of America, and the American Association of Petroleum Geologists. The American Geosciences Institute (AGI) has a complete listing of these and other member societies’ position statements on their website. In 1997, AGI published a directory of geoscience data repositories in the U.S., it was initially intended to be revised every two years (Claudy & Stevens, 1997). The directory has not been updated since its creation.

In a 2002 report, the NRC pinpointed two major areas of concern for the preservation of geological collections: metadata and curation. They believe these two facets are critical in enabling trust in the data for reuse by someone other than the original researcher. In their report, metadata is defined as the accompanying documentation that gives one the ability to find and retrieve physical samples and their supporting documents. This documentation “includes information about the age, location, depth, originator, and the date acquired” for an item in a collection (NRC, 2002, p. 9). Additionally, metadata may be used as a digital surrogate for the physical sample. Curation refers to the process of proper management, storage, and access to the geological specimens. Curation includes such steps as appraisal, disposal, and use of data. These

---

3 http://www.americangeosciences.org/policy/policy-positions/membersociety-positionstatements#datapreservation

4 It is worth noting that in the NRC’s definition of metadata, they do not address the information about an object that relates to its authenticity and authentication.
three points are key in the data lifecycle and their context to this research will be discussed further in the literature review.

In 2005, the Energy Policy Act established the NGGDPP under the direction of the USGS (USGS, 2014a). The act is currently being reviewed for reauthorization and has foundations in the work of many organizations including the NRC and the AASG. The purpose of the NGGDPP is to provide for the archiving of geoscience materials, to create a national catalog of these archival collections, and to provide technical and financial support to the organizations which collect these materials (USGS, 2014a). One of the outcomes of this program was the rescue of the Potash samples in Michigan as mentioned in the introduction.

Over the last 13 years, nearly all states (with exceptions like that of Georgia, see below) have engaged in data preservation efforts of some variety to ensure long term access to their collections. This includes taking part of the 2002 study by the NRC, the NGGDPP starting in 2006, and the development of the National Digital Catalog (USGS, 2014b). The National Digital Catalog provides a search interface for accessing and interacting with geological collections, particularly at the state and federal level. It is designed to be similar to a union catalog as opposed to an extensive resource. In 2009, the AASG co-hosted a Geoscience Data Preservation Techniques workshop with the NGGDPP (Steinmetz, Pierce, & Hill, 2009). Other efforts at the state level include the National Geothermal Data System (NGDS). State geological surveys are among the largest contributors to the NGDS, and its catalog contains documents and datasets that will help lead searchers to physical sample resources.

Preservation and access to geological data is becoming more and more topical. Recent presentations at Geological Society of America’s (GSA) 2014 and 2015 annual meeting focused on access to geological data (Copperdock, Cater, & Lehnert, 2015; Deis, 2014; Dunn, 2014;
Johnston, 2014; Hills, 2014; Walker, et. al. 2015; Wyborn, & Evans, 2015; etc.). The Geoscience Information Society and the GSA Geoinformatics Division sponsored a session in 2014 titled “Where in the World? Access and Availability to Geoscience Data”. Talks included topics such as locating ‘old’ data, citing geological data, and challenges faced by librarians and the scholarly communication community in relation to geological data. In 2015 the Geoinformatics Division’s session focused on cyberinfrastructure and data services.

Current earth science preservation efforts are also focused on cyberinfrastructure. This includes such programs as EarthCube, System for Earth Sample Registration (SESAR) and the Data Observation Network for Earth (DataONE), as well as previous initiatives such as GEON. EarthCube is community-led, and has held 24 end user workshops within the various earth science communities to discern their cyberinfrastructure needs (Ramdeen, 2013). SESAR manages registration of International Geo Sample Numbers (IGSN) as well as preserving sample metadata and managing a sample catalog. IGSNs are 9-digit numbers “assigned to specimens and related sampling features such as drill holes or wells to ensure their unique identification and unambiguous referencing of data generated by the study of samples” (SESAR, 2013). iSamples, an EarthCube Research Coordination Network program, “seeks to advance the use of innovative cyberinfrastructure to connect physical samples and sample collections across the Earth Sciences with digital data infrastructures to revolutionize their utility for science” (EarthCube, 2015). DataONE is distributed cyberinfrastructure. According to their principle investigator, Bill Michener, their goal is to work “with our many collaborators to develop a network of data repositories that makes it easy for researchers to preserve, discover, access, and use valuable scientific data” (DataONE, 2014).
In a broader prospective, the Research Data Alliance (RDA) and the International Council for Science’s Committee on Data for Science and Technology (CODATA) also have interest groups working on physical samples and cyberinfrastructure but they are not as far advanced\textsuperscript{56}.

**Needs related to stewardship.**

All collections, especially living collections, need adequate management. In a recent Eos article, Rep. Holt (N.J.) states, “[scientific collections are] not static museum pieces. These collections are used, should be used, will be used, day by day. It is hard to imagine what future researchers would say if we allowed this [material] to disintegrate or disappear” (Showstack, 2014, p. 351). Though state survey collections are valued, the administration of them has changed, especially as they evolved from museums to active collections. These administrative changes have revealed a great need for preservation efforts. As Taylor (1988) asserts (in relation to museum collections), “calling your spare researcher or technician a ‘curator’ doesn’t confer the ability to curate” (p.119). In 2002, the NRC found that geological collections were in disrepair and under-staffed, leading to these collections of legacy data being at risk of loss.

**Case example:**

In 2004, after serving the state for 115 years, the Georgia Geological Survey was closed. While the geological collection was kept, it became at risk for loss of physical materials, digital files, and institutional memory as it was no longer actively maintained. However, it was at risk even before the state survey was closed. Cocker (2005) highlights the lack of archival and technical skills among the staff as a major concern. “The present digital catalog database is an alphabetical file listing and is not searchable by keywords, topics, authors, dates, or subject areas. This digital catalog was developed by people with no technical background, and no input from the geologic staff was considered” (Cocker, 2005, p. 211).

\textsuperscript{5} https://rd-alliance.org/groups/metadata-standards-attribution-physical-and-digital-collections-stewardship.html

\textsuperscript{6} http://www.codata.org/task-groups/management-of-physical-objects
Historically, state survey collections are at high risk for being lost for a wide variety of reasons including the status of state surveys as research institutions in a governmental system. “From the name, state legislators assumed that surveys would exist until the job was completed, then they would go out of business” (Buchanan, 1992, p.63). Funding is also an issue. Early state surveys were originally intended to be short-term endeavors (Boscoe, 2003). As time went on, legislators were “faced with the problem of justifying the spending of public funds. When times were good and money was easy, geological surveys were authorized; when there was a panic, state expenses were curtailed and geological surveys … always slightly suspect as frills by a few hard-headed legislators, were suspended” (Hendrickson, 1961, p.364).

In combination with funding issues, space is currently a concern. The NRC (2002) states, “Although the overall costs of maintaining geoscience data and collections are low compared with those of reacquisition, the amount of money a single repository requires in a short time to alleviate the space problem can be prohibitive” (p.27). The NRC study was inspired by the worry that these collections are at risk of being lost through events such as state surveys being closed or experiencing budgetary issues, companies merging and discarding records or samples, and growing storage expenses. Among state surveys’ concerns were incentives for keeping worthy sample materials and adequate space to store them (NRC, 2002). The NRC report includes a table with information on capacity of state survey repositories – at the time (2002) nearly two thirds of the reporting state surveys had 10 percent or less space left available for growth in collections (NRC, 2002). The situation has only gotten more severe, as recently outlined by Jonathan Arthur, president of the AASG and director of the Florida Geological Survey; “[m]any of this nation’s geological data repositories, most of which are maintained by State Geological Surveys, are now at or near their storage capacity” (H.R. 5066, 2014a).
Budget shortfalls threaten all scientific collections. In researching the field of botany – another collection based science field, Funk (2014) found a trend demonstrating museum budget cuts often lead to cuts in curators and researchers of physical collections. Funk (2014) suggests that this creates additional perils, “collections that are not studied and maintained, even if they are physically well cared for, can become out-of-date and less useful” (p.14). This fits with the views expressed by the NRC in relation to geological collections. “Each time a geological sample or piece of data is allowed to deteriorate, or is damaged, misplaced, or thrown away without assessing its merits, the information it contains and the knowledge it represents are lost” (NRC, 2002, p.9).

While there are a number of challenges, the renewed interest and attention to these issues are leading to success stories as well. The Alaska Geologic Materials Center (GMC), part of the Alaska Division of Geological & Geophysical Surveys, acts as repository for the geological materials collected in Alaska. In 2011, Curator Papp wrote, “continuing to simply maintain the current GMC facility would likely physically jeopardize the material the State has worked so diligently to acquire and preserve”. Their challenges with maintaining geological collections are related to space and climate control issues. At the Alaska GMC, “retrieving core from the unheated storage containers in February is a miserable exercise in which the center’s workers have had to take extraordinary steps such as thawing frozen padlocks and unsticking icy core boxes to get to the samples. In addition, thousands of core boxes at the center have deteriorated over time” (Ragdale, 2015). Papp further explains, “in these freeze-thaw conditions, it’s really cumbersome and awkward handling the samples, and the rocks degrade over time” (Ragdale, 2015, para. 17). But as noted above and seen in Figure 2, deteriorating conditions also leads to loss of metadata and other identifying information found on core boxes. Through the efforts of
many in Alaska, they have made great advances. On July 1st 2015 they opened a new 90,000 square foot facility which should have adequate space and housing conditions, but not all geological collections are as well positioned (AASG, 2015).

**Justification for research**

Under-appreciation of science collections such as geological repositories has led to under-funding and other risks that are placing these collections at peril. As the example in Los Angeles illustrates, these collections serve a very valuable purpose but if time and imminent hazards were not a factor, would they still be as valuable? The costs of replacing a geological collection can be staggering particularly for cores and cuttings collected from oil and gas exploration. Some materials collections cannot be replaced – sites may no longer be accessible due to development, environmental restrictions and hazardous conditions.

Geological collections present many unique and diverse challenges for preservation. Many geological specimens are not fully replicable with digital surrogates. They are costly to collect and store. The role of historic or previously collected data in current and future research is constantly changing. While next steps for preserving legacy data include developing cyberinfrastructure to assist with discoverability of data and ultimately access and use, these steps do not address all of the preservation needs typically faced by those managing a physical collection. As Cocker (2005) and the NRC (2002) suggest, there is a need for proper training in curation and management to help prevent loss of legacy data. Also, further research is needed to see if current efforts are meeting needs of both the data repositories and their user communities, including reporting this information to nonscientific stakeholders in metrics they understand. Continued funding of data rescue projects relies on those not trained as geologists to understand the value of these collections. Currently there is a gap in knowledge of how to address this issue sufficiently.
In the past decade, there has been growing interest in preserving and managing geoscience collections. For example, the USGS’s NGGDPP which is centered on archiving geological data. However, in all of the work being done to preserve these collections for the future, there has been little to no work looking at the needs of users of these collections. Current research in this area is focused on case studies (e.g. Loudon & Laxton, 2007) or through examples of tools and ontologies (e.g. Ramachandran et al., 2004; Sen & Duffy, 2005). Little research has been done using qualitative methods, or from the end user perspective. This is the current approach of the NSF program, EarthCube. EarthCube’s ultimate goal is to develop cyberinfrastructure for the earth sciences and includes state geological surveys among their stakeholders. EarthCube has conducted over 20 end user workshops in order to collect insight directly from the user community but has yet to publish their results (Ramdeen, 2013).

The management of scientific data collections is changing as scientific research evolves and the technology available for accessing these collections continues to develop. Librarians and other information scientists have unique skills that can be used to shape this future landscape. Managing a geological data collection is not just about preservation; it is also about access. Instead of developing new systems for data management, Heidorn (2008) suggests we need to develop infrastructure for managing data as part of existing institutions. Parsons and Duerr (2005) provide an example of a dedicated staff member within their scientific organization who acts as a reference librarian for data, highlighting that this human element is not often recognized by funding agencies or included in best practices like the Open Archival Information System (OAIS) reference model. Weber, Palmer and Chao (2012) connect these institutional needs to the LIS field directly, stating “LIS programs will need to have well-trained graduates ready to fill these positions” as LIS professionals are trained to understand users’ needs, future needs and
other types of information services (Weber et al., 2012, p. 317). Weber, et. al. also mention that, aside from understanding how users share data, we also need “behavioral studies to better understand how data are produced, used, transferred, appraised and reused in a variety of research” (Weber, et. al., 2012, p. 318).

The costs of collecting these materials make preserving them for future use important, but if potential users cannot access these collections, they may be spending time and money replicating existing collections. In order to better serve these users, it is important to understand their information seeking behavior. By understanding their behavior, better tools for providing access to collections can be created. Part of a well-managed collection is proper record keeping. “Without reliable identification and other recordkeeping, a museum’s collections are of little use” (Danilov, 1982, p. 185). It is important to keep proper records for a collection or they lose whatever scientific data they contain as well as the ability to identify a sample. These two resources are strongly tied together, for some geological materials, knowing the location where it was collected and other metadata makes the scientific data contained within the sample valuable; without it the sample is just another rock.

**Study Overview**

This study investigates information seeking behavior in relation to physical samples. Samples refers to physical objects collected for use in scientific research. In particular, objects which have been gathered with the intention of representing geological information. This study focuses on the issue of information seeking behavior and how it relates to access of geoscience data. Geological data was selected as the focus because of the existing concept of science data collections and the tradition of searching within these types of institutions for reuse of existing materials. Participants were recruited from the various state geological surveys which maintain physical sample repositories. They are users coming from outside of the organization which
owns the geoscience collection, as opposed to the user/creators that exist within the state survey. In addition to the interviews, two questionnaires were sent to state geological surveys to create context to the management and resources available for accessing their collections. Results of this qualitative inquiry were used to develop a model of information seeking behavior. Their behavior is influenced both by the types of institutions and the nature of the materials they are looking for during their search.

Information seeking is of particular interest because we do not know how geologists interact with information sources in order to address their information needs. In this study, information need refers to the activity of interacting with repositories to discover physical samples. This might be mechanical or intellectual interactions. We do not know how science data centers support these interactions, especially to outsiders of the organization who might not be familiar with the organizations data processing or what information sources are available. How do these outsiders find out about information or samples? Is there anything about the nature of physical samples that makes searching different or difficult?
CHAPTER 2: LITERATURE REVIEW

The purpose of this study is to better understand the search patterns normally followed by users of geoscience collections when searching for physical sample materials and to develop an information seeking model based on these observations. In this vein, three topics stand out: representation and organization of knowledge; scientific collections and their roles; and information seeking behavior. The first section of this review will look at how information is conceptualized within the field of information and library science and how that conceptualization may be applied to physical scientific samples. Next it will address scientific collections; specifically, why we maintain collections of physical specimen. It will also address a need for awareness and a better understanding of recordkeeping and management of these collections. This will lead into a discussion about the information seeking behavior of scientists and how they have been studied in the past. Each section weaves together literature drawn from information and library science and from the earth sciences. This includes analytical and descriptive review and suggests possibilities for synthesis.

Information in the earth sciences

**Physical objects as information sources in the earth sciences.**

In debating physical objects as information sources, Buckland (1991) states, “The term ‘information’ is also used attributively for objects … because they are regarded as being informative, as ‘having the quality of imparting knowledge or communicating information; instructive’” (p. 351). As discussed in seminal works by Briet (1951), physical objects can be considered documents when given specific context. In the case of an antelope, in the wild it is an animal, in a zoo it is a document. It is the primary source of information, specifically when it is
made an ‘object of study’. Although not traditionally applied to collections in the earth sciences, these terms could easily be applied to physical object specimens such as geological materials. Physical geological materials include objects such as “cores, cuttings, fossils, … rocks”, are considered primary sources of geological information, and “are the foundation of basic and applied geoscience research and education, and underpin industry programs to discover and develop domestic natural resources” (NRC, 2002, p.1). This quote from a 1817 letter to the editor in Philosophical Magazine, highlights the early and invaluable role that information can be gained from physical geological materials, “[A] fossil shell, petrifaction, or mineral is useless to the geologist, unless it be accompanied with a proper description of the stratum, and of the exact place from whence it was obtained: hence it is necessary that a descriptive catalogue should always accompany a collection of geological specimens” (p.269). It represents learned or communicated information by objects or the specific “strata” (or rock unit) in which they were found (Philosophical Magazine, 1817). Brooks, Lister, Eastop, and Bennett (1996) suggest that an object can be used for “external inference”. The condition of the object and the purpose for which it has been maintained can provide additional insight to some historical objects.

McCreacidie and Rice (1999) conceptualize information as resources which assume a sender and receiver of the information they contain. The authors include physical objects, events, sounds, and even smells as unintentional information sources that exist in the environment around the searcher. Case (2002) states, “it seems that McCreacidie and Rice are trying to make finer distinctions” (p. 44) with their definition than Buckland. Their typology is more limiting or figurative than that of Buckland. Buckland provides a less constructed typology which may be a better fit for geological samples especially as geological materials are considered to be primary sources of information.
In discussing science data collections, Thessen and Patterson (2011) point out there are many ways to define “data” and it is often not consistently applied. They have two broad types of data – observational and processed, neither of which completely captures the idea of the object as an information source. Object-based information sources imply a different set of needs and a different structure for use and access. Physical objects are typically found in small data collections or in aggregated collections of individual researchers. These collections are defined as resource collections by Cragin and Shankar (2006) and described as field collections by the GMRWG (2015). According to the GMRWG (2015), “geoscience materials can be thought of as having two main components: (1) the physical samples collected by research personnel and (2) the data generated by analysis of those samples, by instrumentation-based field surveys, and from engineering research” (p. 18). The latter associated materials will inevitably need to be connected to physical objects and data sets, first related to the search process and then potentially as metadata to add in search and discovery.

Physical materials can be taken and stored for future access; they typically do not change over time. Most well preserved and managed geological samples will not decay or change over time. But some specimens may be more dynamic, such as volatile chemicals and fragile fossil specimens. For the most part however, the information contained in these materials is stable on an everyday timescale. As such, data they hold can be reused, often applied beyond the original intended purpose for which the materials were gathered. The information contained in these materials is “the history of processes that operate on the Earth today and in the past and provide insights that lead to improved prediction of hazards, both immediate and long term” (NRC, 2002, p. 1). As Baru (2007) states, “Earth science data are valuable—they are expensive to collect and data sources may be lost (due to urban development, changes in permitting regulations, etc.)—so there is interest in ensuring that existing data are used to the fullest extent possible” (p.114).
Jackson (1999) calls geological collections “primary source materials” which can contain “valuable information about geological localities, the collectors of material, relationships between collectors, institutions, museums and places of learning, and cultural and social factors (Jackson, 1999, p. 431). Jackson (1999) also highlights the heavy use of historical catalogs for geological specimens but often the lack of actual use of specimen by historians. He discusses this as not a lack of citation in finished research, but a lack of interest in these collections as resources of information to this audience. Jackson (1999) believes this might be the source of misperceptions on the value of collections to historians. He also cites poor curation of collections as a barrier to access, which may lead to the lack of citations.

Furuta, Marshall, Shipman, and Leggett (1996) discuss the nature of physical objects in scholarship and the idea that digital surrogates can be used to ease limitations on access; digital surrogates can be used to provide greater access and awareness of physical objects through databases and digital libraries. Physical objects may contain information that have the potential to be captured through digitization. However, digital surrogates do not fully replace physical objects – some views or representations cannot be replicated digitally (Furuta et al, 1996). The challenge is in properly capturing the terminology, schemas and other metadata in order for scientists to successfully retrieve items. For earth science objects, these connections or discovery points might be seen as chemical testing, secondary analysis, reports, and other associated data.

In a 1975 discussion on managing borehole samples and their associated data, Ekstrom, Wirstam, and Larsson (1975) focus on the development of a computer system. In describing the development of these systems, drill cores and borehole data (physical rock specimens collected while drilling wells, typically oil wells) are associated with other information sources such as
documents with general information that may include location, depths, geologists who worked on the well, etc.; and geological information that may include petrography, tectonics, chemical, joint, or borehole deviation data (Ekstrom, Wirstam, & Larsson, 1975). These data and other materials are connected and may be different from published reports and other formal information sources. This is supplemental information about the specific physical samples and can enrich the usefulness or help with discovering the samples based on criteria in these materials.

**Geological data.**
Earth science is a complicated field of study. It encompasses a number of subdomains from geophysics to geology. Each of these subdomains is complex and has its own history regarding how data is collected, perspectives by which they process these data, and ultimately the methods used to transform raw data into knowledge. Technology has changed the way scientists have addressed research questions and information/knowledge creation. Now scientists are concerned with the ‘grand’ questions or challenges (Ma, Carranza, Wu, & Van Der, 2011; Stewart et al. 2010; Yang, Li, Xie, & Zhou, 2008). These are complex research questions that cross domains, disciplines and require collaborations from “experts from different science, technology, and management domains” (Yang et al., 2008, p.273). They address such topics as “food and water shortages, dependence on foreign oil, and climate change” (Stewart et al., 2010, p.40).

Specifically, geoscience data is used in resources exploration and management; urban development; climate change; water quality; and hazard mitigation (Ma, Carranza, Wu, & Van Der, 2011, p.2).

According to the NRC (2002), geological resources can be used for hazard assessment; basic and applied scientific research; discovery, assessment, and enhanced utilization of national resources; as well as education and public awareness. Table 1, taken from appendix D of the
NRC’s 2002 report, is a list of examples of materials in geological collections. This report also contains a list of derived and indirect data which, in some cases, are developed from examination of the materials listed in Table 1.

Table 1
*Examples of materials in Geoscience Collections*

| Auger samples                  | Rock cores          |
| Fluid samples (oil, gas, water) | Rock cuttings       |
| Geochemical powder samples     | Sediment cores      |
| Hand samples (incl. geotechnical, rock and mineral) | Sidewall cores |
| Ice cores                      | Thin sections and polished sections |
| Paleontological samples (micro/macro) | Type stratigraphic sections |

Note *(Taken from NRC, 2002, p. 98)*

The materials in Table 1 are physical objects and scientific samples. They may collectively represent a single object or a set of objects. For example, rock cuttings are materials from below the surface which are collected at intervals from a drill rig. As the rig drills below the surface, the driller will capture a selection of materials returned from the drill bit and record the interval below the surface they were taken. The complete set of rock cuttings includes the series of sample bags from each of the intervals collected. This set is referred to as a sample, while the individual intervals are subsamples. Alternatively, someone may also take material from a specific interval for analysis, and this would also be a subsample. The relationship between sample and subsample is hierarchical. The information about the parent (the set of cuttings) is also assumed by the subsample, or offspring (see Figure 3). For example, the geographical location at the earth’s surface (latitude and longitude) from which the samples were collected would be carried forward to the various subsamples. The subsamples may be analyzed and the results would not necessarily be transferable between offspring samples, but would be
considered properties of the parent. When searching for samples, researchers may be looking for information related to the parent or the offspring.

Representative of knowledge.
Knowledge generation is important and is reliant on a number of things; however, not everyone has the same framework for understanding knowledge. Much of the information related to tacit knowledge is hard to communicate, “It is assumed that users of geological maps have in common a geological training and background that enables them to interpret the implicit information, but even with such a shared background, there is much room for misinterpretation and misunderstanding” (Loudon & Laxton, 2007, p.331). Particularly in technological systems, one might need more context and provenance to understand where the knowledge is coming from. These systems should “ensure all data and knowledge is recorded using common, documented standards to promote wider interoperability, and data and knowledge exchange” (Howard et al., 2009, p. 825).
In the earth sciences, frame of reference makes an important difference. Knowing the location where a marble was collected tells us different information (contextually and scientifically) about its geological history. Additionally, the point of view of the original researcher can affect data; for example, if knowledge is created or used by a petroleum geologist versus a civil engineer. Additional information is needed to create a common background and understanding of how the knowledge should be interpreted and how it might be applied.

According to Loudon and Laxton (2007), knowledge can be generated based on predetermined procedures and/or “more flexible, holistic” procedures when conducting geological field work (p. 321). To capture these steps, the specific provenance is recorded (or should be) and becomes the metadata used in defining and reconciling information sources. This leads towards the need for a “shared framework and standardized ontologies” to “avoid pointless and unnecessary variation” (Loudon & Laxton, 2007, p. 321). More specifically, Balestro et al. (2013) state there is a need in the earth sciences to “allow sharing and spatially discovering [of] maps, and support [for] interchange of data amongst different systems, experts and communities” (p.254). Given this need for sharing and discovering, one must ask, how are these materials being organized? The organization of these materials impacts search and system design.

**Information organization.**

As mentioned previously, Earth Science is a very diverse field with a wide variety of subdomains. Each have different uses, perspectives, collection procedures, and publishing processes related to similar geological data. For example, within geology, sedimentary geologists and structural geologists use similar terminology (limestone, bedding, etc.) but each has their own definition. This causes issues with reliability, and searching for data across subdomains. This may affect how knowledge is represented in various databases and cyber systems, as well as the development of metadata standards and subdomain-specific ontologies.
In developing knowledge management frameworks, there are three different types of information, explicit, implicit, and tacit. These are defined in Table 2 in the context of earth science data. Knowledge becomes increasingly difficult to formalize as it progresses from explicit to implicit to tacit. The example above with sedimentary and structural geologists illustrates a mixture of implicit and tacit knowledge. Within a specific organization, a community of practice may develop where there is a shared understanding of these frameworks and processes. These practices might not be captured explicitly. Frameworks should “ensure knowledge capture is carried out in the context of prior information, i.e., it builds on and augments prior information, without re-inventing it, [to] ensure that the data and knowledge capture process is verifiable, repeatable and auditable” inside and out of a specific organization (Howard et al., 2009, p.826).

Table 2

<table>
<thead>
<tr>
<th>Three types of knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td><em>Explicit</em></td>
</tr>
<tr>
<td>Conventional</td>
</tr>
<tr>
<td>Digital*</td>
</tr>
<tr>
<td><strong>Implicit</strong></td>
</tr>
<tr>
<td><em>Knowledge that has been recorded, communicated or articulated in some tangible way</em></td>
</tr>
<tr>
<td><em>Knowledge that has been adequately described in a form that is machine-processable</em></td>
</tr>
<tr>
<td><em>Knowledge that is capable of being communicated or articulated, but is yet to be made explicit</em></td>
</tr>
<tr>
<td><em>Knowledge that cannot be articulated, but is acquired and exchanged by experience-based learning (rather than verbal instruction)</em></td>
</tr>
<tr>
<td><strong>Geoscience examples</strong></td>
</tr>
<tr>
<td>Textbooks, indexes; Procedure manuals; Papers, reports, map explanations; Maps, sections, diagrams; Field notes, data files, ephemera; Recorded field or lab observations/data</td>
</tr>
<tr>
<td>Digital 3D and process models; Computer code; Framework, ontologies, metadata; Embedded in digital field support; Hypertext sequences, workflows; GIS, spatial models; Databases</td>
</tr>
<tr>
<td>Unrecorded field observations; Trains of thought supporting models or interpretations e.g. what data has been considered or ignored, and why; Certainty/uncertainty re. observations/models</td>
</tr>
<tr>
<td>Experience in analysis – e.g. seismic/petro-physical interpretation; Spatial judgment – e.g. geology of underground or open pit mine design</td>
</tr>
</tbody>
</table>

*Increasing difficulty to formalize

*Digital represents the capture of information using cyber infrastructure to turn implicit information into explicit.

Note Table based on definitions in Loudon & Laxton, 2007 and Howard, Hatton, Reitsma, & Lawrie, 2009.
This is particularly important with physical objects where tacit and implicit information must be captured, and in most cases, cannot be determined by conventional methods. Implicit decisions the geologist made when collecting the sample (for example, why that particular piece was selected over another) maybe be lost if not recorded. Once captured, we need to be able to manipulate this information. Howard et al. (2009) state that some “geological information either lies latent in a scientist’s mind or is expressed in a software-specific language, being able to express the semantics of that knowledge in ontologies is critical to support the future of geological research” particularly “in a grid environment” (p.833). This information is implicit but may include some tacit knowledge as well. Howard et al. (2009) elaborate further by stating:

“An ontology-based future for the World Wide Web will enable greater access to tacit and implicit knowledge in shared geoscience knowledge bases and the wider web community. By automating data discovery and conditioning tasks, ontologies will also help GSOs to unlock and harness the considerable knowledge assets within their traditional paper records and archives. Sustained investment and international collaboration is needed to capture this valuable intellectual capital and to continue development of the cyber-infrastructures required for wider knowledge exchange and exploitation” (p.834).

Implicit knowledge such as the geologist’s information processes when going from field collection to decision making should also be captured in a framework (Loudon & Laxton 2007). This knowledge may not be associated with a digital environment (electronic data sources, software or analysis tools, etc.) but instead just the researcher’s notes and mind, which makes it harder to automate/capture. Implicit knowledge involves uncertainties and hypothesis making which is constantly changing as new data is added. This level of information is not always captured in the finished product, or even outside of the geologist’s personal notes or memory. This can make it difficult for those outside of the community of practice to search for data.
Objects are often treated as surrogates for the geological processes that formed them, and explicit knowledge used as evidence. An example given by Loudon and Laxton (2007) is of siliceous grit overlaying granite. They both constitute the same rock material, but represent different depositional processes. It is important to capture tacit knowledge such as this reasoning behind final outcomes—the “links between objects, processes and interpretations” - in order for data to be reused and shared (Loudon & Laxton, 2007, p. 323). This requires a shared ontology to “connect information to the underlying reasoning” and develop a “semantic interoperability” (Loudon & Laxton, 2007, p. 323).

**Why do we keep scientific data collections?**

There are many reasons why we maintain collections of scientific data. As Wilson (2007) states, “Collections are non-renewable resources. Many specimens now existing in museum collections would be impossible to collect again due to destruction of sites or habitats… Specimens now in collections may also be irreplaceable today due to restrictions on collecting certain groups or in certain places, restrictions that did not exist when the specimens were collected. Collections are cost-effective” (p.48). For example, scientists may wish to reuse materials for purposes beyond which they were originally collected. But for Perutz, a molecular biologist who came of age as a scientist during WWII, maintaining scientific collections were important to keep scientists accountable and to allow other researchers to better understand the process on which new discoveries were based. Perutz (1989) wrote “True science thrives best in glass houses, where everyone can look in. When the windows are blacked out, as in war, the weeds take over; when secrecy muffles criticism, charlatans and cranks flourish” (p.xvi). Perutz suggested that it takes creativity to conduct scientific research—scientists must think and be original much the way artists are respected for their processes for creating art. In this way collections can be used as inspiration. By reviewing the work of others, or what has already
been collected, perhaps a scientist might develop new ideas. This reinforces the idea that scientific collections can be used for a wide variety of purposes beyond the original intent. “Against a backdrop of disappearing habitats, species extinctions and the destruction of sites of geological and paleontological significance, the specimens in natural science collections have become [a] nonrenewable resource of vital importance to science and society” (Duckworth, Genoways, & Rose, 1993, p.1).

Lee, Bell, and Sutton (1982) highlight the importance of having voucher specimens in museums. Voucher specimens “physically and permanently document data” and comprise the “reliability, accuracy, and ability to repeat otherwise good research” (Lee, Bell, & Sutton, 1982, p.5, iii). Similar concepts exist in the earth sciences, called type or holotype specimens. These are the single example of a specific taxon which will be used in identifying future specimens, often the first to bear the name when a new taxon is established (Duckworth, Genoways, & Rose, 1993). Yates (1987) supports this by explaining that voucher specimens “insure that identifications can be verified or changed if necessary and allow historical comparisons to be made by works in the future” (p. 10). Lee et al. (1982) point out two uses for voucher specimens “1) verifying the identity of the organism(s) used in the study, 2) by so doing, ensures that a study which otherwise could not be repeated can be accurately reviewed or reassessed” (p.5).

A voucher specimen, however, is incomplete without accompanying documentation; specifically “a complete set of notes characterizing the site from which it was taken and the conditions under which it was collected” (Lee, Bell, & Sutton, 1982, p. 15). Lee et al.’s (1982) claims emphasize the need to maintain original specimens as opposed to focusing on digital surrogates or digital alternatives for long term reference. While surrogates can help provide discovery and access, they cannot be used to perform verification.
In addition to theoretical purposes, scientific collections exist for many practical reasons. This is supported by the National Resource Council (2002) in relation to geological samples: “old core can produce new knowledge. Existing geoscience data and collections may be viewed both with new eyes and with new technologies” (p. 12). They can provide information for economic or trade purposes, provide insight into the past, help with managing the quality of our environment, and inform stakeholders such as governments or the public on issues relating to food and agriculture, public health and safety and national security. In addition, they may contain materials with specific intrinsic value – priceless objects that cannot be replaced. Collections may also have a number of unanticipated uses. Changes in technology might allow researchers to ask and answer new questions from existing specimens (NSTC, 2009, p. 2). NSTC (2009) builds on this use by providing a list of functions of permanent scientific collection:

A. As ‘vouchers’ from earlier observations or findings.
B. As standards. Some specimens become permanent references that must be retained for future comparison known as ‘Type specimens’.
C. As sources of specimens for biological research, conservation, and food security.
D. As repositories for rare objects.
E. Sources of ideas and study specimens for education and training.
F. Some agencies proactively collect samples for future analysis or experimental use in line with their missions. (p. 11)

It is important to maintain collections, not just for new unthought-of research, but to confirm previous work. Jackson (1999) states, “A fundamental tenet in science is the need for viable checking and reproducibility of results. Re-analyses may not be undertaken for some time after the original research, but require preservation of the original material worked on in order to be of any value” (p. 423). As Lane (1996) states, work is “vouched and documented by the specimen placed in natural history collections” (p. 537). These specimens act as reliability and validity checks on the work being produced. Heidorn (2008) points out science is based on
theories, and theories are created based on replicable data. If the data is inaccessible and the theory cannot be replicated, scientific results would be unsubstantiated. “The availability of the data behind experiments helps to insure scientific integrity by keeping the process open to external evaluation” (Heidorn, 2008, p. 286).

The NRC state in their 2003 report, the government’s “first role [in relation to scientific data collections] is as an unbiased national source of science and information” (NRC, 2003, p. 25 [emphasis in original]). These collections represent irreplaceable materials, “some of the sources of these collections are now reclaimed, flooded, or otherwise inaccessible” (NRC, 2002, p. 53). The report warns that private-sector sources of the information are often proprietary and not made available to the public. Occasionally, private organizations discontinue their collections because of the significant costs, specifically weighed in relation to the benefits to the individual organization. This is in contrast to the Federal government that serves a larger population where, “from society’s perspective the aggregate benefits would justify the costs” (NRC, 2003, p. 25).

In addition to Federal organizations, there are a number of state agencies such as state geological surveys which “play an important role in generating and disseminating information related to mineral resources” (NRC, 2003, p. 25).

**Scientific collections of information**

Scientific collections are managed sets of data and associated metadata used to conduct scientific research. They are maintained by a single unit within an organization. Many researchers may deposit materials, and the data may have been collected for many diverse purposes. These collections generally have a theme, for example, supporting the mission of the parent organization. They are organized and made accessible to researchers to support their scientific research. The definition for scientific collections in this paper will draw from the description written by the National Science and Technology Council’s Interagency Working
Group on Scientific Collections (2009) related to object based scientific collections; the GMRWG (2015) – specific to collections of geological samples; and Cragin and Shankar’s (2006) definition of scientific data collections (SDC). The NSTC’s description focuses on the management of physical object based collections - specifically preservation and cataloging for long term access such as research. Specific items are excluded from scientific collections; they “do not include art or historical objects, collectibles, or the books and documents that are stored in libraries and archives. The specimens in scientific collections were acquired as objects for scientific study, not for their aesthetic or market value as collectibles” (NSTC, 2013, p. 11). This is echoed by the GMRWG (2015), which views digital sources as a different collection and is not addressed in their report; instead they focus on physical collections, specifically repositories. “Historically, the term “repository” is indicative of a distinct physical location where samples or documents are housed and curated” (GMRWG, 2015, p.6).

Cragin and Shankar (2006) suggest, “SDCs may include data (raw and processed), metadata, annotation and other supplemental information, visualization and analysis tools, links to other knowledge bases, bibliographic data, etc.” (p.186). While documents and books do typically reside in libraries and archives, these items may also include important metadata which is needed to document scientific collections and the connection between the collection and the library or archive is not clearly defined. This may include implicit information which needs to be explicitly documented for physical objects to facilitate search and discovery. Further research is needed to identify where the distinctions typically appear within organizations when these facilities (libraries and archives) are maintained separately.

The GMRWG (2015) defines their SDC from the view point of science data within a single organization. We may consider state geological surveys to be examples of such SDC’s. First,
they broadly define material collections as “a set of specimens that have been brought together on the basis of some common characteristic” (GMRWG, 2015, p.12). This is then detailed as a hierarchy consisting of 5 levels, as seen in Figure 4. At the center is the individual sample. Multiple samples are gathered for a field collection. Field collections from similar “disciplinary, temporal, or geographic parameters” are aggregated as research collections (p. 12). Directed research collections may contain selected samples from various field collections. At the top level of the hierarchy are general collections which “exhibit similarities of sample type or geography” (p. 12).

According to Cragin and Shankar, there are three types of SDCs: 1) research collections – small or individual scale; 2) resource collections – community driven but lacking stable funding; and 3) reference collections – large-scale, serving science and education, diverse user groups.
These scale like variations in data collections are often simplified as small and big data. These collections serve a wide variety of needs both from a creator and a user perspective.

The GMRWG (2015) takes a different view on repositories, with three categories related to the level of curation or long term planning associated with the collections: 1) active repositories – “a permanent facility that assumes responsibility for the long-term storage and maintenance of a collection (or collections) of related materials”; 2) ephemeral repositories – smaller, temporary collections which “allow ready access to samples during the course of research”; and 3) inactive repository – where materials are stored but “neither curator nor management plan exists” (pp. 11). The structure and role of state geological surveys’ collections vary from state to state and may represent any of these levels, if not all.

Information behavior

In the previous sections we identified information in the Earth Sciences, discussed the importance of reuse of existing materials, and reviewed how these collections are organized. Building upon these concepts, we now turn to the behavior of users of these materials. When introducing the concept of information behavior, McCreadie and Rice (2002) define four basic terms: 1) Information, 2) Information need, 3) Information seeking, and 4) Information behavior. Information is defined as “any difference perceived in your environment or within yourself”; information need is a “recognition that your knowledge is inadequate to satisfy a goal”; information seeking “is a conscious effort to acquire information in response to a need or gap in your knowledge”; and information behavior “encompasses information seeking as well as the totality of other unintentional or passive behaviors that do not involve information seeking, such as avoiding information” (Case, 2002, p.5). Figure 5 is a visual representation of the relationship between information behavior, information need, and information seeking based on the definitions given by Case (2002).
Information seeking implies that there is a goal – either a decision which must be made or some other purpose for using information. Seeking continues until either a deadline has passed, the need for information no longer exists or the need is satisfied by the information found. Wilson (1999b) stated that information seeking is “the purposive seeking for information as a consequence of a need to satisfy some goal.”

There are many models of information seeking behavior. Most well-known are Dervin’s model of sense making, Kuhlthau’s model of the information search process, and Ellis’s model of information seeking strategies (Wilson, 1999a). Wilson (1999a) reviewed these three models in the context of the broader concept of information behavior, and his own macro model of information seeking behavior. Within the concept of information behavior, Wilson places Kuhlthau and Ellis’s model in the area he calls active search. Dervin’s model is considered a model of methodology and does not as easily find a fit in this generalized model. Wilson
(1999a) found that Kuhlthau and Ellis’s models of information seeking are complementary, and proposes the possibility of merging the two models. Ellis focuses on modes of exploration while Kuhlthau looks at the associated feelings and thoughts (Wilson, 1999a).

Returning to Wilson’s (1999b) earlier mentioned definition of information seeking, he notes that, “in the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web)” (para 6). In the same discussion, he also includes the term information search behavior. This distinction is not made by Case (2002), but Wilson (1999b) defines it as

“The ‘micro-level’ of behaviour employed by the searcher in interacting with information systems of all kinds. It consists of all the interactions with the system, whether at the level of mechanical interaction (for example, use of the mouse and clicks on links) or at the intellectual level (for example, adopting a Boolean search strategy or determining the criteria for deciding which of two books selected from adjacent places on a library shelf is most useful), which will also involve mental acts, such as judging the relevance of data or information retrieved”

Wilson, 1999b, para. 7

Another term used by Wilson (1999b) is information use behavior, which “consists of the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts which involve, for example, comparison of new information with existing knowledge” (para 8). This subtle distinction between seeking, searching and use of information is not found in Ellis’s model as his has a more narrowed focus. Ellis is not interested in how the user applies the information they have gathered after the seeking event has ended.

Ellis’s research has focused on scientists looking for print materials and has not yet been expanded to include physical objects or their surrogates. These are distinctly different activities,
with their own motivations. As Wilson noted above, Ellis focused on process and not motivation. As systems and cyberinfrastructure are developed to provide access to research data, their focus mirrors our traditional concept of searching for print materials and as such seems to suggest a good fit.

**Ellis’s model of information seeking.**

Ellis developed a series of basic models of information seeking behavior by researching the information seeking of chemists, physicists, research scientists and engineers (Ellis, 1993; Ellis, Cox & Hall, 1993; Ellis & Hagen, 1997). These participants were interviewed in relation to their searches for print materials and other published reports. From these interviews Ellis developed a set of 6-8 categories (varying by discipline) of scientists’ search behaviors. It is possible that many of the concepts applied to information seeking for print materials could be applied to the seeking of physical materials, but this is speculation and justifies the need for research. Without research in this area, one cannot be certain that the traditional methods for providing access to information sources would be applicable to collections of physical objects.

Table 3 shows the categories developed by Ellis and his various co-authors which are similar regardless of terminology. These categories are based on the activities each group undertakes and as seen in the 6th category, “Differentiating” etc., this does not always exactly match.
Table 3.

*Categories for Ellis’s Models*

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Scientists</td>
<td>Starting</td>
</tr>
<tr>
<td>Chemists</td>
<td>Starting</td>
</tr>
<tr>
<td>Physicists</td>
<td>Initial Familiarisation</td>
</tr>
<tr>
<td>Engineers and Research Scientists</td>
<td>Surveying</td>
</tr>
</tbody>
</table>

Note. *The categories are organized by how they match up based on activities. Source compiled from definitions in– Ellis, 1993; Ellis, Cox & Hall, 1993 and Ellis & Haugan, 1997*

*These are two separate categories but both map to differentiating and source prioritization.

**These categories do not map to any other categories.

While Ellis presents his model as a set of categories or stages of behavior, Wilson (1999a) offers a potential diagram of Ellis’s model (Figure 6). In order to create the flow of the model, Wilson notes a number of assumptions. Starting and ending are endpoints to the model, with browsing, chaining, and monitoring being search processes. Differentiating is a way of filtering this information before extracting and verifying would likely occur. Wilson (1999a) suggests this presentation of Ellis’s model has implications for information retrieval system design.
Definitions of Ellis’s categories.
Because the findings of this study will later be compared to the search behaviors discovered by Ellis, his definitions of each of these behaviors are provided here.

Starting/Initial Familiarization/Surveying: “activities characteristic of the initial search for information” (Ellis, Cox, & Hall, 1993, p.359). Starting activities include speaking with personal contacts and using formal search tools in order to develop a base understanding of the information need. This was often followed by Chaining/Chasing: “following chains of citations or other forms of referential connection between materials” (Ellis et al., 1993, p.359). Chasing includes backward as well as forward chaining. It is also important to remember that even though the categories often lead from one to another, they do not have to be processed sequentially and can be repeated.

Browsing: “semi-directed searching in an area of potential interest” (Ellis et al., 1993, p.359). Unlike chaining or chasing which can be very structured, browsing is looking with the hope of finding something useful from the literature. This may include shelf browsing in a library or information gathered at a conference (Ellis et al., 1993, p.361). Extracting is a more focused activity than browsing. Extracting: “systematically working through a particular source to locate material of interest” (Ellis et al., 1993, p.359). Extracting was not always found to be
significant (Ellis et al. mention that extracting was not particularly important to chemists and was minimally important to physicists) and often was employed to pull relevant information from materials found during an initial or familiarization process (Ellis et al., 1993, p.364).

**Monitoring/Maintaining Awareness:** “maintaining awareness of developments in a field through the monitoring of particular sources” (Ellis et al., 1993, p.359). Monitoring is a regular activity and can include watching recent publications – journal articles and books, as well as maintaining personal contacts. Ellis et al. (1993) note that, while chemists often monitor specific fields or topics within chemistry, the chemists and social scientists also had more general interests (p.363).

**Differentiating/Source Prioritization/Distinguishing /Filtering:** “using differences between sources as filters on the nature and quality of the material examined” (Ellis et al., 1993, p.359). Note the subtle differences in the definitions – Distinguishing is “when information sources are ranked according to their relative importance based on the respondents’ own perceptions” and Filtering is “the use of certain criteria or mechanisms when searching for information, to make the information as relevant and precise as possible” (Ellis, 1997, p. 399). In all three cases, the researcher has their own criteria or set of standards for identifying differences and ranking sources.

Chemists have two additional categories, verifying and ending. **Verifying:** “activities associated with checking the accuracy of information” (Ellis et al., 1993, p.359). Verifying was mentioned by social scientists, but as a minimal task associated with chaining (Ellis et al., 1993, p.364). It is easy to see how this category is both important and connected to other processes. Chaining can be a process for verifying information gained in a search, but is slightly different when it is not used to gather new information. **Ending:** “activities characteristic of information
seeking at the end of a topic or project, for example, during the preparation of papers for publication” (Ellis et al., 1993, p.359). It seems like this should be a common category to all, but if a project is not considered to be the end of a greater research goal, it is possible that some researchers do not see their information needs as ever ending.

Physicists have one additional category, that of locating. Locating: “encompasses the activities engaged in when actually finding the information” (Ellis et al., 1993, p.359). It can be considered a sub category (Ellis et al., 1993, p.358). It may be an important category for information seeking when considering physical materials and as such is included in this review.

**Criticism and limitations of Ellis’s model.**
Järvelin and Wilson (2003) found that Ellis’s model is descriptive rather than explaining the stages of information seeking. Understanding these motivations may help determine how the categories interact. There is also a need to study the impact of technological changes on the type of categories of search exhibited by searchers.

In previous studies relating to geologists’ information seeking, barriers were a reoccurring theme. Many of these barriers were related to technology of the times or ownership of the materials. Finding out what behaviors scientists exhibit when they are unable to access materials would be worth investigating. Hallmark (1992) studied geologists’ information seeking for publications from state geological surveys. She found that “geoscientists depend almost exclusively on one another to access state survey publications” which left librarians and databases like GeoRef out of the conversation (Hallmark, 1992, p.204). Bichteler and Ward (1989) mention examples of barriers such as the lack of access to full text and the inability to find translations of foreign articles. The concept of barriers is also closely related to what Downs-Rose (2009) mentions in her work, in relation to the unsuccessful search. Case (2002) lists several myths Ellis’s model seems to accept relating to information, one such myth is that
“there is relevant information for every need” (p.8). Particularly in the sciences, one would think that there may not be an information source available that solves an information need at the level needed and therefore leads the researcher to create the new information themselves through their work.

There are also problems with answering the question, “What is information?” As discussed in previous sections, information can be many different things. In the majority of the works cited in this paper, information seeking was assumed to be oriented to literature and other written knowledge that could be found by reading or visiting the library. This orientation does not address the needs of researchers who need to find information that might be contained in physical items. This information may have already been gathered by someone else. It may also be waiting to be documented (through scientific analysis), and only the raw materials discoverable through metadata related to its acquisition are available. Sometimes this information is accessible through traditional methods like publications, but as mentioned by Joseph (2001), some fields are not particularly interested in publishing results. The data collected by petroleum geologists may remain hidden or difficult to find without knowing how that field works and who to talk to. These are patterns which have the potential to become behavioral categories that do not currently fit into Ellis’s model.

One must also consider ways of thinking about information when developing patterns and categories. Case (2002) mentions uncertainty. Ellis and his co-authors address this concept in their study of chemists with their verifying of information, but there is not further discussion of how to determine if a source is trustworthy and how to reduce doubt. Trustworthy can refer to authority of the source or authority of the contact that lead you to the source. Surely all scientists have criteria to evaluate work beyond differentiating and prioritizing sources? Hauck et al.
(2001) emphasize the way geologists think about information in a spatial sense, and Joseph (2001) discusses the way geologists use visual aids to communicate information and when searching. How do these ways of thinking about information affect geologists’ search patterns?

Järvelin and Wilson (2003) criticize Ellis’s model for being descriptive in nature, but as demonstrated by Hallmark (1992), Bichteler and Ward (1989), and Downs-Rose (2009), geologists face a number of barriers while searching and we don’t know what they do when they encounter these barriers. At this early stage, a description or understanding of which behaviors they do exhibit would be useful. Case’s (2002) concerns that there might not be a solution to every search need and his question about uncertainty are valid but may be addressed during data collection. There is much value in Ellis’s approach to information seeking and his model. In particular, his use of qualitative methods as a way of understanding users’ behavior as opposed to testing the effectiveness of existing systems. Current systems may not support searching for physical samples.
CHAPTER 3: METHODS

Qualitative research requires “an intuitive sense of what is going on with the data; trust in the self and the research process; and the ability to remain creative, flexible, and true to the data all at the same time”

Corbin & Strauss, 2008, p. 16.

Research question

This research is guided by the following question: What is the information seeking behavior of geologists when searching for physical samples? The topic was investigated using multiple data collection techniques as part of a mixed methods approach. This includes two phases. The first phase focuses on the status of state geological collections. It involves two questionnaires and analysis of documentary evidence. The questionnaires were sent to staff at state geological surveys and are used to describe the current standings of the surveys’ sample repositories. The second phase focuses on geologists’ information seeking practices. It involves interviews with geologists and analysis of additional documentary evidence. The interviews were conducted with scientists who are not affiliated with these repositories; the sampling protocol was informed by the results of the questionnaires. The interview data was analyzed using a coding process augmented with memo writing. Interpretations from the questionnaires were used in combination with the results of the interviews to develop a model of information seeking.

The understanding of information seeking in this study is influenced by the methods used by Ellis (1987) in his model development, thought it does not fully duplicate his methods. In his dissertation research, Ellis (1987) took a qualitative approach to understanding information retrieval so the data gathered would not be limited to the constraints of existing systems. Instead
of approaching the question from the view of an existing information retrieval system, Ellis (1987) started with the scientists themselves, and their activities related to information seeking in order to capture their point of view and information seeking as a whole. Ellis (1987) conducted semi-structured interviews using a grounded theory approach and eschewed quantitative methods focused on system testing or hypotheses. This study is employing similar methods in the second phase; data were collected in semi-structured interviews and a review of documentation. Ellis interviewed researchers at a university as participants for his study. Instead of focusing on a particular department or institution, this study looks at a type of individual – researchers looking for physical samples. The questionnaires in the first phase were used to understand the institutions in which these researchers might search, and also to act as recruitment sites for the interview phase.

This is not a replication, Ellis’s sampling methods are different as is the subject matter. This study looks at processes that individuals use to solve their information needs. In this work, I am investigating information seeking behavior related to data. Data have different characteristics and access points than formally published resources. These access points are currently evolving. Information resources used may include personal contacts, reviewing data citations in publications or other literature, and browsing databases. The questions used to interview the scientists are designed to illuminate scientists’ interactions with these resources. The goal of this research is to identify elements of search behavior and to develop a model of information seeking for data-related information needs.

To document search behavior from the user perspective, it is important to find the right population. This study focuses on data from the domain of geology, in particular those geologists who use physical samples in their research. Recruitment sites were restricted to state
geological surveys as they are a distinct type of facility which acts as a service and research institute within the state government system. They are similar to libraries in that function, and as such tend to have part of their mission framed in establishing and maintaining collections as well as providing access to their diverse data holdings. Focusing on geological data repositories as recruiting sites allowed relevant subdomains to emerge from the data. This also prevented assumptions that might have come from seeking out users in a general call or by targeting domains. This is supported by Ellis’s process - in his research on social scientists, Ellis (1987) specified that he did not want to do ‘stereotyping’ by looking at subdivisions. He believed that information retrieval systems do not have differences related to subdomain. With data, the issue of subdomain is now being considered in connection to metadata. This supports semantic searches based on domain differences in information retrieval systems.

Before beginning data collection, the study protocol was submitted to the University of North Carolina’s Institutional Review Board for approval. It was determined to be exempt from further review. The questionnaires, recruitment script, consent form, and interview guide will be discussed in detail below, and can be found in the appendices (E, F, G, H and I).

**Phase one: Questionnaires**

‘Descriptive survey’ – understanding existing conditions.

Little is known or formally documented about the users of physical, geological data, specifically, the user within the state geological survey network. There was a need to describe the current standing of these networks before recruiting participants for the interview phase. In order to address this deficit, a series of questionnaires were distributed to the state geological surveys as the first phase of this study. As suggested by Hank, Jordan, and Wildemuth (2009),

---

7 These subdomains were later targeted to expand data collection in order to address limitations such as researchers who were not aware of or chose not to interact with state geological surveys.
this phase of the study used an existing question set as its base. A study conducted by the USGS’s NGGDPP in 2007 was identified and used as such. The NGGDPP encouraged states taking part in their first round of funding to complete this study; 34 states participated. The questionnaire has two parts. The first section covers governance and management of the collections and was to be filled out by the state geologists. The second section includes specific questions about individual collections and was to be filled out by the collections manager. In the second section, user demographic information was collected for each type of collection housed by the survey. Before including any questions from their study, the NGGDPP program coordinator was contacted, and authorized permission to adapt it for the current study.

It was important to re-administer the questionnaire, as opposed to relying on the results from the 2007 study, for this current project. While invaluable, at the time this study began, it had been 8 years since the NGGDPP study was implemented. Technology has changed, and more importantly, the NGGDPP has made a tremendous impact on preservation and documentation of state collections. These advances would affect the information gathered on a basic level, and it is possible the resources available to enable access and discovery as well as the level of use of these collections may have changed. Finally, the results of this study were never published and are currently not completely available to the public.

The question set was not re-used but instead, it was used as a base for the current study. The scope of the initial question set was much wider than needed for this research. It also placed a significant time burden on the participants. During the first year of the NGGDPP, state geological surveys were awarded grants to conduct preservation efforts and to complete an inventory of their collections. The NGGDPP also required participants to complete a

---

8 Example of current methods for arranging a visit to state repositories can be found in appendix B.
questionnaire, as part of a study they were conducting on preservation needs. While it would be worthwhile to fully recreate this questionnaire, as my study is volunteer based I did not want to overwhelm participants or discourage participation due to the time commitment. Additionally, some of the questions in this study were related to the inventory and preservation planning process of the grant. These activities are not being duplicated, and as such, participants may not have been prepared to answer questions on these topics.⁹

**Description of changes to question set.**
The original NGGDPP questionnaire was developed using Microsoft FrontPage. While web-based, it was not created using software dedicated to this type of research. Therefore, I adjusted some of the questions based on advanced techniques available in Qualtrics. For example, I was able to create a matrix question to replace several individual multiple-choice questions.

For practical purposes, the NGGDPP question set was separated into two different questionnaires, by the role of the intended respondent (state geologist or repository manager). State geologists were asked administrative questions about their institution and repositories and asked to identify a staff member best suited to discuss operations of their physical sample repositories. The study questions were also adjusted to more clearly address multiple repositories at a single institution and multiple collections within those repositories.

After close review, some questions were combined as they captured identical or very similar information. For example, one of the original questions asked about collections being both documented and cataloged. From the responses, it was not clear participants understood a distinction between these terms. The question was changed from “What estimated percentage of

---

⁹ The results of the NGGDPP study have not been published. Initially the NGGDPP was unable to find the results or the original question set. I had a copy of the original survey (from my own participation in this project) but some questions were missing. Through networking, we were able to track down the full question set and responses.
the collection is documented (metadata) and catalogued?” to “What estimated percentage of the collection is documented (metadata, catalogs, etc.)?” for clarity. Another series of questions asked participants to “Please rate the current status of your State Survey in the following areas” without defining ‘rate’ or the expected type of response. These questions were adjusted. For example, the request to rate the phrase “Collections staffing levels are appropriate” was changed to “Are the staffing levels suitable for your collection’s needs? Yes, No, Maybe”. Finally, a few open-ended questions were changed to multiple choice or ‘select all that apply’ questions. For example, the question “What is the geographic scope of the collection?” was changed to “What is the geographic scope of the collection? Please select the area which describes your general collection or typical item in your collection”, with a list of responses to choose from. In the results from the NGGDPP study, the open responses have similar themes and indicate an open-ended question is not necessary. The revised questionnaires were reviewed by relevant community members, including representatives from the state geological survey community, members of the AASG, and other outside experts.

**Questionnaire distribution.**

As a way of establishing trust, Dillman (2011) suggests getting sponsorship for a survey by “legitimate authority” (p.15). This will help to increase response rates. As such, I engaged the AASG and the Earth Science Information Partners (ESIP) as sponsors of this project in order to encourage participation of the various state geologists. The AASG has played an active role in the NGGDPP. Dr. Jonathan Arthur, state geologist for Florida and outgoing president of the AASG, has recently advocated for the reauthorization of the NGGDPP program by Congress (H.R. 5066, 2014a). Dr. Arthur agreed to assist with my research by offering the support of AASG in recruiting and encouraging participation in this study. The question set was reviewed by the AASG’s Data Preservation Committee: Dr. John Steinmetz, State Geologist of Indiana
and the chair of AASG’s Data Preservation Committee, along with his associate Rick Hall, prior to distribution. In discussing the study and considering the population, this review reinforced the idea that the original question set might be burdensome to the expected participants. It also supported removing questions which were outside of the scope of my dissertation. While they are worthwhile and important questions, they are not critical to my work at this time. After their review, Dr. Arthur sent a brief email to the AASG community introducing this research, with links to a more detailed description online (see appendix D).

The questionnaires were then distributed by the researcher by direct email to each of the 50 state geologists or equivalent representative\(^\text{10}\) (see appendix E for both questionnaires). Each state geologist was emailed a unique link so that participation could be tracked and reminder messages send out to delinquent participants. As recommended by Hank, Jordan, & Wildemuth (2009), 1-3 weeks were allowed for data collection. This first questionnaire was used to identify which states have repositories as well as to ask for the state geologists’ support in using their facility as a recruitment location for the interview process. Of the 50 contacted, 35 states responded to the first questionnaire. Of those, 30 state geologists provided contact information for those who would answer the second questionnaire. Specifically, they were asked to identify the representative within their organization who could best answer questions such as accessing resources within their collections and visitor demographics.

The 30 state geologists identified 39 representatives for the second questionnaire\(^\text{11}\). The second questionnaire was then sent to these identified contacts, with 29 responses which represent 25 different state surveys. The gathered data was used to develop a description of the

\(^{10}\) This includes Hawaii. They do not have a state geological survey or repository. The solicitation was sent to the Commission on Water Resource Management as direct by the AASG.

\(^{11}\) 8 state geologists provided contact for 2-3 individuals.
user population, which in turn was used to develop a sampling protocol for the interview process. Information gathered includes types of visitors (academic, government, and industry), and types of use (teaching and research). These results also identified potential recruitment sites.

**Data analysis.**
The questionnaires were analyzed using descriptive statistics as well as qualitative methods. The questionnaires provide a base description for the current standing of state geological surveys. This includes how many have collections, what types of physical samples they have, how they provide access, and the types of visitors. Descriptions and percentages for types of visitors were created from the questionnaire data, as demonstrated in the example from Utah (as seen in Figure 7) and can be found in the results section.

![Figure 7. Utah Core Research Center Usage](image)

*Figure 5. Laine, Dempster, & Bowman; 2009*

**Review of existing documents.**
“Existing documents or traces of physical evidence may be a source of data about people’s information behavior” (Wildemuth, 2009, p. 158). During this phase of the study I also reviewed documents related to access to collections, materials related to the sample collections and repository management, and annual reports for various surveys represented in this study. The state surveys serve as an inroad and allows the ability to triangulate between results found in
the questionnaires and the interviews. The documentation was used as background information and to support findings from the questionnaires.

**Phase two: Interviews**

**Question development.**

In semi-structured interviews, the researcher has a topic guide with generalized questions which “give the interviewer considerable freedom to adjust the questions as the interview goes on and to probe far beyond a particular respondent’s answer to predetermined questions” (Luo & Wildemuth, 2009, p. 233). Interviews were chosen for the second phase of this study because this is a way of collecting data “led by the participant with the researcher not taking control of the process until it comes to the analysis and interpretation” and because the data desired is too complex to be answered easily in something like a questionnaire (Picard, 2013, p. 195).

In developing the interview topic guide, a list of three major topics were developed, with preferred phrasing for questions considered (Luo & Wildemuth, 2009). The topics are introductory questions, search related questions, and general questions. The introductory questions addressed the researcher’s background and their research interests relating to physical samples. This was followed by questions about their searching at repositories and finally, general questions about their search practices. The interview guide contains four types of questions as suggested by Berg (2001): “essential”, “extra” (rephrasing to come at issue from another angle), “throw-away”, and “probing” (elaborate) (p. 85-86). The questions were designed to encourage the interviewee to share details about their search process. This can be difficult as scientists do not always view their habits as being as structured as the term ‘process’ implies. The interview guide needed to include questions which were broad enough to allow for personal perception and ones which were directed to focus on search behaviors in that process.
Once I had developed my initial interview guide, I consulted the interview questions Ellis used in his dissertation (Ellis, 1993). I reviewed the connection between the questions he asked, and the categories developed, as well as the ordering and style of his questions. Then I revisited my own questions and reordered them based on flow and rephrased them to be more about behavior than a specific task. I did maintain a set of questions about task and role, however, as criticism of Ellis’s model is that he doesn’t include these two facets in his work. These two points may have an impact on behavior. Instead of removing them, I moved them to a more appropriate place in the interview guide. Finally, I added some questions from his interview guide to my own, to capture the individual’s search style in a way that is not always reflected in relation to a specific task. These additions include questions such as training an assistant, as well as closing questions about specific tools and services.

When I reordered my questions, I organized them into three sections as described above. The first establishes the participant’s research interests and broad search processes. The second has the participant walk me through a specific recent search, successful or unsuccessful. The final section was used only if the topics did not come up while discussing the participant’s search process. It includes questions about searching tools and services used by the participants, outside of those at the state geological surveys.

Before starting the data collection process, the interview guide was pretested with two experts and three sample participants (Luo & Wildemuth, 2009). This includes members of the AASG, state geological survey staff, and various geologists with whom I am familiar from my professional experience. The interview guide can be found in appendix G. This guide was used to lead conversation and ensure “consistency across different interview sessions” (Zhang & Wildemuth, 2009, p.223).
Sample.
The sampling frame for this phase of the study is broadly described as individuals who have contacted a state geological survey to investigate physical samples in their collection. This includes individuals who may have contacted the repository and determined that the repository did not have materials that would fit their need. It is specific to state geological survey repositories and does not include other geological repositories such as those held in private, academic, museum, or governmental collections (e.g., Integrated Ocean Drilling Program (IODP)\textsuperscript{12}; USGS\textsuperscript{13}, National Lacustrine Core Facility (LacCore)\textsuperscript{14}). The individual’s broader information seeking practices were also of interest and were addressed in the interviews. When sampling, any demographic characteristics such as domain, profession, years of experience, etc., were not considered, since that would assume the prevalence of such characteristics in the population. Instead, demographics of the user community will be reported among the results.

Initially I worked with managers at the repositories to recruit visitors who might fit the sample protocol. Specifically, these include individuals who have contacted the survey to inquire about samples within their collections. This includes individuals who followed up with a visit to view the collections, and those who did not. Next, communities of interest were contacted through email and word of mouth for additional participants. These communities were chosen because they included researchers who use physical geological materials in their work. This includes specific subdomains of geology (e.g., sedimentary geologists), and member organizations for individuals who use cores and cuttings in their research.

\textsuperscript{12} http://www.iodp.org/

\textsuperscript{13} https://geology.cr.usgs.gov/crc/

\textsuperscript{14} laccore.org/
The first step in locating participants was to give a recruitment message (see appendix F) to repository managers (or other designated contact) from state geological surveys and ask that they send it out to their patrons. Recruiting from repositories was critical in locating participants who fit the sample protocol. During the interview process I kept three things in mind: 1) understanding the perspectives of the participants, 2) identifying relationships between events, and 3) looking for points of tension, conflicts or data that do not fit (Janesick, 2000). I realized that this process of recruitment has the potential to exclude researchers who may use physical samples but were unaware of these types of institutions or how to access them. This also represents a type of information seeking behavior that might not have been satisfied by the resources and access points provided by the state surveys. To address this, recruitment was expanded to included individuals who elected not to interact with state geological surveys, instead opting for other repositories. Specifically, after the 13 interviews with researchers who had contacted state geological surveys were complete, the sample was broadened to include researchers who are aware of the collections at state geological surveys, but have never searched within them. Emails were sent to various user communities such as attendees of the 2016 Continental Scientific Drilling Coordination Office (CSDCO) science planning workshop and the Sedimentary Geology Community listserv. This resulted in 2 additional interviews. This expansion was needed to fill in gaps in the data which will be discussed further in the results.

**Interview procedures.**

Once a study participant had been identified, they were sent additional information on what it means to be part of the study, confirmation of qualifications, consent information (appendix I), and instructions on how to continue, should they wish to take part in the study. Interviews took place over the phone, via ‘Go-to-Meeting’ at the participant’s convenience or in person; participants were able to choose the time and location for the interview. Participants
were asked if they had any questions and to give verbal consent before the interview process began. As suggested by Lincoln and Guba (1985), at the end of the interviews, participants were given summaries in order to conduct member checking of the validity of statements.

The interviews were recorded using a digital recorder to assist with note taking and for transcription (Johnson, 2002). Recording prevented what Zhang and Wildemuth (2009) refer to as the researcher being tempted to “pad out” the interviewee’s statements with the researcher’s own thoughts. The audio recordings were sent to an online vendor Rev\textsuperscript{15} for transcription.

Data collected in the interviews were stripped of identifying content, for example, removing identifying information or specifics about industry practices where applicable. Following each interview, memos were written to assist with capturing as much data and reflection as soon as possible (Lofland et al, 2006; Zhang & Wildemuth, 2009). Data collection stopped once saturation had been achieved. Saturation was evident when participants began repeating search practices and concepts which influenced their search behavior or which confirmed earlier participants’ statements. No new information was emerging from the interviews.

**Data analysis.**

Data was reviewed to create descriptions of emerging patterns, processes, relationships and themes. This involved multiple iterations of coding and memos (reflexive writing). Analysis was influenced by Miles and Huberman (1994), who suggest the following sequence for analysis: 1) coding, 2) reflection in memos, 3) sorting and shifting materials to find relationships, patterns and themes, 4) isolating commonalities and differences before additional

\textsuperscript{15} https://www.rev.com/
data collection, 5) creating a small set of generalizations, and 6) comparing the generalizations to form constructs or theories.

**Coding process.**
Boeije (2010) defines coding as a technique “in which relevant parts of data are indicated and labeled” and explains that these codes evolve into conceptual categories (p. 9). I began by coding the data inductively. This was very productive. Ellis (1987) suggests using card sorting to help develop categories while Lofland et al. (2006) suggest diagraming as another option. These two methods were employed as a follow up to the inductive coding. They were used to ensure all considerations of codes were explored. The codes in this study were developed by emerging themes in the transcriptions and observations of potential themes from my various memos. These codes were used to develop analytic or abstract interpretations of the data. During the coding process, I was thoughtful of verifying, testing and validity (Miles & Huberman, 1994). Coding involved breaking down statements into units for analysis and identifying concepts embedded in each unit, essentially using the statements from participants as data. These codes represented an explanation for the themes emerging from the data as recommended by Wasserman, Clair and Wilson (2009).

After creating the initial set of codes, I moved them all into one document and began sorting the codes into groups based on similarities. These similarities included the searchers motivations, expected outcome, or by the types of search actions completed. I compared the groups of codes to look for overlap and potential for merging the groups into a larger ones. Once I had a final set of groups, I developed a short description of each group in a memo. In this memo I made sure to highlight the differences between similar groups and why I felt the need to create such divisions. I then looked at the quotes I had coded in the first pass to reconsider if the code and code group were still appropriate and if new codes or code groups should be developed.
Particular attention was given to frequently occurring codes, and codes which seemed particularly relevant to or telling about the search process.

Continuing the analysis of the initial sample, after grouping the codes, the interview text was reviewed to see if there were any areas that were unclear or needed to be expanded. The questions and concepts from the memos and notes created from reviewing other documentation were matched with interview memos to see if there were any major questions or concerns not addressed. For those areas, more data was gathered through additional interviews with new participants. The participants were recruited by conversations with my contacts, with the goal of finding individuals with experience in the needed areas. In addition, I considered how the behaviors exhibited relate to one another (in relation to other points raised and questions asked) and where possible connections could be made with additional data.

I found one major area where more data was needed. I needed data from individuals who considered state surveys but went to another institution. Why? What do the other institutions offer that these state surveys did not – did they not have the materials; do they not have easy access? Was it just that is not how the researcher was trained? As mentioned earlier, I expanded my sample beyond individuals who had contacted state surveys.

During the expanded data collection, again, all data (interview text, memos, etc.) were coded as they were collected. The codes created were organized into similar categories and reviewed a second time to determine if any codes could be combined to create a single higher-level concept. After a second pass, the codes were organized into the final set of categories.

**Memo writing.**

While conducting the research, I created three different types of memos to assist with my data capture and analysis. I wrote memos during the initial interview process, as I was developing codes, and ultimately as I developed categories for my behavioral model. Memos
contain the “logic of analysis” of categories and properties, and the exploration of their interrelationships (Darkenwald, 1980, p. 74). As Wasserman et al. (2009) note, memos are one place to document gaps and other concerns from initial sampling. In this case, memo writing led to the decision to expand the sample to include those individuals who opted not to utilize state geological surveys. Memo writing was also an important part of the process of coding data; Glaser and Strauss (1967) encourage the researcher to “stop coding and record a memo on your ideas” (p. 107). They believe prolonged coding can lead to conflicts and questions about resulting categories.

My memos during the interviews included reflection on a participant’s answers and my own related thoughts. I later added to the memos. In these instances, I considered not just this interview, but points that had come up in other interviews, in conversations (either about my research or on the topic), and other related ideas. As I began to code, I also wrote memos to capture my reasons for identifying a code, how it might be related to other codes and themes, and concerns I had (limitations, how it might be expanded with data collection, etc.). The final memos were created as I developed my model, identifying categories of behavior represented by the coded data.

**Further data analysis.**

In addition to coding and memo writing, I created a set of index cards, each card representing a participant. After each interview, I would create a card which included basic information about each participant (subdomain, career level, profession, etc), themes that had emerged, and unique characteristics about their search process. I would sort these cards and use them to help me consider the relations and patterns emerging from the data. I used these in combination with the codes and memos to guide additional data collection.
**Finalizing codes.**
Taking the categories created from the codes and memo writing, I gathered them with other thoughts and generalizations, and reconsidered the categories. I reviewed the definitions and decided to combine some categories. Other categories I grouped together to reconsider the boundaries between them and strengthen their definitions. Based on my considerations, I developed a final set of categories and began to consider their relationships and how they might develop into a model or theory. Some of the categories, in particular division of labor, influenced the development of the structure of the model. The final version of the model can be seen in the discussion.

**Trustworthiness of the findings.**
Trustworthiness is important in qualitative research. In both phases of this study, three methods were used to ensure the trustworthiness of the findings: member checking, peer debriefing, and triangulation.

**Member checking.**
Janesick (2000) defines member checking as allowing participants to review the materials created, such as memos and transcripts. These checks allow the researcher to address limitations in qualitative research such as researcher bias. It can also provide emphasis on results the researcher may have missed. Lincoln and Guba (1985) advise giving participants summaries at the end of the interview in order to check the validity of statements. During data collection, I would repeat back to participants summaries of their responses in order to validate my interpretation. I also analyzed the data as it was gathered, and would review my preliminary findings with new participants.

**Peer debriefing.**
Peer debriefing involves working with an impartial colleague to get “a valuable ‘second opinion’ on the meaning of data, proposed categories, and the emerging theory” (Barber &
Walczak, 2009, p. 6). Creswell and Miller (2000) state that peer debriefing is “the review of the data and research process by someone who is familiar with the research or the phenomenon being explored” (p. 129). Through participation in community events, I was able to call on outside experts to walk through my various research materials. This included repository managers of different but similar institutions, and researchers who use physical samples as a primary facet of their work but relied on collections other than those at state geological surveys. I used this opportunity to verify findings outside of the study sample, and to review search practices of participants and potential behavior categories.

**Triangulation of data.**

Triangulation is the use of multiple sources to observe the research issue (Flick, 2004). In this phase of the study, multiple sources of data were used to answer the research question. Existing documents were reviewed to complement the information gathered in the interviews and represent data created independent of the researcher (Shenton, 2013). The list of potential documents included information which would be available to and recommended or provided by the study participants, which provides the link between these materials and their information seeking behavior (Wildemuth, 2009). This includes key literature and reports relating to geoscience collections (e.g. NRC 2002 report, “Geoscience data in peril”; AASG publications\(^\text{16}\); NGGDPP technical reports; state survey’s annual reports; and other public documents) and information gathered from state survey websites. Additional materials were collected from participants – this includes resources from other organizations which were used to conduct searches for data. This documentary evidence helped with triangulation between data points, to

\(^{16}\) The AASG publications include the yearly State Geologist Journal, which includes reports from individual surveys. Each year, the state geologist’s issues a report to the American Associate of State Geologists (AASG) with the current standing of their organization which includes some limited information about the collections.
fill in topic areas which interview data could not fully cover\(^\text{17}\), and to draw stronger conclusions (Wildemuth, 2009). This triangulation was also used to verify data from the interviews (Shenton, 2013).

\(^{17}\) This varies from the approach used by Ellis (1987), who noted that he did not use triangulation as suggested by Glaser and Strauss (1967) as the data was rich enough and it was not feasible to carry out that level of work, leading him to believe this step was unnecessary.
CHAPTER 4: RESULTS

Phase one: Questionnaires
Two questionnaires were developed as part of this study. The purpose of this phase was to capture a descriptive view of the institutions in which patrons might search, and to identify institutions that might act as recruitment sites for the interview stage. The questionnaire was divided into two parts, one aimed at organization administrators, and the other at repository managers. This division was to capture a holistic view of repository management. The questions in the study are relevant to information seeking, as they describe the conditions of the collections and the access points to them that a typical user would experience. In a complex manner, these facets might shape seeking behavior – either in limitations or in creating habit/structure. The questions also address the types of visitors and uses of the collections to better understand the user population. The next two sections address the results of these two questionnaires.

Results.
In the fall of 2015, a questionnaire was sent to state geologists or equivalent representatives for all 50 states. Thirty-five states responded. Of the 35 participants, three respondents answered that they did not have sample repositories and were not eligible to complete the study. Two participants did not fully complete the questionnaire. After review, the partial responses were included in the analysis as the skipped questions were not required and did not affect their eligibility. At the end of the questionnaire, participants were asked to identify

---

18 This includes Hawaii. They do not have a state geological survey or repository.  
an individual at their institution who was knowledgeable of their sample repository (or repositories in states which have more than one), and were best suited to answer additional questions related to them.

In response to this request, 30 state geologists shared contact information for their repository managers. Identified participants were emailed a link to the second questionnaire. Of those contacted, 29 individuals responded representing 25 states (three states had two respondents). Some states have more than one collection, each with a different curator (collections might be located in different locations/cities). Of the 29 individuals who responded, one did not complete the entire questionnaire and was removed. Of the 28 individuals who completed the questions, 20 representatives agreed to assist with recruitment for interviews.

Participants of this stage of the study represented all six regional sections of the Geological Society of America (GSA). Membership sections are based on geographic and geological boundaries and therefore represent different types of sample collections. For example, members of the Cordilleran section are located along the Pacific coast and may focus on collecting data relating to earthquake activities, whereas the South-Central sections may have a focus on geological concerns relating to oil and gas research. The membership regions can be seen in Figure 8.
State Geologists.

State geologists were asked to describe their repositories from an administrative perspective. There were five categories of questions. They covered 1) institutional requirements (including legal mandates), 2) staffing, 3) budgets, 4) current and future needs for their facilities, and 5) accessibility of their collections.

At the beginning of the questionnaire, state geologists were asked if they took part in the 2007 survey by the USGS. As discussed in previous sections, the 2007 survey was used as the basis for the questions in this study. Fifteen had taken part in the 2007 study, 2 had not, and 16 were not sure (N=33). There was a high level of uncertainty among respondents. This may be due to the passage of time and potential staff changes. This question will be useful in future research.
Institutional requirements.

Participants were asked if they have institutional requirements for their collections (N=31). They were asked about three different types of requirements: legal requirements, everyday policies, and long-range plans. Legal requirements may mean the geological survey is obligated by law to maintain certain materials. These types of requirements may be found in states where the survey acts in a regulatory role or where there are economic or environmental interests related to the materials being collected. Many states require companies conducting oil and gas exploration, or drilling injection wells, to deposit a portion of resulting materials within the survey’s collections. This may include physical samples, or supplementary materials such as well logs. In addition to requiring deposition, a state may have a legal mandate to retain and provide access to the collections, as discussed in the introduction.

As shown in Table 4, while 16 states have a legal requirement for the acquisition and preservation of collections, only 11 have requirements to document these materials, and 10 to provide access to them. This is countered by the surveys maintaining their own internal guidelines for acquisition (n=21), documentation (n=22), preservation (n=21), and use (n=24). Two thirds of the state surveys have collection guidelines which supports the idea that they are information institutions, and as such are ideal locations to find information seekers for the second part of this study. Responses related to long range plans for these collections were mixed and inconsistent with legal and everyday policies. One expected outcome from this study are recommendations to these information institutions on how to better provide access to their collections. These would be long term goals. The inconsistent handling of collections may be related to the uncertainty of what is a document and what is not. It may also be related to the different levels of involvement of non-scientists in the legal guidelines. As mentioned in the
introduction, there are many variations on how state geological surveys are structured and the levels of involvement by their states legislators.

Table. 4

*Requirements, policies and plans related to collections.*

<table>
<thead>
<tr>
<th></th>
<th>Acquisition of collection materials</th>
<th>Documentation of collection materials</th>
<th>Preservation of collection materials</th>
<th>Use of collection materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Not Sure</td>
<td>Yes</td>
</tr>
<tr>
<td>Legal requirements</td>
<td>16</td>
<td>15</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Everyday policies</td>
<td>21</td>
<td>9</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Long range plans</td>
<td>14</td>
<td>16</td>
<td>1</td>
<td>16</td>
</tr>
</tbody>
</table>

*Staffing.*

Twenty-one of the 32 states have full time staff for their collections. The number of staff dedicated to the collections ranges from 0.1 to 7 employees, directing 5 hours to 1120 hours per month on the effort. Three institutions do not have any collection staff and another has only 5 staff hours per month dedicated to the collections. State surveys reported an average of 168 staff hours per month, with an average of 3.25 staff members for their collections. When asked if this staffing was suitable for their needs, 16 states said no, 14 said yes, and 2 were not sure.\(^{19}\)

Staffing issues are relevant to the research questions of this study as the interview data suggest searchers rely heavily on collection staff to assist with their interactions at the state surveys. The implications of staff and searching will be discussed in the following sections.

---

\(^{19}\) Before drawing conclusions about staffing issues, the data needs to be considered with other factors (such as size of the collections and level of need).
**Budget.**

The percentage of the institutional budget dedicated for collections ranged from zero to 20 percent (N=31). Nine states responded that they had no budget for this area; one state added that they make-up for a lack of funding with grants and other external funding. Of the 31 states, only 8 felt they had a sufficient budget dedicated towards their collections.

**Facility needs.**

State geologists were asked on a four-point scale how well the facilities that house their collections fit their institutional needs (N=31). These questions were in regard to internal needs as opposed to outward facing. Fourteen respondents said their facilities were in need of improvement (as opposed to adequate or appropriate). Only two of the 31 felt that their current and future needs were being met. Eighteen said their storage space was in need of improvement, 10 felt their current storage needs were being met, only 2 felt their facility had adequate storage space to fit future needs. No state geologist felt the documentation of their collection met their future needs. Five stated their documentation met current needs, 11 stated it was sufficient but not complete, and 16 stated it was in need of improvement. See Table 5 for a summary of the results.

Table 5.  
**Facility needs**

<table>
<thead>
<tr>
<th></th>
<th>...is in need of improvement</th>
<th>...is adequate (is sufficient but does not fill all current needs)</th>
<th>...is appropriate (is sufficient and meets current needs)</th>
<th>...meets our current and future needs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>The facility(s) for housing collections (such as warehouses or other storage locations) ...</td>
<td>14</td>
<td>8</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Storage space for collections ...</td>
<td>18</td>
<td>2</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Collection documentation ...</td>
<td>16</td>
<td>11</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>The preservation of our collections ...</td>
<td>14</td>
<td>8</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>
Repositories.
In the first questionnaire, state geologists identified individuals who maintain their physical sample collections. The titles of these individuals vary but may be divided into two different categories. Seventeen were classified as a type of scientist (e.g. Geologic Scientist, Senior Geologist for Stratigraphy and Geomorphology), and 10 were specifically related to collection management (e.g. Core Library Manager, Data Resources Section Lead, and Information Specialist). Four of the respondent’s titles were inconclusive (e.g. Manager, Director).

In the second questionnaire, these collection managers were asked six categories of questions. The first asked them to identify types of physical samples that are part of their collections. For the remaining questions, they were asked to focus only on their core, cuttings, and thin section collections. The scope of these questions included geographical coverage, changes to the collections over time, documentation, access, and use of the collection. ‘Use’ includes types of visitors and their primary purpose for accessing the collection.

Accessibility.
In the final section of the questionnaire, state geologists were asked how outsiders interacted with their sample collections. Outsiders are defined as individuals who do not work for the state survey. Outsiders were chosen as they would not have intimate knowledge of the organizations structure and information management practices. Of the 32 responses, only two states do not allow outsiders to access their collections (See Table 6). Ten states had limitations to what outsiders could access. Next state geologists were asked if their collections had an electronic database, and if electronic data was accessible to outsiders. Sixteen states have complete databases and 13 had some of their materials in databases. Only 8 states have complete access to their electronic data for outside users, 11 have some but not all, while 13 states reported
that their electronic data is not accessible to individuals outside of their organization. This may have an important impact on searching and will be highlighted in the discussion section.

In regard to non-digital materials, 11 states said these materials were accessible to outsiders, 11 had some restrictions on access, and 10 institutions reported that they did not have access to non-digital materials for outsiders. These restrictions may include legal restrictions or could be curatorial limitations. Curatorial limitations may include a lack of organization, knowledgeable staff, or even the physical process of accessing the materials. This will be discussed in later sections as again, it has important implications for the ability of researchers to independently search within the collections.

Table 6.
*Discovery and access (N=32)*

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>Some but not all</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical sample collections are accessible to outside users.</td>
<td>20</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Collections are cataloged in an electronic database.</td>
<td>16</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Electronic data are accessible to users outside of your state survey.</td>
<td>8</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Documents such as card catalogs or other non-digital materials are accessible to outside users.</td>
<td>11</td>
<td>11</td>
<td>10*</td>
</tr>
</tbody>
</table>

Note *one respondent replied ‘not sure’*

*Types of collection holdings.*

As previously discussed, physical samples in the geosciences may represent many different things ranging from cores to hand samples. Cores are subsurface materials which have very defined procedures for capture in the field using dedicated equipment, while hand samples are often loose materials taken from an outcrop. The two communicate different levels of scientific information and have different associated supplemental materials and records. Cores, cuttings (similar to cores, less reliable but also less expensive to collect), and thin sections
(derivative materials of cores) have a wider variety of uses, and will be of interest to a diverse audience of geoscientists. Table 7 demonstrates the diversity of collections held by state geological surveys. While asked to list all of their physical sample collections, the main study questions focus on repositories’ core, cutting and thin section holdings, treating them as one collection\textsuperscript{20}.

Table 7.

*Physical samples at state surveys*

<table>
<thead>
<tr>
<th>Sample Type*</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cores</td>
<td>26</td>
</tr>
<tr>
<td>Cuttings</td>
<td>28</td>
</tr>
<tr>
<td>Core Chips</td>
<td>19</td>
</tr>
<tr>
<td>Sediment samples</td>
<td>19</td>
</tr>
<tr>
<td>Hand samples</td>
<td>18</td>
</tr>
<tr>
<td>Paleontological samples</td>
<td>14</td>
</tr>
<tr>
<td>Thin sections</td>
<td>24</td>
</tr>
<tr>
<td>Other**</td>
<td>11</td>
</tr>
</tbody>
</table>

Note *Definitions of these terms with images can be found in Appendix A.

**Mostly other types of physical materials not physical samples

*Geographical coverage.*

State geological surveys have a mission to serve their state, and that may include maintaining scientific information from the surrounding area. Repository managers were asked to identify the primary geographic region covered by their collection (Table 8). Of the 28 respondents, only 6 had collections that expanded beyond the boundaries of their state.

Table 8.

*Primary geographic region*

<table>
<thead>
<tr>
<th>State only</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi country collections</td>
<td>1</td>
</tr>
<tr>
<td>Localized region of the United States</td>
<td>5</td>
</tr>
<tr>
<td>Collections divided by region or county within the state</td>
<td>3</td>
</tr>
</tbody>
</table>

\textsuperscript{20} It is possible that these collections have separate curation processes, databases, catalogs etc. At the end of the study the repository managers were asked if combining these three collections types presented a challenge to answering the questions. Eight found responded that they found it difficult.
Changes over time.
The collection managers were asked how their collection size changed over time, demonstrating growth or de-acquisition of samples (N=27). They were also asked if their preservation processes and their policies met their current and future needs given these changes in their collections. Twenty-six of the managers reported that their collections were increasing, and only one said their collection was decreasing in size. Eighteen felt their preservation processes were sufficient but did not fit their current needs, while 4 said their collection was in danger of being lost. Four said their preservation processes fit current and future needs. In regard to the development of their collection policies, 11 felt their policies were sufficient but do not fit their current needs. Four managers felt their policies needed immediate attention and revision, while 3 felt their policies met their current and future needs.

Documentation.
An important facet of searching and of collection management is the documentation that acts as surrogates for items in the collection. For cores and cuttings, this may be records detailing the geographical location where the samples were collected. Repository managers were asked how much of their collection is documented, what percentage of that documentation is accessible in an electronic database, and if that database is accessible online.

Documentation ranged from 0 to 100 percent\(^{21}\) (Figure 9). Of the 28 responses, only 4 institutions had less than 50% of their collections documented. The average for these institutions is 76% documented. Of the 27 collections which were documented, 16 had all of their documentation available in an electronic database (Figure 10). Of those institutions which have an electronic database (N=27), 15 have their catalogs available online, 12 do not external access\(^{22}\).

\(^{21}\) One institution reported no documentation of their collection. They also reported it was inaccessible to individuals outside of their organization and is at risk of being lost.
While participants provided links, this list will not be included as it would be too identifying of the institutions participating in this study.
Access.

Next the managers were asked about the access and restrictions on records for their collections, in particular, both the physical and digital information resources that would allow outsiders to search for samples in the collections. Of the responses only one respondent replied, “Information about collections is not accessible to users outside of our organization”. The others were a mixture of accessible online, accessible in person, and accessible by request (for both online and physical records). A summary can be seen in Table 9.

It is important to note that access to records online does not guarantee access to the physical documents. As expected, some institutions had restrictions on access to paper records while digital records were openly available online. However some institutions had the reverse, paper records were accessible but the digital materials were restricted. The various levels of access are listed in Table 9.
Table 9. How does your organization provide outside access to this collection? (Select all that apply)

<table>
<thead>
<tr>
<th>Institutional ID</th>
<th>An online database, catalog, or website open to the public.</th>
<th>Digital catalogs and records only accessible by internal staff. Patrons must request access</th>
<th>Physical catalogs or records, open to public</th>
<th>Physical catalogs or records, only accessible by internal staff. Patrons must request access</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>7</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>9</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>11</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>12</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>13</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>14</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>15</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>16</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>17</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>18</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>19</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>20</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>21</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>22</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>23</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>24</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>25</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>26</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>27</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>28*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17</strong></td>
<td><strong>14</strong></td>
<td><strong>10</strong></td>
<td><strong>13</strong></td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td><strong>61%</strong></td>
<td><strong>50%</strong></td>
<td><strong>38%</strong></td>
<td><strong>46%</strong></td>
</tr>
</tbody>
</table>

*Note* *One institution reported that information about their collections was not accessible to users outside of their organization.*
Visitors and primary purpose.
The repository managers were asked to identify the various uses for their collection (N=283). They were given a list and asked to check all that apply: research, teaching, reference, private sector, other (please specify). While the institutions had different combinations of use, all listed research in their responses. Twenty-two said their collections were used for teaching, 21 for reference, 19 were used for work in the private sector, and 2 selected ‘other’. When asked to clarify what private sector meant for their institution, responses included energy (1), mining (1), and oil and gas/petroleum (4) industries. The responses selecting ‘other’ gave various answers which represented state government agencies.

Respondents were presented with their responses to the previous question, and were asked to identify the primary use of their collection. 61% of the respondents replied that research was the primary use of their collection (see Figure 11). Twenty-five percent said the primary use was by those in the private sector. Figure 12 illustrates the distribution of the number of visitors received by these institutions each year.

---

23 One institution reported “Not sufficiently organized or accessible for any use.”
Summary.
The two questionnaires provide insight into state geological surveys and their positions as science data centers. The conditions at these institutions vary across many dimensions, including size/scope, staffing, policies, and accessibility. These variations are not so great as to make them dissimilar. They serve similar communities, purposes, and share challenges. Regardless of policies and guidelines, there is a need for more consistent methods of access to records and documentation. And with most institutions, there is a need for additional resources, particularly staff. This need is reinforced by the data collected from patron interviews, where they discussed encountering barriers in their search process and relying heavily on institutional staff to locate samples. Recommendations to state surveys based on interview data can be found in the discussion chapter.

Phase two: Interviews
The first phase of this study was aimed at understanding the science data centers which house physical sample collections as represented by state geological surveys. In particular, this study looked to understand how these institutions provide access to these collections and their
limitations. The second phase, and main focus of this study, is to understand the search behaviors of the users of these collections given the unique nature of physical samples objects as ‘information sources’. With the help of the repository managers, patrons of the state geological surveys were recruited for interviews. In order to address gaps in the data, and to verify results, the recruitment process was expanded to include communities known to use physical sample material in their research. From these recruitment processes, 15 interviews were conducted. They represent a diversity of domains, career status, and experiences. The results of the interviews were coded for themes and categories of behavior were developed.

**Participants.**

During the interviews, participants were asked to provide background information. This included their primary research area, position in their organization, career level, and which state surveys and other similar repositories they have visited. Table 10 contains a summary of this information; however the names of the surveys visited have been excluded to maintain anonymity of the participants. Of the participants, 11 were faculty members of various ranks, 2 were graduate students, and 2 were from industry. Six of the researchers were in the early stages of their careers, 6 were mid-career, and 3 identified as late career. Gender identities have been removed from the discussion of participants to ensure confidentiality.
Table 10. Description of participants

<table>
<thead>
<tr>
<th>Career Stage</th>
<th>Participant ID</th>
<th>Domain/research area</th>
<th>Profession</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Career</td>
<td>LN3</td>
<td>Paleogeography/paleoclimatology</td>
<td>Assistant professor</td>
</tr>
<tr>
<td></td>
<td>PP9</td>
<td>Quaternary geology</td>
<td>Assistant professor</td>
</tr>
<tr>
<td></td>
<td>PH12</td>
<td>Tephrochronology</td>
<td>Faculty</td>
</tr>
<tr>
<td></td>
<td>GA2</td>
<td>Geology - geothermal energy</td>
<td>Doctoral Student</td>
</tr>
<tr>
<td></td>
<td>WA3</td>
<td>Geology</td>
<td>Graduate Student</td>
</tr>
<tr>
<td></td>
<td>SK64</td>
<td>Petroleum exploration</td>
<td>Industry</td>
</tr>
<tr>
<td>Mid Career</td>
<td>SA55</td>
<td>Climate change/atmosphere</td>
<td>Assistant professor</td>
</tr>
<tr>
<td></td>
<td>RF13</td>
<td>Paleomagnetism</td>
<td>Associate professor</td>
</tr>
<tr>
<td></td>
<td>G4</td>
<td>Sedimentary Geologist</td>
<td>Associate professor</td>
</tr>
<tr>
<td></td>
<td>SG4</td>
<td>Sedimentary Geologist</td>
<td>Professor</td>
</tr>
<tr>
<td></td>
<td>ZD7</td>
<td>Pre-Cambrian Economic Geologist</td>
<td>Professor</td>
</tr>
<tr>
<td></td>
<td>FC71</td>
<td>Thermochronology</td>
<td>Consultant</td>
</tr>
<tr>
<td>Late career</td>
<td>F45</td>
<td>Sedimentologist</td>
<td>Associate professor</td>
</tr>
<tr>
<td></td>
<td>FF0</td>
<td>Paleoclimatology</td>
<td>Professor</td>
</tr>
<tr>
<td></td>
<td>JK8</td>
<td>Paleolimnology</td>
<td>Professor</td>
</tr>
</tbody>
</table>

Participants visited a range of survey repositories, representing all 6 membership regions as defined by the GSA\textsuperscript{24} as seen in Figure 8.

One late career researcher had visited 7 state surveys, the USGS’s core repository, and a number of international geological surveys. Not all participants had visited a state survey’s repository, though all were aware of and had experiences with their collections. One individual had contacted a repository but did not find the results warranted a visit at this time (long term search agenda). Two researchers’ geographical research areas were in regions covered by state geological surveys, but they chose to rely on other repositories for their research needs (in their words, the other repositories provided better databases and access to supplemental data). Finally, one researcher from industry relied on someone else to make contact with the repositories.

While finding surveys repositories to be a valuable source of research data, they had not

\textsuperscript{24} The Rocky Mountain section was represented by a participant who has accessed a survey of a Canadian province. This section also includes state surveys which were identified but not utilized by two other participants. These participants chose to rely on alternative institutions outside of the state government spectrum instead.
contacted any state surveys themselves. These variations address important needs and behavior
types that round out the search behaviors of geoscientists and will be discussed below.

During the interviews, participants were asked to discuss their research needs related to
physical samples and to provide an example of a recent search. These conversations were
analyzed in order to identify types of search behaviors. In the section below, I will first discuss
the general concepts of information seeking demonstrated by the participants, then review
specific behavior types exhibited with illustrative quotes. This will be followed by a narrative
example to show how these behavior types might flow during typical research. The discussion
chapter will address how these behavior types fit together into a model of information seeking
and their implications.

Results.
General concepts of information seeking.
The information seeking process exhibited by participants is non-linear. There are
examples where it seems to flow from ‘starting’ to ‘stopping’, but there are exceptions and
variations depending on the search purpose. An individual might have a research question in
mind, or might be looking to see what areas might inspire a research program. In regard to a
research program, new faculty might be looking to become their domain expert for a region, (a
side effect may be becoming the information gatekeeper for this topic/area if there is not much
work done by any domain). For example, junior faculty talked about being the only expert in
their domain for a given geographical region (ZD7). This was often connected to browsing or
keeping current on possible information sources without following up to determine if they were
of current value. This will give them a research platform to work on for years – both their own
research and projects for students.
Starting often began with a research question or a geographical region of interest. The researcher might have a question that can be answered with physical samples, or they might be looking to see what samples are available in order to develop a research objective. For example, one of the scientists uses specialized equipment in their lab to analyze samples that they or others had collected (GA2). When searching, this participant looks for existing materials from researchers in other domains. Given their domain differences, these researchers likely do not have access to or training with the participant’s equipment. The other researchers provide the participant access to their samples, and in exchange, they provided these researchers with the results of the equipment’s analysis. This is one example of negotiating access to data exhibited by participants.

Starting may begin at a geological survey. This might mean looking at the repository website and other online resources, or contacting the manager/curator to see what they have in their collection. A researcher might have a geographical region in mind or perhaps a certain rock formation25. Participants felt many current systems varied greatly in their content and search support, which highlighted a variety of limitations. Understanding how to search in each system was a skill one had to learn (ZD7). Some institutions’ databases are lacking searching features and metadata to assist with discovery. This leads to a prioritization of facilities (dependent on time and search specifications). It also meant searchers had to learn different search methods for each institution and could not search across institutions’ collections. Some participants who chose not to visit state geological surveys explained they could find the samples they needed more easily in other places because of better databases. This includes the USGS’s National

---

25 See Appendix A for definition of formation.
Digital Catalog and databases such as iDigBio which include metadata for many different collections.

Databases and other catalogs were not limited just by search mechanics, but also by the variety and depth of the metadata they contained. Often institutions had additional (supplemental) information about their physical samples (e.g. white papers, lithological descriptions, publications, maps, non-conventional materials such as guidebooks, well logs, geophysical logs, chemical analysis, and other data) but those that afforded better searching had systems that connected these information sources. When absent from the system, searchers used the curators’ expert knowledge. Two participants praised state surveys which allowed them to download KLM files (files used in mapping and visualization) of the survey’s collection. This allowed them to perform customized searches themselves.

A researcher might also start by contacting their personal network (a step that gets repeated as more information is needed). This was a limitation for early career researchers. As one mid-career researcher and educator pointed out, their students do not have the extensive social network they themselves do and, as such, may not know who to contact to find information. It is important for a researcher to be aware of who else is working in their area (domain and geographical region) of interest. This might include researchers from academia, state and federal government, and industry. These individuals can provide information on locating samples, and sometimes may provide access to samples that they manage or own.

Personal or social network was different than personal contact. For example, many participants mentioned a lack of detailed citation information in publications that would allow them to locate the repository where samples referenced are archived. Participants might follow up with an email or phone call to the publication’s authors. A similar tactic was taken when
online searches of state repository holdings were inconclusive. The next step would be to call or email the curator or manager of the collection. A step further might be to arrange a visit to search in person.

Participants were asked about training a research assistant or student in how to search. For students, learning these methods began at the graduate level. As one educator noted, the skills of searching for cores is very similar to searching for a book in a library (LN3). Another participant reflected on how searching for cores was similar to training that currently exists in the curriculum. It’s like “learning how to keep a lab notebook and all that, and learning how to use a library. This is just like a library for geology” (G4). There was some question about whether teaching these skills should be moved earlier in the curriculum (i.e., the undergraduate level). As pointed out by a number of participants, this has its own challenges given the vast amount of information currently required for the undergraduate degree. “Because you have such a limited time window with them, and the skill set of the student can be so variable that you usually end up trying to get the materials ready, and have them ready to go, and so I don't usually have the undergrads looking around for research sites or core materials or anything like that” (G4).

Learning to search for physical samples was often treated as an apprenticeship. Students learn by working with their advisor to locate samples for research. They might be given a task such as evaluating samples already identified or given the contact information for the repository curator and told to find samples.

This brings up the concept of ‘division of labor’. For example, an individual from industry said that ‘division of labor’ meant they hand off their search needs to another staff member in order to locate relevant samples. This is also encountered in the education process
where faculty members give students samples or show them where to find samples without the student having to go through the search process themselves.

Searching did not stop with the discovery of cores or cuttings. Once potential materials are located, they must be evaluated for use. This might include checking the quality and accuracy of the metadata associated with it that either led to discovery or might be used in analysis. It might also have led to the need for more samples to verify the information found. Searches often ended for a number of reasons. The researcher might have felt they had enough samples gathered to analyze and answer their research questions. Alternatively, they might have a time constraint. If no viable samples have been found after a certain level of searching, they might stop and evaluate their research objectives. There are many factors to consider.

**Behavior types.**

The introduction above provides a sampling of the various search behaviors demonstrated by participants and how they might relate to one another. Below is a complete list of the behaviors identified. Each is discussed in detail along with specific examples from the interviews that illustrate the behavior in context. These behaviors were analyzed for commonalities and categories of search behaviors were developed. These categories were used to develop a model of information seeking, which will be addressed in the discussion chapter.

**Starting.**

For the participants, searching for physical samples is motivated by the desire to conduct research. For some, it starts with wanting to answer a specific research question (F45, RF13) or interest in a geographical region (LN3). Some researchers might not have a specific question in mind (GA2). There is often an initial step that starts the process. This might be searching online (SG4, WA3, RF13), reviewing the literature (RF13), emailing the state geological survey in the region (G4), or looking for personal contacts who might have samples available for use (F45,
Staring a search is also influenced by the type of samples a researcher is searching for. As one participant noted, outcrop samples answer a different question than cores; in particular, outcrop samples typically have signs of weathering whereas cores are “fresh” (LN3). Starting is not always a formal process, but more of a number of possible first steps.

Maintaining awareness.
Part of the search process is an everyday activity that involves keeping current with changing information in the field. Several participants mentioned the value of this activity, referring to it as maintaining awareness or keeping an eye out for new information that might change or impact their current search or research goals. Two participants noted that part of keeping current was knowing who else works in your geographical area or domain (ZD7, WA3). Ways of keeping current included being familiar with current publications (FC71, G4, PH12), attending national meetings (G4, ZD7), being aware of repositories that maintain relevant collections (ZD7, G4, PH12), as well as keeping up with news releases, current updates to databases, recent conference papers, and reports from oil and gas companies (PH12, G4, FC71).

Browsing.
Some searchers brought up the concept of browsing as having a different meaning than the concept of maintaining awareness/keeping current. It might mean an unmotivated search through current conference abstracts to look for items of potential interest (PP9), or asking about a broader, unrelated topic while conducting a search at a survey’s repository (PH12). As one participant mentioned, shelf browsing is part of what they do as geologists (RF13). A final variation might be considered happenstance: while looking at various resources the researcher might discover something else of value not relevant to their current task (PP9).

Division of labor.
Division of labor is about separating specialized tasks and saving time. As both students in the study said, they were on a time constraint with finishing their degree (WA3, PH12). One
said that they would have disregarded a particular portion of their study plan if samples had not been readily made available to them by their advisor and their advisor’s connections (PH12). This indicates searching requires a significant time cost and would take away from learning and/or completing their degree. A participant from industry (SK64) said they did not actually conduct searches for cores. Instead they hand the task over to others within their organization. One participant familiar with industry said that the individuals conducting the searches for the industry researchers are not considered ‘scientists’ (JK8). The scientist’s time is better spent on specialized tasks such as evaluating the samples.

**Social network.**
Participants often referred to their social networks as valuable resources. As mentioned in the discussion on maintaining awareness, it is important for these researchers to know who else is working in their sub domain or geographic area of interest. This can be leveraged for many different aspects of information seeking. The social network is a tool for locating cores, for tracking down missing information, for validating information already gathered, and for solving information problems that are not currently supported by literature or online tools (SG4, GA2, PH12, PP9, LN3).

Use of a social network can be intentional – a participant is starting a search and their first thought might be to reach out to the people they know. When asked ‘how do you find out about cores’, one participant responded, “it’s a classic story of people I know, buddies from grad school. … another set of cores I’m working on… was collected by my PhD advisor” (G4). Using one’s personal network doesn’t have to be in the form of a directed inquiry. As the same participant noted, “I had done some pilot work on outcrops, just buzzing around and sampling some outcrops in the region, and then I presented that work to the geologists at the survey, and
they said ‘Here, use this core. That’s much better.’ They’re right because then I don’t have to [drive around and look at outcrops]” (G4).

One participant relied on their social network to get resources otherwise not accessible (GA2). In their subdomain, researchers often have their own personal database of known samples and data. This researcher was given a database by their advisor and had also gathered others from the labs they had worked in as an early career researcher. This gave the participant a good starting off point and was an advantage over others who might not have had such connections.

One participant shared an example of how they used their social network to find a core they couldn’t have collected themselves (LN3). “I heard through the grapevine … a natural gas company had drilled a core and they messed up … they actually missed what they were trying to get. So they had this core that they weren't going to do anything with”. When asked how they got access to it, the participant leaned on their extended network. “So we were doing work at the Kentucky state survey and they mentioned… they didn't have [relevant cores] there. They knew that a few companies were drilling holes actively right now through that set of rocks”. The researcher had a student interning at the company mentioned. They contacted the student and asked if they could find out what the company had and who to contact next. “I emailed their head of geology and she was really supportive, actually, and made all the arrangements [to give us access to the cores]” (LN3).

Additionally participants mentioned using listservs to find information. Appendix J has two examples of these types of requests. In the first, the researcher has a sample that they need more information about. In the second, the researcher is looking for samples that fit a set of criteria and is unsure where to look for resources beyond the ones they have already tried.
Visiting/Personal contact with the repository.

As part of locating samples and verifying scientific value during a search, participants discussed the need to visit the repository holding the cores or the need to make personal contact with the curator/repository manager. One researcher was expecting to find more information in person, as there was not enough online (WA3); another wanted to be there in person in order to browse (RF13). Many researchers wanted to talk to the curator (PH12, ZD7, WA3, PP9, G4, RF13, FC71) to assist with their search. The curator is “critical to searching” (PH12). Curators provide context to the samples (G4) and were able to assist when online catalogs or other portals failed the searcher (PP9). One participant said the core manager was a better retrieval system than a database as they could help with things that don’t show up in inventories (SG4). The curator was mentioned as the researcher’s “primary point of contact” (ZD7) and the individual who can authorize the researcher “to get in the room with the samples”. As one participant noted, “That would be a classic case of, I'd have to email the authors, and see if they know where the cores are, and if there's any willingness to share material” (G4).

Physical samples are information resources. While metadata might lead one to find a valued object, it cannot tell the researcher the complete range of information the object represents. As such it may be necessary for the researcher to inspect the samples in person before determining if they satisfy the search need (WA3, ZD7). As one participant noted, it is better scientifically to visit the samples as there is less risk of them being damaged in transit (LN3).

Chasing.

Chaining involves following a reference in order to find additional information. In traditional views, it is in regard to publications, following a citation in one article which leads to a source or expanding of information in another article or book. It is supported by a structure of
citations and publications. Chaining is a way of finding related materials and broadening the search to include more information beyond the view presented where it was cited. In the interviews, participants talked about a behavior similar to chaining, but it did not always follow traditional pathways, and in some cases, it cannot given the limitations of the way we cite data and samples in research. As such, this behavior is identified as chasing, a more loosely structured version of chaining (as discussed in the literature review). When looking for physical samples, chasing might involve reading a report and contacting the author (G4, WA3, ZD7, JK8), or asking the curator of a collection for help finding cores similar to the ones they have already found (PH12, FC71). In some cases they are chaining off of nontraditional sources such as reports, records, core descriptions, the name of a gas field, etc. (PH12, WA3).

For example, one participant mentioned an experience where they were looking at a specific core at a repository (FC71). The curator noted they were interested in certain formation. Based on that information, the curator suggested another core that was from the same rock unit but located 4 miles away. It was done on the fly, and it took the expertise of the curator to be able connect the two cores. The participant had not seen it in their online search. This might be considered using one’s social network or personal contact to find information, but it also relied on the curator connecting information from one core to another.

*Negotiating.*

The concept of negotiating came up many times by participants during their search process. This might have to do with gaining access to samples or finding the appropriate samples. Three researchers mentioned giving authorship for access to needed samples (RF13, F45, LN3). This may be a demonstration of respect for the original effort of collecting the samples or the act of trading authorship for information. Another participant mentioned trading
access to samples for the materials they generated like photographs of the cores or lithological logs (ZD7).

Other participants mentioned negotiating as a means of working around restrictions such as proprietary information (FC71, LN3). This might mean negotiating what you can do with the samples like ‘destructive sampling’- the process of collecting samples to be used in laboratory analysis which results in the destruction of the samples (F45, LN3). As one participant mentioned negotiating as a way of trading information about samples with researcher from another subdomain, “the other advantage [to negotiating or collaborating] is the geologists are there with the interpreted cores and they are happy to help me navigate my way through the very complex geology so I can provide that geochemistry and meteorology data to them” (G4). It should be noted that negotiating is different but related to the category ‘limitations’, discussed in the next section.

Validating/verifying/Evaluating quality.

As one participant stated, a primary step in searching is evaluating the trustworthiness of the data you find, checking to see if it is reliable (GA2). This could mean reviewing the supplemental data associated with the samples, which might have errors or be mis-identified (F45, RF13); verifying the source of the records, if created by a student worker or non-geologist one might question completeness of the work (SG4); or checking the chemicals in the samples to see if they may have degraded from when first published (ZD7). This is in addition to evaluating the data for fit with your work (GA2, LN3, F45) or reviewing the data in order to correlate with other materials collected (WA3) or even using the data to calibrate a model (RF13). The available supplemental information also affects one’s decision to follow up from an online

26 Geochemistry and meteorology are processes that involve lab analysis on samples.
search to look at samples in person (GA2) and may require talking to someone from the repository on the phone (FC71).

State geological surveys are science data centers which ingest materials from many different sources. As discussed in the literature review, this includes required deposition and donation. One participant spoke of their own experience of creating photographs and logs for cores in a survey and giving the materials back for others to use (LN3, also see F45, SG4). This is a reason to evaluate the quality and completeness of the data before trusting it in one’s own work – as another participant noted, there might be multiple sources of data for one core and there is a lack of standards for most data capture (F45). Standards affect searching and quality and will be addressed in the discussion section.

**Stopping.**
When searching, not all avenues may be exhausted before the researcher decides to stop their search. For example, one participant was collecting data for their dissertation (PH12). The purpose was to complete the project as well as possible, and there were limits on the amount of time that should be spent on information seeking. This is different from running out of resources, where a decision had to be made about saturation of information that fit the research needs. More funds or time could have been found, potentially, if they were needed to ensure the quality of the research. The participant noted that they found some information was missing. The participant did not investigate the missing information further at this time, but said they may return to the search later as a potentially new research endeavor (PH12). This implies multiple levels of searching, immediate need, and potential future use. Another participant echoed this idea, saying that they might have found adequate information for the current task, but they would keep looking for additional information (SG4). In this second case, the participant didn’t find their desired core for 20 years. “Back in the 90s, when I was working on my dissertation, I felt
like I did a really great job looking for the core but I never found that core. I [did] find it 20 years later” (SG4). As another participant described, you don’t wait to decide to keep going or not until all the data is collected, viewing it as an iterative process (GA2). As the work goes on, one is constantly reassessing the situation. This is in contrast to another participant, who said you decide if it is enough research when you have the core you are looking for, or when you determine that it no longer exists (either destroyed or discarded) (LN3).

**Other considerations during searching.**

*Limitations.*

There were a number of limitations that prevented participants from conducting their ideal search and which impacted their search behavior, such as lack of time (G4, RF13, JK8). One might race to find information within a time frame, ignoring harder-to-track data (FC71). If you were not sure of the potential value, you may not follow up on leads due to time (GA2). Students only have so much time to complete their degree (LN3). Funding was also a limitation (G4). One participant, a contractor, was only given so much funding for their work (F45). One of the surveys they visited charged for collecting a sample from their cores, so the researcher had to determine how many different samples they needed to answer his research question and stay within costs. While this was a limitation on their ability to access samples, the researcher did not feel like their work was diminished in quality because of it.

Barriers came in different forms. As one participant noted, they were unable to access the materials they needed due to a poor relationship history between their organization and the state survey (F45). Another researcher said they were a low priority to the person who controlled access to the samples (G4). Alternatively, information providers in other domains, for example industry, might intentionally withhold materials from others (F45). This might be for
their own proprietary research goals and not because of a legal restriction (PH12, FC71). Private industries do not want their competitors to know about their data (FC71).

The biggest hurdle was tracking down samples (LN3). There might be so little information that you don’t know what you are working with (FC71) or the information is not clear in regard to background or source of the samples (WA3, G4). Supplemental information sources about samples might not actually be connected to the samples/sample databases (WA3, GA2). The digitization might not be high enough quality to evaluate the samples (ZD7). The information, “it's just a big digital pile, so you just kinda just search it, and hope for the best” (ZD7) or it might be fragmented (WA3). The information about the collections might not be organized or searchable (PP9). The current publication model does not support locating samples in a repository (FC71). The information might change over time; for example, formation names and definitions might change as new knowledge is discovered (G4).

Sometimes samples were missing, either discarded due to space limitations (PP9), or the curators were not sure where they were in the warehouse (PH12). The samples might not be stored properly and will have weathered or degraded (RF13). Repositories might have narrowly focused collections (PP9) and their databases will be filtered by their own interests (GA2) and as such likely do not have what the researcher is interested in (FF0). One research employs destructive sample (destroying the samples to run chemical analysis) in their research, and not all facilities allow this (LN3).

Distance might also be a reason to not follow up on results: the samples must be viewed in person and the warehouse might be too far (PH12, ZD7). As mentioned earlier in visiting, digital surrogates such as photographs, CT scans, etc. do not replace the physical object and one still needs to visit in person to evaluate the samples for use (LN3). This requires a time
commitment for both traveling to visit the cores and laying them out (G4). One researcher would need a very large grant in order to travel to South America or Australia to view cores for their research. Finally, one participant mentioned that new researchers might be afraid of calling a state survey or they might not know that they can call to get help finding samples (P99). This view was supported by another participant who felt most geologists were not aware of the resources at state geological surveys (FC71).

Limitations were not always viewed as a negative. One participant said limited information meant the topic area or samples were not heavily studied (WA3). One researcher’s personal limitations meant they only worked with previously collected samples and might lead to partnerships with other researchers to do analysis (G4).

**Differences from other organizations.**
During the interview process, participants discussed various intuitions and the differences from state geological surveys. In some examples, the other organizations provided more information – either more detailed records (depth intervals), associated information (photos, associated data), or better databases/user interfaces (F45, GA2, ZD7, LN3). This includes industry institutions, the USGS, universities, core labs, environmental organizations, and programs such as IODP (SG4, FC71, LN3, GA2, F45, ZD7, WA3). The mission or purpose of these institutions leads to different priorities on access. Beyond variation in institution type, one participant mentioned that international organizations have more resources (specifically staff) dedicated to solving these issues (SG4).

**Education/future needs.**
Participants were asked about education and training as a way of understanding their search process but some of their responses illuminate an existing need. Many were not formally trained in searching or in regard to resources that support searching (LN3). Education on
searching does not begin until graduate school (G4), undergraduates are given samples and not trained in how to find them themselves (ZD7). Others were given samples at the masters’ level with no training at a full level (WA3). They might get training by being sent to core repositories to work (G4) or learn through doing – shown the resources and left to explore (RF13).

A number of participants mentioned the future, either tools that would help searching or changes in the community that would be necessary. How do we make gathered knowledge more discoverable (GA2)? There is a need for policies to support giving back information (ZD7) as there is useful information not being collected (GA2, FC71). There is a need for support in the development of sample registries (WA3). The location represented by samples is important to answering a researcher’s questions; one participant said this was the most important facet (ZD7). It would be useful to be able to search spatially, seeing dots on a map representing sample collection points (PP9). Participants wanted to see databases that allowed them to search, like Google Earth (ZD7). They would also like to see more metadata (like formation level) and the ability to search by secondary information sources (G4, F45). Supplemental materials may be created by the users, but are often not integrated into catalogs afterwards in order to make them accessible to future researchers (F45). These concepts of education and future needs will be addressed in the discussion section.

**Example of search behaviors.**

Below is an example of the search process from one of the researchers in this study, researcher GA. Researcher GA demonstrates a number of search behaviors: starting, social network, personal contact, browsing, visiting, and maintaining awareness as well as a number of limitations and future needs. These behaviors may repeat as needed and do not follow a linear pattern. Their search is affected by their particular research area/domain which has specific
requirements. Here the participant’s research area is described followed by an example narrative taken from their interview.

Researcher GA is an early career scientist. They are faculty at a small teaching university. They oversee undergraduate students and have a research lab. Researcher GA studies tephrochronology\textsuperscript{27}, which involves analyzing physical geological data. Researcher GA has contacted a few state geological surveys, but never actually visited one in person. One of the challenges in tephrochronology is limited or poor-quality databases, and lack of training in their use (Lowe, 2011). Databases in tephrochronology can act as reference collections. They may be used for correlation and also allow access to and re-analysis of existing samples (Lowe, 2011). Evidence of this need can be seen in the example below.

\textit{Search process.}

Researcher GA is looking for places that might have samples related to their particular research interest of tephrochronology. Specifically, researcher GA is looking for tephra deposits which are evidence of an event and are determined by the chemical traces left behind in samples. Their first step is to look for people conducting research in the geographical area where such an event is known to have happened. Researcher GA finds individuals conducting research in that area who are from a different domain. In talking to these researchers, they find out that there were cores drilled in that geographical area, and where the cores are stored. Next, they contact the repository which houses these cores to see if they can access them. In addition to the specific

\textsuperscript{27}“Tephrochronology is a unique stratigraphic method for linking, dating, and synchronizing geological, palaeoenvironmental, or archaeological sequences or events. … Tephrochronology in practice requires tephra deposits to be characterized (or ‘fingerprinted’) using physical properties evident in the field together with those obtained from laboratory analyses. Such analyses include mineralogical examination (petrography) or geochemical analysis of glass shards or crystals using an electron microprobe or other analytical tools” (Lowe, 2011, p.107).
cores identified by the other researchers, researcher GA plans to “shelf browse” at the repository when they visit, to see what else can be discovered (GA2 15:30).

The repository has multiple storage locations. Before visiting, researcher GA emails to find out if the cores are available\textsuperscript{28} and what the repository knows about them. If the visit to the repository does not identify any cores useful for their work, researcher GA would do comparisons on published data. Researcher GA can use the new information they have gathered so far in their search to help. Researcher GA considers a number of questions -- Where else might the event of interest be reported? Do the authors have samples available that might be referenced? If any citations are discovered, their next step is to contact the advisor on the project. Sometimes the samples used for testing have been discarded, but the original core that was sampled is still there. Now researcher GA must consider the time and resources needed to go visit the new repository to look at the core. It may involve travel across country, and there will be some limitations on researcher GA’s available funding and schedule (GA2 20:00). They may have to wait for the right time to arise, like a conference in the region. There might be restrictions on the grant funding of their project (lack of skilled workers to send as researcher GA only has undergraduate students; the need to meet publishing deadlines). That is ok, because there is long term value in the materials, and their searching does not always have an immediate need (GA2 22:30).

Overall, researcher GA finds searching is hindered because there are no universal databases which focus on tephras and the type of event they study. There are small regional or project-based databases, but not any that are comprehensive for North America. In order to address this, researcher GA’s PhD advisor kept their own records (a spreadsheet) of known

\textsuperscript{28} This may refer to many things: cores not embargoed, not out on loan to another facility or researcher, the ability of the repository to locate and provide access to the cores, etc.
publications on this event type. The material gathered is filtered by the advisors own personal interests and the list is not easily queried (GA2 24:30). Another way of working around this lack of a global database is to leverage one’s contacts. If a researcher in GA’s domain comes across samples that they do not know much about, they will email other researchers in their domain and ask if they have any information about that data (“shopping around your unknowns to your colleagues”) (GA2 26:00). Ultimately, researcher GA asks, “Where do we put the data? How do we make discoverability [an option]?” (GA2 27:30). Discoverability should allow searchers to find cores collected by researchers in a different domain, and facilitate reusing cores for a different purpose than that for which they were originally collected.

**Summary.**
Participants’ responses suggest that searching is an important part of conducting research with physical samples. During their interviews, participants demonstrated a wide range of search behaviors. Some behaviors were dependent on one’s role within an organization, suggesting a division of labor in the research team. Many behaviors related to a researcher’s knowledge of the domain – knowing where to look, who to talk to, and determining the quality of the information found. Searchers encountered a number of limitations including difficulty in gaining access to materials as well as the variations in databases and information resources at the various state geological surveys.

**Comparisons across research stages**
Results from the questionnaires (phase 1) and the interviews (phase 2) were compared in order to better understand how the two sets of data relate. Specifically, the results were compared to see if the communities served by the state geological surveys were represented by the participants in the interview stage. This included looking at the primary use of the state geological survey collections, and the representation of various geographical regions by both
communities. While geographical regions do not guarantee match in geological interests, as shown in the results from repository managers, the collections often have a state or regional theme. Interview participants also relied on the geographical coverage of a collection as an important criterion in selecting data providers.

**Use of collections.**

Data collected in the first phase of this study included demographic information about the user population at state geological surveys. The makeup of the interview participants represents these characteristics. In their questionnaire responses, state geological survey managers reported 4 categories of use for their collections: research, teaching, reference, and private sector. From these categories, managers reported the primary use of their collections was for research (61%) with the private sector (25%) next highest (see Figure 11). In the interview stage of this study, the majority of the 15 participants use the collections for research (n=13), with two participants representing the private sector (see Table 10). Among those representing research, many also used the collection for teaching (LN3, SG4, GA2, G4). For example, participant G4 uses his research with geological cores to teach undergraduate students. The students assist with lithological descriptions, sample collection, and other interactions at state geological surveys. The final category, reference, was mentioned by participants who used samples to validate other research (LN3, WA3).

In addition to these demographic questions, repository managers were asked about frequency of visitors to their institutions. This information was not used to influence the number of participants interviewed in the second phase. The method used in this study did not dictate a specific sample size; instead it required saturation of information gathered\(^{29}\). The number of

---

\(^{29}\) In some research, one might wish to capture a particular sample size for study participants in order to fully represent the population being studied. However, the large variance in reported visitors to the state geological
visitors per year reported by the state surveys varied from no visitors (n=1) to more than 10,000 (n=1). The majority of institutions (n=11) reported an average of 5-40 visitors per year (see Figure 12). There are many variables that affect the number of visitors to these institutions which may include diversity of their collections and the services they provide. These attributes will be discussed in relation to the data collected and may suggest further research is needed.

**Regions represented.**
Not all state surveys mentioned by participants took part in both questionnaires. Approximately half of surveys mentioned are represented. However, like the interview participants, the participating state surveys represent all of the major GSA regions (see Figure 8).

Direct comparison of statements from interview participants and the data gathered from repository managers at the overlapping institutions will be considered in the discussion section.

**Analysis of other materials**
In order to conduct triangulation with the data describing the current standing of state geological surveys and the search behaviors of their users, supplemental data was collected relating to searching at various data centers. This includes differences in search offerings of other databases (e.g., Integrated Digitized Biocollections (iDigBio))\(^{30}\), a report from a state survey as to how they plan to address these issues in the future, observations at a science planning workshop for scientists interested in continental and lacustrine cores, and a review of services provided by other science data repositories.

In reviewing these various resources, they each reinforce and verify the findings of this research. iDigBio provides services participants thought might be valuable in the future for state surveys did not allow for an accurate estimate of the population size. This variance and uncertainty supports the decision to use a method which relied on data saturation rather than sample size.

\(^{30}\) https://www.idigbio.org/
geological surveys’ collection catalogs. One of the state surveys provided a ‘data at risk’ report, a self-assessment of the condition of their collections. The report suggests actions which will better support search and discovery which align with participants’ needs. The community discussions observed at the CSDCO science planning workshop confirm the practices and support needs described by interview participants. And finally, LacCore, the National Lacustrine Core Facility associated with the CSDCO, provides services they see are valuable to their users, which match with the behaviors exhibited by participants in the interview stage.

**Databases.**

During the interview stage, participants shared various resources they use in their searches for research data. One participant highlighted iDigBio as a tool that provided exemplary search tools and contained a rich amount of information which allowed them to better discover samples (FF0). This resource was reviewed as part of the triangulation process.

iDigBio “is the national resource for digitized information about vouchered natural history collections”.\(^{31}\) iDigBio has 95,033,746 Specimen Records\(^ {32}\). Their mission statement includes the objective to “deliver appliances that integrate and package existing digitization technologies in a manner that enhances and/or simplifies the user experience”. They offer integrated data sources, which “provides access to millions of records about neontological and paleontological specimens curated at museums and other institutions in the US. Records might include information about the specimen, when, where and by whom it was collected, the institution providing the data, images and other media related to the specimen”.\(^{33}\) This statement

---

31 https://www.idigbio.org/about/project-scope

32 as of 4.6.2017

33 https://www.idigbio.org/portal/tutorial
includes a number of important search criteria mentioned by interview participants. For example, there is an articulated need for supplemental data which would help a searcher evaluate resources before determining a visit to a repository is warranted.

**State survey Data at risk evaluation.**
During the course of data collection, one of the state surveys shared with me a report they created on the topic of data at risk. The report is based on guidelines provided by the USGS with the view of evaluating at risk, legacy collections using the Information Product Data System (IPDS) guidelines for documents and recordkeeping. This report identifies a collection at risk and includes recommendations for providing better preservation and access to the collection.

The primary need listed in the report is that there is not a data portal for their repository. The author\(^\text{34}\) states in their report that, while the repository’s website “lists the evidence of a repository” and contact information for the sample manager, “a potential user needs to navigate through the website to find [their] contact information and contact [them] directly to access information about the repository”. This matches with experiences described in the interviews by researchers.

The report recommends outcomes which would support users more easily knowing what samples the repository houses, providing access to a wider level of associated metadata about the samples, and using ArcMap to allow users to view the sample distribution spatially. It also recommends including images of samples so that users may better evaluate the collection before deciding to visit and use the repository. These are all critical needs identified by participants during the interview stage. These advances would support participants’ searching in a way that

\(^{34}\) The state survey and author are not named in order to maintain confidentiality.
is more aligned with their search behaviors. It is encouraging to see this repository is aware of these issues and identifying steps to address them.

**Community discussions.**
In the fall of 2016, the Continental Scientific Drilling Coordination Office (CSDCO) hosted a workshop entitled, ‘Scientific Drilling/Coring and Earth-Life System Evolution’. The goal of the workshop was “to identify and prioritize the compelling science drivers, drilling/coring targets, strategic frameworks, and timelines focusing on continental sedimentary basins for the establishment of paleorecords in the coming decade”\(^{35}\). The workshop brought together a community of individuals whose research needs are focused around cores drilled on land and in lakes (as opposed to those which work with ice cores). I attended this meeting in order to observe discussions related to information needs. This was an illuminating opportunity to see the researchers discuss their projects at multiple levels, with attendees reporting on current, past, and upcoming projects.

Many attendees discussed information seeking and information needs. This discussion confirmed and provided a larger context to behaviors and needs found in the interview stage of this research. Attendees voiced concerns with discovering and providing access to data across subdomains, working with partners in industry to discover and share research data, limited knowledge available in supplemental information such as drilling logs, limited publications discovered in searches (unclear if due to lack of publications or lack of search tools to enable discovery), and how information is gathered about existing samples before determining a new site for drilling a new hole and conducting new core collection. Some attendees cited state geological surveys as the sources of cores used in analysis. At the end of the workshop, breakout

---

\(^{35}\) https://csdco.umn.edu/about/science-planning/workshop-november-2016
groups were created to discuss future needs. One group reported the need for more repositories open to deposition of cores and other samples collected from funded projects (in particular academic research). Another group discussed needs such as databases to house information related to individual researchers’ data and better guidelines for data preservation/management.

**Services.**
The workshop mentioned above was sponsored in part by an alternative science data center which curates cores and cuttings, LacCore. LacCore is the National Lacustrine Core Facility and houses “cores, grab samples, subsamples, and other materials,” provides laboratory tools for sample analysis and field equipment for the collection of new samples, and is affiliated with a number of other labs which offer expanded sample analysis. LacCore offers a wide variety of services and focuses on providing samples to researchers. From a LacCore brochure:

“Need Samples?
- We disburse thousands from our collection annually
- Technicians available to fill your requests, or
- Visit to pull samples yourself
- Huge cost savings over retrieving replicate materials”

(Infrastructure for Continental Scientific Drilling and Coring, n.d.)

In considering it from the view point of services and their effects on searching, these services reflect practices and support behaviors described by participants in this study. This includes providing technicians to help with search request and sample capture, and access to laboratory equipment which better facilitates an individual’s ability to assess samples for use in research.

---

36 http://lrc.geo.umn.edu/laccore/facilities.html
CHAPTER 5: DISCUSSION

Information seeking

As outlined in the literature review, there are many different ways of modeling information seeking behavior. When deciding on a research direction for this study, I decided the style of Ellis’s (1993) model, along with his follow up versions with various co-authors (Ellis, Cox & Hall, 1993; Ellis & Haugan, 1997), would best suit this topic. His models focus on research scientists of different domains, specifically on their search and not the larger concept of behavior. In Ellis’s various studies, subjects discussed their search practices in relation to their work as researchers. His model is primarily a list of behavior categories as opposed to a structured diagram. The model created in this study is a variation on this approach. Ellis (1993) did not consider the task or the position of their subjects within their organization.

Of the behaviors identified in this study, four overlap with Ellis’s categories. The categories are browsing, chasing, maintaining awareness, and validating/verification. Browsing is defined in a similar manner; however, chasing in this study includes following nontraditional citations such as references in reports which do not contain structured references. Maintaining awareness in this study better fits Ellis’s variation for chemists, in that the geologists are maintaining awareness across their more general interests as opposed to just keeping tabs on regular activities (Ellis et al., 1993). Additionally, the categories of verifying and ending (stopping in this study) are also similar to the categories observed in chemists (Ellis et al., 1993). Looking past the categories, Ellis did not have a structured diagram for his model of information seeking behavior. Wilson (1999a) proposed a diagram (Figure 6) which separates extracting,
verifying and ending from the other behaviors. This differs from the model developed in this study in a number of ways (discussed below). First, extracting was not observed in this study. Second, observations in the data, such as division of labor, led to the separation of starting and stopping from the majority of the search behaviors. Finally, verifying, which is listed here as validating/verification, is not separated out from the other search behaviors. The data did not suggest that this was a different activity that only occurred later in the search process.

Variations on his models include optional behaviors such as the categories of starting and ending. In this work, these two behaviors are not only optional, but may also be decided by someone other than the individual conducting the search. This decision is based on the concept of ‘division of labor’, which emerged from the data. This includes one’s supervisor or coworker when working in a team environment. Faculty and students interviewed in this study discussed scenarios, where the faculty member has given search criteria to a student, and will ultimately decide when the search has been successful, i.e., will end. The student conducts the search at a repository to which they have been directed by the faculty member, and will report back to discuss their findings. This process is used by the faculty to help train the student, either as part of a master’s student’s training or with the student in a research assistant position working for the faculty member (which may be undergraduate or graduate students).

One participant from industry described a situation where tasks were specialized within their organization. When samples are needed for research, the task of searching is directed to a different employee. The researcher will decide what criteria are needed for the samples, and will decide when the search will end. But the actual process of searching is done by another employee. These data influenced the design of the model but also suggest future research
directions. The topics of task and role should be further explored in relation to information behavior related to samples.

From the results of the interviews, a model was developed to visualize the broader information seeking process. The model, shown in Figure 13, has 3 parts. The first part of the search process represents ‘starting’ behavior. Before starting a search, researchers assess the purpose of their search tasks. This assessment includes considering if they are working on a timed search or an ongoing research project. These decisions may impact how long they search and when they might stop. At the start of the search process, researchers may have specific research questions in mind, or they may be looking more broadly with the hope of finding an information need that they might investigate. This starting behavior includes reviewing literature to identify research opportunities (e.g., samples available, questions not yet answered, etc.). While this is a search behavior, it also may occur as part of the starting phase. At the start of a search, the researcher may be limited by a geographical area – either one which they can be funded to visit or those which are supported by science data centers. Science data centers representing specific geographical areas would have existing samples available for study. Finally, some scientists who use existing samples have specific conditions for which the samples must be maintained in order to be viable for their work (e.g., paleomagnetism). Starting may begin by assessing the types of samples which may be available (and accessible) in existing repositories.
The second section of the model represents the bulk of the search behavior, both active and passive information seeking. This section of the model is nonlinear, and each category of searching behavior is optional and may be repeated or skipped as the search dictates. These broad categories represent different activities a searcher will conduct while looking for resources, in this case physical samples, see Table 11 for definitions of the behaviors. The categories may influence each other, for example, browsing may lead to personal contact (e.g., with a curator or a colleague). An individual looking through a collection may decide they need more information to evaluate the value of a set of samples. They may wish to contact the individual responsible for sample collection or for metadata creation (an example of visiting/personal contact). During this conversation, other samples may be recommended and so on.
Table 11.  
Definition of behavior categories

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting</td>
<td>Initial step that starts the search process. This includes developing a research question.</td>
</tr>
<tr>
<td>Maintaining awareness</td>
<td>Keeping current with changing information in the field such as attending national meetings or being aware of which repositories are used by projects of interest.</td>
</tr>
<tr>
<td>Browsing</td>
<td>Unmotivated search or unstructured searching, with the hopes of finding something of interest. This includes activities like shelf browsing.</td>
</tr>
<tr>
<td>Division of labor</td>
<td>Separating out specialized tasks. For example, a student may be asked to conduct searches so their advisor can do more specialized analysis.</td>
</tr>
<tr>
<td>Social network</td>
<td>Knowing who else is working in one’s domain or geographical area of interest. This can be useful in tracking down missing information or suggesting ways of accessing resources.</td>
</tr>
<tr>
<td>Visiting/personal contact</td>
<td>Needing to visit the repository or make contact with repository staff. This may be to get additional information beyond that which is available online or to get expert guidance.</td>
</tr>
<tr>
<td>Chasing</td>
<td>Finding related materials and broadening one’s search, not always tied to traditional citations in publications.</td>
</tr>
<tr>
<td>Negotiating</td>
<td>Gaining access to or finding samples. This includes agreeing to share information created with the researcher supplying the samples.</td>
</tr>
<tr>
<td>Validating/verification</td>
<td>Evaluating the trustworthiness of the data you find or checking to see if it is reliable. This includes looking for errors or misidentifications, or checking to see if samples have degraded over time.</td>
</tr>
<tr>
<td>Stopping</td>
<td>Ending the active search. Time and funding constraints for grants or degree programs may mean a search will end before all available information is exhausted.</td>
</tr>
</tbody>
</table>

Categories such as ‘social network’ and ‘personal contact’ may have similarities, but they represent different ways of finding and accessing information. For example, one relies on their social network to suggest possible repositories, publications, or other researchers working in
one’s topic area. This is different than needing to contact the curator of a collection or author of a paper in order to obtain information they might have that is missing from a database or publication. In each case, it may be a lack of answers from existing infrastructure which lead the researcher to speak with another individual but they also represent different concepts and motivations. Similarly validating/verifying the accuracy of information provided with samples may result in personal contact, but looking for missing information may be different from confirming the quality of the information provided. The various categories support one another and are influenced by the searchers’ needs.

The final section of the model is stopping. Stopping may represent a completion of the search process; information discovered while searching leads the researcher to decide to end their search. This includes cases where the researcher has found the samples needed to answer their research question. It also includes cases where the researcher has limitations on their search, such as time and funding and must stop. This was noted by the independent contractor and one of the students. While both participants felt they found enough data to complete their research, their motivation for stopping was not data saturation but these limitations. A search does not always have to be successful to lead to an ending. This includes discovery that samples might not exist that suit the searchers’ needs or that the samples they were searching for were destroyed. During the search process, as researchers assess their data, they may determine they need to return to actively seeking information as described in the active information section of the model.

The first and last section of the model, starting and stopping, are optional. As previously mentioned, the concept of ‘division of labor’ emerged from the data and influenced the decision to separate starting and stopping from the body of the model. In these instances, determining the
scope of the task and what conditions might shape how one started the search were evaluated by a researcher (either a senior researcher or a faculty member) and the search task was handed off to a student or another employee. The student/employee was given a set of search qualifications and tasked with completing steps from the information seeking stage. In addition, not all searches are so formalized as to have a clear start or assessment of the task goals, and conversely, they may not have a definitive ending. One participant pointed out that they spent 20 years looking for a set of cores. Their search never formally stopped. As such, these two parts of the model are outlined in dashed lines to illustrate their optional status.

**Seeking.**

In Wilson’s (1999) discussion of information seeking, he reminds us that during one’s information seeking they may interact with many different systems. The interview participants spoke of this as one of the obstacles they encountered during their searching, having to learn a new system and way of searching in that system for each repository. There are no standard records for physical samples and no common set of metadata across these systems. In contrast, if we were to think of a library catalog record for a book or publication, they may vary slightly in the information they provide but there is a basic description that is common both across these records and the systems which provide access to them.

A record for a book includes information such as title, author, year of publication, and an abstract. In contrast, a record for a core sample may include data type, geographical location, depth intervals, and the rest varies. This may include a name of a core or title of the project for which it was captured, but it doesn’t typically include a summary of what the sample represents. Some collections may link to lithological descriptions, which may include rock types, formations, or fossils present in the sample, but this is not standard. Information like these types
are used by researchers to understand what is in the sample and if it might be relevant to their needs.

With physical samples, there are of course similarities shared across the records, but each institution includes a different set of information and provides access to them in different ways. Also, given the nature of the samples (as objects where much of the associated information is tacit or implicit), most systems are not currently designed to support all relevant information about a sample. This puts an additional burden on the searcher to follow up with the curator or author to gather additional information. This is a skill participants said they had to learn and might influence their decision to not search at a particular institution.

The USGS is working on this issue with the development of their National Digital Catalog\textsuperscript{37}. Through their NGGDPP, they have been providing funding to state geological surveys to inventory their collections and to import minimal metadata about the collections into the National Digital Catalog. There are 13 fields, 7 of which are mandatory\textsuperscript{38}. The mandatory categories are: title, abstract (human readable description), data type, supplemental information (e.g., where to locate the sample, may be a URL), coordinates, and date of the record. This catalog will act as a low level ‘union catalog’ for state geological repositories. It does not solve all the problems faced by participants given the minimal metadata, but it does help point them in a direction to start their search. They still need to learn the mechanics of searching within each individual state survey’s collections.

\textsuperscript{37} https://data.usgs.gov/datacatalog/

\textsuperscript{38} https://datapreservation.usgs.gov/docs/uploaded/NGGDPP\%20Metadata\%20Preparation\%20Guide\%202017.pdf
Social networks and personal contact.
The most frequently used search process by interview participants was relying on their social network to recommend or locate samples. This was for a number of reasons including the nature of the samples, the infrastructure relating to samples, or both. In some instances, participants found it was easier to articulate their search needs to other experts. Current search tools do not support the expert level semantic searching needed by researchers, and not all information needed to conduct a search was either digitized or linked. Repository managers’ self-reports corroborate these concerns. For example, managers reported that their collections are not fully digitized; on average repositories had 76% of their collections documented. Only 15 of the 27 repositories had their catalogs accessible online. This blends two different issues: having all the needed information in a searchable form, and being able to access it.

State geologists reported that surveys have on average 168 hours per month of staff time dedicated to collections. That is the equivalent of one full time employee, but does not describe the expertise of the individual available and their ability to assist with an outsider’s search needs. This equivalent staff member is made up of time from full time or part time staff, students and volunteers. With these hours, half of the states felt they did not have suitable staffing needs for their collections. When staffing is limited and searchers must depend on institution staff to access collections, this is problematic. Fourteen institutions said that their digital records were accessible only by staff members, as seen in Table 9; thirteen reported their physical records were accessible only by staff members. If a researcher wanted to search within these collections they would need to request access, and work with a staff member at the institution. State geologists reported other critical demands of their managers’ time beyond user requests, such as curation and preservation of their collections. Only one of 32 institution reported that the preservation needs for their collection met their institution’s current and future needs. Curators
and repository managers have many demands on their time; it is easy to understand why searchers might report they are a low priority to those who control access to collections. These limitations can impact their search behaviors and complicate the search process.

**Secondary data.**

During the interviews, participants expressed a desire to have better access to secondary materials related to the cores and cuttings in collections. Secondary is defined as not primary metadata related to the samples. This includes data derived from samples (e.g., logs, scans, chemical testing, etc.), supplemental materials (e.g., reports about the project which facilitated the sample collection, etc.), and resulting publications (e.g., maps, white papers, journal articles, etc.). As noted in the literature review, the NRC (2002) states that documentation for physical samples “includes information about the age, location, depth, originator, and the date acquired” (p. 9). During the questionnaires, state geologists and repository managers were asked about ‘documentation’ of their collections. This question was not fine grained enough to capture these concepts of secondary and supplemental data related to the samples in their collections. The interview participants are pushing this definition to include a broader set of information. Follow up is needed to better understand their opinions and priorities regarding these materials.

Participants mentioned prioritizing institutions based on access to these materials as they help with discovery of samples as well as evaluating their usefulness.

Accessing secondary data is not as simple as digitizing these materials. In order to leverage these materials to the fullest, ontologies and semantic tools would also be needed. These tools connect terms based on their relationships, and they allow concepts to be associated with a term. By building these relationships into a system, it can facilitate automated connections between data and supplemental data. Consider subject headings in library catalogs. They link records which have similar terms associated with them automatically. If we look at
the record for a book tagged with a subject heading relating it to the American civil war, we can follow the subject heading to find other resources that also have that subject heading. Applying this strategy to geological data could help searchers to find samples that have similar characteristics. Imagine searching for a core, and then being able to find all other cores which have similar log results or similar lithological descriptions.

**Resources by region.**
When participants searched for samples, the geographical region was an important part of the starting process. Repository managers reported that their collections typically represented their state – a defined geographical region. If searchers were considering a region, going to the repositories in that state would be a good first step. But how does one keep current on all of the possible repositories in a region? In 1997, AGI addressed this need by publishing the first edition of their *National Directory of Geoscience Data Repositories*. It is a listing of 124 repositories organized by state. It was created to allow users to locate geological data and AGI intended to update this listing every 2 years (Claudy & Stevens, 1997). The document has never been updated, and the online version was recently removed from their website. This is a simple tool that is missing that might be an easier fix than developing standards and ontologies.

**Education in searching.**
During the interviews, participants were asked how they might train someone in searching. The question led to an unexpected result: participants were not sure when this should occur. Two of the junior faculty members were surprised by the question and had not considered formal training in searching, or conversely in data management practices. One said it did not come up because they typically supervise undergraduate students. This led to a discussion of when this skill set should be taught to future researchers. The topic was brought up with later participants. There is no formal part of one’s geology education which includes these practices.
Students typically learn by watching others, experiencing the search process as an apprentice, and taking the lead on the process during their doctoral training. Participants questioned if this type of learning should be more formalized, and if it should be moved earlier into the educational program. Undergraduate students have the potential to be left at a loss, needing to learn on the job if the skill is required.

Part of training in searching for samples is understanding the record creation process. This includes what is captured in a record and what is not. Perhaps part of the problem with searching and lack of information in records is a poor understanding of data management practices, i.e., knowing what should be captured to enable sharing and access by future researchers outside of the initial research team. This connects with the recommendation for additional training for repository managers and geologists in searching and data management practices, but in either regard this would be a social change for the field. It begs the question, what is the role of organizations like AGI in enabling access to scientific data? This must come from the society level which includes organizations such as GSA, AGU, AGI, among others.

**Connection to science data centers**

A recent perspectives piece published in *Science* magazine (McNutt, Lehnert, Hanson, Nosek, Ellison, & King, 2016), discusses the problem of accessing samples and data from what they call the ‘field sciences’. In particular they focus on the lack of formal citations for data in publications, instead reporting that authors in these domains will typically include the note “data and samples available upon request” (McNutt et. al, 2016, p. 1024). This is in contrast to other domains where the nature of the data means they can be deposited online and cited with a DOI. Persistent identifiers like DOIs help connect data and samples to their provenance information. In particular, the persistent identifier can be cited in literature to create a connection between the samples and the publication. They can also be used to connect supplemental materials to the
samples from which they were derived. These may be combined with semantic tools, which layer concepts and relationships to make information more discoverable.

During the interview process, participants shared their thoughts on the value and limitations of existing tools; they also suggested improvements. Many felt that the current online catalogs and databases offered by state geological surveys were limited. These limitations might not be the result of a lack of intent from the surveys, but may be based on where the data came from and what information was collected about those data. As science data centers, the state surveys’ collections are conglomerates of materials collected from multiple researchers for multiple purposes (see Figure 4). That means their databases and catalogs are limited by the information provided with the samples at the point of deposition.

There are currently no standards for the minimum information that should be captured along with a sample at the point of collection. While the USGS recommends minimum metadata for repositories, this is not pushed down to the individual researcher gathering samples out in the field. Sixteen of 31 state geologists reported they had a legal mandate to maintain their collections. This means that even if samples arrive with missing or limited metadata, the repository must still ingest it into their collections. It may be a time-consuming task to fill in these metadata gaps. What are the challenges faced by curators of SDCs and how does that impact the quality and richness of their catalogs? This is a topic which should be addressed in future research.

One recommendation to address these issues would be for the USGS, or other leaders in the field, to put an emphasis on expanding the documentation process, either in practice (e.g., peer review) or through education. In the long run, this would increase outsiders’ unassisted access to records and take the burden off collection staff. However, budget issues may also be a
limitation. Twenty-three of the 31 state geologists reported they did not have adequate budgets for their collections; nine state geologists reported they had no budget for their sample collections. With limited funding, there may be other considerations that are more pressing than documentation. For example, 18 state geologists reported needing more space, and the NRC (2002) found storage issues put collections at risk for loss.

Another consideration is the expertise of the staff managing these collections. Of the 27 responses, 17 of the repository managers had titles that were related to geological responsibilities and not curation. Searchers found this expertise in geology to be invaluable for assisting in their search, but the searchers may not have the proper training in data management practices needed to implement some of the tools recommended. This includes an understanding of the value and role of ontologies, semantics, and database design. Even with tools that satisfy the searchers’ immediate needs, existing systems cannot evaluate the quality of the resources or understand the domain expertise needed by searchers. The curators will continue to remain an invaluable resource. This area should be further explored in future research.

As Cocker (2005) suggests, there is a need for proper training in curation and management to help prevent loss of legacy data. While next steps for preserving legacy data include cyberinfrastructure as it assists with discoverability of data and ultimately access and use, it does not address all of the preservation needs typically faced by those managing a physical collection. One must look forward as well and think of data currently being collected, or that which will be collected in the future.

In order to develop successful infrastructure for accessing physical geological data, systems must address a number of barriers including: “1) resource registration; 2) resource discovery; 3) information interoperability; 4) services interoperability; 5) analysis and
processing; and 6) publishing” which will allow for “enhanced representation of geoscientific knowledge, and … enhanced tools for processing the knowledge during data integration activities” (Brodaric & Gahegan, 2006, p.2). By developing collection curation and management systems which have interoperable data, semantic search capabilities and greater knowledge capture, scientists can focus their energies on answering those ‘grand’ science questions instead of wasting valuable time processing and translating data.

While the USGS is working on a National Digital Catalog, there is still a need for federated catalogs with richer metadata. Searchers are already using tools which fit these needs (see the discussion on iDigBio in chapter 4 as an example). To borrow from the world of libraries, we might consider an example like WorldCat which searches across the holdings of many institutions. A tool like this would require a much richer set of standards for metadata. This would encompass those collecting samples and creating initial metadata records and collection managers who would be overseeing a larger set of related records. However, this does not address supplemental data.

These challenges are not unique to state geological surveys, but may apply to all organizations which house physical samples and other types of scientific collections. This includes individual department collections and regional repositories. Preservation includes maintaining the connection between digital surrogates and physical samples. These hybrid collections require new and sometimes advanced training. Organizations’ preservation efforts should include continual training and collaboration between science data curators/managers/etc. and those with Library and Information Science (LIS), archives, or museum experience. Not only will domain scientists need assistance with perspectives on preservation, but information professionals should be trained in the unique challenges of curation, preservation, and
stewardship duties in the science realm. As will be discussed in the next section, physical samples as scientific objects have different needs than standard objects. Training of information professional needs to reflect these considerations.

In a 2005 report, Microsoft suggests that in the future there will be a need for domain experts who have computer science expertise and training in curation. Interestingly the report neglects to mention these skills can be found in such fields as LIS or archives (Emmott & Rison, 2005). In their report on preserving natural science collections, Duckworth, Genoways, and Rose (1993) highlight the need for education in managing physical specimen collections as they are vital to preserving science collections. They also outline the various knowledge areas needed for curators in this role, which include many LIS skills.

**Nature of the data**

In this work, physical specimens such as cores, cuttings, and thin sections are referred to as samples. As outlined in Table 2 from the literature review, there is a difference between information and knowledge. Samples provide researchers with information that is used in the generation of knowledge. With samples, that information might be tacit or implicit, and not often captured in a formal way. The sample is needed to validate or elaborate on the information that might be transformed into knowledge. This includes an individual’s observations about the samples when they were collected that might not have been entered into formal records but may have influenced any knowledge created. The scientific language used by geologists has also been known to cause problems with preservation and access (Aldrich, 1976; Jordan, 1976).

While samples may be information sources, they cannot be directly searched in the same way one can look for a print resource. Instead, one must search within records created to represent the samples. These records provide clues and suggestions to what type of information might be extracted from the samples. While samples and print resources may have metadata
acting as surrogates, there are universal standards for how one creates a surrogate for a print resource. Even though there are some standards for physical objects, rarely do they encapsulate all of the types of information that could be made explicit. These current standards cannot represent the implicit information held by geological samples. With a journal article you can scan an abstract, look at key words, etc., but with a sample, unless there is rich supplemental information about the sample, this is not the case (and even then, it might not be exhaustive for one’s research goals or domain). As discussed by the various interview participants, searching for samples requires much more work in order to narrow the candidate of possible samples and then to evaluate it for potential fit. This includes viewing the samples, conducting chemical testing, talking to the researcher who collected the sample, reviewing supplemental data associated with the sample such as geophysical logs, etc. While this information takes on a mixture of forms, it is not machine process-able or readily available. It forces the searcher to employ additional search behaviors beyond those required by print materials.

**Limitations of this study**

One limitation is this study is that it focuses on a single type of institution. This gives a narrow view of the search behavior that may not be extendable to other types of repositories (department, industry, or private collections). There is value in this approach, however, as it created a context to the search behavior, particularly given the diverse nature of physical samples and the tools used in accessing them. Including a wider spectrum of repository types would have created more diversity in the results which may have made them less conclusive. Another limitation to this study was the pool of participants. Industry and private sector patrons were difficult to recruit into this study. They represent 25% of the user population (see Table 10) and are underrepresented in this study with only two participants representing this group. In pretesting the questionnaires, this potential issue came up in discussion with repository
Managers felt industry patrons would be less likely to take part in this study as their search practices may be seen as proprietary. Managers and industry researchers may have a tenuous relationship.

In describing qualitative research, Boeije (2010) compares the steps between data collection and writing up results to a black box. This is the part of the process that creates “vagueness” and “suspicions about the credibility of the conclusions” (Boeije, 2010, p. xi). He states that the lack of detail when writing up study results can also lead to problems testing the validity of the work. Trustworthiness was built into the study design, memo writing and triangulation were used during data collection while debriefing was used during analysis in order to ensure the quality of the outcomes.
CHAPTER 6: CONCLUSION

In the field of geology, physical specimens such as cores, cuttings, fossils, and rocks are primary sources of data; they represent “the foundation of basic and applied geoscience research and education, and underpin industry programs to discover and develop domestic natural resources” (NRC, 2002, p. 1). Science data centers, such as the state geological surveys included in this study enable discovery, access, use, documentation, and preservation of physical scientific data throughout their lifecycle. This study investigated the search behavior of geologists while looking for physical samples at these science data centers.

In the first phase of this study, 35 state geologists and 28 repository managers responded to questionnaires. This data captured an overview of these science data centers and their handling of collections of physical samples. While there were variations in practices across these institutions, they were overall similar in nature, with similar issues. This includes the need for additional resources (e.g., additional staff and support for wider documentation) to create smoother access to collections by patrons. These findings were reinforced by the participants in the second phase of the study. In this phase, 15 individuals were interviewed about their search experiences. Responses highlighted various search behaviors as illustrated in Figure 13. Most prominent of these behaviors was the importance of personal contact. This includes visiting repositories to evaluate samples and reaching out to one’s personal network or known experts to facilitate discovery. Other results include discussions of division of labor, questions regarding the need for training and when this might occur in the education process, and challenges and
barriers encountered during the search process (e.g., negotiating access to materials, learning a new database structure for each institution, etc.).

The results of this study suggest a number of recommendations. This includes 1) developing infrastructure which allows researchers to search between data and supplemental information, and to leverage the relationships between the two, 2) to further explore the nature of task and role in relation to searching for physical samples, especially in regards to education, 3) the development of training systems for managing physical sample collections, particularly related to facilitating access and use of these collections, 4) developing standards for metadata creation related to physical samples in order to support use and reuse, and 5) developing tools which would, in large part, remove the burden from the searcher of having to rely on personal contact while searching.

These recommendations can be approached from a practical standpoint on three different levels. The most basic would be for state geological surveys and other science data centers which provide access to physical samples to create guides on how to search within their collections. In information and library science, librarians create tools called libguides, or subject guides. These are themed around a topic and include information about resources available both in the library and online (databases, books, etc.). Science data centers could create guides that would help outside users navigate their collections, understand the depth and breathe of their holdings, and how to use the tools they provide. A more complicated approach to these recommendations would be to develop community standards, this would include metadata standards for physical sample records and standardized systems which provide access to metadata. This approach would require the various organizations to act together to develop such standards, and to work together in implementing them. A more complicated approach to user
needs and the recommendations listed above, would be to develop ontologies and interactive cyber systems such as dynamic, federated catalogs. This would require funding and coordination.

**Future research**

As mentioned in the discussion, there is a lack of interconnectivity between various repositories in the geosciences. Physical samples are scattered across institutions. What kind of burdens or barriers does this create for scientists in determining where to deposit their samples for long term preservation, and where to search in order to discover new resources? Distributed repositories for samples suggest the value of persistent identifiers. If standards existed for sample citation, would it impact search behavior? LacCore has offered to partner on this research by providing access to their user community.

One assumed theme that did not come up in the interview process was the generation of new data as a search process. It was an assumption in this study that researchers might stop searching and gather data themselves if they did not find what they were looking for in existing collections. While this still may hold true, it did not directly come up during the interviews. This topic could be explored in future research.

If the data from the NGGDPP’s 2007 survey can be accessed, it would be beneficial to compare the results with an updated survey. This could provide insight into the effects of the evolution of cyberinfrastructure on user access, and the impact of the NGGDPP on these collections.

Other possible research based on this study might investigate the question, “Are there differences in domain practices or occupational training that affect search behavior?” From this research questions may arise indicating there may be different search practices in different occupations which affect the way geologists search for information. This includes the major
settings identified in the literature - industry, academia and governmental organizations. These differences may be learned ‘on the job’ as opposed to being fundamental to education in geology.

While there is a clear need for ontologies and standards, how these might be developed and implemented is not clear. There is some work on this topic coming out of the structural geology community and they have encountered challenges given the nature of the data. This is a large area to explore.

Finally, the model developed in this study may be compared to those of Ellis’s model of information seeking behavior. In this study, Ellis’s categories were used as sensitizing concepts (e.g., probing based on them during the interview). Any categories not represented could be investigated in follow-up projects as the purpose of this study is to create an original model of the process of searching for physical sample materials.
APPENDIX A: GLOSSARY

Terms taken from the AGI’s Glossary of Geology (2005). Images belong to the author except where noted.

**Chip sample:** A series of chips of ore or rock taken at regular intervals across an exposure.

**Core [drill]:** n. a full diameter core is a cylindrical section of rock, usually 5-10 cm in diameter and up to several meters in length, taken as a sample of the interval penetrated by a core bit, and brought to the surface for geological examination and/or laboratory analysis.

**Core sample:** One or several pieces of whole or split parts of a core, selected for analysis; a sample obtained in coring.

**Formation:** A persistent body of igneous, sedimentary, or metamorphic rock, having easily recognizable boundaries that can be traced in the field without recourse to detailed paleontologic or petrologic analysis, and large enough to be represented on a geologic map as a practical or convenient unit for mapping and description; the basic cartographic unit in geologic mapping.

**Fossil (paleontological sample):** n. Any remains, trace, or imprint of a plant or animal that has been preserved in the Earth's crust since some past geologic or prehistoric time; loosely, any evidence of past life. adj. Said of any object that existed in the geologic past and of which there is still evidence.
**Hand specimen (sample):** A piece of rock of a size that is convenient for megascopic study and for preserving in a study collection.

**Sediment:** (a) Solid fragmental material that originates from weathering of rocks and is transported or deposited by air, water, or ice, or that accumulates by other natural agents, such as chemical precipitation from solution or secretion by organisms, and that forms in layers on the Earth's surface at ordinary temperatures in a loose, unconsolidated form; e.g. sand, gravel, silt, mud, till, loess, alluvium. (b) Strictly, solid material that has settled down from a state of suspension in a liquid. In the singular, the term is usually applied to material held in suspension in water or recently deposited from suspension. In the plural, the term is applied to all kinds of deposits, and refers to essentially unconsolidated materials.

**Thin section:** A fragment of rock or mineral mechanically ground to a thickness of approximately 0.03 mm, and mounted between glasses as a microscope slide. This reduction renders most rocks and minerals transparent or translucent, thus making it possible to study their optical properties.

**Well cuttings:** Rock chips cut by a bit in the process of well drilling, and removed from the hole in the drilling mud in rotary drilling or by the bailer in cable-tool drilling. Well cuttings collected at closely spaced intervals provide a record of the strata penetrated. Syn: cuttings; drill cuttings; well samples.
Coring processes (illustrated).

Modern technology for scientific drilling: the basic elements

Coring processes (illustrated).

Figure 8. Drilling scheme

ICDP projects address a whole host of geological targets from deep to shallow, from tectonically simple to complex, and under very different pressure and temperature conditions. Modern technology ensures all these targets can be reached, even if they lie at 12 km depth! Having said that, costs rise exponentially with depth and degrees of difficulty, so detailed and careful planning is prerequisite.
APPENDIX B: PROTOCOLS FOR ARRANGING A VISIT TO STATE COLLECTIONS

Examples existing protocols for arranging access to state geological survey collections.

Alaska.


1) The Alaska Geologic Materials Center (GMC) is open to the public during normal state working hours (Monday – Friday, 8:00 am to 4:30 pm, except State holidays); arrangements for visits after normal working hours are at the discretion of the curator upon sufficient advance notice;

2) All materials, processed materials, and well data reports at the Alaska GMC are available to the public for examination; i.e., no material at the GMC is confidential;

3) It is recommended that the examiner make reservations for any planned Alaska GMC visits.

Illinois.

https://www.dnr.sc.gov/geology/repo.htm

“Visitors must register at the Geological Samples Library office to obtain permission to study the collections. Samples Library staff will assist visitors with sample retrieval and layout. Samples are not loaned, but selected sampling of the collection is permitted with prior approval. No fee is charged for studying samples, and no appointment is required; however, we request that you call 217-333-3567 at least a day ahead of your visit, so that we can retrieve the samples you want and set up a work space for you.

If samples are used for thesis research, students are required to provide a bound copy of their thesis, any thin sections made from the collection, and copies of any slides or photographs made of the samples. Failure to comply may result in the student's institution losing rights to use the facility.”

Indiana.

https://scholarworks.iu.edu/dspace/bitstream/handle/2022/1002/D09.pdf?sequence=1

“This catalog is an index of the cores filed in the core library of the Indiana Geological Survey. The holdings in the library may be used by the public for research and for the development of a better understanding of the stratigraphic sequence of Indiana. About 700 cores are filed in the collection; all but six are from holes drilled in Indiana (tables 1 and 2). More than half of these cores were obtained with equipment operated by the Geological Survey, and the rest were donated by the oil and mineral industries. Many of the cores have been chipped and added to the sample library of drill cuttings, which is a separate facility containing nearly 13,000 sample tests. The cores are split in half by a core splitter that provides a fresh unaltered surface for examination. All cores are stored in numbered cardboard boxes that correspond to an internal file system. Facilities for examining the cores include examination space, microscopes, and lamps. Also available in the core examination area are a microfiche file of individual-well data and a microfiche reader that may be used in conjunction with the core-examination. Cores may be
examined at the core library from 8:00 a.m. to 5:00 p.m. (Eastern Time) Monday through Friday. By arrangement small samples may be obtained from the cores for specialized research involving destructive analysis. After-hours examination or loan of cores is not permitted. A map showing distribution of the cores within Indiana is available upon request. Persons wishing to use the core facility should contact the Petroleum Section, Geological Survey, 611 North Walnut Grove, Bloomington, IN 47405. Telephone 812 335-5412". 
APPENDIX C: INITIAL EMAIL AND QUESTION REQUESTS

Email to state geologists.

Email sent to AASG list (Approved/edited/sent by previous president of AASG, Dr. Jon Arthur)

Subject line: Data Preservation Survey - AASG/USGS supported study

Greetings all!

I would like to make you aware of a study of state geological survey sample collections and repositories. The researcher, Sarah Ramdeen, is a former employee of the Florida Geological Survey and is currently a PhD Candidate at the University of North Carolina at Chapel Hill.

The purpose of this study is to better understand state survey core and cuttings collections and to assemble information about management, access, and use of these collections. The results will be reported back to the AASG and the USGS National Geological and Geophysical Data Preservation Program.

Ms. Ramdeen will be asking you as State Geologist and your Survey’s sample repository manager (or similar position) to take part in a brief questionnaire on this important topic. The questions in this study are based on the 2007 survey conducted by the USGS and have been reviewed by John Steinmetz, former chair of the AASG’s Data Preservation committee, and John Yellich.

Please stay tuned for the email from Ms. Ramdeen. For more information about her project, see https://ramdeen.web.unc.edu/dissertation password: aasg or you may contact her at ramdeen@email.unc.edu.

Thanks,

Jon

Jonathan D. Arthur, Ph.D., P.G.
Past-President, AASG

Follow up email to individual state geologists, to be sent by Sarah
Re: Data Preservation Survey - AASG/USGS supported study

Hello ${m://FirstName}

I am contacting you as ${e://Field/EmbeddedDataB} for the ${e://Field/EmbeddedDataC}. I am conducting a study related to core and cuttings sample libraries or repositories held by State Geological Surveys.
To take part in this study, please follow this link: ${l://SurveyLink?d=Take the Survey}

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

You will be asked a series of questions that will take between 5-10 minutes to complete. Please complete this questionnaire within the next two weeks (before October 6th).

As previously mentioned by Dr. Jon Arthur, the purpose of this study is to better understand state survey core and cuttings collections, specifically information about management, access, and use of these collections. Once completed, the analyzed results and data collected in this study will be presented to the AASG and the USGS National Geological and Geophysical Data Preservation Program.

This research is being conducted as part of my dissertation. A brief overview of this project is available online at [http://ramdeen.web.unc.edu/dissertation/](http://ramdeen.web.unc.edu/dissertation/) password aasg. If you have any questions about this research, please contact me.

Thank you very much,
Sarah

Sarah Ramdeen Doctoral Candidate
School of Information and Library Science
University of North Carolina
ramdeen@email.unc.edu

Email to repository managers/contacts listed in the first questionnaire.

Email to individuals suggested by state geologists in first questionnaire.
Re: Data Preservation Survey - AASG/USGS supported study

Hello ${m://FirstName}

I am contacting you as ${e://Field/EmbeddedDataB} for the ${e://Field/EmbeddedDataD}. You have received this email because your ${e://Field/EmbeddedDataC} passed along your contact information and supports your participation in this research. I am conducting a study related to core and cuttings sample libraries or repositories held by State Geological Surveys.

To take part in this study, please follow this link: ${l://SurveyLink?d=Take the Survey}

Or copy and paste the URL below into your internet browser:
${l://SurveyURL}

You will be asked a series of questions that will take between 15-25 minutes to complete. Please
complete this questionnaire within the next two weeks (before October 28th).

The purpose of this study is to better understand state survey core and cuttings collections and to assemble information about management, access, and use of these collections. The results will be reported back to the AASG and the USGS National Geological and Geophysical Data Preservation Program. The questions in this study are based on the 2007 survey conducted by the USGS.

This research is being conducted as part of my dissertation. A brief overview of this project is available online at http://ramdeen.web.unc.edu/dissertation/ password aasg. If you have any questions about this research, please contact me.

Additionally, I will be attending GSA this year if you are interesting in meeting in person.

Thank you very much,
Sarah

Sarah Ramdeen
Doctoral Candidate
School of Information and Library Science
University of North Carolina
ramdeen@email.unc.edu
APPENDIX D: STUDY DESCRIPTION ON WEBSITE

Introduction.

As part of my dissertation research, I am conducting a study related to core and cuttings sample libraries (also known as repositories) associated with state geological surveys. This work is being completed as part of my dissertation research. In addition, John Steinmetz, outgoing Chair of the Association of America State Geologists (AASG) Data Committee has reviewed and supports the questions in this study. This study aligns with work conducted by the United State Geological Survey (USGS)’s Data Preservation program – the National Geological and Geophysical Data Preservation Program (NGGDPP).

The first part of this project is meant to replicate and update a study conducted by the USGS’s NGGDPP in 2007. The results of the USGS’s study have not been published. They may also be out of date given advances in technology and current preservation efforts. As such, this current research is important.

This research has also been funded by an Earth Science Information Partner’s (ESIP) Funding Friday award. This funding was given in order to support research related to preservation and stewardship of physical earth science data and developing a community for those interested in this topic.

Research Objectives.

- To gather information about the size, scope, and standings of state geological survey’s sample collections. (Informative for AASG and NDDGPP)
- To determine demographic information about the users of state geological survey sample collections. (Informative for AASG and NDDGPP)
- To create a model of information seeking behavior of users when they are searching for physical geological data. (Informative for Information science and data science research)

Part one: State geologists and repository managers.

The first part of this study consists of two questionnaires. One set of questions is directed to the state geologists and the other is intended for their sample library or repository manager (or other related position). The questions in this study are based on a 2007 survey conducted by the USGS as part of the Data Preservation program (NGGDPP). All information gathered in this study will be de-identified – specific institution names or other identifying information will not be published.

How to participate.

Questionnaires.

The first questionnaire will be sent out to the state geologists. It will be online, and will include questions to identify a contact who will complete the second questionnaire. This contact (your repository manager) will be asked if your institution will act as a recruitment site for the second
part of this study. The second part is described below, and will focus on interviews with patrons of these collections.

State geologists.

This first questionnaire includes questions about the overall content, governance, management, mission and characteristics of the state geological survey in which the collections are contained. This section should be completed by the state geologist or his/her representative. It will take between 5-10 minutes to complete.

Repository managers.

The second questionnaire includes specific questions about individual sample collections within the state. This section is intended to be completed by the relevant collection or repository manager. It will take between 15-25 minutes to complete depending on the size and scope of their collection.

Viewing the questionnaires in advance.

Below are PDF versions of the two questionnaires which will be used in this study. This includes all possible questions. Based on your responses when answering the live questionnaire, some questions may not appear.

State Geologists questionnaire

Repository managers questionnaire

Timeline.

- Mid-September - Initial email sent by Jon Arthur to State Geologists
- Mid-September - Questionnaire sent to individual state geologists by Sarah Ramdeen
- Early October - State geologist questionnaire closed and results analyzed
- Late October - Questionnaire sent to repository staff as identified by state geologists
- Early November - Repository manager questionnaire closed and results analyzed
- Early November - Preliminary results shared with the AASG at GSA
- Early January – additional results presented at the ESIP Winter Meeting.
- Spring 2016 - Final results to be included with my dissertation findings.

Part two: Patron Interviews.

How to participate: Recruitment.

During the first part of this study, I will ask the state geologists to give permission for their collection manager (or equivalent position) to work with me to recruit participants from their user community. Instead of collecting contact information of these users from the managers, I will provide the managers with a recruitment statement to pass along to potential participants. The statement will include background information on the study, qualifications for participation,
and the topics to be covered in interview. It will also include information on how to contact the researcher to take part in the study. Participants of this phase of the study will take part in a 45 minute semi-structured interview over the phone.

**Expected outcomes.**

Results of the questionnaire will be presented in a report to the AASG. They will also be used to develop a sampling protocol for the second part of this study and for the completion of my dissertation research.

**Expected dissertation related results.**

There are three planned outcomes from this research:

- Developing a model for understanding information seeking\(^{39}\) for seeking physical samples.
- My research will influence the design and development of cyberinfrastructure systems. Cyber systems link users to physical information objects; they enumerate processes for data access and include methods for capturing various forms of knowledge.
- My work will underscore the need for the training of information professionals and domain scientists to prepare them for curation, preservation, and stewardship duties in the science realm.

**Questions.**

If you have any questions about this study, the participation process, or the expected results, you may contact me by email. Additionally, I will be attending the Geological Society of America’s annual meeting in Baltimore, the American Geophysical Union’s fall meeting in San Francisco, and the Earth Science Information Partner’s winter meeting in DC. If anyone is interested in an in person meeting, please contact me.

**Interviews – background.**

**Introduction.**

The purpose of this research study is to build a basic model of the information seeking behaviors of scientists who use physical geological samples in their research. And to better understand how individuals access these materials. These materials includes cores, cuttings, thin sections, fossils etc.

**Participants are being recruited from State Geological Surveys. Inquires at the survey do not need to be successful (i.e. result in a visit or acquisition of samples) in order to take part in this study.**

---

Starting at the 26 minute mark of [this video](#), is a brief overview of this project. You may wish to watch from the beginning for context of my work with the work currently being conducted by the USGS.

To take part in the study, you would agree to a **30-45 minute** phone interview. During your interview you will be asked questions about your recent contact with and/or visit to a State Geological Survey’s sample repository. Questions will focus on your practices and the resources used in searching for physical sample materials.

Before beginning the interview, I will share with you a consent form which outlines your rights. This includes the following points:

- You may choose to skip any question at any point in time if you choose not to provide an answer.
- The questions presented by the researcher may lead to follow up questions but you may choose not to answer at any time.
- You may also choose the time for the call. If you opt to meet in person, you may also pick the location.
- With your permission the researcher will record the interview to facilitate note taking. You may choose to not be recorded and still take part in the study.
- All data collected will be reported as aggregated results and all raw data will be saved in a secure password protected location accessible only to the researcher.
- You will not receive any compensation for taking part in this study.

**Interview guide.**

The following is a list of **possible** questions that may be used to guide the conversation during the interview process. Please note, **this is just a guide and not all questions will be asked**. New questions may arise that are not part of this list, or will be added in the future based on data collected during initial interviews. The data gathered in this interview process will be de-identified and is being used to create a generalized model of information seeking behavior.

**Introductory questions**[^40]
**Search related/repository specific questions**
**General questions**

[^40]: Expandable lists. For complete list of questions see appendix G.
Example of information seeking model.

Categories for Ellis’s Models

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Scientists</td>
<td>Starting Chaining Browsing Extracting Monitoring Differentiating Ending</td>
</tr>
<tr>
<td>Chemists</td>
<td>Starting Chaining Browsing Extracting Monitoring Differentiating Verifying** Ending</td>
</tr>
<tr>
<td>Physicists</td>
<td>Initial Familiarisation Chasing Maintaining Awareness Source Prioritisation Locating**</td>
</tr>
<tr>
<td>Engineers and Research Scientists</td>
<td>Surveying Chaining Browsing Extracting Monitoring Distinguishing /Filtering* Ending</td>
</tr>
</tbody>
</table>

The categories are organized by how they match up based on activities. Source compiled from definitions in– Ellis, 1993; Ellis, Cox & Hall, 1993 and Ellis & Haugan, 1997

*These are two separate categories but both map to differentiating and source prioritization.

**These categories do not map to any other categories.
APPENDIX E: QUESTIONNAIRES

The following questionnaires will be distributed online using Qualtrics. The first will be emailed to all state geologists. The second will be sent out to those states which qualify for the study, i.e. maintain a sample collection of some kind.

State Geologists.

Introduction.
The Association of American State Geologists is supporting a study to investigate the users of state geological survey sample collections and repositories. The first set of questions is directed to you, the state geologists. The second set is intended for your survey’s sample repository manager(s) (or other position related to management of physical samples). At the end of this questionnaire you will be asked to provide contact information for them.

The questions will take between 5-10 minutes to answer. You may skip any question, at any time. All information gathered in this study will be de-identified. Specific institution names or other identifying information will not be reported.

For a more detailed description of this study, please see https://ramdeen.web.unc.edu/dissertation. Password: aasg

For questions or concerns about this study, please contact the principal investigator, Sarah Ramdeen via email at ramdeen@email.unc.edu or you may contact the UNC IRB and Office of Human Research Ethics at irb_questions@unc.edu

By clicking ‘Next’ you consent to taking part in this study. Participation is voluntary. You may skip any question at any time.

IRB # 15-1243
Sarah Ramdeen
Doctoral Candidate
School of Information and Library Science
University of North Carolina
http://ramdeen.web.unc.edu/

Questions.
What is the name of your geological survey? This information will only be used for determining participation, it will not be reported with the data.

In 2007, the National Geologic and Geophysical Preservation Program of the USGS sent out a version of this questionnaire as part of their first wave of grants. Did your survey complete that questionnaire?

☐ Yes
☐ No
☐ Not sure
Does your state survey have a sample library, repository or other similar collection? Samples include cores, cuttings, core chips, hand samples or other physical geological materials.

- Yes
- No

If No Is Selected, Then Skip To End of Survey

I Planning
For these next questions, consider your survey's sample library, repository or other similar collection.

<table>
<thead>
<tr>
<th>I.1 Does your state survey have any legal requirements for:</th>
<th>Yes</th>
<th>No</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Documentation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Preservation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Use of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I.2 Does your state survey have everyday policies and procedures for:</th>
<th>Yes</th>
<th>No</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Documentation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Preservation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Use of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I.3 Does your state survey have a long-range plan for:</th>
<th>Yes</th>
<th>No</th>
<th>Not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Documentation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Preservation of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Use of collection materials</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
II Staff

II.1 Approximately how many people at your state survey work in preservation, curation, and/or management of physical sample collections?

<table>
<thead>
<tr>
<th>Position Type</th>
<th>Number of employees (FTEs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Time</td>
<td></td>
</tr>
<tr>
<td>Part Time</td>
<td></td>
</tr>
<tr>
<td>Volunteer</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td></td>
</tr>
<tr>
<td>Other (specify)</td>
<td></td>
</tr>
</tbody>
</table>

II.2 For each position type, how many hours per month combined do these employees work in the area of preservation, curation, and/or management of physical collections?

II.3 Are the staffing levels suitable for your collections needs?
- Yes
- No
- Not sure

III Budget

For the next few questions, consider your most recent fiscal year.

III.1 What percentage of the state survey’s annual operating budget was designated for the care and management of physical sample collections? Please provide an estimate.

III.2 Was the state survey’s annual budget for care and management of collections sufficient?
- Yes
- No
- Not sure
### IV Collection needs
For the following questions, please select the choice which best completes the sentence when describing your collection.

<table>
<thead>
<tr>
<th></th>
<th>is in need of improvement</th>
<th>is adequate (is sufficient but does not fill all current needs)</th>
<th>is appropriate (is sufficient and meets current needs)</th>
<th>meet our current and future needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>The facility(s) for housing collections (such as warehouses or other storage locations) ...</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Storage space for collections ...</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Collection documentation ...</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>The preservation of our collections ...</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

### V Discovery and Access
For the following questions, please select the choice which best describes your collection.

<table>
<thead>
<tr>
<th></th>
<th>yes</th>
<th>no</th>
<th>some but not all</th>
<th>not sure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical sample collections are accessible to outside users.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Collections are cataloged in an electronic database.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Electronic data are accessible to users outside of your state survey.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Documents such as card catalogs or other non-digital materials are accessible to outside users.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
VI Wrap-up
The next part of this study should be completed by the individual(s) in charge of managing or overseeing the core/cuttings collections, library, or repository for your state survey. They will be emailed with a link to another questionnaire.

In order to communicate with this representative, please provide their contact information below. If needed, space has been provided for more than one representative. If you are the relevant contact person, please fill in your own information below.

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Email address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representative 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representative 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representative 4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion
Thank you for completing this questionnaire. By clicking the >> you will submit your responses.

For more information about this study, please see the study website https://ramdeen.web.unc.edu/dissertation (password: aasg) or contact the PI via email at ramdeen@email.unc.edu.
Repository managers.

Introduction.

You have received this questionnaire because your state geologist passed along your contact information and supported your participation in this research. Depending on the diversity of your collection, the questions will take between 15 - 25 minutes to answer.

All information gathered in this study will be de-identified. Specific institution names or other identifying information will not be reported.

For a more detailed description of this study, please see https://ramdeen.web.unc.edu/dissertation. Password: aasg

For questions or concerns about this study, please contact the principle investigator, Sarah Ramdeen via email at ramdeen@email.unc.edu or you may contact the UNC IRB and Office of Human Research Ethics at irb_questions@unc.edu.

By clicking ‘Next’ you consent to taking part in this study. Participation is voluntary. You may skip any question at any time.

IRB # 15-1243
Sarah Ramdeen
Doctoral Candidate
School of Information and Library Science
University of North Carolina
http://ramdeen.web.unc.edu/

Questions.

p.1 What is the name of your geological survey? This information will only be used for determining participation, it will not be reported with the data.

p.2 Please indicate the type of physical sample collections your institution holds. Include only collections that are a permanent part of your institution’s holdings.

- Cores (1)
- Cuttings (2)
- Core Chips (3)
- Sediment samples (4)
- Hand samples (5)
- Paleontological samples (6)
- Thin sections (7)
- Other (for multiple collections, separate entries with a semicolon). (8)

p.3 How extensive are your collections? (Rough estimates are acceptable, exact figures not required).OPTIONAL - please skip this question if it will be time consuming to answer.
You have indicated that you do not have any physical sample collections. Your institution is not eligible to participate in this study. If you have reached this screen incorrectly, please restart the survey by clicking your original reference link again.

Thank you.

For questions or concerns about this study, please contact the principle investigator, Sarah Ramdeen via email at ramdeen@email.unc.edu

The next set of questions is focused on your core, cuttings, core chip, and/or sediment sample collections.

Please consider all of these materials as a collective when answering these questions instead of individual collections.

I.1 What is the geographic scope of the collection? Please select the area which describes your general collection or typical item in your collection.
   - Collections divided by region or county within the state (1)
   - State only (2)
   - Localized region of the United States (3)
   - Representative of the wider United States (4)
   - Multi country collections (5)

I.2 Over the past five years, has your collection changed in size?
   - The collection has increased (1)
   - The collection has stayed the same (2)
   - The collection has decreased in size (3)

I.3 How well preserved is this collection?
   - The collection is in danger of being lost. (1)
   - Preservation is sufficient but does not fit all current needs. (2)
   - Preservation is sufficient and meets all current needs. (3)
   - Meets current and future needs. (4)

I.4 How well developed are the policies for this collection? (Based on the resources/funding available to you.)
   - The collection policies needs immediate attention and revision. (1)
   - Policies are sufficient but do not fit all current needs. (2)
   - Policies are sufficient and meet all current needs. (3)
   - Policies meet current and future needs. (4)

I.5 What estimated percentage of the collection is documented (metadata, catalogs, etc.)?
I.6 What estimated percentage of the collection documentation (metadata, records, catalogs) is accessible through an electronic database?

I.7 Is this database available openly on the Web? If so, please provide a link.
- Yes (1) ________________
- No (2)

I.8 How does your organization provide outside access to this collection? (select all that apply)
- An online database, catalog, or website open to the public. (1)
- Digital catalogs and records only accessible by internal staff. Patrons must request access (2)
- Physical catalogs or records, open to public (3)
- Physical catalogs or records, only accessible by internal staff. Patrons must request access. (4)
- Information about collections is not accessible to users outside of our organization. (5)

I.9 How is the collection used? Select all that apply.
- Research (1)
- Teaching (2)
- Reference (3)
- Private Sector (please specify) (4) ________________
- Other (please specify) (5) ________________

I.10 From the uses you selected, what would you say is the primary purpose of this collection?

II For the next set of questions, please consider your most recent fiscal year.

II.1 What is the estimated number of people who visited the collection over the course of the year?

II.2 How frequently do you get visitors in person to your collection?
- Daily (1)
- 2-3 Times a Week (2)
- Once a Week (3)
- 2-3 Times a Month (4)
- Once a Month (5)
- Less than Once a Month (6)
- Never (7)
II.3 Over the past 5 years, what is the long-term trend of usage of the collection?
☑ Increasing (1)
☑ Remaining constant (2)
☑ Decreasing (3)

II.4 Who are the outside users of the collection? (Please select all that apply)
☑ K-12 students and/or educators (1)
☑ General Public (2)
☑ University Students (3)
☑ Regulatory Agencies (4)
☑ Other Government Agencies (5)
☑ Professional Researchers (6)
☑ Private Sector (7)
☑ Other (8) ____________________
☑ No entry is permitted to this collection by outsiders. (9)

II.5 Over the past 5 years, have there been any changes to the variety of patrons who use this collection?
☑ Increasing - broader variety of user types (1)
☑ Remaining the same (2)
☑ Decreasing - narrowing focus of user types (3)

Q44 Was it difficult to answer these questions when considering them as a single collection instead of individual collections?
☑ yes (1)
☑ No (2)

Conclusion

In the second part of this study, I will be interviewing users/patrons of state geological survey collections from such communities as identified in this series of questions. If you chose to take part, I will ask you to pass along a recruitment message to anyone you think might be willing to take part in this study. I will NOT ask you for contact information of your patrons/users.

c.1 May I contact you about the second part to this study? If so, please provide your email address.
☑ Yes, please contact me about potentially taking part in the second half of this study. (1)
__________________
☑ No, my organization does not wish to take part. (2)
c.2 Thank you for completing this questionnaire. By clicking >> you will submit your responses.

For more information about this study, please see the study website: please see https://ramdeen.web.unc.edu/dissertation. Password: aasg
APPENDIX F: RECRUITMENT SCRIPT

Recruitment email – to repository managers

Re: Recruitment for educational study on information seeking

Thank you for agreeing to assist me in recruiting for the next stage of my research. In order to recruit participants, I am providing you with a brief message and I ask that you forward it along to visitors and potential users of your collections. This includes individuals who may have contacted you about your collections but not actually visited in person. For this study collections are defined as cores, cuttings, thin sections, fossils, and their associated data.

By you forwarding this email to your patrons, it will ensure the privacy of your visitors and will not require you to reveal names or email addresses to me.

Please let me know if you have any questions about this process. Your time and assistance is invaluable to me. I greatly respect the work you both do to support scientific research at the INSITUTION NAME.

Thank you again.
Sarah

Sarah Ramdeen Doctoral Candidate
School of Information and Library Science
University of North Carolina
ramdeen@email.unc.edu
http://ramdeen.web.unc.edu/

******************************************************************************

Recruitment message
******************************************************************************

Based on your previous communication with the INSITUTION NAME, you are being asked to take part in a study for educational research. The purpose of this research study is to better understand how scientists search for and access physical research materials such as cores, cuttings, thin sections, fossils, and their associated data.

To take part in the study, you would agree to a 30-45 minute phone interview with the researcher.

During your interview you will be asked questions about your recent contact with and/or visit to this state geological survey’s repository. Inquires at the survey do not need to be successful (i.e. result in a visit or acquisition of samples) in order to take part in this study. Questions will focus on your search practices and the resources used in locating physical sample materials for your research.
To take part in this study, please contact the researcher directly by email: Sarah Ramdeen ramdeen@email.unc.edu.

For more information about this study, please see her website:
https://ramdeen.web.unc.edu/dissertation/interviews-background/
Password: interview

Thank you.

*******************************************************************************
Message I will send to them in response to their inquiry
*******************************************************************************
Hello INSERT NAME

Thank you for your email and interest in this project. I am conducting interviews with researchers to better understand how scientists search for physical objects such as cores, cuttings, thin sections etc.

The interview will take between 30-45 minutes. In order to schedule a time, please send me a list of two or three possible hour time blocks in the next two weeks (INSERT DATE RANGE) which work best for you. To assist with scheduling, here is a link to my calendar (times are in EST)
https://calendar.google.com/calendar/embed?src=ptlakncip1dj5p23f775uvm6l4%40group.calendar.google.com

If you chose to take part, your interview will be conducted using GoToMeeting (accessible by phone or computer). Once we have a time scheduled, I will send you the call in information.

Finally, attached is a consent form. It outlines your rights (to skip questions, stop at any point in time etc.) and how I will protect your privacy. Please review it before our conversation, and let me know if you have any questions.

Thank you again for your interest! I look forward to speaking with you in the future.
Sarah

Sarah Ramdeen Doctoral Candidate
School of Information and Library Science
University of North Carolina
ramdeen@email.unc.edu
http://ramdeen.web.unc.edu
APPENDIX G: LIST SERVE RECRUITMENT

RE: Seeking stories about searching for core and cuttings at state geological surveys

*****************

Do you work with cores, cuttings or thin sections? Have you ever contacted a state geological survey to find these materials in their collections?

*****************

For my dissertation, I am researching how scientists search for physical geological materials such as cores, cuttings and thin sections. Participation involves a 30 minute interview about your experiences searching at state geological surveys.

To take part in this study, please contact me: Sarah Ramdeen ramdeen@email.unc.edu.

Inquires at state geological survey’s do not need to be successful (i.e. result in a visit or acquisition of samples) in order to take part in this study. Questions will focus on your search practices and the resources used in locating physical sample materials for your research.

For more information about my study, please see my website:
https://ramdeen.web.unc.edu/dissertation/interviews-background/

Please share this message with any individuals or communities that might be interested in participating.

Thank you!
Sarah

Sarah Ramdeen Doctoral Candidate
School of Information and Library Science
University of North Carolina
ramdeen@email.unc.edu
http://ramdeen.web.unc.edu
APPENDIX H: INTERVIEW GUIDE

Obtain consent.
*Verbal consent will be obtained from the participant.*
Earlier I emailed you a consent form. Do you have any questions about the information contained in the form?
During the interview I will be asking you a series of questions. Please note that your name and institution will not be associated with your answers in any materials resulting from this study. And any identifying materials will be destroyed, including the recordings once they have been analyzed.
Do you have any questions about the research study or the risks involved? Do you consent to be part of this study? Do you consent to being recorded during this study?

Interview guide.
*The following questions will be used to guide the conversation during the interview process. Not all questions may be asked. This list of questions will be used to guide the discussion and interview. New questions may arise that are not part of this list, or will be added in the future based on data collected during initial interviews. The data gathered in this interview process will be de-identified and is being used to create a generalized model of information seeking behavior.*

**Introductory questions – determining research interests**

- What field do you work in (describe from sampling protocol – identify specific domain/subfield/industry)?
- Describe your research area (in relation to your work with physical samples – may include clarifying questions to determine what role physical samples play in their work. What are you planning on doing with the material/knowledge you have gained?)
- What is your role within your organization (manager, researcher, team leader, decision maker etc.)?
- What types of projects do you work on that require physical samples?
- How frequently do search for existing sets of physical samples in your work?
- How important do you feel existing collections are to your work?
- How do you begin a project related to these topics? What is your first step?
- How do you keep up to date with developments relating to your research topics?
- How do you keep up to date with developments in your field?
- How would you approach the task of moving on to a new topic but in a closely related area?
- How would you approach the task of inducting a research assistant or research student into working with physical samples?
- Could you identify key ideas, authors, to send a research assistant or research student to?
- How would you approach the task of moving on to a topic in an area about which you knew nothing?
- What is the most difficult problem you experience in looking for material or keeping up to date?
- What criteria do you employ when assessing whether to follow up with material?
Search related/repository specific questions

- Let’s talk about a specific recent search. First, a search which you were successful in finding the information you need. To follow up – do you have an experience of a search where you felt you were unsuccessful? Can you tell me about that search process?
- What type of research were you conducting? Does that change the way you search?
- What were you looking for? Did you have a specific research project in mind or was this exploratory in nature?
- How did you find out about this repository?
- Have you used this particular repository before (the one recruited from)
- What resources did you use to find things before you arrived? For example (if needed): Recommendation from a colleague; Survey’s website/catalog; USGIN; National catalog; Publication citations; Etc.
- Who did you talk to while there (repository staff, other geologists etc.)
- Did you look at anything while there besides samples to help guide your search (like publications, catalogs etc. at the survey itself)
- Did you find samples to examine during your visit? Did you find samples that fit your information need?
- What problems did you encounter while searching 1) that are specific to the survey and 2) that are specific to searching for physical samples?
- How do you determine that you are successful in your searching?
- What would you do next after visiting the survey?
- Did visiting the survey satisfy your need for information?
- If it is not satisfied, where else would you look for information?
- Would you look to other repositories? And if so, where are they found (industry, federal government, private collections, academia etc.)
- Did you consider or look at other repositories before visiting the one that they were recruited from?
- When do you decide you might want to collect your own samples? Is this typically part of the process?

General questions (to be asked if the specific topics are not covered in the discussion above – tied to task and may not have come up. Here they are generalized beyond this specific search).

- What are the main sources of information for your work?
- Are there any sources which are of particular importance?
- Are there any distinctions between the sources or the material which are of particular importance to you?
- Which is the most important type of information source: books, journals, reports, conference proceedings, newspapers, etc.
- Which are the principal ways you have employed, or intend to employ to publish your own results?
- If it is intended to publish the results in journals are these the same ones as those followed?
- Do you follow up references cited in material consulted?
- How do you decide which references to follow and when to stop?
- How do you decide which references to cite in your own work?
- What is the most difficult problem you experience in looking for material or keeping up to date?
- What criteria do you employ when assessing whether to follow up material?
- Have you ever made use of any online databases to find samples? (Examples: DataONE, National Catalog, USGIN, repository specific examples).
- Did you find them useful? Why/why not?
- Have you ever used a publication to find physical samples? (Looking for data cited in an article to specifically reuse, for same or different research).
- Did you find it useful? Why/why not?
- Are there other tools or services you have used that have not been mentioned to discover data resources?
- Did you find the results useful? Why/why not?
- What limitations there are in this type of work (i.e. searching for physical samples to assist with research)?
APPENDIX I: INFORMATION AND CONSENT

Before the interview, participants will be sent the following consent form to review. Consent will be captured verbally during interviews, and recorded.

University of North Carolina at Chapel Hill
Consent Form Version Date: 6/10/2015
IRB Study # 15-1243
Title of Study: Information seeking behavior of scientists when searching for physical geological data
Principal Investigator: Sarah Ramdeen
Principal Investigator Department: School of Information & Library Science
Principal Investigator Email Address: ramdeen@email.unc.edu

What are some general things you should know about research studies?
You are being asked to take part in a research study. To join the study is voluntary. You may refuse to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. You may not receive any direct benefit from being in the research study. There also may be risks to being in research studies.

Details about this study are discussed below. It is important that you understand this information so that you can make an informed choice about being in this research study.

You should ask the researchers named above, or staff members who may assist them, any questions you have about this study at any time.

What is the purpose of this study?
The purpose of this research study is to build a basic model of the information seeking behaviors of scientists who use physical geological samples in their research and to better understand how they access these materials. This will provide a foundation for future research into the users of physical samples. The results of this study will include information about how participants look for existing physical samples collections, what resources they use to assist their searches, and how they react when they encounter barriers in their search.

You are being asked to be in the study because you have previously visited or contacted a state geological survey inquiring about physical sample materials when conducting work or research.

Are there any reasons you should not be in this study?
You should not be in this study if you do not use physical samples in your work or research.

How many people will take part in this study?
There will be approximately 15-30 people in this research study.
**How long will your part in this study last?**
Responding to questions in this study will only take between 30-45 minutes of your time.

**What will happen if you take part in the study (Individual)?**
If you take part in this study you will be part of a phone interview. During this interview you will be asked a series of questions. Afterwards you will be asked a few descriptive questions about your background in geology. You may choose to skip any question at any point in time if you choose not to provide an answer. The questions presented by the researcher may lead to follow up questions but you may choose not to answer at any time. You may also choose the time for the call. With your permission the researcher will record the interview to facilitate note taking. You may choose to not be recorded and still take part in the study.

**What are the possible benefits from being in this study?**
Research is designed to benefit society by gaining new knowledge. You will likely not benefit personally from being in this research study.

**What are the possible risks or discomforts involved from being in this study?**
There may be uncommon or previously unknown risks. You should report any problems to the researcher.

**How will your privacy be protected?**
In order to protect your privacy, no identifying information will be collected during this study. All data collected will be reported as aggregated results and all raw data will be saved in a secure password protected location accessible only to the researcher.

Participants will not be identified in any report or publication about this study. Although every effort will be made to keep research records private, there may be times when federal or state law requires the disclosure of such records, including personal information. This is very unlikely, but if disclosure is ever required, UNC-Chapel Hill will take steps allowable by law to protect the privacy of personal information. In some cases, your information in this research study could be reviewed by representatives of the University, research sponsors, or government agencies (for example, the FDA) for purposes such as quality control or safety.

**What if you want to stop before your part in the study is complete?**
You can withdraw from this study at any time, without penalty. The investigators also have the right to stop your participation at any time. This could be because you have decided that you do not wish to answer the questions posed in this study or if you do not have the time to complete the study.

**Will you receive anything for being in this study?**
You will not be receiving anything for taking part in this study.

**Will it cost you anything to be in this study?**
It will not cost you anything to be in this study.
What if you have questions about this study?
You have the right to ask, and have answered, any questions you may have about this research. If you have questions about the study, complaints, or concerns, you should contact the researcher listed on the top of this page.

What if you have questions about your rights as a research participant?
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, or if you would like to obtain information or offer input, you may contact the Institutional Review Board at 919-966-3113 or by email to IRB_subjects@unc.edu.

Participant’s Agreement:
Do you have any questions about the study or your rights? Do you consent to taking part in this study?
APPENDIX J: EXAMPLE OF SEARCH BEHAVIOR

Below are two examples of requests for information sent by a researcher sent to a list serve.

Example one.
Dear colleagues,

If anyone can help to identify where this fossil might originally be from it would be most useful (photo enclosed). Please reply to me off-list.

The specimen shown in the enclosed photo is about 76cm wide in its longest axis and is a slab of rock preserving multiple trackways of trilobites.

The specimen was originally in Wigan and Leigh College Museum which closed down a few years ago. They did not have any information associated with it when it was passed on to Manchester Museum, where it now resides. There is an assumption that it was probably from Wales originally. If anyone has seen something very similar and has the provenance details the information would be most welcome.

Thank you for your time, it is appreciated.
Example two.

Hi

I’m looking for specimens of ‘black lithics’ (e.g. lignite, jet, oil shale, cannel coal, manjack) from the Caribbean, Venezuela, Columbia and Guyana to use as comparative reference material for a study of prehistoric Caribbean carvings.

Ideally I'm looking for material that I could loan for non-destructive analysis using X-radiography/micro-CT and XRF. If I could also take (very) small samples for FTIR that would be even better.

I’m aware of specimens at the Natural History Museum, London, and the Oxford University Museum of Natural History, but I’d be really grateful if anyone could let me know of any other collections (however large or small)

Many thanks,
REFERENCES


Showstack, R. (2014), Geological Data Preservation Program Receives Bipartisan Support, Eos Transactions. AGU, 95(39), 350


Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PloS one, 8*(7), e67332.


