DIGITAL CURATION THROUGH THE LENS OF DISCIPLINARITY

THE DEVELOPMENT OF AN EMERGING FIELD

A Dissertation presented to the
Faculty of the Simmons College School of Library and Information Science

In partial fulfillment of the requirements for the Degree of Doctor of Philosophy

By Patricia Beasley Condon, BA, MA, MLIS

©2014
Patricia Beasley Condon
All rights reserved
DIGITAL CURATION THROUGH THE LENS OF DISCIPLINARITY

THE DEVELOPMENT OF AN EMERGING FIELD

A Dissertation presented to the

Faculty of the Simmons College School of Library and Information Science

In partial fulfillment of the requirements for the Degree of Doctor of Philosophy

By Patricia Beasley Condon, BA, MA, MLIS

©2014
Patricia Beasley Condon
All rights reserved

Jeannette Bastian, PhD, Chair                  Dec. 16, 2014
Date

Ross Harvey, PhD                             17 December 2014
Date

Martha Mahard, DA                             19 Dec 2014
Date

Nancy McDovern, PhD                          Dec 16, 2014
Date
DIGITAL CURATION THROUGH THE LENS OF DISCIPLINARITY
THE DEVELOPMENT OF AN EMERGING FIELD

Patricia Beasley Condon
Simmons College School of Library and Information Science

ABSTRACT

Digital curation is both maturing within the information disciplines and becoming increasingly embedded in practice. We are observing an increase in employment opportunities, education and training, and research in the area of digital curation. However, it is still unclear how and where the transmission of this knowledge and skills set fits within higher education. The purpose of this study was to explore the development of digital curation as an academic field by conducting an in-depth analysis of how this area is evolving. The research questions addressed were: Is digital curation emerging as an autonomous discipline? Where does digital curation fit within the educational landscape?

The methodologies employed were scoping the literature, content analysis of published literature in the area of digital curation, and interviews with individuals engaged in the area. In this dissertation, the conceptual model put forth by D’Agostino (2012), which views a discipline as the interaction of ten elements that characterize a discipline interpreted within a framework of “shallow consensus” was used. Five key themes emerged from the data analysis: terminology, collaboration, multiple discipline engagement, education, and areas of professional and scholarly focus. Findings suggest that digital curation has not emerged as an autonomous discipline, but does meet several of the criteria to indicate its potential for emergence. Although education for this area fits well in coordination with the information disciplines, skill development is important across all domains. This study provided markers for gauging the progress of digital curation as an emerging field.
ACKNOWLEDGEMENTS

I would like to thank my dissertation committee for their support throughout this project. I thank Jeannette Bastian for chairing the committee and guiding me through the doctoral program; Martha Mahard for the kindness she showed me; and Nancy McGovern for our many conversations even when they were not directly about this dissertation. I give a special thanks to Ross Harvey for his mentorship, candor, and thoughtfulness throughout this project and my time at Simmons.

A sincerest thank you to my interview participants, each of whom was very generous with her/his time and thoughtful in her/his remarks. Not only were our conversations central to this research, I enjoyed having them and hope they will continue in the future.

I truly appreciate the doctoral fellowship from the IMLS-funded Building the Future of Archival Education and Research Initiative and the financial support provided by the School of Library and Information Science at Simmons College.

Thank you to the faculty at SLIS, the AERI community, and my fellow doctoral students at Simmons for moral support and engaging discussions: shout-outs to Eliot Wilczek, Heather Soyka, and Katherine Ruffin. Thank you to Jennifer Andrews, Deanna Beattie, Lauren Wallace, and the doctoral program committee for all the behind-the-scenes support that makes things like this happen. Thank you to the staff at SLIS, the College’s IRB, Simmons Beatley Library, and UNH Dimond Library.

Thank you to my friends – if you’re reading this, that’s you. Thank you to my wonderful family who have always provided me with love and a good laugh: Jim and Gail; Katie, Tim, Emma, Oliver, and Elliot; Jimmy and Sacha; Lee and Tony; Atticus, Harriet, and Winston; and Miles and Hester. Most of all, thank you to Zack Irons for his patience, fortitude, and encouragement during this adventure.

Each of you has shared with me a unique and valuable perspective, which has inspired the pages that follow.
# TABLE OF CONTENTS

ABSTRACT ................................................................................................................................. I
ACKNOWLEDGEMENTS .............................................................................................................. II
TABLE OF CONTENTS ................................................................................................................ III
LIST OF TABLES ........................................................................................................................ VI
LIST OF FIGURES ....................................................................................................................... VIII

## CHAPTER 1: INTRODUCTION .................................................................................................. 1

STATEMENT OF THE PROBLEM ............................................................................................... 2
PURPOSE OF THE STUDY AND RESEARCH QUESTIONS ....................................................... 3
CONCEPTUAL FRAMEWORK ..................................................................................................... 4
OVERVIEW OF RESEARCH DESIGN ....................................................................................... 7
  Limitations ............................................................................................................................... 9
  Delimitations .......................................................................................................................... 10
SIGNIFICANCE OF STUDY ...................................................................................................... 11
SUMMARY OF CHAPTERS ........................................................................................................ 13

## CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY ...................................................... 15

RESEARCH DESIGN .................................................................................................................. 16
METHODOLOGY ........................................................................................................................ 18
  Scoping the Literature ......................................................................................................... 20
  Description of the Sample .................................................................................................... 23
  Interviews .............................................................................................................................. 23
  Description of the Sample .................................................................................................... 25
  Interview Procedure and Protocol ...................................................................................... 26
  Ethical Considerations .......................................................................................................... 27
  Content Analysis .................................................................................................................. 29
  Description of Sample ......................................................................................................... 29
  Data Collection ..................................................................................................................... 33
  Trustworthiness ................................................................................................................... 33
  Pilot Studies .......................................................................................................................... 35
DESCRIPTION OF DATA ANALYSIS ....................................................................................... 37
SUMMARY ................................................................................................................................. 40

## CHAPTER 3: UNDERSTANDING ACADEMIC DISCIPLINES ................................................ 41

ACADEMIC DISCIPLINES ........................................................................................................ 41
  Toward a Definition of Discipline ....................................................................................... 42
THEORIES OF DISCIPLINARITY ............................................................................................ 45
  Disciplines through the Lens of Cultural and Social Theories ............................................. 46
Metaphors of Disciplinarity .................................................................................................................. 49
Frameworks for Studying Disciplinary Differences .............................................................................. 51
Emerging Disciplines ............................................................................................................................ 60
Criteria in the Literature ........................................................................................................................ 64
On Establishing Definitive Criteria ........................................................................................................ 67
Conceptual Framework for this Research ................................................................................................ 70
Summary .................................................................................................................................................. 73

CHAPTER 4: CONTEXTUALIZING DIGITAL CURATION .............................................................................. 75
Foundations of Digital Curation ............................................................................................................. 75
Archival Community ................................................................................................................................ 76
Library and Information Science Community .......................................................................................... 79
Research Communities ............................................................................................................................ 84
The Preservation and Curation Communities .......................................................................................... 90
The Preservation Community .................................................................................................................. 90
The Curation Community ......................................................................................................................... 94
Summary .................................................................................................................................................. 95

CHAPTER 5: DIGITAL CURATION IN THE LIS/ARCHIVE CURRICULA .................................................. 97
Emerging Roles and Responsibilities ....................................................................................................... 97
Identifying Core Competencies ............................................................................................................... 101
Curriculum Development ....................................................................................................................... 104
Continuing Education and Professional Training .................................................................................. 110
Pedagogy .................................................................................................................................................. 113
Outlook .................................................................................................................................................... 115
Summary .................................................................................................................................................. 117

CHAPTER 6: FINDINGS AND DISCUSSION: PART I ................................................................................ 118
Theme One: Terminology ....................................................................................................................... 120
Terminology: Scoping the Literature ....................................................................................................... 121
Terminology: Interviews .......................................................................................................................... 126
Terminology Explored ............................................................................................................................. 129
Theme Two: Collaboration ....................................................................................................................... 133
Collaboration: Content Analysis ............................................................................................................ 134
Collaboration: Interviews ......................................................................................................................... 138
Summary .................................................................................................................................................. 142

CHAPTER 7: FINDINGS AND DISCUSSION: PART II .............................................................................. 143
Theme Three: Multiple Discipline Engagement .................................................................................... 143
Multiple Discipline Engagement: Content Analysis Subset ................................................................. 144
Multiple Discipline Engagement: Interviews .......................................................................................... 148
# Multiple Discipline Engagement Explored

## Theme Four: Education

**Education: Interviews**

**Education Explored**

**Summary**

## Theme Five: Areas of Professional and Scholarly Focus

**Research Agendas**

**Summary of Findings and Discussion**

## Chapter 9: Conclusion and Future Research

**Summary of Research**

**Research Questions**

*RQ1: Is digital curation emerging as an autonomous discipline?*

*RQ2: Where does digital curation fit within the educational landscape?*

**Implications of Findings**

**Discussion of Limitations**

**Conclusion**

**Future Research Directions**

## References

## Appendix A: Scoping the Literature

**Search Terms**

**Databases and Search Fields**

## Appendix B: Interview Guide

## Appendix C: Coding Schemes

**Fixed Codes for Content Analysis**

**Final Emergent Codes from Interview Transcripts and Content Analysis**

## Appendix D: Content Analysis Bibliographies

**Conferences Presentations**

*International Journal of Digital Curation*  

*Top 5% Most Cited (Scopus and Web of Science)*
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table 1.1</th>
<th>Conceptual framework: Summary of elements and criteria that comprise a discipline as a cultural system (D’Agostino, 2012)</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Scoping the literature: Summary of the methodology</td>
<td>22</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Interviews: Description and identification of interview participants</td>
<td>25</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Interviews: Interview protocol</td>
<td>28</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Content Analysis: <em>International Journal of Digital Curation</em>, 2006-2012</td>
<td>30</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Content Analysis: Paper and poster presentations from published proceedings, 2001-2012</td>
<td>31</td>
</tr>
<tr>
<td>Table 2.6</td>
<td>Content Analysis: The top 5% most frequently cited articles from Scopus and Web of, 2001-2012</td>
<td>33</td>
</tr>
<tr>
<td>Table 2.7</td>
<td>Trustworthiness: Comparison of criteria for assessment of rigor (Krefting, 1991; Guba, 1981).</td>
<td>34</td>
</tr>
<tr>
<td>Table 2.8</td>
<td>Pilot interviews: Participants</td>
<td>37</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Representative areas and issues commonly found within disciplines (Phillips, 2003; Phillips, 2005).</td>
<td>70</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Conceptual framework: Summary of the elements that comprise a discipline as a cultural system (D’Agostino, 2012).</td>
<td>72</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>Conceptual Framework: Elements of a discipline mapped to the themes emerging from the findings</td>
<td>119</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>Scoping the literature: Search results</td>
<td>121</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>Content analysis sample: Authorship of the collaborative articles that were authored by multi-discipline teams</td>
<td>136</td>
</tr>
<tr>
<td>Table 7.1</td>
<td>Content analysis subsample (top 5% most cited): Publication titles for articles in the information field (n=53)</td>
<td>146</td>
</tr>
</tbody>
</table>
Table 7.2  Content analysis subsample (top 5% most cited: The number of articles citing references from publications of different fields of study)

Table 7.3  Content analysis subsample (top 5% most cited): Articles (n=170) from Scopus and Web of Science (2001-2012) by search term and field of journal

Table 7.4  Examples of terminology used for programs in higher educational institutions

Table 9.1  Summary of disciplinary elements for digital curation and markers of progress in the field.

Table A.1  Scoping the literature summary: Article databases used in scoping the literature
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1</td>
<td>Comparison of conceptual frameworks of disciplinary differences</td>
<td>59</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Scoping the literature sample: Number of articles per search term by year</td>
<td>123</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Scoping the literature sample: Number of articles per search term by year normalized as percentages</td>
<td>123</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Content analysis sample: Articles by number of authors</td>
<td>134</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Content analysis sample: Multi-authored articles by type of collaboration</td>
<td>135</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Content analysis subsample (top 5% most cited): Percentage of articles by the field of the journal in which they were published</td>
<td>145</td>
</tr>
<tr>
<td>Figure 8.1</td>
<td>Content analysis sample: The six areas of professional and scholarly focus and their relationships</td>
<td>168</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Digital curation is both maturing within the information disciplines and becoming increasingly embedded in practice. Defined broadly in this study, digital curation involves the active lifecycle management of data for current and future use (the definition of digital curation is discussed in more detail in the Delimitations section of this chapter, the Curation Community section in Chapter Four, and in the Terminology section in Chapter Six). As the creation and use of digital information become ubiquitous, it is increasingly acknowledged that data require lifecycle management for their long-term accessibility and preservation. Data are defined by the United Kingdom-based Digital Curation Centre (DCC) as “any information in binary digital form” including databases, simple digital objects such as text or image files, and complex digital objects such as websites (Digital Curation Centre, n.d.a). This definition encompasses materials such as digital cultural heritage; research and scientific data in the natural, physical, and social sciences; and business, government, and personal digital information.

The area of digital curation has emerged rapidly within professional and research practice, more rapidly than higher education curricula or professional training have been able, for the most part, to establish coherent approaches to instruction. As practitioners and researchers develop fundamental best practices and procedures for the lifecycle management of data, the educational and skill requirements for professionals in digital curation and related areas continues to be defined and refined (Kim, Warga, & Moen, 2013; Osswald, 2013). There has been progress in developing core curriculum models for higher education and professional training programs. Most notably, comprehensive curriculum frameworks were developed by two international projects, the Digital Curator Vocational Education Europe project (DigCurV, http://www.digcur-education.org), and the Digital Curation Curriculum Project (DigCCurr,
http://www.ils.unc.edu/digccurr) at the University of North Carolina at Chapel Hill, both which concluded in 2013.

Within archival studies, and library and information science (LIS) programs, there are expanding degree-granting and continuing education programs offering advanced certificates, concentrations, courses, and training in the area of digital curation. In fact, I found no less than 10 million dollars in grants has been awarded from 2010 to 2014 by the Institute of Museum and Library Services (IMLS) through the Laura Bush 21st Century Librarian Program for projects in the United States that are specifically designed to improve training opportunities and build curricula for continuing and higher education programs in curation and management of materials in a digital environment (http://www.imls.gov/recipients/grantsearch.aspx). This sum only includes projects related to the development and improvement of education and training for digital curation and related areas. There were numerous other funded projects involving implementation and applied research.

**Statement of the Problem**

Some indicators suggest that digital curation is emerging as an area distinct from other fields or disciplines. There have been advances in best practices, visible growth in higher and continuing education curricula offering both coursework and training in digital curation, and considerable grant funding for applied research, educational programming, and practical implementation in digital curation and related areas. In fact, a 2011 article published in the *International Journal of Digital Curation* (Higgins, 2011), proposes that digital curation has emerged as a new discipline. In the article, Higgins outlines the emergence of a digital preservation infrastructure in the United Kingdom starting in the mid-1990s, the steady increase of international preservation development projects and agendas, and the growth of education and
training opportunities in digital curation. Higgins suggests that “through the iterative workshop and agenda setting process” (p. 84) digital curation has made great strides in establishing itself as a new discipline.

However, identifying digital curation as a newly emerged discipline may be premature. Becoming a discipline requires a certain degree of embeddedness and institutionalization within higher education. Digital curation is at least equally embedded in practice. Although there is a visible growth in educational programs, it is still unclear how the education for this knowledge and skills set fits within the educational landscape. The answer to this uncertainty hinges on whether digital curation is emerging as, for example, a discipline in its own right, a specialization within the information disciplines, or as a set of interdisciplinary best practices employed by those who work with digital assets. No study has provided an in-depth analysis into this matter. Insight gained from such a study will be useful for stakeholders, including educators and practitioners, seeking to understand the environment in which the knowledge transfer of competencies related to digital curation takes place.

**Purpose of the Study and Research Questions**

This dissertation research is designed to explore the educational landscape of digital curation and investigate whether the area of digital curation is emerging as an independent discipline in its own right. This study focuses on identifying themes that emerge from the academic discourse and scholarly communication, and understanding the level of consensus among educators, practitioners, and researchers about how they view this domain. Its purpose is to explore the character, development, and education of the area of digital curation by conducting an in-depth analysis of how this area is evolving.
The research questions addressed in this study are:

RQ1: Is digital curation emerging as an autonomous discipline?
   a. In what ways do interviewees describe the relationship of digital curation to areas within the information disciplines?
   b. Are there indicators that suggest digital curation is emerging as a discipline in its own right?

RQ2: Where does digital curation fit within the educational landscape?
   a. To what extent and in what ways do interviewees express the role of education in the knowledge transfer of digital curation practices, skills, and theories?

**Conceptual Framework**

Academic disciplines are the interaction of intellectual, organizational, and social elements. In this research, I use the definition of discipline presented by Del Favero (2003). Academic disciplines are defined as:

Providing the structure of knowledge in which faculty members are trained and socialized; carry out tasks of teaching, research, and administration; and produce research and educational output. Disciplinary worlds are considered separate and distinct cultures that exert varying influences on scholarly behaviors as well as on the structure of higher education (p. 10).

This definition captures the complexity of a discipline and highlights the influence disciplinarity has on how members of the discipline are educated, socialized, and perform their research and teaching responsibilities. In addition, it emphasizes the institutionalization of disciplines and the impact that differences among disciplines have on the structure of higher education.
Emerging disciplines are characterized by organizational instability, a lack of coherent discourse, and a struggle for legitimacy within higher education. Governance and leadership are not clearly defined. The intellectual, research, and teaching domains “often sit uncomfortably between existing academic institutional boundaries” (Kimbell, 2009, p. 3). The power and social structures of emerging disciplines are viewed as unstable, marked by a lack of consensus and shared norms (Maguire, Hardy, & Lawrence, 2004).

There exists an extensive body of research on academic disciplines. The process of convergence, differentiation, fragmentation, and specialization of disciplines into subdisciplines and new disciplines is a complex process involving intellectual, organizational, and social changes. Criteria for identifying that a field has emerged as a new discipline are not well-established; however, several researchers have put forward useful sets of criteria. Although the criteria vary in presentation, they provide similar elements of assessment that take into account institutional and intellectual aspects that aim to get at the dynamic qualities of a discipline.

In this dissertation, I use the conceptual model put forth by D’Agostino (2012) that views a discipline as a cultural system: the interaction of elements that characterize a discipline interpreted within a framework of “shallow consensus” (p. 341). In D’Agostino’s model, a discipline is made up of an intellectual field of inquiry and institutional infrastructure. D’Agostino derived from the literature a set of ten elements that comprise a discipline within these two domains (Table 1.1). This set of elements “represents, if nothing more, a good starting point for further enquiry, implying, as it does, a checklist of questions which we might raise about a discipline or ‘wannabe’ discipline” (p. 337). The ten elements each contain more granular observable criteria. In presenting these elements, D’Agostino is “claiming that there is wide acceptance of, commitment to, or use of, these elements by a community of enquiry whose
existence as a community is, in turn, based on this consensus in attitudes, values and activities” (p. 340).

Table 1.1

*Conceptual framework: Summary of elements and criteria that comprise a discipline as a cultural system (D’Agostino, 2012)*

<table>
<thead>
<tr>
<th>Element</th>
<th>Institutional Infrastructure</th>
</tr>
</thead>
</table>
| **Academic Organizational Unit** | Hires and promotes staff  
|                                | Allocates workload  
|                                | Reproduction of discipline through graduates  
| **Curriculum in Higher Education** | Instruction of disciplinary knowledge  
|                                | Identifies graduate attributes  
|                                | Engagement with threshold concepts  
|                                | Signature pedagogies  
| **Professional Association**   | Conferences  
|                                | Awards and prizes  
|                                | Scholarly journals (supported by the professional association)  
|                                | Code of ethics  
| **Publishers**                 | Scholarly journals  
|                                | Reference books  
|                                | Textbooks  
|                                | Research monographs  
| **Recognition in Classificatory System** | Recognition within a classificatory system, such as bibliographic classification scheme  
| **Social Networks**            | Patronage networks  
|                                | Gatekeepers who control access to opportunities  
|                                | Audiences for dissemination of work  

**Intellectual Field of Inquiry**

| Accumulated Knowledge | Agreed intellectual domain  
|                       | Research agenda  
|                       | Characteristic methodology  
| Discursive Community  | Shared field-specific terminology  
|                       | Norms for scholarly writing  
|                       | Accepted ways of displaying competencies  
| Narrative of the Field’s Development | Heroes, leaders, and founding figures  
|                                | Ceremonies that support community-building  
| Styles of Subjectivity | Value criteria for assessing work  
|                       | Concept of rigorous research  
|                       | Demeanor that identifies members of discipline  

D’Agostino contends that although we can define a discipline in terms of it meeting the criteria, we must also contextualize the development of the discipline historically and accept that members of the discipline may have varying interpretations of what each element embodies. D’Agostino uses the phrase shallow consensus to denote the varying degrees of agreement that members of the same discipline share about their discipline or aspects of their discipline. Shallow consensus permits a definition of disciplines using rigid criteria and still tolerates internal variation within the discipline. It is within this social framework of shallow consensus and the historical context of the area’s development that we understand how the elements of a discipline came to be organized and represented as they are. Chapter Four provides an historical narrative of the development and emergence of digital curation. This narrative contextualizes the themes and their corresponding disciplinary elements that are presented in the Findings and Discussion chapters.

**Overview of Research Design**

This dissertation employs a qualitative multimethod research design. The methodologies used in this dissertation were scoping the literature, content analysis of published literature in the area of digital curation, and interviews with individuals engaged in the area. First, I conducted a broad, comprehensive, and systematic survey of the literature related to the area of digital curation. Next, I conducted content analysis using a standardized coding instrument on a selection of the literature surveyed (n=460). Then, I conducted fourteen interviews with individuals identified from scoping the literature. Lastly, the units of analysis from the content analysis and interview transcripts were coded using an open coding framework in NVivo qualitative analysis software.
The initial step in the data collection process was a broad and systematic scoping the literature related to the area of digital curation. This survey was designed to be comprehensive in order to gain a holistic representation of the literature associated with the development and current practice of the area of digital curation. There were four objectives of this survey. First, I used the survey to broadly describe and characterize the literature relating to the area of digital curation. Second, I used the results of the survey to delineate the population from which the sample was taken for the content analysis. Third, I used the results of the survey to identify participants for interviews. Lastly, I used the literature that was collected during the survey to compile a narrative of the development of the area and conduct a literature review on education and curriculum development in the area of digital curation.

From scoping the literature, I identified three subsets of literature for the content analysis. The content analysis probes in depth the International Journal of Digital Curation (n=237; 2006-2012), conferences presentations (n=53; 2001-2012; search terms digital curation and digital stewardship), and the top 5% most frequently cited articles in Scopus and Web of Science (n=170; 2001-2012; search terms: data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation, and digital stewardship). For the content analysis, a total of 460 articles, editorials, conference posters and presentations, and research papers were coded. The data from the content analysis were recorded using a standardized spreadsheet coding instrument.

From scoping the literature, I identified participants for the interviews. Three selection criteria were applied to identify the participants: interviewees must have displayed active engagement in the area of digital curation discourse and community within the past year; interviewees must have been involved or working with digital materials in some capacity related
to digital curation for at least five years; and collectively, participants had to make up a diverse, international panel. Fourteen participants were interviewed, at which point no new or relevant information was emerging from the interviews. The interview protocol was based on a set of predetermined open-ended questions that were flexible enough to allow for follow-up questions, probing, and pursuing topics that naturally arose during the interview but structured enough to address the object of this research. Interviews were conducted via telephone (3) or Skype (11) and each took approximately 60 minutes.

Data collection, analysis, and discussion of the research questions were structured using the conceptual framework as a guide. Interview transcripts and the abstracts from the content analysis samples were coded using NVivo. The coding scheme included both emergent (data-driven) and pre-defined (research- and theory-driven) codes. For the analysis, I mapped the emergent codes to the pre-defined codes that were defined by the conceptual framework. This was in addition to the fixed bibliometric coding scheme that was used for the content analysis as mentioned above.

Limitations

There are four limitations that have the greatest potential impact on the quality of the data and their effectiveness at addressing the research questions. First, purposive sampling was used in this study rather than a randomized sampling method. Non-probability sampling restricts generalization of the findings because the sample is not necessarily representative of the population. Purposive sample is common in qualitative research designs. Second, the sample for the content analysis is limited by the materials available. Discourse in the area of digital curation includes community-based publications and grey literature. They are more difficult to locate via conventional searching methods and not accessible in citation databases. This has the
potential to impede the selection of units of analysis for the content analysis as well as the
selection of possible participants for the interviews. Third, in the library and information science
fields publishing trends differ between scholars and practitioners (Schlögl & Stock, 2008;
Haddow & Klobas, 2004; Powell, Baker, & Mika, 2002), and this difference may very well be
present in digital curation publications. Additionally, publishing trends differ among disciplines.
Because the sample selection for both the content analysis and the interview participants relies
heavily on published materials, these differences have the potential to impact the sample. Lastly
all interviews were conducted in English and samples for the content analysis were limited to
articles or presentations written in English. These limitations are discussed in more detail in
Chapter Nine.

Delimitations

I have chosen to investigate whether digital curation is emerging as a discipline rather
than whether it is, for example, a community of practice (Lave & Wenger, 1991; Wenger,
McDermott, & Snyder, 2002), an emergent profession (or semi-profession) (Abbott, 1988;
Freidson, 1999; Bastian & Yakel, 2005), or a field of research (Fensham, 2004). Several key
criteria for digital curation to be considered a profession are visibly missing, specifically the
existence of a professional association and community sanction either by licensure or educational
requirements (Chatterjee & Stevenson, 2008). Researching digital curation as an emergent
discipline allows for its professionalization and status as a community of practice or field of
research to be at least tangentially addressed in this study. Additionally, focusing on
disciplinarity embeds the discussion of digital curation within higher education.

I have chosen to approach this research using the term digital curation as designation of
an area, which is sometimes defined or referred to using terms such as data curation, data
stewardship, digital preservation, and digital stewardship. Each of these terms expresses a particular nuance in their meaning and use, which I explore in this study. I define digital curation broadly as involving the active lifecycle management and preservation of data for current and future use. Using a broad definition allows me to capture the breadth and depth of the area without getting distracted by ambiguous terminology. The year 2001 was selected for scoping the literature and content analysis because that is the year that the term “digital curation” was coined (Beagrie, 2006). This date is significant if we assume that Kuhn’s concept of paradigm and lexicon bears some relevance – the existing taxonomy was not adequate to explain the phenomenon and therefore a new term was coined to account for the changes that were occurring. Additionally, although the relationship between digital curation and digital libraries, digitization, or electronic records management will be explored, articles about these three topics without some reference to the larger vision of digital curation are not included in the analyses.

Significance of Study

Reflection on the scholarly and societal position of emerging fields is a familiar theme. Some fields such as communication, cultural studies, nursing, and women’s studies have lengthy histories of reflexivity about the nature of their status as an independent discipline within higher education often emphasizing that their field possesses standard characteristics of a discipline (see During, 2006; Craig, 2008; Boxer, 2000). More established fields, although perhaps still new compared to the traditional natural and social sciences, often focus on exhibiting the historical development and legitimacy of their discipline rather than on demonstrating that standard elements of a discipline are present (see Bender & Schorske, 1998; Lemaine, MacLeod, Mulkay, & Weingart, 1976).
The information disciplines (museums studies, archives and records management, and library and information science) are not immune to this reflection. For the information disciplines, this reflexivity about the nature of their position within higher education often stems from the challenges of education for a practical or professional field in general. Library science has long struggled with not only its identity as a discipline within higher education but also with its identity as a profession (Bates, 2012; Bates, 2010; Audunson, 2007; Goode, 1961). Archives and records management are also poised between expanding as a profession, or semi-profession, and an academic discipline (Shepherd, 2012; Bastian & Yakel, 2006; Cox, 1986). Indeed, Bates (1999, 2012) argued that “information science need(s) to be seen as a different type of discipline, in comparison to the usual array of disciplines” (Bates, 2012, p.2). She used the term “meta-discipline” to describe the information disciplines.

The area of digital curation faces a similar quandary. Digital curation may be regarded as an emerging discipline, a specialization within an existing discipline, a broadening of a current field such as preservation, or a set of interdisciplinary best practices. Depending on how digital curation is perceived by its community and how it is positioned within higher education and practice determines important features of the education and training of digital curation practices, skills, and theories. Aspects such as the environment in which it is taught, for example continuing education and training workshops or degree-granting higher education programs; which disciplines or fields are most aligned with the area; and who is responsible for instruction and who needs to receive instruction are ascertained by understanding how the area is maturing. This study provides insight into the development of the area of digital curation.

The impact of having a field identified as an independent discipline can have many significant advantages. For example, it increases the stability and legitimacy of the field within
higher education institutions; it allows for greater potential to obtain financial resources; and it provides the discipline with the autonomy to develop curriculum, train students, and select doctoral students for the continuation of the discipline. This study is useful for stakeholders who seek to understand the current educational landscape of digital curation, and its position among the information disciplines and within higher education.

Additionally, this research contributes to the literature about the emergence of disciplines and furthers our understanding of the educational complexities accompanying that potential emergence. Although the focus of this study is an in-depth analysis of digital curation, it approaches the analysis uniquely by first applying a recently published model of disciplinarity and then honing in on what the current identity of digital curation means for knowledge transfer.

**Summary of Chapters**

Chapters One and Two introduce the study and research methodologies. Chapter Two (Research Design and Methodology) explains the research design and methodologies employed in this study. Additionally, the procedures for data collection and analysis, descriptions of the samples, and issues of trustworthiness are explained.

Chapters Three, Four, and Five provide background and contextual information that situate the findings and discussion. Chapter Three (Understanding Academic Disciplines) discusses the previous literature related to theories of disciplinarity. The aim of this literature review is to provide a context for understanding disciplines, exploring what it means to be an emerging discipline, establishing criteria for identifying disciplines, and providing support for the conceptual framework used in this study. Chapter Four (Contextualizing Digital Curation) provides an historical perspective on the emergence of digital curation. The function of this chapter is to give a context for the findings of this study by situating the observed disciplinary
elements within the development of the field. Chapter Five (Digital Curation in the LIS/Archive Curricula) presents a review of relevant literature about education and curriculum development in the area of digital curation, embedding the discussion within higher education.

Chapters Six, Seven, and Eight (Findings and Discussion, Parts I, II, and III) present the key findings and discussion. Five major themes emerged from the data analysis: terminology, collaboration, multiple discipline engagement, education, and areas of professional and scholarly focus. The five themes were mapped to the elements of the conceptual framework, and were used to support the conclusions presented in the last chapter.

Chapter Nine (Conclusion and Future Research) addresses the research questions, draws conclusions based on the findings of this study, and offers future research directions.
CHAPTER 2: RESEARCH DESIGN AND METHODOLOGY

In this chapter I describe the research design and methodology used in this dissertation. This study employed a qualitative multimethod research design. The methodologies employed were scoping the literature, content analysis of literature related to digital curation themes, and interviews with people engaged in the area of digital curation. I developed the coding instrument and interview protocol to collect data about the elements that characterize a discipline based on the criteria established by the conceptual framework that was identified from the literature review (see Chapter Three).

Principles and practices related to the lifecycle management and long-term preservation of data are increasingly relevant for information professionals and researchers working with digital assets. We are observing an increase in employment opportunities, education and training, and research in the area of digital curation. Some indicators suggest that digital curation is more than a set of practices and skills; rather it is emerging as an autonomous discipline. The purpose of this study is to explore the academic position of digital curation and what it means to be an emerging discipline. Insight gained from this study will be useful for stakeholders seeking to understand the position of digital curation among the information disciplines and within higher education.

The research questions addressed in this study are:

RQ1: Is digital curation emerging as an autonomous discipline?

a. In what ways do interviewees describe the relationship of digital curation to areas within the information disciplines?

b. Are there indicators that suggest digital curation is emerging as a discipline in its own right?
RQ2: Where does digital curation fit within the educational landscape?

a. To what extent and in what ways do interviewees express the role of education in the knowledge transfer of digital curation practices, skills, and theories?

This chapter is organized in three main sections. The first section, Research Design, discusses the study’s qualitative multimethod research approach. The next section, Methodology, details the strategies of inquiry and data collection techniques employed in this study, as well as descriptions of the samples. The final section of the chapter, Data Analysis, focuses on the treatment and analysis of the data.

**Research Design**

The first part of this study comprised literature-based investigations. A review of relevant literature was conducted to provide a context for understanding academic disciplines, explore what it means to be an emerging discipline, and establish criteria for identifying disciplines (Chapter Three). This literature review engages material from multiple disciplines, referencing publications across many domains. Then, the database that was compiled from the results of scoping the literature was used to inform a narrative of the development of the area of digital curation (Chapter Four) and a literature review about digital curation education and curriculum development (Chapter Five). Together, these three chapters provide a backdrop on which the findings, discussion, and conclusion are contextualized.

The second part of this study comprised primary data collection using a qualitative multimethod research design. Qualitative research is suitable for the exploration and holistic description of complex phenomena. The aim of qualitative research is to “make sense of, or to interpret, phenomena in terms of the meanings people bring to them” (Denzin & Lincoln, 2005,
Compared to quantitative studies, samples sizes in qualitative research tend to be smaller and purposively selected. In qualitative approaches, the data are less structured and data collection is more flexible and inductive than in quantitative strategies (Guest, MacQueen, & Namey, 2011). Data collected is usually text, images, or sounds (i.e. not numbers) and analysis requires “close reading” to uncover contextual meaning and identification of patterns. Findings tend not to be generalizable but rather potentially transferable to different settings.

In a qualitative multimethod approach, the researcher collects, analyzes, and integrates data from more than one qualitative strategy of inquiry or data collection technique within a single study. In multimethod research, the investigator does not mix qualitative and quantitative approaches but employs more than one data collection method or more than one strategy of inquiry from within the same type of approach, qualitative or quantitative, to address the research questions (Tashakkori & Teddlie, 2003). This differs from mixed method research designs, in which the researcher combines qualitative and quantitative strategies of inquiry and data collection techniques (Creswell & Clark, 2007; Tashakkori & Teddlie, 2003).

The rationale for using a multimethod research design is to improve the trustworthiness, or reliability and validity, of the study through triangulation; to offset the weaknesses of each individual method; and to provide a more complete picture of the research problem. Triangulation involves collecting data from multiple sources or using multiple methods to ascertain if the various forms of data converge to support similar findings (Kopinak, 1999). This technique is used to improve confidence in the findings and support the credibility, dependability, and confirmability of the research (Krefting, 1991; Jonsen & Jehn, 2009). Employing multiple research methodologies or the collection of multiple forms of data also acts to offset the weaknesses and balance the strengths inherent in the individual approaches. For
example, content analysis describes what text reveals, but may not explain the underlying motivation or behavior in the patterns observed; whereas interviews provide explanations of the meanings behind actions. Used together, the researcher is able to observe not only what patterns emerge but why the patterns are emerging. Another example is that interviews involve interpersonal relationships which may be challenging to replicate or unintentionally bias the data, whereas content analysis is designed to be unobtrusive and potentially replicable. Lastly, a single method is often not sufficient to capture the trends and details of a complex phenomenon such as the potential emergence of digital curation as an autonomous discipline. Using a multimethod approach provides more variety of detail and different levels of perspective in the analysis (Kopinak, 1999).

Methodology

This research employed three methodologies: scoping of the literature, content analysis, and interviews. Scoping studies are a broad stroke survey of the literature. According to Arksey and O’Malley (2005), there are four different reasons for conducting scoping studies: “to examine the extent, range, and nature of research activity; to determine the value of undertaking a full systematic review; to summarize and disseminate research findings; or to identify gaps in the existing literature” (p. 21). Scoping studies are distinct from systematic reviews in that they do not usually assess the quality or examine each study in depth. Rather scoping studies summarize the extent of the domain by mapping the breadth of the literature (Levac, Colquhoun & O’Brien, 2010).

Content analysis has its foundation as a research methodology in mass communication. Krippendorff (2013) defines content analysis as “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use” (p. 24). Elo
and Kyngäs (2008) build on that definition and identify that “the aim [of content analysis] is to attain a condensed and broad description of the phenomenon, and the outcome of the analysis is concepts or categories describing the phenomenon” (p. 108). Using this methodology, inferences about patterns and themes are made through the examination of the record (White & Marsh, 2006).

For content analysis employed in this study, qualitative data is collected and findings are presented both qualitatively and quantitatively. There is a tendency to present content analysis as either a qualitative or quantitative research methodology. This dichotomy is debated in the literature (see Krippendorff, 2013; Zhang & Wildemuth, 2009; Elo & Kyngäs, 2008; Neuendorf, 2002; Holsti, 1968). Krippendorff (2013) questions the value of distinguishing between quantitative and qualitative content analysis: “Ultimately, all reading of texts is qualitative, even when certain characteristics of a text are later converted into numbers” (p. 22). Creswell and Clark (2007) consider it a “gray area” and explain that in a typical content analysis a researcher collects qualitative data and analyzes the data both qualitatively and quantitatively (p. 12).

The interviews consisted of focused, semi-structured interviews with individuals engaged in digital curation and related areas. The aim of the interview is to understand behaviors, experiences, facts, and meanings from the point of view of the respondent. Kvale and Brinkmann (2009) lay out ten key features of interviews including that the interviews are a social interaction; they are focused on the research topic but flexible in their structure; the interview is conducted in a manner that encourages description of specific situations rather than general opinions and strives to reduce ambiguity through follow-up and clarification; and the researcher is open to new and unanticipated themes rather than imposing predetermined categories onto phenomena (p. 30-32).
Scoping the literature, content analysis, and interviews are suitable methodologies to address the research questions of this study. As described in detail in the next chapter, disciplines are the interaction between intellectual, organizational, and social structures. Two of the essential mechanisms of this interaction are communication and consensus. Several theorists view scholarly publications as facilitating and mediating disciplinary communication (Stichweh, 2001; Hyland, 2004). Hyland (2004) explains:

In academic fields this means that texts embody the social negotiations of disciplinary inquiry, revealing how knowledge is constructed, negotiated and made persuasive. Rather than simply examining nature, writing is actually seen as helping to create a view of the world. This is because texts are influenced by the problems, social practices and ways of thinking of particular social groups. In other words, discourse is socially constitutive rather than simply socially shaped; writing is not just another aspect of what goes on in the disciplines, it is seen as producing them. (p. 3)

Consensus, or the level of agreement among scholars within a discipline, is an underlying factor of most of the conceptual frameworks for modeling our understanding of disciplines and emerging disciplines. Scoping the literature provides a summary view of the area’s discourse. Content analysis uncovers patterns and themes found in scholarly communication while interviews reveal how these patterns are understood by members of the community. These methodologies expose communication patterns and levels of consensus among scholars.

**Scoping the Literature**

The initial step in the data collection process was a comprehensive survey of the literature relating to the area of digital curation. There were four objectives of this survey. First, I used the survey to broadly describe and characterize the literature relating to the area of digital
curation. Second, I used the results of the survey to delineate the population from which the sample was taken for the content analysis. Third, I used the results of the survey to identify participants for interviews. Lastly, I used the literature that was collected during the survey to compile a narrative of the development of the area and conduct a literature review on education and curriculum development in the area of digital curation.

The process implemented for scoping the literature was adapted from the method for performing systematic literature reviews (Jesson, Matheson, & Lacey, 2011) and was designed to be comprehensive, methodical, and transparent (see Table 2.1 for a summary of the approach). This survey was broad and inclusive in order to gain a holistic representation of the literature associated with the development and knowledge base of the area of digital curation. The survey included practical and scholarly articles, books, conference papers and presentations, reports, and white papers that discuss concepts and practices related to digital curation. Literature searches were conducted primarily using vendor libraries and databases available through Simmons College Library, and supplemented with the journals listed on the Digital Curation Centre website (http://www.dcc.ac.uk/resources/curation-journals). An exhaustive list of search terms was used to conduct this survey. See Appendix A for the literature databases, search fields, and complete list of search terms used in scoping the literature.
Table 2.1

Scoping the literature: Summary of the methodology

<table>
<thead>
<tr>
<th>Dates of Data Collection</th>
<th>October 2013 – November 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td></td>
</tr>
<tr>
<td>Describe and characterize the literature of the area</td>
<td></td>
</tr>
<tr>
<td>Delineate the population from which the sample is taken for the content analysis; identify samples for content analysis</td>
<td></td>
</tr>
<tr>
<td>Identify participants for interviews</td>
<td></td>
</tr>
<tr>
<td>Inform narrative of the development of the area and literature review about digital curation education</td>
<td></td>
</tr>
<tr>
<td><strong>Accessible Population</strong></td>
<td></td>
</tr>
<tr>
<td>Published articles, books, conference presentations, reports, and white papers that discuss concepts and practices related to the area of digital curation, accessible through (see Appendix A for more details):</td>
<td></td>
</tr>
<tr>
<td>Targeted journal searches (<a href="http://www.dcc.ac.uk/resources/curation-journals">http://www.dcc.ac.uk/resources/curation-journals</a>)</td>
<td></td>
</tr>
<tr>
<td><strong>Dates Range</strong></td>
<td>2001-2012</td>
</tr>
<tr>
<td><strong>Search Terms</strong></td>
<td></td>
</tr>
<tr>
<td>Total of 98 search terms (see Appendix A for a complete list)</td>
<td></td>
</tr>
<tr>
<td>Main search terms</td>
<td></td>
</tr>
<tr>
<td>data archiving</td>
<td>digital archiving</td>
</tr>
<tr>
<td>data curation</td>
<td>digital curation</td>
</tr>
<tr>
<td>data preservation</td>
<td>digital preservation</td>
</tr>
<tr>
<td>data stewardship</td>
<td>digital stewardship</td>
</tr>
<tr>
<td><strong>Workflow</strong></td>
<td></td>
</tr>
<tr>
<td>1. Log on to vendor’s library, database, or journal website</td>
<td></td>
</tr>
<tr>
<td>2. Set search field parameters</td>
<td></td>
</tr>
<tr>
<td>3. Enter one search term/phrase into database search box</td>
<td></td>
</tr>
<tr>
<td>4. Record data in a spreadsheet: date of search; vendor’s library, databases, and/or journal website; search fields; search term; number of articles retrieved; date range of retrieved items; and comments</td>
<td></td>
</tr>
<tr>
<td>5. Select all items retrieved and export content including abstract to EndNote</td>
<td></td>
</tr>
<tr>
<td>6. Create groups in EndNote for each search term; file retrieved items in relevant group</td>
<td></td>
</tr>
<tr>
<td>7. Repeat for each search term</td>
<td></td>
</tr>
<tr>
<td>8. Repeat for each vendor’s library, database, or journal website.</td>
<td></td>
</tr>
</tbody>
</table>
**Description of the Sample.** Scoping the literature consisted of two parts. The first part was a comprehensive, systematic search across 50 databases using 98 search terms (see Appendix A). The result was an EndNote database of bibliographic information for the 50,476 articles that were retrieved. The results included books, conference proceedings, dissertation and theses, journal articles, and reports. Excluded from the results were newspaper articles and patents. The date coverage for the entire database was 1957-2014 and comprised articles across all domains. The relevancy of the results varied among the results of different search terms and even within the results of the same search term. The EndNote database was used to inform the literature-based investigations, the selection of samples for the content analysis, and identifying interview participants.

For the second part, eight key search terms were isolated from the results of scoping the literature and the date range was limited to 2001-2012. The key search terms were: *data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation,* and *digital stewardship.* For this subset, 5,579 items were retrieved and a more detailed analysis was conducted.

**Interviews**

Purposive sampling was used to select interviewees who offered a particular expertise, insight, or perspective concerning this study’s research purpose. Weiss (2008) explains that many interview studies use “panels of knowledgeable informants” rather than samples that represent the population (p. 17). The panels are made up of “people who are uniquely able to be informative” and “view [the] topic from different perspectives or who know about different aspects of it” (p. 17).
The participants for this study were selected to provide a diverse array of outlooks on this research. The panel of interviewees was designed to contain an international representation; a variety of roles (for example, educators, practitioners, or participants in research and development); a mixture of disciplinary backgrounds; and a range of topical interests in the area (for example, biocuration, data publishing and sharing, or digital preservation).

Fourteen purposively selected interview participants were identified first through suggestions made by my dissertation committee and then supplemented by authors that emerged from the literature survey results. Selection of interview participants had to meet three criteria:

1. Participants must have displayed engagement in discourse, education, or research related to digital curation within the past year. For example, potential participants must have published research in a journal or presented research at a conference related to digital curation; taught coursework, continuing education, or training workshops related to digital curation; or have participated in a collaborative project related to digital curation. This criterion was established to ensure that interviewees were active members of communities engaged in digital curation and related areas.

2. Participants must have been involved or working with digital materials in some capacity related to digital curation for at least five years. This criterion was established so that the interviewees would have a longitudinal view of the area.

3. Collectively, participants had to make up a diverse, international panel. The area of digital curation is often assumed to be a global and multi-discipline challenge (McGovern & Skinner, 2012; Constantopoulos et al., 2009; Rusbridge et al., 2005). This criterion was established to account for this assumption. Diversity was achieved
by including participants with different nationalities, disciplinary backgrounds, professional roles, and types of organization of employment.

**Description of the Sample.** The interview participants presented diverse and international perspectives on this study’s area of interest. Fourteen people engaged in the area of digital curation were interviewed. Table 2.2 identifies the interview participants by country of employment and type of organization in which they were employed.

Table 2.2

*Interviews: Description and identification of interview participants*

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Country</th>
<th>Type of Organization Employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewee 01</td>
<td>Australia</td>
<td>Archive/Library</td>
</tr>
<tr>
<td>Interviewee 02</td>
<td>Australia</td>
<td>Archive/Library</td>
</tr>
<tr>
<td>Interviewee 03</td>
<td>Continental Europe</td>
<td>Education, Research, and Development</td>
</tr>
<tr>
<td>Interviewee 04</td>
<td>Continental Europe</td>
<td>Higher Education (Computer Science Faculty)</td>
</tr>
<tr>
<td>Interviewee 05</td>
<td>United Kingdom</td>
<td>Education, Research, and Development</td>
</tr>
<tr>
<td>Interviewee 06</td>
<td>United Kingdom</td>
<td>Education, Research, and Development</td>
</tr>
<tr>
<td>Interviewee 07</td>
<td>United Kingdom</td>
<td>Data Archive</td>
</tr>
<tr>
<td>Interviewee 08</td>
<td>United Kingdom</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
<tr>
<td>Interviewee 09</td>
<td>United Kingdom</td>
<td>Higher Education (Computer Science Faculty)</td>
</tr>
<tr>
<td>Interviewee 10</td>
<td>United States/Canada</td>
<td>Archive/Library</td>
</tr>
<tr>
<td>Interviewee 11</td>
<td>United States/Canada</td>
<td>Data Archive</td>
</tr>
<tr>
<td>Interviewee 12</td>
<td>United States/Canada</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
<tr>
<td>Interviewee 13</td>
<td>United States/Canada</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
<tr>
<td>Interviewee 14</td>
<td>United States/Canada</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
</tbody>
</table>
Interviewees were employed as administrators, educators, and practitioners in eight different countries. There were five female and nine male interview participants. All participants had been involved in their area related to digital curation for at least seven years and had a minimum of a Master’s level education. Half of the participants held PhDs. The backgrounds of the interviewees included archaeology, archival science, biology, computer science, information technology, library and information science, and physics. Interviewees were employed at funding agencies; national, private, and regional archives or libraries; physical and social science data archives; research centers; and universities. Interviewees identified themselves as occupying roles such as administrator, advocate, archivist, biocurator, data curator, digital archivist, educator, preservation specialist, and project manager.

**Interview Procedure and Protocol.** All of the interviews followed the same procedure. Potential participants were sent an email inviting her/him to participate in the study. The email introduced the researcher and the purpose of the study, and described the details of the interviewee’s participation. A copy of the informed consent form was included as an attachment with the invitation email. Interviews were scheduled during December 2013 and January 2014 for those individuals who replied to the email expressing interest in participating. Interviews were conducted via telephone (3) or Skype (11) and each took approximately 60 minutes. Two days before the scheduled interview, the participant was emailed a meeting reminder with the interview guide attached for her/his review. All interviews were recorded using an Olympus DM-620 digital recorder to help ensure accurate transcriptions. Transcriptions of the interviews were outsourced to a local service in Boston, MA who signed a confidentiality agreement. All correspondence and interviews were conducted in English.
The focused, semi-structured interviews were guided by a written interview protocol (see Appendix B). The interview protocol was based on a set of predetermined open-ended questions that was flexible enough to allow for follow-up questions, probing, and pursuing topics that naturally arose during the conversation. The interview protocol in this study was developed using a model presented by Wengraf (2001) (Table 2.3). Wengraf identifies a three step model in which interview questions are derived from theoretical questions which are derived from the study’s overarching research question. Wengraf explains that “the [theoretical questions] are formulated in the theory-language of the research community, and the [interview questions] are formulated in the language of the interviewee” (p. 62). The assumption of the model is that the interview questions provide the data for the researcher to address the theoretical question, which taken together assist the researcher in answering the research question. Additionally, some of the interview questions were adapted from research discussed in the literature review (see Fensham, 2004; Phillips, 2003).

**Ethical Considerations.** This research was approved by the Simmons College Institutional Review Board (study number 13-028) on October 24, 2013 and was conducted to meet high ethical standards. As the researcher, I completed the Simmons College human subjects research investigator web-based training for minimal risk research through the CITI (Collaborative Institutional Training Initiative) Program. To protect the confidentiality and privacy of the interviewees, each is only identified by her/his geographical region and general professional role. Additionally, informed consent was obtained from all participants.
<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Subquestions</th>
<th>Interview Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is digital curation emerging as an autonomous discipline?</td>
<td>In what ways do interviewees describe the relationship of digital curation to areas within the information disciplines?</td>
<td>What do you consider to be the key concepts or principles and practices that drive this area? Do you think these are relevant to analog materials – if yes in what ways?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What do you consider to be the salient models, perspectives, or theories that influence this area?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do you think this area fits within the information disciplines? Why or why not?</td>
</tr>
<tr>
<td></td>
<td>Are there indicators that suggest digital curation is emerging as a discipline in its own right?</td>
<td>What term do you use when you are discussing this area? How do you define this term and how is it different or similar to other terms that people use?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How do you engage, communicate, and stay up-to-date with the community?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Would you consider the area practice-driven, theory-driven, or something else?</td>
</tr>
<tr>
<td>Where does digital curation fit within the educational landscape?</td>
<td>To what extent and in what ways do interviewees express the role of education in the knowledge transfer of digital curation practices, skills, and theories?</td>
<td>What are your thoughts about the educational opportunities available in this area?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In what ways do you envision the educational landscape of this area maturing?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Broadly speaking and based on your experience, what types of knowledge and skills do you think are important for professionals in this area to possess?</td>
</tr>
</tbody>
</table>
Content Analysis

Content analysis is a systematic process. The content analysis conducted in this study follows the procedure outlined below (see Leedy & Ormrod, 2010; Zhang & Wildemuth, 2009; White and Marsh, 2006; Neuendorf, 2002). First, based on the study’s research questions, the content to be analyzed was identified. The population and sampling method were defined, the unit of analysis was determined, and the sample was drawn. Next, categories of inquiry and coding scheme were established and the codebook was prepared. The coding scheme was piloted, assessed for consistency, and the coding process was adjusted as necessary. Then, records were systematically reviewed and data were recorded according to the coding scheme. The data were then summarized, patterns were identified, and research questions addressed, and findings from the content analysis were related to previous research or conceptual framework.

Description of Sample. Purposively selected subsets of the literature were investigated in depth using content analysis. A sample was selected that represented factors that characterize the emergence of disciplines. Three subsets were used as samples: the International Journal of Digital Curation (n=237; 2006-2012); conference papers and posters retrieved in the literature survey using the search terms digital curation and digital stewardship (n=53; 2001-2012); and the top 5% most highly cited articles in Scopus and Web of Science using the main search terms data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation, and digital stewardship (n=170; 2001-2012). The year 2001 was selected as the start date because the phrase “digital curation” was coined in that year (Beagrie, 2006, p.4). The unit of analysis for the content analysis was individual articles, conference papers or posters, editorials, research papers, and reports. For the content analysis, 460 units of analysis were coded (see Appendix D for bibliographies of the content analysis samples).
The *International Journal of Digital Curation (IJDC)* was selected because journals “provide authorized outlets for contributions to the literature of the field” (D’Agostino, 2012, p. 335). Although several journals publish articles and research relevant to digital curation topics, the *IJDC* is the first journal dedicated to publishing articles, news, and research on the curation of digital assets and related issues. In discussing the role of journals in scholarly communities, Solomon (2007) notes that “a hallmark of a discipline’s coming of age is the establishment of a new journal: in essence, staking out the intellectual territory of the new field” (p. 3). *IJDC* was established in 2006 by the Digital Curation Centre (DCC), a leading organization in the United Kingdom that provides expertise in data management, and digital preservation and curation activities. The *IJDC* is an online, biannual, open access journal published twice annually, with the exception of three issues in 2009 and one in 2010 and 2006. The journal publishes peer-reviewed research papers and non-peer-reviewed general articles (see Table 2.4 for details of the sample). Submission is open, but many published items were first presented at the International Digital Curation Conferences.

Table 2.4

*Content Analysis: International Journal of Digital Curation, 2006-2012*

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refereed Papers</td>
<td>115</td>
</tr>
<tr>
<td>Non-Refereed Articles</td>
<td>104</td>
</tr>
<tr>
<td>Editorials</td>
<td>18</td>
</tr>
<tr>
<td><strong>Total Number of Items Coded</strong></td>
<td><strong>237</strong></td>
</tr>
</tbody>
</table>

(8 years, 15 issues)

Conferences are another sanctioned channel of communication and contribution within disciplines. They support both formal and informal network-building, make public current and
emergent research, and “foster in individuals an affiliative identity with the discipline” (D’Agostino, 2012, p. 334). Conference paper and poster presentations were selected from the results of scoping the literature (see Table 2.5 for details of the sample). The sample concentrated on presentations that were retrieved using the search terms digital curation and digital stewardship to create a manageable sample size that honed in on content created by communities that specifically employ this terminology. Focusing on these two search terms provided a narrower sample that was mostly influenced by the information fields and to some degree the technology fields.

Table 2.5

| Content Analysis: Paper and poster presentations from published proceedings, 2001-2012 |
|---------------------------------|------|
| Presentations                   | 49   |
| Poster Presentations            | 4    |
| Total Number of Presentations Coded | 53   |

Twenty-nine conferences were represented in the sample. The International Conference on Theory and Practice of Digital Libraries (formerly the European Conference on Research and Advanced Technology for Digital Libraries, ECDL), IS&T Archiving Conference, Joint Conference on Digital Libraries (JCDL), and the iConference, accounted for almost half of the papers in the sample.

Highly cited articles were selected because citations are another means of communication and network-building, a measure of validation of research findings or ideas, and are an indication of the level of contribution within a field (Lucio-Arias & Leydesdorff, 2009; Bornmann, de Moya-Anegón, & Leydesdorff, 2010). Citations can be useful indicators of trends in scholarly
communication and the diffusion of disciplinary knowledge “by paying homage to pioneers, identifying original publications, providing background reading and alerting readers to forthcoming works” (Meho, 2007, p. 3). The most frequently cited articles germane to the area of digital curation were identified in Scopus and Web of Science. These included articles and research papers, conference presentations, and editorials that had been cited in those databases seven or more times (see Table 2.6 for details of the sample). Limiting the sample to the top 5% allowed for a manageable sample size that was also large enough to include articles from an array of disciplinary perspectives.

Using the key search terms data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation, and digital stewardship, a total of 3,383 unique items were retrieved. Of these, the 215 that were cited seven or more times, according to the citation counts in the databases, were collected. Forty-three of those articles were determined to be not substantively relevant to the area of digital curation and two of the articles were in German. The remaining 170 (the top 5% most cited articles) included 127 unique journal or conference titles. The articles represented a range of disciplines and subdisciplines within the broad fields of archival studies, museums studies, and library and information science; computer science, engineering, information technology, and mathematics; and the life, physical, and social sciences.
Table 2.6

Content Analysis: The top 5% most frequently cited articles from Scopus and Web of Science, 2001-2012

<table>
<thead>
<tr>
<th>Category</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles, Research Papers, Reviews</td>
<td>159</td>
</tr>
<tr>
<td>Editorials</td>
<td>4</td>
</tr>
<tr>
<td>Conference Papers</td>
<td>4</td>
</tr>
<tr>
<td>Other (Introduction to Issue; Letter, White Paper)</td>
<td>3</td>
</tr>
<tr>
<td>Total Number of Items Coded</td>
<td>170</td>
</tr>
</tbody>
</table>

**Data Collection.** Data collection for the content analysis took place over the four-month period of December 2013 through March 2014. A coding instrument was developed for the content analysis. The coding instrument consisted of a codebook and corresponding data collection form. The codebook was drafted to record definitions of codes and to ensure consistent coding (see Appendix C). A corresponding standardized data collection form created in Microsoft Excel was used to record data about each unit of analysis. The data collection form consisted of basic bibliometric variables about each unit of analysis including data about the publication source, authorship, disciplinary perspective, and subject focus. Each unit of analysis was examined in turn. In many cases, it was possible to assign values in the coding instrument from the abstract or associated metadata. In most cases, it was necessary to skim part or all of the publication to assign certain codes.

**Trustworthiness**

Ensuring the trustworthiness of qualitative research is comparable to ensuring the objectivity, reliability, and validity of quantitative research. Guba (1981) proposed four criteria for assessing the accuracy and consistency of qualitative research that correspond to the more
broadly understood concepts used for quantitative research (see Table 2.7). Strategies for establishing trustworthiness are addressed widely in the literature (see Shenton, 2004; Krefting, 1991; Lincoln & Guba, 1985; Guba, 1981).

Table 2.7

*Trustworthiness: Comparison of criteria for assessment of rigor (Krefting, 1991; Guba, 1981).*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Qualitative Approach</th>
<th>Quantitative Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth value</td>
<td>Credibility</td>
<td>Internal Validity</td>
</tr>
<tr>
<td>Applicability</td>
<td>Transferability</td>
<td>External Validity</td>
</tr>
<tr>
<td>Consistency</td>
<td>Dependability</td>
<td>Reliability</td>
</tr>
<tr>
<td>Neutrality</td>
<td>Confirmability</td>
<td>Objectivity</td>
</tr>
</tbody>
</table>

There are several ways that I addressed trustworthiness in this study. Notably, this research was designed as a multimethod study employing three distinct research methods and collecting data from multiple types of sources. As mentioned above triangulation increases the trustworthiness. Additionally, for the content analysis, intracoder reliability was assessed and a standardized coding instrument was used. For the interviews, sessions were recorded, transcribed, and returned to interviewees for review.

Intracoder reliability was assessed for a portion of the content analysis. Intracoder agreement is established when one coder makes the same coding decisions about a record at two separate times. Intracoder agreement is effective for evaluating changes in the coding process over time that may result from, for example, the coder’s expectations of the data, maturation of the coder’s thinking, or fatigue (Neuendorf, 2002; Weber, 1990). For the purposes of this study a simple percent agreement was calculated. In April 2014, after coding was completed, I
recoded the first twenty documents that I had coded in December 2013. Intracoder agreement was ninety-nine percent.

Employing a standardized coding scheme for data collection with pre-defined classifications increases the trustworthiness of the analysis and provides a means of replicability (White & Marsh, 2006). A standardized coding scheme, complete with codebook and spreadsheet-based data collection form, was used for one part of the content analysis. The second part of the content analysis was coded using NVivo qualitative analysis software using a combination of emergent and a priori coding. The a priori codes were based on D’Agostino’s (2012) model of a discipline.

There is a complex relationship between the interviewer and interviewee and my presence had an effect on the respondents and how they engage with the phenomenon of interest. I used the same interview guide for all interviews. Although it was flexible enough to allow for tailored follow-up questions, the main set of questions was always addressed. It is important to understand and acknowledge the role I perform in collecting and analyzing the data. In the qualitative interviews, I act as the data collection instrument. To address issues of bias and trustworthiness I recorded interviews for accurate transcriptions; I took notes during the interviews to record contemporaneous annotations; and I wrote memos documenting the research process throughout (Shenton, 2004). Additionally, each participant had the opportunity to review, comment on, and make corrections to the transcripts of her/his interview. It was these edited transcripts that were used for the final coding and analysis.

**Pilot Studies**

The coding instrument and interview protocol were both piloted on small samples prior to data collection for the full study. The coding instrument for the content analysis was piloted to
assess the feasibility and scale of the coding categories and sample size. The interview protocol was piloted to assess the clarity and wording of interview questions. The pilots also helped identify areas with the potential for bias. Both the interview protocol and the coding instrument were refined based on the experiences of the pilot studies.

One of the disadvantages of content analysis is that it can be time-consuming depending on the sample size and complexity of the coding scheme. This was addressed in the pilot study. I piloted the coding instrument for feasibility and scale using the 2010 issue of the *International Journal of Digital Curation* which contained sixteen items. Adjustments were made to the coding instrument based on the outcome of the pilot. First, an attainable and effective sample size was able to be established for the final study. Second, two changes were made to the coding instrument. Originally, authorship codes included author name, affiliation, role, discipline, and country for the first three authors. However, professional role and discipline were not straightforwardly distinguishable in most cases. Those two codes were eliminated. Also, originally, research methodology codes were going to be collected for research design, methodology, and data collection technique. These codes were replaced with an open-coded methodology comments option.

The interview protocol was piloted on three participants (see Table 2.8). Pilot interviews were conducted in a similar manner as the research interviews. The main distinction was that for the pilot interviews, participants were encouraged to comment on the interview questions. I was interested in her/his answers to the questions, and also interested in observations about the clarity of the questions, if they seemed leading in any way, and the overall coverage of the interview. Based on the pilot interviews and discussions with the interviewees, the interview guide was refined and revised. The revised interview protocol was submitted to the Simmons College IRB.
and the changes were approved. The pilot interviews were coded using NVivo to assess the coding procedure; however, the data were not included with the research interviews for the final analysis.

Table 2.8

Pilot interviews: Participants

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Country</th>
<th>Type of Organization Employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot Interviewee A</td>
<td>Australia</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
<tr>
<td>Pilot Interviewee B</td>
<td>United States/Canada</td>
<td>Higher Education (Archive/LIS Faculty)</td>
</tr>
<tr>
<td>Pilot Interviewee C</td>
<td>United States/Canada</td>
<td>Archive/Library</td>
</tr>
</tbody>
</table>

Description of Data Analysis

Interview transcripts and the abstracts from the content analysis samples were coded using NVivo qualitative data analysis software. This was in addition to the bibliometric coding scheme that was used for the content analysis as described above. Coding the units of analysis from the content analysis and the interview transcripts together within the same coding scheme effectively merged data analysis for a more comprehensive presentation and interpretation of the findings. NVivo provides a platform that facilitates the coding process and improves the ability to systematically extract, manage, organize, and visualize data. All content was coded in the interview transcripts. For the units of analysis from the content analysis, abstracts were coded and in the absence of an abstract, introductory sections were coded. From NVivo and the coding instrument for the content analysis, descriptive statistics of the samples and qualitative excerpts were extracted.
NVivo was used to code both emergent and *a priori* codes. The primary coding scheme was based on emergent, data-driven codes that grew organically out of the data. A copy of the coding scheme is found in Appendix C. In addition, an *a priori* coding scheme was based on the criteria set forth in the conceptual framework used in this study (D’Agostino, 2012). This set of criteria, which was derived from previous research, was used because it provides a comprehensive and detailed list of elements that comprise a discipline. The emergent codes were mapped to the *a priori* codes for analysis.

Data coding and analysis was an iterative process. I moved progressively from unstructured, emergent themes to the refinement of those themes and the development of categories to the mapping of those categories to *a priori* codes. This coding process allowed for a granular analysis of themes emerging in the literature and the identification of how those themes correspond and constitute a discipline. The process for coding was largely adapted from Hahn (2008).

For the initial coding, I conducted an unstructured thematic analysis by identifying data-driven themes and patterns that emerged from the interviews and content analysis sample. For example, the following excerpt from an article in *IJDC* was initially coded as *DCU Digital Curation Process Model*:

> In this paper we review two recent models for digital curation introduced by the Digital Curation Centre (DCC) and the Digital Curation Unit (DCU) of the Athena Research Centre. We then propose a fusion of the two models that highlights the need to extend the digital curation lifecycle by adding (a) provisions for the registration of usage experience, (b) a stage for knowledge enhancement and (c) controlled vocabularies used by convention to denote concepts, properties and relations. (Constantopoulos, et al., 2009, p. 34)
This excerpt also from *IDJC* was coded as *Preservation Network Model*:

We describe the components of a Preservation Network Model and go on to show how it may be used to plan preservation actions according to the requirements of the particular situation using illustrative examples from scientific archives. (Conway, Matthews, Giaretta, Lambert, Wilson, & Draper, 2012, p. 3)

This excerpt from an interview transcript was coded as *DELOS Digital Library Reference Model*:

*Interviewee 08:* The reference model for digital library management system used to be called DELOS Reference Model, is quite an interesting model for this as well. It looks more at digital libraries, but it does look at the whole aspect of how people gain access and so on to this material, so obviously keeping it for something.

The second round of coding was more focused, in which I explored the initial coding and aggregated codes into more general categories. The above coded excerpts were combined with other similarly coded items. General categories were created for types of models which grouped the above excerpts into the codes *Digital Curation Models, Preservation Models,* and *Digital Library Models,* respectively.

The final round of coding consisted of refining the coding scheme into manageable final codes. I collapsed, expanded, revised, and finalized the coding categories using the categories from the second round as subcodes, when appropriate. The codes described above, for example, were aggregated together into a single final code named *EmergentCode_Models*.

Lastly, to thoroughly explore the themes that emerged from data analysis and address the research questions, I mapped the finalized coding categories with the *a priori* codes established in the conceptual framework. The *EmergentCode_Models* code was aligned with the disciplinary element “body of accumulated knowledge” from the conceptual framework (D’Agostino, 2012). This final code informed one of the criteria that characterizes that element, specifically, the recognition of conceptual and explanatory models, frameworks, and theories.
Summary

In this chapter, I discussed the research design and methodologies employed in this dissertation, as well as actions taken to address ethical considerations and issues of trustworthiness. This dissertation employs a qualitative multimethod research design. The methodologies employed in this dissertation were scoping the literature, content analysis of published literature in the area of digital curation, and semi-structured interviews with individuals engaged in the area. This study focused on identifying themes that emerged from the practical and scholarly discourse, and understanding the level of consensus among educators, practitioners, and researchers about how they view this domain. The content analysis data was recorded using a standardized coding instrument. Units of analysis from the content analysis and interview transcripts were also coded together using emergent codes in NVivo. Emergent codes were mapped to a priori codes derived from the conceptual framework.

In the next chapter, I present a review of relevant literature related to theories of disciplinarity. The aim of this literature review is to provide a context for understanding academic disciplines, explore what it means to be an emerging discipline, and establish criteria for identifying disciplines. The literature review provides support for the conceptual framework used in this study.
CHAPTER 3: UNDERSTANDING ACADEMIC DISCIPLINES

In this chapter I review relevant literature about academic disciplines. There exists an extensive body of research on disciplinarity from the fields of education, history of science, philosophy, and sociology of knowledge and science that examines the differences between and relationships across academic disciplines as well as the internal dynamics within individual disciplines. The function of this literature review is to provide a context for understanding academic disciplines, exploring what it means to be an emerging discipline, and establishing criteria for identifying disciplines. This chapter also serves to provide support for the conceptual framework used in this study.

This literature review is organized in three main sections. The first section, Academic Disciplines, briefly provides an historical context for disciplinarity and then explores the definition of a discipline. The next section, Theories of Disciplinarity, reviews cultural and social theories and metaphors that have been employed for understanding disciplines, and conceptual frameworks for classifying disciplinary differences. The final section, Emerging Disciplines, focuses on how new disciplines emerge and criteria for identifying them.

Academic Disciplines

Although there is a long history of academic disciplines as a means of organizing knowledge for teaching and learning in higher education, the structure of the modern university system that emerged in the nineteenth century changed the function of disciplines and the role of the academic researcher. The discipline evolved from being a storehouse for disciplinary knowledge to a complex social system of disciplinary knowledge production. Since the nineteenth century, “the discipline has functioned as a unit of structure formation in the social system of science, in systems of higher education, as subject domain for teaching and learning in
Beginning in the nineteenth century, higher education experienced an evolution of its intellectual, organizational, and social structure that had significant impact on disciplinarity. A shift from predominantly instruction and dissemination of knowledge to an increased emphasis on knowledge production and discovery through original inquiry promoted the establishment of research universities (Rothblatt, 2001; Shumway & Messer-Davidow, 1991; Metzger, 1987). As original research in the university setting increased, disciplinary specialization increased. In part, specialization was a response to the growth of scientific and technical knowledge and the emergence of the social sciences (Strober, 2006; Wray, 2005; Shumway & Messer-Davidow, 1991). Organization of universities into departments, which were largely determined by discipline or subject matter (Shumway & Messer-Davidow, 1991; Faricy, 1974), further encouraged specialization in both research and teaching (Shils, 1979). Disciplinary specialization necessitated improved communication, such as the formation of discipline-specific associations and journals, among geographically-dispersed researchers whose focused interests were aligned (Shumway & Messer-Davidow, 1991; Metzger, 1987). These developments, among others, transformed higher education into a modern university system which supported the increase in number and diversity of academic disciplines and their subsequent subdivision and proliferation throughout the twentieth and into the twenty-first centuries.

**Toward a Definition of Discipline**

In their most basic form, disciplines are fields of knowledge that organize the teaching and learning of practical skills and theory. Modern usage of the term “discipline” combines the meanings of “discipline,” which is historically associated with the learning of practice and
exercise, with “doctrine,” which is historically associated with the teaching of abstract theory (Craig, 2008; Shumway & Messer-Davidow, 1991). In tracing the etymology of the term “discipline,” Shumway and Messer-Davidow (1991) concluded that “to call a field a ‘discipline’ is to suggest that it is not dependent on mere doctrine [or abstract theory] and that its authority does not derive from the writings of an individual or a school, but rather from generally accepted methods and truths” (p. 202).

In current use, however, academic disciplines are more than fields of knowledge, but “are bound up in complex ways with organized scholarly professions and academic departments” (Craig, 2008, p. 10). Becher and Trowler (2001) expounded on this complexity:

The concept of an academic discipline is not altogether straightforward…The answer [of whether an academic field is a discipline] will depend on the extent to which leading academic institutions recognize the hiving off in terms of their organizational structures…and also on the degree to which a freestanding international community has emerged, with its own professional associations and specialist journals…International currency is an important criterion, as is a general though not sharply defined set of notions of academic credibility, intellectual substance, and appropriateness of subject matter. (p. 41)

In this excerpt, Becher and Trowler highlighted three aspects of the concept of discipline: community, intellectual qualities, and organizational structure.

Some definitions of academic disciplines emphasize their role in the organizational and institutional structure of higher education. Possibly the most visible effect of the differentiation of a discipline is the expansion and growth of academic departments and the variety of degree-granting programs in universities (Becher & Trowler, 2001; Evans, 1995; Clark, 1983).
Universities are organized into departments, which are for the most part defined by discipline. This has led some to believe that “a discipline is at bottom nothing more than an administrative category” (Jencks & Riesman, 2002, p. 523). However, the manner in which an institution groups disciplines and departments into colleges and schools can vary among institution. Craig (2008) emphasized the institutional function of disciplines by asserting that an “ill-defined set of academic units and professional groups” (p. 10) has displaced the association of a discipline with learning of practice and exercise.

Other definitions focus on the intellectual qualities of academic disciplines. Simply put, disciplines are “the tools, methods, procedures, exempla, concepts, and theories that account coherently for a set of objects or subjects” (Klein, 1990, p. 104). In his exploration of whether physical education had reached disciplinary status, Henry (1976) defined an academic discipline as “an organized body of knowledge collectively embraced in a formal course of learning… the content is theoretical and scholarly as distinguished from technical and professional” (p. 13). As in this definition, theory and research are often privileged over the practical application of skills.

Still other definitions highlight the community and social construction of academic disciplines. A community is a group of individuals that shares something in common, such as backgrounds, culture, identity, or interests. Bueger (2012) explained that “disciplines are communities organised by a distinct set of shared practices, vocabularies and institutions” (p.101). These shared practices have significance within the context of the group. Bueger further explained that disciplines “give meaning and legitimacy to academic practices such as writing and presenting research, reading or teaching; they are a means to evaluate one’s status, and give intelligibility to distinct claims to knowledge” (p. 101).
In addition to the three components of disciplines highlighted above, disciplines are also defined in terms of power relations, in particular as social control over knowledge or boundary constructions around domains. Perhaps most associated with expressing disciplines in terms of power is Foucault who described a discipline as “a system of control in the production of discourse” (Foucault, 1971/1972, p. 224). The goal of a discipline, as a social structure, is to “control a body of knowledge by subjecting it to systematic organization” and “shape the consciousness and conduct of those who seek to apply this body of knowledge” (Winters, 2010, p. 4889). The power relations within and among disciplines are also a struggle to demarcate knowledge and research domains. In writing about the contentious relationship between women’s studies and disciplinarity, Boxer (2000) explained that a discipline “serves to identify both a body of knowledge and a process of constructing boundaries, of regulating and shaping that body, as well as the training required to participate in producing the knowledge housed in it” (p. 121).

In this research, I use the definition presented by Del Favero (2003) because it combines the various aspects of disciplines discussed above. Academic disciplines are defined as:

Providing the structure of knowledge in which faculty members are trained and socialized; carry out tasks of teaching, research, and administration; and produce research and educational output. Disciplinary worlds are considered separate and distinct cultures that exert varying influences on scholarly behaviors as well as on the structure of higher education. (p. 10)

**Theories of Disciplinarity**

Exploring the definition of academic disciplines provides an understanding of the complexity of the concept, but does not uncover how the different elements interconnect to form
and sustain disciplines. Disciplines have layers of interrelated structures: the research community operates within the structures of the discipline; a discipline interacts with other disciplines; the discipline exists within the larger higher education system; and the disciplinary and higher education systems function within society. All of these areas have been explored in the literature.

**Disciplines through the Lens of Cultural and Social Theories**

Bourdieu (2001/2004) offered this definition of academic disciplines based on elements of a broader cultural and social theory:

A discipline is defined by possession of a collective capital of specialized methods and concepts, mastery of which is the tacit or implicit price of entry to the field. It produces a ‘historical transcendental’, the disciplinary habitus, a system of schemas of perception and appreciation (where the incorporated discipline acts as censorship). It is characterized by a set of socio-transcendental conditions, constitutive of a style. (p. 65)

There are three interacting and essential concepts in this definition: field, capital, and habitus. Bourdieu (1993) defined field as “a separate social universe having its own laws of functioning independent of those of politics and the economy” (p. 162) and “where, in accordance with its particular laws, there accumulates a particular form of capital” (p 164). An academic discipline qualifies as a field. Bourdieu provided three fundamental forms of capital: “economic capital, which is immediately and directly convertible into money...; cultural capital, which…may be institutionalized in the form of educational qualifications; and social capital, made up of social obligations (‘connections’)” (Bourdieu, 1997, p. 47). Other examples of capital in academic disciplines include awards, committee appointments, grants, and publications. Habitus has been defined in many ways, but one way to understood it is as “embodied history, internalized as a
second nature and so forgotten as history” (Bourdieu, 1990, p. 56). Fleshing this definition out, it is past learned assumptions, behaviors, and dispositions through socialization that structure actions in the present. As it concerns academic disciplines, the habitus of scholars is their opinions and perception of academic activities associated with their discipline. These combine to produce a discipline that is characterized by particular behaviors and conduct.

The broader cultural and social theory supporting Bourdieu’s definition of discipline is practice theory. Simply defined, practice theory is “a theory of the relationship between the structures of society and culture on the one hand, and the nature of human action on the other” (Ortner, 1989, p. 11). Practices are actions, beliefs, skills, and tacit knowledge that occur within a political context. Practice theory is interested in exploring the relationship of the action of individuals to the whole system of culture and how these practices produce, reproduce, alter, and perpetuate a culture (Ortner, 1989; Schatzki, 2001). Bourdieu understood practice as the interaction and relationship between capital, field, and habitus.

Working within the framework of practice theory, Knorr Cetina (1999) concluded that disciplines can be described as epistemic cultures. Knorr Cetina defined epistemic cultures as “those amalgams of arrangements and mechanisms…which, in a given field, make up how we know what we know. Epistemic cultures are cultures that create and warrant knowledge” (p. 1). Culture here is defined as “the aggregate patterns and dynamics that are on display in expert practice and that vary in different settings of expertise” (Knorr Cetina, 1999, p. 8). Knorr Cetina conducted ethnographic research in high-energy physics and molecular biology laboratories. The research explored the diverse components and strategies for arranging, constructing, and valuing knowledge within the group, emphasizing the attitudes, practices, and processes of knowledge production. The practices of knowledge production are contextualized within the
microcosm of an epistemic culture (or knowledge setting) which can explain variation among disciplines. The usefulness of discussing a discipline in terms of epistemic cultures establishes a context for seeing disciplines as “knowledge-in-action” (Knorr Cetina, 1999, p. 3) and not limited to static organizational structures.

Influenced by Niklas Luhmann’s theory of social systems, Stichweh described disciplines as communication systems. Communication and the transfer of knowledge is a key issue in understanding the structure of a discipline. Stichweh (2001) explained that “the emergence of the scientific discipline is equivalent to the invention of new communication forms specific of disciplinary communities” (p. 13728). Stichweh viewed scholarly publications as facilitating and mediating disciplinary communication, influencing knowledge production and exchange, and reifying a discipline’s identity. As noted above, one of the major impacts of disciplinarity during the nineteenth century was specialization, the division of labor and subsequent differentiation and fragmentation of disciplines into subdisciplines and new disciplines.

“Specialization is, first of all, an intellectual orientation. It depends on a decision by individual scientists to concentrate on a relatively small field of scientific activity, and…one needs a social context supporting it” (Stichweh, 2001, p. 13728). Stichweh, therefore, equated the growth of disciplines in the nineteenth century with the emergence of scientific communities. Scientific communities are bounded social systems and “function as the infrastructure of communication systems” (Stichweh, 2001, p. 13728). It is a discipline’s publications, which build on and incorporate prior publications and encourage reactions and more publications, through which the disciplinary communication process sustains and maintains the discipline.
Metaphors of Disciplinarity

The concept of boundaries is often associated with the nature of disciplines. Star and Griesemer (1989) asserted that boundaries act as interfaces that allow for communication and exchange between communities and across different social worlds. Galison (1997) considered these boundaries as “trading zones” where people within disciplines can exchange knowledge and mutual understanding. Galison’s research focused on how different types of physicists collaborate on large projects. Galison (1997) explains:

My question is not how different scientific communities pass like ships in the night. It is rather how, given extraordinary diversity of the participants in physics – cryogenic engineers, radio chemists, algebraic topologists, prototype tinkerers, computer wizards, quantum field theories – they speak to each other at all. And the picture… is one of different areas changing over time with complex border zones that sometimes vanish, coalesce, and even burgeon into quasi-autonomous regions in their own right. (p. 63)

Galison’s trading zone metaphor was inspired by the work of linguistic anthropologists who study the cultural use of pidgins and creoles. These boundaries between sub-disciplines are regions of collaboration, communication, and interaction.

Other research, however, has viewed boundaries not as means of communication and knowledge production but as disciplinary divisions. Disciplinary divisions define academic behaviors, intellectual and knowledge domains, and worldviews. These divisions act as boundaries that are constructed to facilitate the expansion of knowledge domains, monopolization of authority and resources, and protection of autonomy over professional activities. “The boundaries of the discipline are protected by a more or less codified, strict and
high cost of entry; they are more or less sharply defined, and sometimes at issue in struggles with the neighboring disciplines” (Bourdieu, 2001/2004, p. 65).

Gieryn (1983) introduced the term “boundary-work” to help articulate efforts to demarcate science from nonscience. Boundary-work is the intentional characterization of science as distinctive and “constructing a social boundary that distinguishes some intellectual activities as non-science” (Gieryn, 1983, p. 782). However, the concept of boundary-work is not limited to the differentiation of science from non-science and has been appropriated into other discussions, including theories of disciplinarity.

When the point is to establish or protect a discipline, boundaries mark it as a territory to be possessed by its owners, not appropriated by others, and they indicate the relations it may have with other disciplines. But these same boundaries may be redefined if the discipline is attempting to expand into new territory. When the point is to regulate disciplinary practitioners, boundary-work determines which methods and theories are included, which should be excluded, and which may be imported. (Shumway & Messer-Davidow, 1991, p. 209)

The notion of boundaries can be used to construct community and identity by creating an insider that is different from an outsider and reinforcing exclusion (Li, 2009, p. 94).

Becher and Trowler developed the metaphor of culture throughout their research on academic disciplines (Becher, 1981; Becher, 1987; Becher, 1994; Becher & Trowler, 2001). Borrowing concepts and a cultural perspective from anthropology, Becher and Trowler expressed disciplines as the relationship between academic “tribes,” the “distinctive cultures within academic communities,” and their disciplinary “territories,” academic ideas and knowledge (Becher & Trowler, 2001, p.3). Disciplines have “recognizable identities and
particular cultural attributes” (Becher & Trowler, 2001, p. 44). These cultural attributes include boundaries, communication, and power and social structures. Disciplines are made up of internal boundaries separating subspecialties, and external boundaries separating disciplines from each other. Each discipline has unique power and social structures that include department and university standings, gatekeepers, publications, recognition, social networks, and tenure. There exists formal and information communication, and modes of interaction including publication and research collaboration. For Becher and Trowler (2001), “academic cultures and disciplinary epistemology are inseparably intertwined…disciplinary knowledge forms are to a large extent constituted and instantiated socially…their constitution has a reciprocal effect in the culture from which they spring” (p. 23). Becher and Trowler define culture as “sets of taken-for-granted values, attitudes and ways of behaving, which are articulated through and reinforced by recurrent practices among a group of people in a given context” (p. 23). Disciplinary behavior, knowledge, and values are contextualized within the disciplinary community.

**Frameworks for Studying Disciplinary Differences**

Within the literature, there is a variety of conceptual frameworks for classifying the variation among disciplines. Academic disciplines differ markedly in their intellectual, organizational, and social structure. The intent of these frameworks is to articulate and account for the distinctions of administration, culture, departmental organization, faculty instruction and research, socialization, student learning, and worldview among the diverse academic disciplines. Researchers have developed frameworks of disciplinary variation, which include continuums, descriptive typologies, and models, based on an array of dimensions. Examples of these include typologies based on Biglan’s hard and soft schema (Storer, 1967; Biglan, 1973a; Biglan, 1973b), the Holland Theory of Occupational Classification (Smart, Feldman, & Ethington, 2000),
learning styles (Kolb, 1981), and organizational structure (Collins, 1975; Whitely, 1984; Fuchs, 1992); and models developed to classify disciplinary differences in terms of their level of codification (Zuckerman & Merton, 1971), consensus (Hagstrom, 1964; Hargens, 1975), and paradigm development (Lodahl & Gordon, 1972). These conceptual frameworks, and others, have been used extensively in the literature to investigate difference among disciplines such as department administration and organization; faculty publication trends, research practices, and teaching; and socialization.

In 1996, Braxton and Hargens conducted a comprehensive review of research on frameworks for analyzing disciplinary differences. In their book chapter, the authors identify eleven analytical frameworks and provide a thorough review of the five that earned the most attention by other researchers. A 2011 article by Jones updated the work of Braxton and Hargens; one additional framework was identified and relevant literature since their publication was reviewed. The findings of the Braxton and Hargens review, echoed by Jones, led them to conclude that “differences among academic disciplines are profound and extensive. Their manifestation range from global characteristics, such as disciplinary structural patterns, to individual scholars’ everyday teaching and research experience” (Braxton & Hargens, 1996, p. 35).

As Braxton and Hargens (1996) convey, the classification of academic disciplines has existed as long as disciplines have existed and disciplinary differences between researchers and practitioners have often been noted. Early efforts of classification centered on creating a hierarchy of disciplines. Efforts in the early twentieth century focused on understanding and comparing the differences of progress in disciplines. Beginning in the 1960s, research in this
area grew and more dynamic classification frameworks were developed, with an emphasis on the sociological.

Most of these conceptual frameworks are discussed in relation to the tenets that Kuhn sets forth in 1962 in *The Structure of Scientific Revolutions* about progress in scientific fields (Braxton & Hargens, 1996). Kuhn argued that scientific progress occurs when an “anomaly” is discovered that is irreconcilable within the current “paradigm” in a period of “normal science”; this causes a period of “crisis,” which leads to a “revolution” in which there is a change from the former paradigm to a new paradigm that can account for the anomaly; normal science is then resumed under the new paradigm. A classic example of this paradigm shift is Copernicus’s model of a heliocentric universe displacing Ptolemy’s model of a geocentric universe in the 1500s (Kuhn, 1970).

Fleshing out Kuhn’s paradigm model reveals how it applies to conceptual frameworks for understanding disciplinary variation. Kuhn defined paradigms as “universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners” (Kuhn, 1970, p. viii). A paradigm can be understood as a shared set of beliefs, facts, methods, theories, taxonomy, and values that a community of scientists agrees satisfactorily describe their understanding of the world. “Men whose research is based on shared paradigms are committed to the same rules and standards for scientific practice. That commitment and the apparent consensus it produces are prerequisites for normal science, i.e., for the genesis and continuation of a particular research tradition” (Kuhn, 1970, p. 11).

In Kuhn’s model, a scientific discipline progresses through a pre-paradigm period, and cycles through a stable normal science period and a crisis period. The pre-paradigm period exhibit paradigms but there is not yet consensus about disciplinary taxonomy, techniques, and
theories and the period “is regularly marked by frequent and deep debates over legitimate methods, problems, and standards of solution, though these serve rather to define schools than to produce agreement” (Kuhn, 1970, p.47-48). The period of normal science is the usual state of scientific research operating under a shared paradigm and taxonomy; it is “research firmly based upon one or more past scientific achievements, achievements that some particular scientific community acknowledges for a time as supplying the foundation for its further practice” (Kuhn, 1970, p. 10). An anomaly occurs when there is a deviation or inconsistency from what is expected based on the current paradigm that explains normal science (Kuhn, 1970). The discovery that the current paradigm and taxonomy does not adequately explain a phenomenon can lead to a crisis, which is a period of exploration, experimentation, and the emergence of new scientific theories. This crisis culminates in scientific revolution “in which an older paradigm is replaced in whole or in part by an incompatible new one” (Kuhn, 1970, p. 92) and normal science resumes under the new paradigm and taxonomy.

Shifting from understanding progress in scientific disciplines, Lodahl and Gordon (1972) applied Kuhn’s concepts to exploring the academic behaviors of research and teaching in higher education. Lodahl and Gordon suggested that Kuhn’s concept of paradigm could be understood as the “degree of consensus or sharing of beliefs within a scientific field about theory, methodology, techniques, and problems” (Lodahl & Gordon, 1972, p. 58) and that disciplines with high paradigm development have greater consensus. They further contended that the level of paradigm development of an academic discipline is predictive of particular activities, attitudes, and characteristics, and specifically “that the greater consensus of fields with high paradigm development would facilitate teaching and research activities as compared to fields with low paradigm development” (Lodahl & Gordon, 1972, p. 68). Faculty from 80 university
departments, representing the disciplines of chemistry, physics, political science, and sociology, responded to a survey in which they ranked the disciplines “in terms of consensus over theory, methodology, and training” (p. 58). Lodahl and Gordon concluded “that the structure of knowledge lends predictability to teaching and research, and that scientific fields with more developed paradigms have more structure and thus more predictability than fields with less developed paradigms” (p. 66).

While the paradigm development model focuses on behavior dimensions of scholarly activities, codification describes variation of disciplines’ body of knowledge. Introduced by Zuckerman and Merton (1971, 1972), codification is “the consolidation of empirical knowledge into succinct and interdependent theoretical formulations” (1972, p. 303). Zuckerman and Merton (1972) explained:

In [less codified disciplines], scientists must get command of a mass of descriptive facts and of low-level theories whose implications are not well understood. The comprehensive and more precise theoretical structures of the more codified field not only allow empirical particulars to be derived from them but also provide more clearly defined criteria for assessing the importance of new problems. (p. 303)

The first study that Zuckerman and Merton conducted to advance the concept of codification is concerned with journal rejection rates across disciplines. They concluded that the degree of variation is in part due to differences in codification of knowledge. Disciplines that are highly codified, such as the physical sciences, have lower average rejection rates, whereas disciplines that are less codified, such as humanities, have much higher average rejection rates (Zuckerman & Merton, 1971). In a second study, they argued that because of the compact structure of knowledge in highly codified disciplines, researchers are able to make significant contributions
earlier in their career than researchers in less codified fields (Zuckerman & Merton, 1972). Codification can be further expressed as “the extent of agreement on standards of scholarship” (Zuckerman & Merton, 1971, p. 77) with highly codified fields displaying more agreement or consensus.

One of the more influential frameworks is Biglan’s three-dimensional model based on the hard/soft typology which incorporates both knowledge and behavior domains. According to Braxton and Hargens, Storer (1967) was among the first to argue the potential use of the hard/soft classification to explain disciplinary variation. Storer was “interested in the differences in the qualities of social relationships in the different sciences [and] in the ‘atmospheres’ or ‘moods’ that characterize different fields of science, and with how such differences produce visible consequences” (p. 75). Storer explained that “our use of the terms ‘hard’ and ‘soft’ is somehow related to these aspects of scientific disciplines, and that these are in turn related to the organization of knowledge within these different fields” (Storer, 1967, p. 75). However his work did not garner much attention outside of his own field of sociology of science (Braxton & Hargens, 1996, p.3). It was Biglan who, in two seminal articles (1973a, 1973b), took up the hard-soft dimensions and produced the schema with which most are familiar (Braxton & Hargens, 1996). In fact, Jones (2011) found it to be the most used classification schema for examining disciplinary differences since Braxton and Hargens’ review in 1996 (p. 16).

Biglan found that three dimensions adequately characterized academic subject matter and the structure and output of university departments: hard/soft, pure/applied, and life/nonlife systems. Using faculty perception of the similarities and difference of disciplines, Biglan (1973a) distinguished the important characteristics, or dimensions, that characterize academic subject matter. In a corresponding study, Biglan (1973b) employed the results of the first
analysis to explore the relationship between subject matter and department organization. The strongest correlation was found in the hard/soft dimension, or “the degree to which a paradigm exists” (Biglan, 1973a, p. 204). The second strongest correlation was the pure/applied dimension or “the degree of concern with application” to practical problems over theory (Biglan, 1973a, p. 204). The third dimension, which is rarely used (Nelson Laird, Schwartz, Shoup, & Kuh, 2005), is the life/nonlife, or the “concern with life systems” or inanimate objects (Biglan, 1973a, p. 204). Biglan concluded from these studies that these three dimensions effectively classify academic disciplines according to faculty perception of subject matter, as well as other aspects of academic behavior and organization such as social connectedness, commitment to teaching and research, and scholarly output.

Becher and Trowler expanded on Biglan’s hard/soft and pure/applied cognitive dimensions, adding the social dimensions of urban/rural and convergent/divergent. Becher and Trowler’s research was based on multiple studies of in-depth interviews with academics in 12 disciplines (Becher, 1981; Becher, 1987; Becher, 1994; Becher & Trowler, 2001). Through this research, Becher and Trowler refined their four-dimensional taxonomy for mapping the patterns found in academic tribes and territories and explored the connection between disciplinary culture and the culture of higher education organizations (Becher & Trowler, 2001). Becher and Trowler explained that researchers characterized as urban “select a narrow area of study, containing discrete and separable problems, where their rural counterparts typically cover a broader stretch of intellectual territory in which the problems are not sharply demarcated or delineated” (p. 106). They explained that disciplines that are convergent tend to be “tightly knit in terms of their fundamental ideologies, their common values, their shared judgments of quality, their awareness of belonging to a unique tradition and the level of their agreement about what
counts as appropriate disciplinary content and how it should be organized,” while disciplines that are divergent tend to “lack a clear sense of mutual cohesion and identity” (p. 59).

An undercurrent of the conceptual frameworks discussed above is the concept of consensus, or the level of agreement among researchers. Consensus as a framework for disciplinary variation was initially taken up by Hagstrom (1964, 1965), who focused on attitudes about, and degree of, collaboration and corroboration in research. Ten years later Hargens (1975) built on Hagstrom’s approach focusing on dimensions of social solidarity, with one dimension being the level of agreement about beliefs, norms, and values with respect to research within their discipline. High level of consensus is associated with disciplines with more developed paradigms while low consensus fields have less developed paradigms. Del Favero (2003) suggested that the use of paradigm development and consensus are “interchangeable as they describe a common dimension of disciplinary fields – the extent of agreement on structure of inquiry and the knowledge it produces” (p. 11). She also indicated that the model of high and low consensus has for the most part displaced codification because consensus “has been determined to be a function of codification” (Del Favero, 2003, p. 10). Biglan’s hard/soft model is based on the degree to which a paradigm exists and is supported by research that sought consensus among faculty. There is clearly a correlation between Becher and Trowler’s convergent/divergent dimension and high/low consensus. Additionally, in a Postscript for the 1970 reprint of The Structure of Scientific Revolutions, Kuhn, influenced by the work of Hagstrom (1965), wrote: “If this book were being rewritten, it would… open with a discussion of the community structure of science” (p. 176).

To appreciate the relationship between some of these models, Figure 3.1 provides a visual comparison of four of the conceptual frameworks for understanding disciplinary variation.
presented in this chapter. High consensus disciplines correlates with Biglan’s hard typology, existence of a developed paradigm, and highly codified disciplines. This suggests that these disciplines, often described as the natural and physical sciences, would have structured knowledge bases with clearly developed high-level theory and agreement of the disciplines models, problems, techniques, and values. On the other hand, low consensus correlates with Biglan’s soft typology, less developed paradigms, and less codified disciplines. These disciplines, often described as the social sciences and humanities, exhibit less agreement on the knowledge base, methods, and theories that define their discipline. Characteristically, high consensus disciplines would have, for example, higher journal acceptance rates, more co-authorship of articles, and increased use of teaching and research assistants than low consensus disciplines.

Figure 3.1. Comparison of conceptual frameworks of disciplinary differences.
Emerging Disciplines

The emergence of new disciplines and fragmentation of current disciplines into subdisciplines are complex processes involving intellectual, organizational, and social changes. Early investigations attributed disciplinary differentiation and specialization to changes in the social and organization structure of science (Wray, 2005). Ben-David and Collins (1966) argued that the desire to establish a new occupational and social role, and the benefits from that role, significantly influenced specialization. Other research argued that as the growth of knowledge increases exponentially, for effective research to occur researchers must carve out a manageable area in which to focus their attention. Later studies considered the impact of new scientific instruments and technology on specialization (Wray, 2005). For example, Law (1973) argued that X-ray crystallography was established in part due to the development of new instrumentation for studying the molecular structure of crystals. Researchers also began to understand the complexity of specialization and take into account the effects of conceptual and epistemic changes in addition to organizational and social changes (Lemaire, MacLeod, Mulkay, & Weingart, 1976; Wray, 2005).

Kuhn’s model of disciplinary specialization privileges conceptual and epistemic changes. In response to criticisms of his work and changes in his own thinking, Kuhn further developed his model. Kuhn (1990) allowed that:

After a revolution there are usually (perhaps always) more cognitive specialties or fields of knowledge than there were before. Either a new branch has split off from the parent trunk, as scientific specialties have repeatedly split off in the past from philosophy and from medicine. Or else a new specialty has been born at an area of apparent overlap between two preexisting specialties. (p. 8)
Kuhn’s original model of scientific progress noted that a discipline progresses through a pre-paradigm period, and cycles through periods of stable normal science with a shared paradigm and periods of crisis when the paradigm is inadequate to address current problems; after a revolution the old paradigm is replaced with a new paradigm. Kuhn refined the concept of paradigm, first by introducing the more encompassing term “disciplinary matrix” and later replacing the term paradigm with the concept of lexicon changes (Kuhn, 1970, p. 182; Kuhn, 1990). A result of this development included the mitigation of the punctuated nature of scientific revolutions by allowing for more gradual and subtle changes and specialization.

Law (1973) categorized disciplinary specialties in three types: subject matter, technique or methods based, and theory based. Subject matter specialties are research activities around a particular set of problems or topical areas in which a variety of methods and theories may be applied. Technique or methods based specialties employ a particular research methodology or use of particular scientific equipment or a specific practice to address any range of problems. Law categorizes X-ray crystallography as a technique-based specialism. Theory-based specialties focus on theory development as their main research activity and subdisciplinary concern (Law, 1973). Clark (1983) summarized the labyrinth of specialization within disciplines:

The major disciplines are extensively subdivided. For example, physics is broken down into such major subdisciplines as optics, mechanics, fluids, nuclear physics, and elementary particle physics – the latter dividing still further into cosmic ray physicist, who study natural particles, and high-energy physicist, who use accelerators. These major subfields, in turn, contain more specialties…The division of labor accounts for large differences in originality and type and degree of competition. (p. 35-36)
Abbott (2001) developed a conceptual framework for the growth within disciplines based on patterns of change. Abbott used his discipline of sociology as a case study to illustrate that disciplines differentiate in fractal-like patterns. Fractals are repeating geometric patterns. Abbott explained that conflict between existing dichotomies within disciplines, such as pure vs. applied or qualitative vs. quantitative, causes a split with the dominant group persisting. These dichotomies, however, repeat themselves at each level of differentiation and throughout the discipline in space and time. Abbott (2001) explained:

> On the one hand, [the fractal cycle] generates perpetual change. Old ideas are always being thrown out. Intellectual autocracy is perpetually overthrown. On the other, it produces perpetual stability. The new ideas are always the old ideas under new labels. The new people are always the old people in new roles. (p. 26)

Although Abbott’s central concern was the development of social sciences, these fractal distinctions also explain specialization within a discipline and its differentiation into subdisciplines or new disciplines, but this only occurs when the size of the discipline and resources allow for it (Abbott, 2001).

Metzger (1987) developed a process he termed “subject parturition” to illustrate the differentiation and fragmentation of disciplines. Subject parturition occurs when a new field develops from an older discipline or disciplines. The new field gradually achieves independence as a new or sub discipline. The new discipline or specialty integrates into higher education curriculum through a process Metzger called “program affiliation” and becomes accepted and valued through “subject dignification.” Lastly, “subject dispersion” occurs as the new academic discipline or specialty continues to grow and spreads beyond its original location of acceptance (p. 128-131).
Clark (1983) reasoned that external factors, namely the heterogeneity of students and the diversity of labor markets, work in conjunction with the effects of research and scholarship in higher education to lead to differentiation within disciplines and the creation of organizational structures to support them. He provides several generalizations to explain the process and impact of differentiation: Because universities have historically been organized based on disciplines as a means of grouping academic work and interest, differentiation can cause significant changes to the organizational structure of higher education. When groupings of different or specialized interests are formed, disciplines are divided between those who want stability through the maintenance of old interests and those who introduce the new interests. There are power struggles between those vested in the old structure and those seeking change; the outcome of the struggle is determined by which group has more influence, legitimacy, and power. Influence, legitimacy, and power depend on recognition by the wider community that there is value in the activities and ideologies of the group. Clark contended that “subject fragmentation has arguably become the source of ever-growing system complexity” and further noted that it has a larger impact than external factors such as employment outlook (Clark, 2002, p. 326).

In a comparative case study of the development of six academic disciplines, Hashem (2007) elaborates on four factors that facilitate their emergence and independence. The model involves the interaction of the external factors, namely “societal pressures in the form of public demands, including crises that call for specialized expertise” and “state intervention through funding and legislation;” and internal pressures, namely “the academic resourcefulness of the field that is aspiring to become independent” and “the institutional dynamics of the field, including competition, organizational expansion, and professional consolidation” (p197).
Hashem concluded, based on the patterns that emerged from the case studies, that academic resourcefulness was the most predictive factor of the field’s successful independence.

**Criteria in the Literature**

In a 1978 article, Thornton discussed the disciplinary nature and potential of the field of American Indian studies. Since Thornton was unable to locate any set of published criteria for assessing disciplinarity, he based his analysis on the four general criticisms found in the literature about the “disciplinary shortcomings” of American Indian studies (Thornton, 1978, p. 11). He categorized these criticisms as lacking a distinct methodology, abstract concepts, a unique area of concern, and scholarly traditions. Some argue that disciplines have a characteristic methodology associated with research: for example, ethnography is a methodology particular to anthropology. Disciplines are often concerned with abstract concepts, such as gravity, force, and mass in physics. Disciplines are often bounded areas of study with minimal overlap with other fields: American Indian studies is often a subject in anthropology and other social science. Thornton found the requirement of scholarly traditions and a unique area of research and teaching to be the most essential component of a discipline.

Thornton (1978) explained that the typical pattern of an emerging discipline is for the intellectual qualities to emerge, such as an area of interest or new problem, followed by the structural elements, such as courses, faculty, and programs of study, needed to support them (p. 10). However, he argued that American Indian studies took the opposite approach. American Indian studies programs were created in university departments to meet political and social mandates. The intellectual structure continued to take shape and mature, but was typically contained within the boundaries of other existing disciplines. Thornton viewed the lack of unique intellectual and scholarly areas of research in American Indian studies to be the most
valid criticism keeping it from achieving disciplinary status but provided the beginning of a research agenda to help remedy this.

Some research on the disciplinary nature of a field uses criteria established for identifying professions. While published criteria for gauging academic disciplinary status are limited, the literature on accepted criteria for classifying professions has a long history (Chatterjee & Stevenson, 2008; Flexner, 1915; Goode, 1957; Greenwood, 1957; Roberts & Donahue, 2000). There are examples in the literature of using the criteria for defining a profession to also define disciplinarity (Bird, Welsch, Astrachan, & Pistrui, 2002). Although some features of disciplines and professions seem to overlap, disciplinarity and professionalization are two distinctive phenomena.

Richardson (2008) analyzed the development of accounting history as a discipline in the context of a “professional project.” Professional project is a conceptual model for understanding the actions that shape the professionalization of an occupation or “an attempt to translate one order of scarce resources – special knowledge and skills – into another – social and economic rewards” (Larson, 1977, p. xvii). Richardson adapted the concept of professional project to examine the disciplinary project of accounting history by focusing on the strategies and activities that, in retrospect, made possible its emergence as a discipline. He summarized the strategies as “making history relevant” by connecting accounting history to other activities or disciplines already considered legitimate; “making history visible/controversial” by defining boundaries around the intellectual domain and identifying exemplars in accounting history; and “institutionalizing accounting history” through the establishment of associations and programs of study that provide the necessary support (p. 251).
For Richardson (2008), accounting history had emerged as an academic discipline because it had an institutional structure identified “by the existence of specialist journals and academic associations,” an intellectual structure identified “by the existence of a bounded body of knowledge,” and a social structure identified “by the existence of a critical mass of academics that self-identify as members of the discipline” (p. 267). However, the disciplinary project, he concluded, was weak in three areas: “integrating accounting history into the curriculum; ensuring that accounting history is recognized and validated in tenure and promotion standards; and, demonstrating the relevance of accounting history to the process of accounting standard-setting” (p. 268). He noted that providing courses and programs of study in an area is a key factor in emerging as a discipline.

Other research based the investigation of disciplinary status on a definition of discipline. As noted above, and made evident in this discussion, the definition of discipline is multifaceted and comes in several varieties. Yet, the definitions often group defining elements of a discipline as intellectual, organizational, or social. This is illustrated in two different studies, one on the disciplinarity of consumer behavior (MacInnis & Folkes, 2010) and the other on digital media (Sterne, 2005). The studies used two different definitions of discipline. MacInnis and Folkes established a definition that focused on the intellectual and social elements, while Sterne used a definition that privileged organizational properties. However, both sets of authors organized their analysis to address the intellectual, organizational, and social properties of their respective field. Both studies concluded that neither of their respective fields met the definition of discipline.
On Establishing Definitive Criteria

In the postscript to the 1970 reprint of The Structure of Scientific Revolutions, Kuhn introduced the concept “disciplinary matrix” to replace his notion of paradigm. Disciplinary matrix is a more encompassing term used to distinguish a set of group commitments, principles, and practices that hold an academic community together and accounts for “the relative fullness of their professional communication and the relative unanimity of their professional judgments” (p. 182). Kuhn did not provide an exhaustive list of the components that make up the disciplinary matrix. He provided four components as examples: symbolic generalizations, or scientific theories that can be expressed formulaically; metaphysical paradigms, or abstract theories; values; and exemplars (p. 182-187). Kuhn drew his conceptual model from the field of physics and therefore applying the model to the humanities and social sciences reveals some biases. As is typical in Kuhn’s work, the example components of the disciplinary matrix focused primarily on conceptual and intellectual aspects of the discipline.

In their 1976 edited volume of studies about the emergence of new disciplines, Lemaîne, MacLeod, Mulkay, and Weingart presented six “problematic spheres” designed to provide areas of focus for research on the development of new disciplines. The editors provided examples of questions that can be useful to explore each of the six areas. They contended that “every sphere can, at least in some instances, influence appreciably the course of scientific development” (p. 14). More specifically, inquiry into these spheres assess the “rate of scientific development” or the rate at which information increases in a field; the “direction of scientific development” or the particular area of intellectual investigation as opposed to another area; and the “content of scientific development” or findings, principles, and theories (Lemaîne, MacLeod, Mulkay, and Weingart, 1976, p. 14).
The problematic spheres can be summarized as internal intellectual processes, internal social processes, external intellectual factors, immediate institutional context, specific economic and political factors, and diffuse social influence. New fields typically develop from previous research or advances of scientific or technical knowledge. The internal intellectual and social processes spheres map the intellectual origins of an emerging field. The editors recommended questions about research methods, problems, and techniques and questions about communication, dissimilation of research, research community, and social position of actors. The communication of new scientific or technical knowledge is not always an internal process within a research community, but can be influenced by external intellectual factors as well as the immediate institutional context in which the research is produced. To address these spheres, the editors provided questions about the relationship between the external community groups, the institution, and the research community and questions about the economic, organizational, and social structure of the research institution. Although the editors argued that it is “unsatisfactory to claim broad changes in the structure of society have determined the course of scientific development in any simple, direct, or uniform fashion” (Lemaine, MacLeod, Mulkay, & Weingart, 1976, p. 11), there is merit in understanding the indirect impact of economic or political factors on research, including policy, research problems, and support. The final problematic sphere that the editors presented was concerned with the reception of the emerging field on social interest and values. The editors recommended questions that addressed the relationship between researchers and the community, including recruitment and the moral and social value of the field (Lemaine, MacLeod, Mulkay, & Weingart, 1976).
Institutionalization is a key element in establishing a new discipline. It is through higher education that the discipline perpetuates itself. Krishnan (2009) provided a list of six characteristics of academic disciplines:

1) disciplines have a particular object of research (e.g. law, society, politics), though the object of research may be shared with another discipline; 2) disciplines have a body of accumulated specialist knowledge referring to their object of research, which is specific to them and not generally shared with another discipline; 3) disciplines have theories and concepts that can organize the accumulated specialist knowledge effectively; 4) disciplines use specific terminologies or a specific technical language adjusted to their research object; 5) disciplines have developed specific research methods according to their specific research requirements; and maybe most crucially; 6) disciplines must have some institutional manifestation in the form of subjects taught at universities or colleges, respective academic departments and professional associations connected to it. (p.9)

Krishnan noted that not all disciplines will demonstrate all six of these characteristics; however the more that the discipline has, the more successful it is likely to be at growing and sustaining itself.

In a presentation that explored the designation of emergency management as a discipline, Phillips (2003) outlined nine elements “representative of areas and issues commonly found within disciplines” which “also reflect core issues often found in accreditation” (para. 5). To explore these elements of disciplinarity, Phillips posed several questions related to each key element; examples of some of the questions she proposed are presented in Table 3.1. Phillips developed the criteria from her experiences creating new programs in both women’s studies and emergency medicine. Although strikingly different fields, Phillips contended that “both are
breaking new ground in new ways, both face similar questions of content, terminology, student markets and outcomes, and faculty roles – not to mention questions of what we should teach and how we should teach it” (para. 4).

Table 3.1

*Representative areas and issues commonly found within disciplines (Phillips, 2003; Phillips, 2005).*

<table>
<thead>
<tr>
<th><strong>Naming the field</strong></th>
<th>What is the right term/phrase to call this discipline/field of study/area?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Defining the field</strong></td>
<td>How is [the field] defined?</td>
</tr>
</tbody>
</table>
| **Key concepts and core curriculum** | What are the key concepts that have driven the field? How have the key concepts (assuming they exist) been defined?  
Is there an implicit core curriculum? |
| **Evolution of the field** | What is the history of the field? What are the implications of this history? What more can/should we add? |
| **Theories** | What are the models, perspectives, paradigms, theories and philosophies that influence this field?  
How are they used/not used in classrooms and/or in practice? |
| **Methods** | What/which methods should be taught to students in this field? |
| **In practice** | What should be the role(s) and relationship(s) between academia and industry/practice?  
Where do we stand on the question of accreditation of … programs? |
| **Student Outcomes Assessment** | What are students doing with their degrees and certificates?  
What do employers think of … graduates? |
| **Faculty Roles** | What are traditional faculty roles? Do those roles apply to emergency management?  
How can we grow professional development opportunities for educators? |

**Conceptual Framework for this Research**

From a synthesis of the literature, D’Agostino (2012) derived the most comprehensive list of elements and corresponding criteria that compose and are supported by a discipline.
D’Agostino (2012) views a discipline as a cultural system: the interaction of elements that characterize a discipline interpreted within a framework of “shallow consensus” (p. 334). Additionally, interpretation of the disciplinary elements needs to be situated within the context of the development of the field. The historical context illuminates how and why the elements of a discipline came to be organized and represented as they are. Discipline as a cultural system is reminiscent of other theorists who have described disciplines in terms of a cultural metaphor (Geertz, 1983; Bauer, 1990; Becher & Trowler, 2001). D’Agostino’s notion of shallow consensus supports not only theories of disciplinary variation but also growth and specialization of disciplines (Hagstrom, 1964; Kuhn, 1970; Lodahl & Gordon, 1972; Hargens, 1975).

D’Agostino divided the elements that constitute a discipline into institutional aspects and intellectual aspects. Distinguishing between these two broad categories is common. Abbott (2001) differentiated between the social structure between disciplines, which includes institutional organization, and the cultural structure within disciplines, which includes intellectual structures. Becher and Trowler (2001) used the metaphors tribes, for academic culture, and territories, for their intellectual domain. During (2006) identified these two aspects of disciplines as “subdivisions of knowledge and institutional forms of life” (p. 266). A summary of the ten elements D’Agostino compiled is presented in Table 3.2. The disciplinary elements each contain more granular observable criteria that further define and refine the element (see Table 1.1 in Chapter 1).
D’Agostino (2012) argued, however, that these elements that constitute a discipline have to be understood within the framework of shallow consensus. D’Agostino contended that although we can define a discipline in terms of it meeting the criteria, we must accept that members of the discipline may have different understandings of what embody the individual elements. D’Agostino termed the latter shallow consensus:

It is actually this kind of consensus or agreement in commitments or values that will characterize the disciplinary communities, I believe. While there may be agreement in the philosophical community that Plato is a significant figure, or that Wittgenstein is, what Plato and Wittgenstein mean to different individuals in the community will vary widely. (p. 341)
However, it is this varying level of detail in agreement on intellectual and institutional aspects of disciplines that facilitate the growth of knowledge, through discourse, innovation, and new ideas. D’Agostino’s shallow consensus permits a definition of disciplines using rigid criteria and still tolerates internal variation within the discipline.

**Summary**

In this chapter I reviewed relevant literature from several fields that explores the nature of academic disciplines. Academic disciplines are often defined in terms of their intellectual, organizational, and social structure, invoking the use of terms such as community, institutional organization, organized body of knowledge, as well as power relations and social control. Because disciplines have layers of interrelated structures, they negotiate a range of relationships which have been explored using cultural and social theories and explained using metaphors.

Researchers tend to agree that disciplines are different in codifiable and predictable ways. Researchers have created classification systems in an attempt to explain these disciplinary variations. The modern classification frameworks reviewed in this chapter are not hierarchical; they do not suggest one discipline is superior to another, rather they indicate that disciplines exist on a continuum of variation.

The development of the modern higher education system in the nineteenth century supported disciplinary growth and specialization. The process of differentiation, fragmentation, and specialization of disciplines into subdisciplines or new disciplines is a complex processes involving intellectual, organizational, and social changes. Criteria for identifying that a field has emerged as a new discipline are not well-established; however several researchers have put forward useful sets of criteria. Although the criteria vary in presentation, they provide similar
elements of assessment that take into account institutional and intellectual aspects that aim to get at the dynamic qualities of a discipline.

In the next chapter I provide a narrative of the development of the area of digital curation. The narrative provides an historical context for understanding the circumstances of competing and shared interests in the emergence of the area of digital curation. The function of the narrative is to provide context for the findings of this study.
CHAPTER 4: CONTEXTUALIZING DIGITAL CURATION

In this chapter I present an historical perspective on the emergence of the area of digital curation. The function of this chapter is to provide context for the findings of this study. In accordance with the conceptual framework and numerous other models of disciplinarity, this chapter provides an historical context for understanding the circumstances of competing and shared interests in the emergence of the area of digital curation. An awareness of how an area developed is essential for understanding how the elements of a discipline, as defined by the conceptual framework, came to be organized as they are.

This chapter is organized in two main sections. The first section, Foundations of Digital Curation, provides a brief historical narrative of the development of this area. The section is divided into three subsections that depict the growth of automation and digital archiving in the archival, library and information science (LIS), and research communities. The second section, Preservation and Curation Communities, focuses on how the archival, LIS, and research communities begin to overlap to form a preservation community, and the early emergence of digital curation as a term for articulating something broader than preservation.

Foundations of Digital Curation

The emergence of area of digital curation as the lifecycle approach to managing and preserving digital assets evolved from a history of uncertainties about the longevity and sustainability of data and efforts for addressing these concerns. Technology progressed rapidly during the Second World War and onward, and consequently the use of computers in business and research increased. These advancements in technology brought great changes, but also with them came large amounts of data threatened by obsolescence, poor management, and our evolving understanding of the digital environment. In response to these new formats of data and
records, research and professional communities responded. Collaborative and cooperative initiatives were a central part of this response.

Interest in the stewardship of digital assets was evident early on in the development of computers. There have been a number of stakeholders interested in the archiving and preservation of digital assets. Many of the earliest endeavors to manage digital assets came from within the research communities themselves, dealing with mounting quantities of data in digital formats. The archive and LIS professions have seen their practice transform: not only has technology changed how they conduct their work, but also the materials they work with are increasingly digital.

Archival Community

In the archival community, initial reflection about machine-readable records focused on appraisal issues and the “recordness” of this form of media. The National Archives and Records Administration (NARA) in the United States determined as early as 1936 that punch cards, from the Bureau of Census, were not records. Satisfied with archiving the original questionnaires, summaries, and final processed data, punch cards were always not retained (Fishbein, 1972; Brown, 1987). Schellenberg disagreed on the grounds that machine-readable records were in fact records, defining records as any “documentary material, regardless of physical form or characteristic” (Schellenberg, 1956/2003, p. 16). However, in his influential 1956 internal bulletin, *The Appraisal of Modern Public Records*, Schellenberg agreed with the conclusion that, for the most part, machine-readable records need not be retained in the archives. He concluded that “punchcards and tape recordings are commonly unusable without resort to expensive equipment” and to be preserved in an archive records should be in a form that makes them easily accessible to users without the need of specialized equipment. NARA routinely opted for
conventional forms of media for permanent archival retention (Rieger, 1966; Fishbein, 1972; Brown, 1987).

In his 1956 bulletin, Schellenberg also made an important comment about the treatment of research data. He noted that, in most instances, raw scientific data does not meet the prerequisites to be maintained in the archive. He states:

They are usually quite voluminous. Often they have attributes that make their further use impracticable. They may be intelligible only to the persons who recorded the data. Like punchcards produced in statistical work, they may be in a form that is difficult to interpret without resort to mechanical or electronic devices. (Applying Tests of Informational Values, Records Relating to Phenomena section, para. 23)

He continues by noting that controlled laboratory research is repeatable and therefore does not have secondary value. Schellenberg makes a notable exception of archiving raw data with geosciences observational data, but otherwise favors retaining records that possess value to more typical archival researchers, such as historians.

Views about the appraisal of electronic records began to change in the 1960s. An impetus for change began with a 1965 report by the Committee on the Preservation and Use of Economic Data submitted to the Social Science Research Council. The report, coined the “Ruggles Report” after Yale University economics professor Richard Ruggles, the chairman of the committee, made recommendations for maintaining economic data for preservation and reuse in a centralized government data center (Ruggles, Miller, Kuh, Lebergott, Orcutt, & Pechman, 1965). Following the Ruggles Report, the Archivist of the United States of America established a four-person Committee on the Disposition of Machine-Readable Records to investigate current practices and recommendations for future policies. This committee, led by Meyer Fishbein,
agreed with the conclusions of the Ruggles Report, acknowledging that electronic media are records, and that NARA has the responsibility to treat them as such. In late 1968, the Data Archives section was established at NARA. NARA received its first electronic record transfer in 1970 (Navale & Cameron, n.d.) and in 1972 issued its first General Records Disposition Schedule for automated records (Brown, 1987). However, funding and commitment to the electronic records program and staffing was tumultuous over the decades that followed (Brown, 1987; Brown 2003).

Slow progress was made by the archival community during the 1970s and 1980s in their approach to managing electronic media. In the literature the debate about appraisal continued, ignited by Dollar’s *Appraising Machine-Readable Records* published in 1978, and evolved into discussions about the applicability of archival principles to electronic media and best practices for their preservation. By the mid-1980s, the first comprehensive appraisal guidelines and practical manual were published for archivists working with electronic media (Naugler, 1984; Hedstrom, 1984). However, a 1994 study on the archival profession in North America concluded that “the archival profession in the United States has not done well in structuring itself to manage electronic records” (Cox, 1994, p. 24). It continued that “despite three decades of activity, few archival institutions have developed programs to administer and preserve these special records” (Cox, 1994, p. 24).

This early discourse laid the conceptual foundation for and shaped the application of archival thinking to approaches to digital media that is evident today. Cook (1991) affirmed that the archival community was moving into a new generation of electronic records archivists, noting that the earlier generation was characterized by being isolated from traditional archivists and focused only on the records of those also interested in the preservation of electronic media,
such as social science researchers. This new generation of archivist recognized that electronic records were becoming an alternative to traditional archival media. “It was no longer a matter of a few records being rendered ‘machine-readable,’ but all traditional media being rendered ‘electronic,’ the media lines being blurred thereby, and the paper backups either disappearing or not even being produced” (Cook, 1991, p. 206). Hedstrom outlined the argument for an electronic records research agenda and the areas that needed to be explored, adding that “consultation and collaboration with experts in other fields will be essential elements of successful research” (Hedstrom, 1991, p. 335). Dollar saw the archival profession and records manager professions reuniting around shared concerns about electronic records, hinting at a lifecycle approach to managing this form of media. He identified three key concerns about electronic records: ensuring integrity and authenticity; early intervention by the archivist in the records lifecycle; and maintaining accessibility and overcoming technological obsolescence (Dollar, 1993).

Library and Information Science Community

Prior to the 1990s, libraries focused their efforts on using computers for bibliographic control of their holdings and automation of library catalogs using online public access catalogs (OPACs). In the early 1960s, the Library of Congress conducted a study to determine the feasibility of automation for cataloging, searching, indexing, and retrieving library materials, and the interoperability between library catalogs (King et al., 1963). As a result of the recommendation and positive outlook of this study, in 1966 the Library of Congress introduced the Machine Readable Cataloging (MARC) format for their own use and use in other libraries (Arms, 2012; Avram, 1975).
Following the introduction of the MARC format, several consortia were formed to help improve access and share resources through integrated library catalogs. In 1967, a group of Ohio university presidents and library directors established the Online Computer Library Center, Inc (OCLC, founded as Ohio College Library Center) and began work on an online integrated library catalog using MARC records. The first library began using online cataloging in 1971; soon OCLC began providing its services to libraries outside of Ohio (Gray, 2013; OCLC, 2012). Another consortium, the Research Libraries Group (RLG), was founded in 1974 by Columbia, Harvard, and Yale Universities and the New York Public Library with a similar vision as OCLC, but RLG expanded their catalog beyond libraries to include archives, museums, and other cultural heritage institutions (Martinez, 1998).

With the formation of integrated catalogs, support for information retrieval among distributed information systems became necessary. In the early 1970s work began on what would become the Z39.50 protocol for cross-database searching. The protocol was designed to address the challenge of interoperability between different systems. Later iterations of the protocol were highly extensible and saw implementation in search and retrieval for abstracting, indexing, and full text databases. Z39.50 became an ISO standard for information retrieval in 1988 (Lynch, 1997; Iltis, 1995).

As social science data archives developed distinctively from traditional information services, gaining control of data files became increasingly important; however the role of libraries remained ambiguous. In fact, a 1967 study by the National Research Council/Committee on Information in the Behavioral Sciences commented:

Some years ago it appeared plausible that maintenance of repositories of behavioral science research data would gradually be brought under the umbrella of the conventional
research library in matters of personnel and financing, after the initial innovation phase. Such a possibility is not, of course, to be ruled out completely for the future, but it is already clear that such a future remains distant. (p. 37)

One pioneer in advocating for increased bibliographic identification and control of social science data files in libraries was Sue Dodd, data librarian at the Inter-university Consortium for Political and Social Research (ICPSR), and active in the International Association for Social Science Information Services & Technology (IASSIST) and the American Library Association (ALA). Dodd advanced a concept of cataloging data files and published manuals for the practical application of these cataloging rules for librarians. One author noted that Dodd “was almost unique in her support of libraries as important resources for data. She believed that libraries wanted to be involved in data services and that they should be involved” (Gray, 2013, p. 17).

With Dodd’s input, the second edition of the Anglo-American Cataloging Rules (1978) included standards describing machine-readable data files and creating bibliographic records for electronic media (Dodd, 1982; Gray, 2013; Stephenson, 2013). Well into the 1980s, however, library involvement in maintaining research data remained unresolved (Dennis, 1970; Gray, 2013).

Attention shifted from automating library catalogs to confronting a changing information environment, and technology’s impact on how libraries provide services to researchers. By the late 1980s and early 1990s, catalog automation was well established, membership to a library consortium allowed for resources sharing, and online journal indexing and abstracting services become more affordable. Demand for digitally available content followed. Outside of the library purview, Michael Hart demonstrated that digitized content was achievable. Project Gutenberg found its ambitious, but gradual, beginnings in 1971 with the Declaration of Independence becoming the first e-book (Hart, 1992; Hane, 2004).
The digital libraries movement began to address some of the challenges and opportunities libraries faced. The earliest efforts of electronic libraries, as they were first called, focused on access to digitized journal articles. Three of the earliest implementations were the Mercury Electronic Library project at Carnegie Mellon University (1989-1992), which provided local access to computer science journals; the collaborative Chemistry Online Retrieval Experiment (CORE) (1991-1995), which digitized articles from journals published by the American Chemical Society; and the University Licensing Service (TULIP) through Elsevier Science (1991-1995), an online delivery service (Tedd & Large, 2005). JSTOR originated as a Mellon grant-funded project at the University of Michigan to determine whether it was feasible to replace print back issues of journals with digitized versions. The successful pilot project was launched in 1995 and soon established a self-sustainable economic model (Guthrie, 2000; Schonfeld, 2003).

In the United States, the major digital libraries initiative focused more on computer science research and technology solutions, such as systems architecture, interoperability, and information retrieval. In 1994, the first phase of the Digital Libraries Initiative (DLI-1) that funded six projects was launched in the United States by the National Science Foundation, the National Aeronautics and Space Administration (NASA), and the Defense Advanced Research Projects Agency (DARPA) (Tedd & Large, 2005). In fact, DLI is attributed as having popularized the phrase digital libraries, after some debate about whether it should be plural or singular and putting aside other popular terms at the time such as electronic library (Arms, 2007; Griffin, 1998). DLI-1 was more interested in developments in computer science and progress in discipline-specific communities, and only tangentially interested in models for improving library services (Dempsey, 2006; Lynch, 2000). Thirty-six more projects were funded by phase 2 of the
Initiative which began in 1998 and included additional funding sponsorship from the National Library of Medicine (NLM), the National Endowment for the Humanities (NEH), and the Library of Congress (Tedd & Large, 2005). With funding sources coming from within the library community, projects began to address issues facing the library.

Several program-based funding activities across the globe set out to facilitate change in libraries. In 1990 the European Commission launched the Telematics for Libraries Programme. The goals of the funding program were to automate library catalogs, network services between libraries, promote innovation, and produce technological tools and services for the library community (European Commission, n.d.; Raitt, 2000; Dempsey, 2006). In 1993, a review of libraries in the United Kingdom was initiated to determine how library teaching and learning services were being impacted by an increase in undergraduates, and how library research services were being impacted by rapidly evolving technologies. One of the key conclusions of the resulting report, known as the Follett Report, was that “the exploitation of IT is essential to create the effective library service of the future” (Follett, 1993, Chapter 1). Based on the recommendations of the report, the Electronic Libraries Programme (eLib) was established and funded 59 collaborative projects in the United Kingdom that were intended to advance the role of libraries and support the use of information technology in teaching and research in higher education (Electronic Libraries Programme, n.d; Raitt, 2000; Dempsey, 2006; Duke & Jordan, Ltd, 2006). The National Digital Library Federation (now DLF) in the United States was established in 1995 as a program of the Council on Library and Information Resources. DLF began as an agreement between the Commission on Preservation and Access, and fifteen research libraries and archives. The primary goal of the DLF was to support library developments and implement a networked, digital library over the Internet, through the adoption
of standards and best practice to support interoperability (Commission on Preservation and Access, 1995).

Interest in and research on digital libraries continued to increase, spanning the fields of computer science, and library and information science. However, the desired outcomes for the fields were different: LIS wanted to improve services with technology, while computer science sought to improve and develop more advanced tools. By the mid-1990s, journals were launched dedicated to publishing research and developments in digital libraries, including D-Lib Magazine in 1995, Ariadne in 1996, and the International Journal of Digital Libraries in 1997. Several conferences began in the mid-1990s too, including the International Conference on Asian Digital Libraries, ACM Conferences on Digital Libraries, IEEE-CS Conferences on Advances in Digital Libraries, and European Conference on Digital Libraries.

Research Communities

Some of the earliest efforts for preserving, sharing, and making accessible digital assets were the establishment of social science data archives beginning in the early 1960s. With an increase in interest and scholarship in the social sciences after the Second World War, coupled with rapidly advancing technology that made processing data more attainable, researchers encountered new data management challenges (Shankar et al., 2014; Dennis, 1970). In fact, the Roper Center for Public Opinion Research, with a collection focused on commercial and news media survey data, was founded as early as 1947 at Williams College in Williamstown, MA. Now housed at the University of Connecticut, it is the oldest social science archive for machine-readable data in the United States (Bisco, 1966; Roper Center for Public Opinion Research, n.d.). A number of the first social science data archives have proven to be sustainable. The Inter-university Consortium for Political and Social Research (ICPSR), founded as the Inter-
university Consortium for Political Research in 1962, was organized as a consortium of twenty-two universities led by the University of Michigan, to promote institutional-level social science data archiving and research methods training for students and faculty. ICPSR now has more than 700 member institutions and continues to be a leader in data stewardship. Not only is ICPSR committed to preservation of and access to social science research, and data management and research methodology education, it was also part of early digitization efforts of historical materials for the archive (Inter-university Consortium for Political and Social Research, n.d.).

The Steinmetz Stichting at the University of Amsterdam, established in 1964, is now merged into the Data Archiving and Networked Services (DANS) of the Netherlands. The Steinmetz Stichting set out to collect, organize, and make internationally accessible social research from the Netherlands and promote social science research and education (Bisco, 1966). DANS upholds this objective by conducting research and training into sustainable digital stewardship, encouraging “researchers to archive and reuse data in a sustained manner,” and providing the infrastructure for them to do so (Data Archiving and Networking Services, n.d.). In response to the growth of social science research in the United Kingdom, the UK Data Archive was established in 1967 as the Social Science Research Council Data Bank. The Archive set out to address “the problem of sharing information about social surveys and the data generated by them” (UK Data Archive, 2007, p. 2). Changes in attitudes and policy about sharing research data in the 1970s increased acquisition of data for the Archive and today its collections continue to grow.

The number of both local and collaborative social science data archives increased. With this increase came efforts to improve collaboration, communication, and interoperability among repositories. Initiated by ICPSR and funded by a National Science Foundation (NSF) grant, the
Council of Social Science Data Archives (CSSDA) was formed in 1962 to support standards development, and encourage collaboration and communication among data archives largely in the United States. CSSDA eventually dissolved in the late 1960s due to changes in the United States funding environment (Adams, 2006; Bisco, 1966). Also beginning in 1962, the International Social Science Council sponsored several Conferences on Data Archives. These conferences resulted in the formation of the Standing Committee on Social Science Data Archives (SCSSDA) whose membership overlapped with the CSSDA. Emerging from SCSSDA meetings, IASSIST was established in 1974 as a professional association to unite the social science research, information science, and computer science communities around data management and use (Adams, 2006; International Association for Social Science Information Services and Technology, n.d.). Following the creation of IASSIST, in 1976 the Consortium of European Social Science Data Archives (CESSDA) was established “to provide seamless access to data across repositories, nations, languages and research purposes” and to “encourage standardisation of data and metadata, data sharing and knowledge mobility across Europe” (Consortium of European Social Science Data Archives, n.d.).

In addition to the data archiving activity in the social sciences, archives and curated databases for research data in the biological and chemical sciences were also becoming prevalent. The Cambridge Crystallographic Data Centre (CCDC) originated in the Department of Chemistry at the University of Cambridge in the United Kingdom in 1965. Through its curated Cambridge Structural Database (CSD) it provided data for the chemical and crystal structures of small molecules for academic and pharmaceutical research. Having garnered a successful international reputation, it became an independent non-profit company in the late 1980s, although it still maintains a close relationship with the Department of Chemistry
In the mid-1970s, collaboration between the CCDC and Brookhaven National Laboratory in New York established the Protein Data Bank archive. In the early years of the archive, the structural biology community was so small that requests for deposits occurred by way of personal communication through existing relationships. As technology advanced, so did the field which witnessed an increase in the amount of available data and shifts in attitudes about data sharing (Berman, 2008). Today, the Worldwide Protein Data Bank (wwPDB) comprises organizations from Europe, Japan, and the United States to form a single freely and publicly available archive of macromolecular structural data (Worldwide Protein Data Bank, n.d.). These early and successful curated databases facilitated international collaborations such as the UniProt Consortium and the Human Genome Project.

The physical science community was influential on a national level at establishing data centers. The establishment of the World Data Centers (WDC) in 1957 played a role in initiating a focus on the preservation and sharing of raw data in discipline-specific data centers (Duerr, Parsons, Marquis, Dichtl, & Mullins, 2004). In 1952, the International Council of Scientific Unions (ICSU) proposed conducting the International Geophysical Year, a coordinated initiative for researchers around the world to take part in geophysical scientific observations. Fifty-two World Data Centers were established to archive the collected data. Although this data was not digital, Duerr, Parsons, Marquis, Dichtl, and Mullins (2004) argue that the Centers set a precedent for distributed discipline-specific raw data stewardship and set the stage for a disconnect from archival preservation and records management. Thirteen of the WDCs were established in the United States, many of which are now housed in National Data Centers (National Research Council, 2003; National Oceanic and Atmospheric Administration, n.d.). The National Space Science Data Center, for example, was established in 1966 as a long-term
archive for NASA space mission data, and also houses the World Data Center for Satellite Information (National Aeronautics and Space Administration, n.d).

The increase in volume of data had a significant impact on data storage, data processing, and archival systems. With the rapid increase in the volume of data as technology advanced, NASA struggled early on with establishing sound data management practices and an effective archival system (Holmes, Poland, & Demmerle, 1967; United States General Accounting Office, 1987). As early as 1964 NASA’s Goddard Space Flight Center began work on an “archival system which could handle the large volumes of data at a substantial saving in cost and with improved performance” (Holmes, Poland, & Demmerle, 1967, p. 2). Although the National Space Science Data Center was established in 1966, in the late 1980s data management and archiving practices at NASA were still insufficient (United States General Accounting Office, 1987). The Government Office of Accountability concluded that “unless it continues to improve its management of space science data, NASA will be hard-pressed to effectively and efficiently handle the massive volumes of data expected from its future mission” (United States General Accounting Office, 1990, p. 3). Working to address this insufficiency, NASA established Distributed Active Archive Centers (DAACs) to store data that were still in active use. These were linked to a centralized search interface for accessing data remotely (Duerr, Parsons, Marquis, Dichtl, & Mullins, 2004).

It was in this environment that NASA initiated a call for a standards development effort for an archiving system reference model. The process began in 1994 with the objective of creating a functional model to improve the archiving and preservation of NASA’s space science data. The scope of the project extended beyond its initial audience. In open forum discussions, contributions came from the preservation community, including participants from the archives
and computer science fields. In 2000, a draft of the Open Archival Information System Reference Model (OAIS) was circulated and became an ISO standard in 2003. The OAIS Reference Model provides a framework and shared vocabulary for discussing the functions of a preservation system and facilitates cross-disciplinary communication. The OAIS Reference Model remains one of the most important digital curation standards and models in use today (Lee, 2010).

The research communities were also leaders in developing digital preprint and institutional repository prototypes. Research communities have long shared findings both through the formal channels of publication, and informal channels of copying and mailing preprints. By the end of the 1980s, emailing preprints was beginning to replace the postal service (Ginsparg, 1994). Within the scholarly community, technology was changing communication in a “remarkable new way that is not only incomparably more thorough and systematic in its distribution, potentially global in scale, and almost instantaneous in speed, but so unprecedentedly interactive that it will substantially restructure the pursuit of knowledge” (Harnad, 1990, p. 342). Physicist Paul Ginsparg initiated arXiv in 1991, an automated distribution archive service for preprint abstracts and papers of physics research. The purpose of the preprint archive was not to replace peer-review journals, but rather “to provide equal and uniform global access to prepublication materials” (Ginsparg, 2004). Today, arXiv has expanded to include more research fields and is hosted by Cornell University Library, and is just one example of a preprint repository. These discipline-specific and institutional repositories were transforming scholarly communication, and provided a means of scholars to share research and manage, to some extent, their digital assets (Jones, 2006).
In the late 1990s, the Open Access Movement found strength with the growth of the World Wide Web and was coupled with the emergence of the Open Source Initiative (OSI) and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). These informed the development of institutional repositories, and continue to inform the development of technologies in this area and cyberscholarship practices (Van de Sompel & Lagoze, 2000). Open source and OAI-PMH compliant institutional repositories software began to emerge in the early 2000s: EPrints in 2000 out of the University of Southampton in the United Kingdom (Cervone, 2004), DSpace launched in 2002 out of MIT (Smith et al., 2003), and the public release of Fedora in 2003 out of Cornell University and the University of Virginia (“History”, n.d).

The Preservation and Curation Communities

The landscape began to change rapidly in the mid-1990s. Technology was transforming how people communicated, accessed information, and approached research. Computer innovations advanced exponentially, coupled with the trend in technology to be faster, smaller, and more affordable. The World Wide Web was conceived of in 1989 by Tim Berners-Lee, a scientist at CERN; in 1993 it was released into the public domain and began to flourish. The key concerns of research communities that drove the early endeavors prior to the 1990s continued to inform and propel the later efforts as interest began to increase. Attention in the archive and library communities shifted from improving their traditional services to learning how to function and stay relevant in this new environment.

The Preservation Community

A key 1996 report acted as a catalyst pushing the digital preservation agenda to the forefront of the discussion in archives and libraries. Previous conversations centered on issues, such as improving access to library holdings, data archiving for scientific progress, and archival
appraisal of electronic records. In 1994, RLG and the Commission on Preservation and Access (CPA) assembled the Task Force on Digital Archiving to investigate “continued access indefinitely into the future of records stored in digital electronic form” (Waters & Garrett, 1996, p. iii). A draft of the publication *Preserving Digital Information* was available for comment in 1995, and the final version of the report was published in 1996. This significant report identified the technological, strategic, and conceptual issues of digital preservation. Importantly, it differentiated between digital archives and digital libraries, stressing that digital libraries tended to focus on access issues, not on the longevity of digital content.

Australia forged ahead as leaders in digital preservation initiatives and research. The National Library of Australia (NLA) established one of the first digital preservation sections in 1995. In 1996, NLA established the web archive, Preserving and Accessing Networked Documentary Resources of Australia (PANDORA), to archive online material relating to Australia and Australians (Webb, 2002). That same year, Brewster Kahle founded the Internet Archive in the United States with a less selective mission of creating an internet library (Lyman and Kahle, 1998). The following year, in 1997, the Preserving Access to Digital Information Safekeeping project (PADI) was established to provide a gateway to digital preservation resources to help the archive, library, and preservation communities remain up-to-date with best practices, developments, and progress in the field (Webb, 2002; National Library of Australia, n.d.).

Developments in Australia informed early thinking about digital curation and preservation in the international community. By the mid-1990s the concept of the records continuum rather than the traditional lifecycle approach to archives and records management practices was formally articulated. In the Records Continuum, the boundary between records
manager and archivist is blurred because management of a record is conceived of as a continuous process across space and time. The model provides an integrated framework for recordkeeping systems in which the stages of a record’s activity, from creation to maintenance and use, transition seamlessly or exist concurrently (Upward, 1994). This perspective informed lifecycle approaches to digital curation, in which early intervention is critical to longevity and the cycle perpetuates continuously with the reuse and transformation of data. In 2002, the National Archives of Australia (NAA) published a Green Paper of their approach to digital preservation. In the paper, the authors outlined the Performance Model, an approach to digital records that treats them as a dynamic action, or a performance, rather than a static object (Heslop, Davis, & Wilson, 2002). The NLA also played the central role in writing UNESCO’s Guidelines for the Preservation of Digital Heritage in 2003.

With these new challenges and the drive for finding solutions came increased funding opportunities for digital preservation projects. JISC was founded in 1993 as an advisory committee to funding bodies in the United Kingdom with the mission “to provide vision and leadership to universities, to stimulate innovation and fund research into technology developments for the benefit of all universities and colleges” (JISC, n.d.). JISC managed the funding for the eLib program that funded research in digital libraries in the 1990s and also co-funded the Digital Curation Centre (DCC) in 2004. In 1996, the Institute of Museum and Library Services (IMLS) was established by the Library and Museum Services Act, combining the Institute of Museum Services and the Library Programs Office. At the same time IMLS was created, the Library Services and Technology Act (LSTA) was also enacted, earmarking grant funds for technological innovations, resource sharing, and interoperability (United States 108th Congress, 2003). As part of this, the IMLS funded a large number of cultural heritage
digitization projects in the late 1990s and early 2000s (Institute of Museum and Library Services, 2003).

An early and influential funding body was the European Commission, which launched their research and development framework programmes in the early 1980s to improve innovation and collaboration among European researchers. As noted above, support for the library community occurred as early as 1990, from the Third and Fourth Framework Programmes, with the launch of the Telematics for Libraries Programme. Funding continued to increase throughout the later programmes, focusing on research and development projects for digital preservation. Between 2001 and 2011 the European Commission awarded over 90 million Euros for projects related to digital preservation (Strodl, Petrov, & Rauber, 2011; European Commission, 2001; Johnson, 1996). These projects included ERPANET (Electronic Resource Preservation and Access Network, 2001-2004), the DELOS Network of Excellence on Digital Libraries (2004-2007), Digital Preservation Europe (DPE, 2006-2009), and Planets (Preservation and Long-term Access through Networked Services, 2006-2010).

The increased recognition of the potential benefits of collaboration in the digital preservation agenda, both nationally and internationally, was not only for research and development but also for community and infrastructure building. In 2000, the Library of Congress entered the digital preservation conversation and launched the National Digital Information Infrastructure and Preservation Program (NDIIPP). NDIIPP was interested in cooperative efforts to strengthen networks and community, build a national infrastructure for digital preservation, and achieve shared solutions (Library of Congress, 2002). The Digital Preservation Coalition was established in the United Kingdom in 2001 as an institutional membership organization, to foster partnerships and cooperative projects. With funding support
and expanding networks, collaborative and interdisciplinary projects, research, and best practices for the preservation of digital records increased through the 2000s.

**The Curation Community**

The phrase “digital curation” was originally coined from discussions about the long-term management and preservation of research data. The first use of the phrase can be traced back to an invitational seminar, *Digital Curation: Digital Archives, Libraries and e-Science Seminar*, sponsored by the Digital Preservation Coalition and the British National Space Centre, held in London in October 2001 (Beagrie, 2006; Beagrie & Pothen, 2001). The term *curation* had already been used in various contexts outside of libraries and museums, including as a critical component of database management and publication in the biological sciences, and it had been introduced as an aspect of the information infrastructure for the United Kingdom E-Science Programme (Taylor, 2001). Discussions about using the term curation in this context were further advanced at a task force meeting in 2002 to discuss a strategy for establishing an infrastructure for maintaining research data in the United Kingdom (Ray, 2012; MacDonald & Lord, 2003).

These meetings culminated in the launch of the United Kingdom’s Digital Curation Centre (DCC) in 2004, jointly funded by JISC and the Engineering and Physical Sciences Research Council’s e-Science Core Programme. The DCC is a collaborative research and development initiative led by the University of Edinburgh, HATII at the University of Glasgow, UKOLN at the University of Bath, and the Science and Technology Facilities Council. The DCC defines digital curation as “maintaining, preserving and adding value to digital research data throughout its lifecycle” (Digital Curation Center, n.d.b). Although the DCC’s definition of digital curation still reflects a focus on research data, definitions used by other initiatives, such as
the University of California Curation Center and the Digital Curation Unit at the Athens Research Center, outwardly reflect a more comprehensive view of digital assets.

Adopting the phrase digital curation was an effort to recognize that longevity and reuse of data combines the activities of digital preservation and data curation, and is an ongoing process influenced by complex cultural, organizational, and technical issues. The phrase attempts to capture the position that caring for digital assets over time involves more than just the technological aspects of preservation or end-of-lifecycle disposition to an archive; and that digital assets require active management from the point of creation and throughout their lifecycle, including adding value through annotation, contextual documentation, and metadata. Digital curation, as a phenomenon, necessitates top-down organizational support and bottom-up cultural changes influencing how individuals manage their data. By consolidating the activities of digital preservation and data curation, and their communities, the area of digital curation attempts to address shared problems and solutions across all digital content types and formats (Beagrie, 2006; Dallas, 2007).

**Summary**

In this chapter I presented a narrative of the historical developments leading up to the emergence of digital curation. Some of the earliest efforts for preserving, sharing, and making accessible digital assets began within the research communities, with social and physical science data archives and curated databases in the biological sciences. Early reflections about machine-readable records at NARA in the United States focused on appraisal issues and the “recordness” of this form of media. Discussions about the preservation of electronic records in the archival literature gradually developed. NARA received its first electronic record transfer in 1970. Prior to the 1990s, library communities focused their efforts on using computers for bibliographic
control of their holdings and automation of library catalogs using online public access catalogs (OPACs). Interest and research in the digital library movement began in earnest in the 1990s. More recently, the case for libraries, especially large research libraries, to play a role in data archiving, curation, and management has been strongly argued.

A key 1996 report, *Preserving Digital Information*, acted as a catalyst for pushing the digital preservation agenda to the forefront of the discussion in archives and libraries. This report identified the technological, strategic, and conceptual issues of digital preservation. With these new preservation challenges came increased funding opportunities for collaborative research and development projects. Coining the phrase “digital curation” in 2001 was an effort to recognize that longevity and reuse of data combines the activities of digital preservation and data curation, and is an ongoing process influenced by complex cultural, organizational, and technical issues. The term emerged from discussions about the long-term management and preservation of research data; however current use of the term digital curation reflects a more comprehensive view of digital assets.

In the next chapter, I review relevant literature about education and curriculum development in the area of digital curation. Roles and responsibilities of researchers and information professionals have changed in response to activities related to digital curation and these new responsibilities require new skills and competencies. The purpose of this chapter is to emphasize the function of education in the discussion of the potential emergence of digital curation as a discipline.
CHAPTER 5: DIGITAL CURATION IN THE LIS/ARCHIVE CURRICULA

In this chapter I review relevant literature about education and curriculum development in the area of digital curation. This chapter situates digital curation curriculum development within higher education, specifically within LIS and archival programs. It also discusses continuing education, professional development, and training in response to the changing roles and responsibilities with regard to activities related to digital curation. The function of this chapter is to emphasize the role of education as a key criterion for identifying an emerging discipline.

This chapter is organized in four main sections. The first section, Emerging Roles and Responsibilities, provides an overview of literature that addresses knowledge and skill gaps and the need for professionals educated and trained in digital curation. The second section, Identifying Core Competencies, is a review of relevant literature that has identified knowledge and skill requirements for activities related to digital curation. The third section, Curriculum Development, examines curriculum development, professional training opportunities, and pedagogical approaches in the area of digital curation. The final section, Outlook, is a review of relevant literature that provides analysis of the implementation and outcomes of these curricular developments.

Emerging Roles and Responsibilities

In 1997, the president of the Commission on Preservation and Access and the Council on Library Resources noted: “The role of the librarian of the twenty-first century will bear little resemblance to that of the typical librarian today - on this there is nearly universal agreement” (Marcum, 1997, p. 35). We have increasingly witnessed new technology-focused responsibilities and IT skills added to traditional roles, new data-centered positions added to library departments, and new specializations added to education programs (Xia & Wang, 2014;
Cox & Corrall, 2013; Goetsch, 2008; Khurshid, 2003). However, several authors have identified gaps in the education of LIS and archives curricula for skills and knowledge related to digital curation. In their chapter on education in *Aligning National Approaches to Digital Preservation*, Davidson, Corrall, Coulbourne, and Rauber (2012) observe that “if education and professional development training in the library and information science sector do not evolve to cover data management and curation, there is a risk that librarians and other information specialists will not be able to contribute appropriately to the management of research data” (p. 276).

Advances in technology have not only influenced how archivists and librarians perform their work but have created new roles and responsibilities that require a new knowledge and skills base. In a comprehensive literature review and content analysis, Allard, Mack and Feltner-Reichert (2005) identified six new requirements of librarians working in institutional repositories: understanding institutional repository software; project planning, management, and implementation; defining collections; metadata guidance and standards; reviewing institutional repository submissions; and training contributors and users of the institutional repositories. The researchers concluded that the role of librarians is adapting to meet the needs of this new information environment. In a 2009 study that looked at job advertisements for positions in academic libraries, Choi and Rasmussen confirmed that there has been a shift in staffing needs towards qualifications in digital collections, services, and technology applications. The researchers note:

> While confirming many anecdotal discussions on new skills needed of library professionals in the digital environment, the result sheds light on the changing requirements for library professionals who are equipped to meet these new challenges.
This change brings new demands for professional skills and expertise which must be met by enhancing professional development, training, and education (p. 466).

Although the primary purpose of the study centered on digital libraries, its scope included other professional library positions with “digital” in the position title, department name, or duties and qualifications (Choi & Rasmussen, 2009).

These emerging roles and responsibilities necessitate clear definitions. In a report commissioned by JISC about career trajectories and education in data management and data science in the United Kingdom, Swan and Brown (2008) identified four distinct roles: data authors or creators, data scientists or specialists, data managers, and data librarians or archivists. The authors recognized that responsibilities for these roles not only overlap but the descriptions of the roles are not consistent within the data community. The authors concluded that the role and career trajectory for data librarians needs to be formalized and defined, along with appropriate curricula (Swan & Brown, 2008). At present, a consensus study is being conducted that delineates the diversity of digital curation roles, career paths, and employment opportunities; the variety of skill sets required; and education and training models. The study, *Future Career Opportunities and Educational Requirements for Digital Curation*, is funded by the Institute of Museum and Library Services (IMLS), National Science Foundation (NSF), and Alfred P. Sloan Foundation (Board on Research Data and Information, n.d.). Although the report was due out the first half of 2014, as of November 2014 it had yet to be made publically available.

Archives and libraries are also faced with establishing their role in the burgeoning cyberinfrastructure, and adapting to changes in scholarship that places emphasis not just on publication of results, but on data as a core output of research. For example, academic libraries are increasingly involved in research data curation and e-science services. Hey and Hey (2006)
remarked that “the e-Science revolution will put libraries and repositories centre stage in the development of the next generation research infrastructure” (p. 526). A number of libraries have expanded their instruction services to include data management training, in the form of library-sponsored workshops as well as collaborative efforts for integrating data management skills into the curriculum (Adamick, Reznik-Zellen, & Sheridan, 2012; Johnston, Lafferty, & Petsan, 2012; Piorun et al., 2012).

Several surveys to date have investigated the role of the library in the institutional repository landscape. A 2009 Australian study found that, for the most part, Australian institutional repositories were the responsibility of the institution’s library and in turn 71.1% were headed by an individual with a library background (Kennan & Kingsley, 2009). A 2010 United Kingdom study found a similar result: 78.6% of staff working in United Kingdom repositories possessed a library and information science background (Wickham, 2010).

Although information professionals are taking on these new roles and responsibilities related to digital curation, their contributions lack recognition within the wider research and professional communities. A 2012 Italian study found that although the repository manager is the most important role in Italian institutional repositories, the positions were often part-time, highlighting a lack of professional recognition. The study’s authors also note that the current higher education curricula did not meet the needs of this new role (Cassella & Morando, 2012). In a 2012 German survey of researchers from eleven disciplines reported by Osswald and Strathmann (2012), the study’s investigators found that there was no clear role defined for library involvement in curation and preservation of research data, although there was a “broad understanding that methodological expertise of libraries and librarians would be helpful to improve research data-related activities of scientists and research institutions” (p. 8). The
authors of the report concluded with a call to redesign LIS curricula to include competencies for
digital curation and preservation of research data.

**Identifying Core Competencies**

The improvement of educational and training opportunities in this area involves
identifying the new knowledge and skills base required for information professionals engaged in
activities related to digital curation. A document published by the Securing a Hybrid
Environment for Research Preservation and Access (SHERPA) project in 2008, and updated in
2009 and 2011, was developed to assist writing job descriptions for institutional repositories
staff. The document identifies two key roles, repository managers and repository administrators,
and eight responsibility and skill areas: management; software; metadata and interoperability
standards; storage and preservation; content, including familiarity with intellectual property
rights; advocacy, training, and support; communication; and current awareness and professional
development. The document does not focus on any particular repository type or position within
a repository, but rather covers the spectrum of abilities, knowledge, and skills needed for the
development and management of any institutional repository, and therefore can be realized in a
combination of various staff roles and external partners (Pryor & Donnelly, 2009; Robinson,
2009; Wickham, 2011).

At a 2008 meeting of the Research Data Management Forum (RDMF) participants
identified a set of core skills needed by people in the data management sector using the four roles
identified by Swan and Brown (2008) as a guide. RDMF is a United Kingdom-based group that
encourages dialogue among the diverse stakeholders engaged in data management activities and
is sponsored by the Digital Curation Centre (DCC) and the Research Information Network
(RIN). Although the group acknowledged some responsibilities belonging uniquely to each role,
there existed a number of essential core skills, including metadata, preservation, and standards development that were needed by multiple roles (Pryor & Donnelly, 2009; Swan & Brown, 2008). This further illustrates that there are not yet clear boundaries between these positions.

Several studies have identified core competencies for professionals engaged in activities related to digital curation in order to directly inform curriculum development. In a 2008 study, researchers from the DigCCurr project team at the University of North Carolina at Chapel Hill conducted a survey of information professionals working in the area of digital curation and preservation to help inform their curriculum development project. Survey respondents identified a strong need to emphasize the functions of preservation planning and implementation in a core curriculum, followed by description, organization, and intellectual control. An emphasis on administration planning, establishing policies and standards, and data management followed closely. Respondents ranked technical skills, practical experience, and interpersonal skills highest for desired professional competencies when filling a job vacancy (Tibbo, Hank, & Lee, 2008). An analysis of job advertisements was undertaken as part of the University of North Texas’ curriculum development project iCAMP (Information: Curate, Archive, Manage, Preserve) to identify skills needed to perform essential job functions associated with digital curation. The research identified seven core competency areas including communication and interpersonal skills; curation and preservation skills; knowledge of curation services, technologies, and tools; management and planning; and familiarity of systems and models (Kim, Warga, & Moen, 2013).

In an effort to extend the data curation curriculum at the University of Illinois at Urbana Champaign to include humanities data, researchers conducted a study to better understand the roles and skill sets of information professionals working with digital humanities data.
Researchers conducted a series of interviews with directors and senior staff of digital humanities centers. Skills relating to interoperability, markup, database design, and metadata were highly ranked by all respondents, and most respondents gave a high rank to project management skills and instruction. Researchers found a range of desired skills with little consensus among respondents (Muñoz, Varvel, Renear, Trainor, & Dolan, 2011).

Other research has focused on identifying competencies for curriculum development for eScience information professionals. Using a combination of interviews and focus groups, researchers at Syracuse University conducted a job analysis study focusing on the core competencies of eScience professionals. Researchers identified a range of skills that centered on three key functional areas: skills related to data such as data curation activities; skills related to things such as maintaining hardware or managing budgets; and skills related to people such as communication and interpersonal skills. These findings translated into ten recommended courses for preparing eScience professionals (Kim, Addom, & Stanton, 2011; Stanton et al., 2011). In 2013 researchers at Wuhan University in China conducted a study designed to improve eScience education in the Chinese LIS curriculum. From an analysis of job advertisements, mostly from the United States, the study found that the key functions of eScience specialists included consultation and reference services for data curation, instruction and training on data curation, and creation of metadata standards. The top required qualifications were soft skills, namely teamwork, communication, and interpersonal skills, followed by familiarity with data curation tools and technologies. The top preferred qualifications were a second advanced degree, knowledge of library and information science, and familiarity with metadata standards (Si, Zhuang, Xing, & Guo, 2013). Recommended coursework differed between the two studies; however researchers from both studies recommended courses in research methods, database
management, introduction to data curation, and overview of the cyberinfrastructure and eScience environment.

**Curriculum Development**

The archival profession has long recognized the impact of technology on archival records and practice, and the need to diversify their skill sets and knowledge base and stay abreast of the changing technological landscape; however response in archival education has been slow. In 1993, at a time when archival education was still developing, Cox noted that although archivists understood the impact of electronic records and automated techniques on the profession, archival educators were remiss in incorporating these areas into curriculum. Stielow (1993) questioned whether the traditional skill set of archivists, such as paleography and concepts of intrinsic or artifactual value, still applied to digital records that were “not fixed in time and place but characterized by mutability” and could be “constantly updated or combined into new records” or “written at diverse times and places by multiple authors” and “copied or transmitted without differentiating them from the originals” (p. 61-62). Stielow advocated for a model of archival education that integrated information technology into the curriculum, rather than offering separate and special courses.

The challenge remains how to establish an archival curriculum that provides core archival knowledge and sufficiently incorporates information technology. Tibbo (2006) found that although there are technology courses available as electives in archival programs, few students are able to take them because LIS core courses and archival requirements, including a practicum, crowd their credit hours. Still, a 2012 Scandinavian study of archivists working with electronic records in local governments found that there is still a need for archival education to address new competencies and increase technology requirements in the curriculum (Kallberg, 2012).
With the recognition of new roles and responsibilities of information professionals and efforts to identify core competencies required for activities related to digital curation, graduate-level and continuing education in these areas needed improvement. In their 2008 JISC report, Swan and Brown note:

Library educators have an important role to play in planning for and delivering appropriately skilled people to meet the latent demand for data librarians to manage the libraries’ potential data curation role. Yet very few library and information science schools currently teach the skills that future data librarians will need (p. 25).

A 2011 study of repository staff in Australia and New Zealand found that for the most part, respondents reported that on-the-job training was their primary method of acquiring needed skills. Further, 83% of respondents reported having never been exposed to digital repositories in their formal education even though 54% had library and information science degrees and 15% had information technology degrees (Simons & Richardson, 2012).

Several cooperative international initiatives were launched to address these curricular concerns in higher education. The International Data Curation Education Action (IDEA) Working Group held four meetings during 2008 – 2009. The working group, organized by the DCC, IMLS, and the School of Information and Library Science at the University of North Carolina at Chapel Hill, was formed to explore international collaboration for the development and progress for data curation education and training curricula (Hank and Davidson, 2009; Ray, 2009). In 2011, JISC sponsored the International Curation Education (ICE) Forum; in attendance were many of the same participants of the IDEA Working Group. The questions that guided the ICE Forum focused on forms of digital curation education needed, current approaches
to education, the skill and knowledge gaps that need to be addressed, and ways that educators and trainers can most effectively share resources (Kilbride & Molloy, 2011).

Several of the efforts to develop higher education digital curation curriculum focused on cooperative and shared outcomes. For example, the second phase of the German project Network of Expertise in Long-Term Storage of Digital Resources (NESTOR, 2006-2009), identified the need for education and training in the area of digital curation. Through partnerships with higher education institutions, mostly in Germany, they provided training as well as support for graduate-level curriculum development. The project partners maintained their collaboration after the end of the project phase and continue to provide education and training opportunities (Schrimpf, 2009). DigCCurr (Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum, 2006-2013) was a two-phase IMLS grant-funded project at University of North Carolina at Chapel Hill in collaboration with the National Archives and Records Administration (NARA). The DigCCurr project set out to accomplish several key tasks: create an openly accessible digital curation curricular framework; provide doctoral education opportunities with the goal to prepare future faculty for research and teaching in the area of digital curation; extend education to the professional community through professional institutes and training; and create an environment for educational networks and community building through regular symposiums, meetings, and the Digital Curation Exchange web portal (Tibbo, Hank, & Lee, 2008; Lee, Tibbo, & Schaefer, 2007; Hank, Tibbo, & Lee, 2010). The digital curation curriculum framework is designed as a six-dimensional matrix to aid in the development of curricula by endorsing core areas of knowledge and skills. The six dimensions, each detailed and supported by research, are: mandates, values, and principles; functions and skills; professional, disciplinary, or institutional/organizational context; type of
resource; prerequisite knowledge; and transition point in the information continuum. The framework was created to be openly accessible, scalable, and extensible (Lee & Tibbo, 2011).

Other projects were designed for more local application, focusing on establishing concentrations in specific masters-level program. One of the earliest specializations within an LIS curriculum emerged out of the IMLS grant-funded Data Curation Education Program (DCEP) at the Graduate School of Library and Information Science at University of Illinois at Urbana-Champaign. The program was awarded funding in 2006 to create a data curation specialization and host annual professional summer institutes (Palmer, Thompson, Baker, & Senseney, 2014). The specialization was later expanded to include coursework in digital humanities data (Renear, Muñoz, & Trainor, 2010). Additional grant funding was obtained in 2011 to support Data Curation Education in Research Centers (DCERC), a program designed to provide students with field experience. DCERC is a collaborative undertaking between the University of Illinois at Urbana-Champaign, the University of Tennessee at Knoxville, and the National Center for Atmospheric Research (NCAR) (Palmer, Allard, & Marlino, 2011). The School of Information at the University of Michigan established a Preservation of Information specialization in 2007 to more directly attend to digital curation education. The specialization assumes a holistic approach to preservation across all media and formats, rather than educating for analog and digital materials separately (Yakel, Conway, Hedstrom, & Wallace, 2011; Yakel, Conway, & Krause, 2009).

In the United Kingdom and Europe, the occurrence of degree granting programs rather than specializations is on the rise. Aberystwyth University in Wales, Kings College in London, and Robert Gordon University in Scotland offer Master’s degrees in Digital Curation. The Humanities Advanced Technology and Information Institute (HATII) at the University of
Glasgow, recognized for its leadership in digital curation and preservation education and research since its establishment in 1997, offers a Master’s degree in Information Management & Preservation (Digital)/(Archives & Records Management). In Italy, a Master in Education, Management and Preservation of Digital Archives (Formazione, gestione e conservazione degli archivi digitali, FGCAD) is offered jointly by the University of Macerata and the University of Padua. A recently established degree granting program at the University of Turin focuses on preparing information professionals for careers in digital curation, an area they approach as encompassing multiple roles and shared responsibilities. The curriculum therefore highlights not only competencies and skills for sustainable digital preservation, but also communication and team-building with information technology specialists, computer scientists, and other stakeholders (Vivarelli, Cassella, & Valacchi, 2013).

Several schools have launched post-graduate certificate programs to address the knowledge and skills needed to supplement current LIS and archive curricula. For instance, the School of Information Studies at Syracuse University offers a Certificate of Advanced Study in Data Science, the School of Library and Information Science at Simmons College offers an online post-master’s Digital Stewardship Certificate, and the School of Information and Library Science at the University of North Carolina at Chapel Hill offers a Graduate Certificate in Digital Curation and a post-masters Certificate in Data Curation. Currently, in its last year of IMLS funding, the iCAMP project at the University of North Texas seeks to develop a competency-based curriculum for their Digital Curation and Data Management Graduate Academic Certificate. The certificate comprises four online courses aimed at three primary audiences: current LIS students, post-master’s LIS professionals, and graduate students in any discipline.

One of the first such certificate programs was the IMLS grant-funded Digital Information Management certificate program at the School of Information Resources and Library Science (SIRLS) at the University of Arizona, in partnership with the Arizona State Library, Archives and Public Records (ASLAPR). The program was designed to be interdisciplinary to reach a diverse audience of information professionals (Fulton, Botticelli, & Bradley, 2011; Botticelli, Fulton, Pearce-Moses, Szuter, & Watters, 2011). In determining how to integrate digital curation competencies into the curriculum, SIRLS opted for a certificate program to signal “its intention to expand the boundaries of a traditional library/archives education, but not to redefine or replace the core elements of the existing master’s degree” (Botticelli, Fulton, Pearce-Moses, Szuter, & Watters, 2011, p. 156).

In addition to curriculum development projects resulting in specializations or post-graduate certificates, a number of existing LIS and archive programs have been adding courses to address the gap in skills related to digital media. For example, both the Library Information and Archive Sciences Department at the University of Malta and the School of Information Management at Victoria University of Wellington, New Zealand have added digital curation courses to their existing curricula to accommodate local needs in their small countries (Dobreva & Oliver, 2013). Further, Bastian and Yakel (2006) conducted a study to identify an archival core curriculum in North America and found that preservation courses were consistently offered throughout the programs evaluated; however, only one program offered a course on digital preservation at that time (Bastian & Yakel, 2006; Bastian, Cloonan, & Harvey, 2011). Early findings of an unpublished update of their curriculum study reported by Bastian, Cloonan, and
Harvey (2011) found that the number of programs offering a course on digital preservation had increased to eleven.

**Continuing Education and Professional Training**

Research supports the need for continuing education and training for professionals in all areas related to digital curation, including digital preservation and data management. A needs assessment survey was conducted in 2010 to define United States national training needs prior to the launch of the Digital Preservation Outreach and Education (DPOE) initiative by the Library of Congress. The survey found that although the majority of respondents had some kind of staff member assigned to digital preservation, only 33% reported having dedicated full or part-time professional staff, while 49% reported that duties were assigned as needed and 14% reported having volunteers doing digital preservation. Twenty-two percent reported having no digital preservation staff. When asked if there were funds available in their organization for professional development or training, 34% reported that there were no funds available. Those reporting that there were no funds available were more likely to be from organizations with small staff sizes (Library of Congress, 2010).

Results of a 2012 survey conducted as a follow-up to the DPOE training webinar found that training needs differed somewhat between research libraries and cultural heritage institutions. The researcher who conducted the survey concluded that the results showed a need for free or low-cost preservation training that focused on content type, specifically born-digital special collections, born-digital records, and digitized collections (DeRidder, 2013). Research as part of the APARSEN (Alliance for Permanent Access to the Records of Science in Europe Network) project in 2012 supported the results of these needs assessments, finding that the key
challenges were limitations in the availability of training in digital preservation and securing the funding to attend (Kilbride, Cirinnà, & McMeekin, 2012).

Continuing education and professional training has developed alongside efforts in master’s level LIS and archive education. Not only is this reflected in the trend of post-graduate certificate programs but other training efforts, such as professional institutes, have been aspects of several higher education curriculum development projects including DigCCurr and DCEP. Many of the continuing education and professional training initiatives have come out of collaboration with professional communities. Opportunities have increased over the past several years with training and workshops sponsored by the Digital Curation Centre, the Digital Preservation Roadshow and the Advanced Practitioner Training course offered by the Digital Preservation Coalition (DPC), the Digital Preservation Training Programme run by the University of London Computer Centre, the Digital Archives Specialist Certificate through the Society of the American Archivists, and the online MANTRA (http://datalib.edina.ac.uk/mantra/) Research Data Management Training for researchers and information professionals.

Several of these continuing education and professional training programs have established curriculum frameworks and instruction models. One of the earliest professional training initiatives was the Digital Preservation Management Workshop and Tutorial, initially developed by the Cornell University Library through National Endowment for the Humanities (NEH) grant funding in 2003 (http://www.dpworkshop.org/). Currently maintained by MIT Libraries, the Digital Preservation Management Workshop is a series of week-long workshops and an accompanying online tutorial. The curriculum framework for the Workshop supports
developing a digital preservation program based on sustainable organizational, technological, and resources infrastructures (Kenney & McGovern, 2003).

The Digital Preservation Outreach and Education (DPOE) program launched by the Library of Congress boasts a mission to “foster national outreach and education about digital preservation by building a collaborative network of instructors and partners to provide training to individuals and organizations seeking to preserve their digital content” (Library of Congress, n.d.). In 2010, DPOE conducted a needs assessment to identify training requirements, and then in 2011 held their first training session. In the DPOE model, the Library of Congress trains professionals to conduct workshops based on the DPOE Baseline Digital Preservation Curriculum, creating a geographically dispersed network of trainers. The Baseline Digital Preservation Curriculum focuses on six central themes: identify types of digital content; select what will be preserved; store selected content for the long-term; protect content from threats; manage stored content; and provide access over time (Library of Congress, n.d.).

The Digital Curator Vocational Education Europe (DigCurV) was a project funded by the European Commission from 2011-2013 to identify and assess training needs and current training offerings, and to establish a curriculum framework from which training programs in digital curation can be developed or compared. The curriculum framework identifies skills and competencies in four categories: knowledge and intellectual abilities, personal qualities, professional conduct, and management and quality assurance. The framework provides three lenses based on types of staff conducting curation activities: practitioners, managers, and executives. The skills and competencies required for each lens differs, allowing training curriculum to be tailored based on the needs of the audience (Molloy, Gow, & Konstantelos, 2014; Moles, 2014; Moles & Ross, 2013).
Pedagogy

The pedagogy for teaching digital curation in LIS and archive programs progressed in conjunction with the development of new courses and curriculum. Similar to efforts in digital library education (Mitchell, 2014a; Mitchell, 2014b; Lin & Abels, 2010; Pomerantz, Abbas, & Mostafa, 2009), digital curation curriculum has sought to integrate technology into the curriculum through hands-on use of applications and tools or project-based assignments in a laboratory setting. At the University of Texas at Austin, the Digital Archiving and Preservation course has incorporated hands-on assignments and project-based assignments since 2000, providing students with direct exposure to digital materials. The learning model established by Galloway has evolved and now the digital archiving program has a dedicated laboratory (Galloway, 2011).

Virtual laboratories have also been established to accommodate online learning and facilitate sharing instructional material among educators. In 2010 the School of Library and Information Science at Simmons College and the Department of Information Technology and Media at Mid Sweden University joined together to create the Digital Curriculum Laboratory (DCL) (Bastian, Cloonan, & Harvey, 2011; Anderson, Bastian, Harvey, Plum, & Samuelsson, 2011; Bastian, Harvey, Mahard, & Plum, 2010). The DCL contains exercises, teaching scenarios, and digital repository applications to form a virtual workspace or “sandbox” in which students can “learn about and experiment with digital tools and materials in digital environments” (Harvey & Bastian, 2012, p. 2). At Simmons, the DCL is used to enhance teaching and learning in several of the courses, most notably in the Preservation Management and Cultural Heritage tracks and the range of digital stewardship courses (Bastian, Cloonan, & Harvey, 2011; Bastian, Harvey, Mahard, & Plum, 2010). The iCAMP project at the University
of North Texas also has created a virtual laboratory for their digital curation and data management curriculum. The virtual lab is a central part of their vision to structure the digital curation curriculum after the teaching hospital model, which is based on engagement, hands-on training, and supervision (Moen, Kim, & Halbert, 2011; Helsing, Lewis, & Warga, 2013).

An internship requirement has played a central function in digital curation curriculum development. The information disciplines have historically relied on internships to address the application and relationship of practical and scholarly aspects of the field (Grogan, 2007). Of archival education, Bastian and Webber (2008) note that “too great an emphasis on internships may reduce archival education to mere training, but with no practical learning at all, students are ill-equipped to enter a workplace that places a high value on experience” (p. 15). Library and information science education programs, including archive and records management, continue to rely on practicum or internship components as part of their curriculum.

Several programs require internships specific to their digital curation specializations. Internships provide the basis for DCERC. DCERC provides graduate students specializing in data curation field experience (Palmer, Allard, & Marlino, 2011). The Preservation of Information specialization at the School of Information at University of Michigan has always required field experience through the School’s Practical Engagement Program. However, identifying internships for digital curation and preservation students in an area that is still maturing is challenging. To adapt to these challenges, faculty at the University of Michigan combined internships with a cognitive apprenticeship model, allowing instructors to co-mentor students while in the field (Yakel, Conway, & Krause, 2009). San Jose State University in California and Victoria University of Wellington in New Zealand have piloted a virtual internship project to support digital curation education and promote an international community
of practice. San Jose State University offers an online master’s degree with an internship option. The virtual internship model opens up new opportunities for students in the digital curation career path (Franks & Oliver, 2012).

**Outlook**

Archival and LIS higher education curricula and continuing education have responded to the new roles and responsibilities of information professionals, and the gap in competencies related to digital curation in the workforce that has been identified. The studies of graduates from new programs related to digital curation, although limited, have provided positive comments about experiences and preparation for employment, and valuable feedback for continued improvements (Palmer, Thompson, Baker, & Senseney, 2014; Gregory & Guss, 2011). However, there is a diversity of potential roles in the digital arena, and the LIS and archive curricula in the area of digital curation are still maturing. A 2012 CLIR report noted that LIS students “must be committed to constructing a data-intensive education for themselves…in order to come out of most existing United States LIS programs with the skills and knowledge necessary to support the needs of data-intensive research” (Keralis, 2012, p. 35).

Apprehension remains as to whether LIS and archive educational programs are responding adequately and quickly enough to prepare professionals for managing digital assets. In a study of course catalogs from 26 iSchools in 2009, only one third offered a digital preservation course. An analysis of the syllabi from those course showed that the focus was on theoretical concepts rather than technology skills (Costello, 2010). A 2012 study that looked at 52 LIS schools in North American found that only one-third offered data or digital curation courses. A closer look at the syllabi showed that topics varied significantly, and no two courses were alike (Harris-Pierce & Liu, 2012). A 2012 study by researchers at the University of Illinois
at Urbana-Champaign examined data curation at 55 iSchools or other LIS schools at the program level, for instance, those that offer advanced certificates, Master’s degrees, or specializations. Of those schools, 11 institutions offered programs that focused exclusively on data curation, data management, or data science; 12 offered programs that had components related to data topics; and 15 institutions offered programs with an emphasis on digital topics, not necessarily research data, relevant to data professionals. Of note, the terminology used to identify and define these different programs varied significantly (Varvel, Bammerlin, & Palmer, 2012). These studies illustrate that there is not yet consistency across higher education curricula for course offerings or course content in the area of digital curation.

There are concerns that functions and practices of libraries and archives are not evolving as quickly as the needs of researchers and users working in the digital environment. These concerns impact the potential roles of information professionals working with digital materials. For example, a 2012 CLIR report found that researchers have a different perception of the library’s role than does the library. The authors of the report observe that “few researchers see the library as a partner, and most of the researchers … regard the library as a dispensary of goods rather than a locus for badly needed, real-time professional support” (Jahnke & Asher, 2012, p. 16). Additionally, in an article about the function of libraries in e-Science, Heidorn (2011) contends that “if libraries do not actively engage in the task, then society may choose to create a new type of institution to curate digital data” (p. 670). Information professionals view themselves as a central component of the multifaceted digital landscape providing a significant and unique expertise and perspective. Effective education and training is a key factor for archives and libraries to remain relevant in the changing technological landscape.
Summary

In this chapter I reviewed relevant literature related to the growth of educational and training opportunities in the area of digital curation. Advances in technology have not only influenced how information professionals perform their work and the materials they work with, but have also created new roles and responsibilities that require new competencies. Gaps in the knowledge and skills base related to the area of digital curation have been observed. In response, several grant-funded projects have aimed to identify core competencies and develop graduate-level curricula. As a result of these efforts, there has been an increase in courses and programs related to digital curation, including post-graduate certificates and specializations in archival or LIS degrees, as well as professional training opportunities. However, concern remains about whether the fields of archival studies and LIS will be able to keep pace with the rapidly expanding area of digital curation.

The next three chapters comprise the findings and discussion. Five major themes emerged from the data analysis. Part one explores the themes of terminology and collaboration; part two explores the themes of multiple discipline engagement and education; and part three explores the theme of areas of professional and scholarly focus and provides a summary of all five themes.
CHAPTER 6: FINDINGS AND DISCUSSION: PART I

This research was designed to investigate whether the area of digital curation is emerging as an autonomous discipline. The purpose of this study was to explore the character, development, and education of digital curation by conducting an in-depth analysis of how this area is evolving. This dissertation employed a qualitative multimethod research design. Data were collected from scoping the literature, content analysis of published literature in the area of digital curation, and semi-structured interviews with individuals engaged in the area. This study focused on identifying themes that emerged from the practical and scholarly discourse, and understanding the level of consensus among educators, practitioners, and researchers about how they view this domain.

The findings and discussion are presented thematically in Chapters Six, Seven, and Eight. Five major themes emerged from the data analysis:

- **Terminology.** Defined as words or phrases used to identify the area of practice and research related to digital curation.

- **Collaboration.** Defined as any instance or mention of two or more people or institutions cooperating, sharing resources, or working together for mutual benefit or to achieve shared goals.

- **Multiple discipline engagement.** Defined as any instance or mention of two or more people or institutions from different disciplines or fields working together, or the combination and application of concepts, methods, practices, or theories from two or more distinct disciplines or fields.

- **Education.** Defined as any comment about educating or training professionals for activities and employment related to digital curation.
Areas of professional and scholarly focus. Defined as a descriptive summary of the main subject matters that characterize the area of digital curation represented in the sample of published literature analyzed in this study.

Table 6.1 shows how these themes were mapped to the elements of a discipline that were defined by the conceptual framework.

Table 6.1

*Conceptual framework: Elements of a discipline mapped to the themes emerging from the findings*

<table>
<thead>
<tr>
<th>Institutional Infrastructure</th>
<th>Theme Four: Education</th>
<th>Theme Two: Collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic Organizational Unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum in Higher Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Association</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publishers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recognition in Classification System</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Networks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intellectual Field of Inquiry</th>
<th>Theme Five: Areas of Professional and Scholarly Focus</th>
<th>Theme One: Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulated Knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discursive Community</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrative of the Field’s Development</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Styles of Subjectivity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This chapter is Part I of the findings and discussion and is organized in two main sections. In the first section, I convey in detail the findings and discussion for the theme of terminology. In the second section, I present the findings and discussion for the theme of collaboration.
Theme One: Terminology

Terminology was the first theme that emerged from the data analysis. The theme of terminology was defined as words or phrases used to identify the area of practice and research related to digital curation. The theme was derived from codes for definitions of relevant terms, language used for collaborative initiatives and educational programs, and terms used to refer to the field, subareas, and related areas.

The terminology theme mapped to the “discursive community” element from the conceptual framework. A discursive community is a group of people who advance and perform common practices of communication such as patterns of citation, norms for scholarly writing, approaches to knowledge claims, and shared terminology. Through socialization within a discipline and assimilation by faculty and graduate students of these common practices, “a discipline becomes a ‘discursive community’ in which participants converse with each other in a common language in the hope of recognition” by others from within the same community (Post, 2009, p. 766)

One criterion of the discursive community element is having shared “field-specific jargon or terminologies” (D’Agostino, 2012, p. 336). Learning “discipline-specific terminologies” assists to indoctrinate faculty and students into a discipline (Embree, 2010, p. 8) and, in turn, sustain the discursive community or discipline. However, shared terminology serves to do more than facilitate communication and enhance disciplinary socialization. Krishnan (2009) contends that “in academia, disciplinary languages are developed at least in part with the goal of protecting knowledge and disciplinary identity from outside infringement” (p. 23). Therefore, a shared terminology creates a sense of community and suggests a unified means of communication and a collective disciplinary identity, and asserts ownership of a subject area.
Terminology: Scoping the Literature

Scoping the literature provided an in-depth look at the use of eight key search terms (data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation, and digital stewardship) in published literature from 2001-2012. Table 6.2 shows the number of articles retrieved per search term and the percentage of the total for which those articles account. The table includes the year in which the search term first appeared in an article from the sample and how many articles occurred before 2001. Of note, the scoping the literature sample does not include articles from the *IJDC* because it is not abstracted or indexed in any of the databases used in the literature survey.

Table 6.2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Archiving</td>
<td>753</td>
<td>13.5%</td>
<td>1971</td>
<td>269</td>
</tr>
<tr>
<td>Data Curation</td>
<td>319</td>
<td>5.7%</td>
<td>1994</td>
<td>5</td>
</tr>
<tr>
<td>Data Preservation</td>
<td>344</td>
<td>6.2%</td>
<td>1975</td>
<td>81</td>
</tr>
<tr>
<td>Data Stewardship</td>
<td>173</td>
<td>3.1%</td>
<td>1993</td>
<td>10</td>
</tr>
<tr>
<td>Digital Archiving</td>
<td>705</td>
<td>12.6%</td>
<td>1976</td>
<td>137</td>
</tr>
<tr>
<td>Digital Curation</td>
<td>305</td>
<td>5.5%</td>
<td>2001</td>
<td>0</td>
</tr>
<tr>
<td>Digital Preservation</td>
<td>2,959</td>
<td>53.0%</td>
<td>1988</td>
<td>220</td>
</tr>
<tr>
<td>Digital Stewardship</td>
<td>21</td>
<td>0.4%</td>
<td>2000</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>5,579</strong></td>
<td></td>
<td></td>
<td><strong>722</strong></td>
</tr>
</tbody>
</table>

Search Fields: Abstract, Keywords, Subject, Title

Exclusions: Newspaper Articles, Patents

Databases/Vendors: DOAJ, EBSCOhost, Emerald, JSTOR, Medline, ProQuest, Sage, Scopus, Web of Science
A closer look at the results of the literature survey reveals trends in the use of each term by year. The findings from scoping the literature clearly convey *digital preservation* as the most frequently used term in the sample from 2001-2012. This trend continued with *digital preservation* making up over half of the sample each year from 2001 through 2010. In 2011 the combination of the other keywords began to outnumber the use of *digital preservation*. Over the twelve-year time period the findings show *data curation* and *digital curation* rise in use; *data archiving*, *data preservation*, and *digital archiving* remain somewhat consistent; *data stewardship* and *digital preservation* begin to dip in use; and *digital stewardship* having little impact.

Figure 6.1 shows the cumulative number of articles per search term retrieved by year. Overall, the figure shows an obvious increase of articles being published in this area and an increase in the diversity of terms used. Figure 6.2 shows the same data, but normalized as percentages to present the relative changes in use. Overall, the figure shows a relative decrease in the use of *digital preservation* and a relative increase in the use of *data* as a qualifier.
Figure 6.1. Scoping the literature sample: Number of articles per search term by year.

Figure 6.2. Scoping the literature sample: Number of articles per search term by year normalized as percentages.
The Digital Curation Centre (DCC) in the United Kingdom was pivotal in the adoption and diffusion of digital curation as a phrase to express the lifecycle approach to long-term access, management, and preservation of both digital cultural heritage and research data. The findings clearly show the term *digital curation* emerge in 2001. After two years with no articles in the sample, it appeared again in 2004, the year that the DCC was established, and has since showed a steady increase in use. Until 2006, all of the articles retrieved using *digital curation* as a search term were either briefs that discussed the launch of the DCC or were authored by people directly involved with the DCC. One exception was a life science article published in 2005 that used an alternative definition of digital curation as meaning the curation of web content such as social media and personal digital information. In 2006, the sample shows authorship slowly began to diversify beyond DCC-associated authors and use of the term expand. Additional support of the phrase digital curation came in 2005 when the Institute of Museum and Library Services (IMLS) in the United States announced a call for grant proposals to support the development of digital curation education programs. Soon after that, in 2007 the Digital Curation Unit (DCU) was launched as a research, development, and education center in Greece.

The findings from scoping the literature also show that the use of *data* as the qualifier appears to increase over the sample period in relation to *digital* as the qualifier. *Data curation*, in particular, has increased at roughly the same pace as *digital curation*. The five items identified that occurred before 2001 comprised three articles on data curation and interoperability of curated databases in the life sciences, and two about data curation and data management for museums. The increased use of *data* as a qualifier implies a growing interest in the management of digital research data, which is not surprising. The research community has long been responsive to the value of data archiving. As early as 2000, a draft of the Open Archival
Information System (OAIS) Reference Model was circulated. It was originally developed to address the swelling amounts of space science data. In 2003, the National Institutes of Health (NIH) published a statement on data sharing and began requiring a data sharing plan with research grant applications. Also in 2003, the National Science Foundation (NSF) Blue Ribbon Advisory Panel on Cyberinfrastructure released their report on the curation, management, and preservation of large data sets in engineering and sciences. Two years later, several multidisciplinary teams were assembled to draft a vision for cyberinfrastructure which was published in 2007 (National Science Foundation, 2007a). Also in 2007, the NSF announced Sustainable Digital Data Preservation and Access Network Partners (DataNet), a solicitation for grant proposals “to develop the new methods, management structures and technologies to manage the diversity, size, and complexity of current and future data sets and data streams” that “will integrate library and archival sciences, cyberinfrastructure, computer and information sciences, and domain science expertise” (National Science Foundation, 2007b). By 2011, the NSF began requiring data management plans with all research grant applications.

The largest increases in articles from year to year over the sample period occurred in 2004, 2006, and 2008. Several events can be detected from scoping the literature, which likely influenced the use of specific terminology or increased attention to certain topics. For example, in 2003, the NIH, a large funder for biomedical research, began requiring data sharing plans with grant applications. The following year, the largest increases in articles were seen in the search terms data archiving, digital archiving, and digital preservation, all practices and concepts germane to data sharing. There was also an increase in the search term digital curation due to the launch of the DCC in 2004. In 2005, the IMLS announced a call for grant proposals for digital curation education programming. The following year, the most noticeable increases were
seen in the search terms *digital curation* and *digital preservation*, coinciding with a number of grant-funded educational programs in the area of digital curation. In 2007, the NSF announced the DataNet program. The increases in 2008 were observed in all of the search terms except *digital curation* and *digital stewardship*, which are more commonly associated with the information fields than the domain sciences (life sciences, physical sciences, and social sciences). The increase included all terms that use *data* as a qualifier. Also, the main activity for the search term *digital stewardship* is in 2010 when the Library of Congress launched the National Digital Stewardship Alliance.

**Terminology: Interviews**

When asked about what and how terms are used in the area of digital curation, many interviewees initially remarked about the general lack of consensus within the community. For example:

*Interviewee 01:* I actually think that’s an area of some confusion, generally, professionally.

*Interviewee 08:* Okay, that is quite complicated, really.

*Interviewee 11:* They are all problematic.

*Interviewee 13:* It is very much in flux.

Interviewees addressed two key properties of the theme of terminology. First, all of the interviewees acknowledged that the different terms used to describe this area or related areas had at least subtle differences in meaning. Many interviewees described digital curation as a blanket or umbrella term that encompasses related terms and subareas. For example:

*Interviewee 05:* Often curation has been used as a way of expressing something that is slightly richer than preservation.
**Interviewee 07:** [Data management is] part of [digital curation], very definitely part of it. I am of the opinion that everything is data as well…If it's on the computer, [it comes under] digital curation basically.

**Interviewee 08:** Digital curation is a wider term in my opinion, and implies managed activities, a large array of different managed activities in the human and technical interfaces. Whereas digital preservation is a much narrower term. Digital preservation is much more specific - technical activities or metadata activities that one might undertake on the data to ensure the long-term preservation.

**Interviewee 11:** I use curation, both digital curation and data curation, and by curation, I mean the enhancement of materials so that they can be understood and used by people both now and in the future…I think digital curation encompasses data curation...Data curation is a very specific term.

**Interviewee 12:** I think digital curation has been used as an umbrella term to encompass a lot of these different concepts together.

Interviewees also distinguished terms by the type and scope of the content being preserved. For instance, one interviewee distinguished specific terminology when working with research data:

**Interviewee 08:** I think research data management or data curation is a kind of a subset of digital curation. I mean research data management is looking out for a particular type of data - research data.

Another interviewee limited the use of digital archiving to the context of electronic records and digital archival material:

**Interviewee 12:** Digital appears to have a broader scope than the electronic record does, and so [the term I use] also depends on the scope of the materials that we are discussing…. [Digital archiving and digital curation] are different in terms of the nature of the creation of the material and the nature of the secondary use of the material.

Some interviewees acknowledged that they often used terms as relatively equivalent expressions, although fully aware that the meaning of the terms are distinct:
Interviewee 04: Usually [I use] digital preservation and digital curation and, although I am aware that curation is a bit more general than digital preservation, I use those two quite interchangeably, as tradition terms.

Interviewee 14: When I am talking about the area, I use the word preservation sometimes, and sometimes I use the word curation. I am probably a little loose about which term I use, but I should probably be a little more precise because I do believe there are, in other words, subtle differences between the two - that preservation is one aspect of curation and curation is the much broader term.

Second, eight of the interviewees commented on using terms that were the most meaningful to their audience. Comments included observations about the use of community-specific jargon and the need to more thoroughly explain concepts in laymen terms for audiences outside of the community. For example:

Interviewee 02: Recently, we have been writing some documents where we have not been talking in digital preservation terms, because it seems to me that other people [outside of the community] just hear us talking in a foreign language sometimes. So, we have been using different language to make our concepts easier to understand.

Most comments focused on using terms that presented a meaningful and targeted message that fit the needs of the audience. In these instances, it was less about the denotations of the terms and more about the connotations as perceived or understood by the specific community. For example:

Interviewee 04: The other area where we got interested in recently…preserving science data, or preserving entire scientific processes and business processes and in that community we use a bit more community-specific terms…So it really has to do a lot with which community we are talking to and what the focus of our conversation is.

Interviewee 05: I tend to use whatever term speaks best to the audience I am dealing with at the present time…But again it is a matter of what the word means to particular people. I don’t get hung up on it, really.
Interviewee 11: So it depends on the audience, right. If I am [talking to information professionals I] describe it [one] way. If I am talking to you as a researcher or someone who is preparing the data, then I describe it a little bit differently…We frame things or try to tailor the message to the audience, and I think that is important because if you just have one generic message, it doesn’t hit home.

Terminology Explored

Although many of the terms related to the area of digital curation are used somewhat interchangeably, at least informally or conversationally, each characterizes a distinct concept or group of concepts. “Digital curation” was originally coined to capture the concept of the lifecycle management and preservation of all types of digital assets including processes and workflows, dynamic and static content, and research data and cultural heritage materials. Indeed, most interviewees explained digital curation as broad in scope and encompassing the actions implied by the other terms. Although digital curation remains comprehensive in its meaning and inclusive of research data, it has not yet been extensively adopted within the research communities. The findings from the literature survey suggested that its use is most commonly found within the information fields, or in connection with the DCC. Digital stewardship is conceptually broader than digital curation and takes into account long-term preservation at the ethical, policy, and social levels; however it has minimal impact at present.

Data curation, on the other hand, was described by interviewees as representing a narrower conceptual terrain than digital curation, and was often viewed as a component of digital curation. For the most part, as used among the domain sciences and from the digital curation perspective, data curation is a fundamental set of ongoing activities for the management and long-term preservation of research data to ensure its quality, understandability, and reusability. Specifically, it involves adding value to and ensuring the integrity of data through annotation, metadata, and associated information integral to the sharing and reuse of data into the future.
In this sense, data curation is also viewed as a vital aspect of data stewardship, which is a comprehensive concept used largely in the domain sciences to denote the overall long-term care, management, and responsibility of research data. However there are some, mostly within the information fields, that define data curation more holistically, suggesting only a subtle distinction from digital curation. That distinction narrows the content focus of data curation to the lifecycle management of research data with an emphasis on the reuse of data for scholarly pursuits (example from content analysis sample: Palmer, Heidorn, Wright, & Cragin, 2007). The former definition applies to specific essential activities within the lifecycle of data, while the latter definition comprises the entire lifecycle.

Archive and its derivatives are contentious terms because they have been co-opted from archival studies and now carry a variety of superficial meanings. For example, in the technology field the word archive is used synonymously with backing-up data. Conventionally speaking, digital archiving and data archiving represent two separate, yet related, and established practices. Cunningham (2008) describes digital archiving as digital curation for electronic archival records. The only distinction between archives and digital archives are the format of the records. Both are complex practices driven by the same archival policies and based on archival theory. Data archiving, on the other hand, is part of a data management strategy, and is associated with the publication, reuse, and sharing of research data. Data archiving involves the long-term preservation and storage of research data for future use, and has become a core principle for scholarly communication in the digital era. Data archives provide professional services to support ingest, preservation, storage, and access of data. Alternatively, the term data archiving is also used to refer to extended storage within a data collection software or instrumentation, and
digital archiving is also used as equivalent to collection and storage of digital assets without regard to long-term preservation.

Although digital preservation, in particular, was most often used interchangeably by interviewees with digital curation, its narrow denotation and the activities to which it pertains were well articulated. Both digital preservation and data preservation are technical terms to describe a set of managed activities to impede obsolescence and keep digital assets accessible, renderable, and usable into the future. Digital preservation applies to all forms of digital assets while data preservation is most commonly used to refer to research data. Digital preservation, like data curation, is considered a set of essential activities within the lifecycle approach of digital curation.

Digital preservation is at the crux of the discussion of terminology for this area. Preservation in general and the issues surrounding technology in particular are of keen interest to the community. Although the limitations of the term’s definition were widely acknowledged, digital preservation has been broadly interpreted and widely employed as evidenced by it being the most commonly used search term in the literature survey sample. This does not necessarily mean that that the community is most interested in the set of activities and technology solutions that are characterized by the phrase digital preservation. But rather it highlights the significance of the emergence of digital curation as a broader conceptual phrase.

Prior to 2001, there was no single term that captured the holistic approach required for the long-term care and sustainability of digital assets. Although the principles of preservation remain largely unchanged between analog and digital materials, the nature of the materials is different and therefore the approach to preservation is different. Digital curation not only provided a term to guide this discussion but also provided a focal point for multi-disciplinary
collaboration and discussion about the sameness that underlies the preservation of all digital assets. In turn, with more accurate terminology available and understandable to diverse audiences, the community was able to more articulately engage in discussion and the development of this evolving practical and research area. Hence, the overall use of *digital preservation* as a catchall phrase begins to decrease as it is eclipsed by more descriptive and precise terminology.

The decision to use one term over another is not strictly determined by definition alone but rather is based on a combination of factors. The findings from the interviews and content analysis suggest that the primary factor for selecting terminology is the audience. Speakers select terms that will be most meaningful and most impactful to their audiences. This underscores the disciplinary differences the findings show in terminology use. Audiences have different backgrounds and worldviews, work with different content type and formats, and have different goals or objectives. With this in mind, audiences project different connotations on terms and those connotations can influence how an audience will respond to concepts, ideas, and practices. Expressing concepts related to digital curation in language that the audience will identify with and that relate directly to the needs of the audience, even if imprecise, appears to be more valuable to the interviewees than promoting a standardized lexicon.

Considering the extent and variety of disciplines that have an interest in concepts and practices related to digital curation, it is not altogether surprising that a common terminology has yet to be established. Each term carries distinctive connotations and denotations, which evoke interpretations that reflect the audiences’ and speakers’ backgrounds. In terms of emerging disciplines, and for that matter even a cohesive field, an agreed upon set of conceptual terms is foundational. What we have seen is a shift from passive archiving of data or preservation of
digital cultural heritage to active lifecycle management of digital assets, from creation to preservation to access and use. The terminology that has emerged reflects these changes, and suggests that at a conceptual level there is consistency across domains even if that consistency is not yet evidenced by a shared vocabulary.

**Theme Two: Collaboration**

Collaboration was the second theme that emerged from data analysis of the interviews and content analysis. The theme of collaboration was defined as any instance or mention of two or more people or institutions cooperating, sharing resources, or working together for mutual benefit or to achieve shared goals. The theme was derived from codes for authorship, collaborative networks and projects, participation in alliances and consortia, and pooling resources.

The collaboration theme aligns with the “social networks” element from the conceptual framework. Networks are supported through formal means such as professional associations, academic departments, or collaborative research initiatives and projects. These promote networks for financial and social support, and the control of access to opportunities. Social networks also include the “audiences for the dissemination of work” (D’Agostino, 2012, p. 335) and informal channels for communication and social interaction.

Social networks preserve formal and informal channels of communication and social interaction within and among disciplines. These networks link people and ideas. Becher and Trowler (2001) explain: “Social interaction, communication of all sorts and the partly socially constructed nature of disciplines that is associated with them are the forces that bind together the sociological and the epistemological, giving shape and substance to the links between knowledge forms and knowledge communities” (p. 104). Social networks build a scaffold for mutual
understanding of concepts, practices, and values through the promotion of community, collaboration, and resource sharing.

**Collaboration: Content Analysis**

Patterns of authorship in the published literature provide information about collaboration among members of the community. For the content analysis, 460 published articles and conference presentations were coded for number of authors and geographic location of authors. Seventy-two percent of articles in the sample were authored by two or more people. Of the 128 articles that had single authors, 15 were editorial introductions to *IJDC* issues. Figure 6.3 shows the breakdown of articles by number of authors.

![Figure 6.3. Content analysis sample: Articles by number of authors](image-url)
Collaborative articles, those with two or more authors, exhibited three general levels of interaction. Collaborations existed between people from within the same institution, from different institutions within the same country, or international. Figure 6.4 presents how many collaborative articles from the sample represented each alliance.

![Figure 6.4](image-url)

**Authorship**

*Figure 6.4. Content analysis sample: Multi-authored articles by type of collaboration.*

Identifying the disciplinary home of authors was not always feasible because of the lack of biographical data included with the articles. Although place of employment and geographic location were almost always included, the author’s role and more detailed background information were sporadic and inconsistent. The latter information was needed to identify the discipline of the author: for example, there are a range of positions that an individual can hold within a library or data archive setting. The most that could be accomplished was to identify the domain in which each author held a job post. With this in mind, of the 332 collaborative articles
only 63 could be tentatively identified as being authored by multi-discipline teams based on the disciplinary setting of the author’s job post. This was determined by the department or institution of employment for each author, additional metadata provided with the article such as position title when available, and targeted internet searches if necessary. Table 6.3 presents how many of these apparent multi-discipline teams had representation from authors who held positions in an information field, technology field, both, or no representation.

Table 6.3

*Content analysis sample: Authorship of the collaborative articles that were authored by multi-discipline teams*

<table>
<thead>
<tr>
<th>Authorship</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one author held position in Library, Archive, or Museum (LAM)</td>
<td>17</td>
</tr>
<tr>
<td>At least one author held position in Technology</td>
<td>17</td>
</tr>
<tr>
<td>At least one author held position in LAM and at least one author held position in Technology</td>
<td>16</td>
</tr>
<tr>
<td>LAM and Technology not represented in authorship</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>63</td>
</tr>
</tbody>
</table>

Patterns and practices of collaboration vary among the disciplines, between subfields, and even from journal to journal. Traditionally, the life sciences and physical sciences tend to collaborate more than the social sciences and the humanities, with the technology fields falling toward the middle of the spectrum. The occurrence of collaboration is rising in the social sciences and humanities including library and information science, and international collaboration is rising across all disciplines (Cronin, Shaw, & La Barre, 2003; Sin, 2011; Franceschet, 2011). Reasons for or against collaboration are embedded into the culture of each discipline and underscore beliefs about competition, scholarship, and knowledge and theory.
production. Collaboration is also influenced by funding opportunities and policies, growing international social networks, and grand challenges that cross disciplinary boundaries.

A cursory survey of the literature shows how this level of collaboration compares to other areas. In a 2011 study of six top ranked, high prestige LIS journals, the researchers found that 64% of published articles had two or more authors. Of the multi-authored papers, 16% were international collaborations (Sin, 2011). An unpublished study that focused on an international selection of archival journals found that only 25% were authored by two or more authors (Conway, Condon, Dorey, Garcia, Lenstra, & Song, 2012). Research looking at authorship patterns in bioinformatics found that 90% of articles were written by two or more people. This is not surprising considering the high level of collaboration found in the life sciences and the interdisciplinarity of the bioinformatics field (Amsaveni, Manikandan, & Manjula, 2013).

In a 2011 study of collaborative authorship in computer science, researchers found that 77% of the journal articles and conference papers were multi-authored. The researchers observed that conference papers were more collaborative. Analyzed separately, 81% of the conference papers and 70% of journal articles had two or more authors (Franceschet, 2011). A co-authorship network analysis of conference proceedings in digital library research community found that 80% of conference papers had two or more authors, with the majority having two and three authors, 29% and 24% respectively. International collaboration accounted for 7% of the multi-authored conference papers (Liu, Bollen, Nelson, & Van de Sompel, 2005). For comparison, in the present study, analyzing the conference presentations from the content analysis in isolation there were 86% that had two or more authors, and 17% of those were international collaborations.
The findings suggest that the patterns of collaboration in the area of digital curation are similar to those in the LIS field and computer science field. They are not as high as those found in the life or physical sciences, or as low as in the archives field or other humanities-oriented disciplines. Analyzing the subsamples individually did not show any stark differences in collaboration practices: the findings suggest a higher percentage of collaborative partnerships in conference presentations, which is a trend found in other fields. Only considering the data for the *International Journal of Digital Curation*, the findings show that 72% of articles had two or more authors (67% if you include the 15 single-authored editorial journal issue introductions) and 15% of the multi-authored articles were international collaborations. For the top 5% most frequently cited subsample, 75% had two or more authors, and of those multi-authored articles, 22% were international collaborations. The content analysis sample included articles from a wide range of disciplines, and therefore comprises a mixture of collaboration practices.

**Collaboration: Interviews**

The frequent occurrence and value of collaboration are assertions often cited within the digital curation community. The level of collaboration suggested from the findings of the content analysis supports the assertions that collaboration occurs frequently in this area. Interview data provided insight on how people within the community articulate the value of collaboration. There was little indication that collaboration was founded on competition as witnessed in other areas (Becher & Trowler, 2001). Rather it exists to produce more effective practical and research outcomes, and cultivate knowledge and social networks.

Interviewees approached the theme of collaboration from two main perspectives. First, collaboration was talked about as the definitive approach to successful and sustainable digital
curation, particularly with respect to the magnitude of the undertaking. In a follow-up email after one interview, the interviewee wrote:

_Interviewee 06:_ I’m not sure we talked much about ‘collaboration’ which is certainly a principle that the [digital preservation] community has always talked about. There is an often-stated principle that preservation is too big a task for organizations to do on their own. That collaboration – at all levels – including internationally – is a sine qua non for effective preservation of the cultural and scientific record.

Another interviewee focused comments on pooling resources for effective technological infrastructure:

_Interviewee 13:_ So if you think of libraries and archives working individually, our capacity to develop new technology is quite limited, and in a way, we kind of have to work together collaboratively to pool our resources in order to get better systems in place, to get a better infrastructure.

Although both examples illustrate recognition of the value of collaboration, interviewees also addressed concerns and limitations of collaboration in practice. _Interviewee 06_ observed: “The politics and the practicalities of collaboration can be difficult.” _Interviewee 13_ believed that “at a content or a collection level, there is still going to be a difference. Things will still break down along collection lines or disciplinary lines.”

In turn, communication and interpersonal skills were highlighted as essential in the area of digital curation, in large part to facilitate collaboration. For example:

_Interviewee 04:_ But I guess something that people would need as a kind of common denominator is first an understanding and appreciation of long-term stability and all those concepts that are the motivation for digital preservation… On the other hand, since the data and the objects that we are talking about is digital it requires basic understandings of IT skills… and to be able to talk to the respective domain specialists across discipline boundaries.
Interviewee 05: You need to be able to communicate at least with people who run IT services, and that ability to communicate with people in many different backgrounds and be able to translate requirements and needs between them, I think is, was really significant, for lots of people in those roles.

Interviewee 08: I think knowing how to talk to IT people as well actually, because, and I have discussed this already how IT people and trained information professionals don’t necessarily understand the same thing. I think if you curate data and you do understand database structures and have some understanding of how the internet works and so on, then you are more likely to be able to talk to the IT people and more able to be taken seriously.

The second perspective interviewees presented on collaboration addressed building communication and social networks with others working in the area. Most comments about collaboration and network building in the area of digital curation highlighted it as an international activity. Some interviewees focused on cultivating informal relationships to sustain channels for collective learning and information exchange. For example:

Interviewee 01: And talking to people of course, which is crucial. Because I have a lot of colleagues in the US and Europe, I have been involved in lots of things over the years, so I have a lot of contacts and friends in the UK and US, I stay in touch with them.

Interviewee 02: From that perspective we have international colleagues, which we stay in touch with through workgroups and listservs. You have a network of people that you know. You try and keep an eye on blogs and listservs and things like that.

Several interviewees discussed the collaborative projects as channels for staying engaged in the broader digital curation community:

Interviewee 04: There are research projects with regular meetings, there are conferences. So, I think meeting and talking to people all the time, that is pretty much what I am doing, hardly anything else…We have a whole bunch of [collaborative] projects running currently…It is a big global community I would say.

Interviewee 07: And then of course, there is pretty much all the projects that I’ve worked on that are externally funded have been collaborative projects with people outside the organization and also outside the country as well.
Practice and research related to digital curation requires a range of skills and therefore involves individuals with different expertise collaborating for more productive outcomes. It was common for interviewees to comment that no single entity or set of skills could effectively address the practical and research concerns of the area of digital curation. It requires multiple perspectives and shared resources to be sustainable. This point of view is also reflected in the themes of multiple discipline engagement and education. The findings from the interviews suggest that the community views collaboration as the most effective way for integrating diverse knowledge and skills bases. At the institutional level, collaboration is seen as a means of combining resources and a prerequisite for stronger infrastructure, interoperability, and shared solutions.

Because of the multiple and distinct communities that converge to form the digital curation community, collaboration or participating in collaborative research initiatives also serves as a means of building social networks, and staying informed with current practice and research. Collaboration was cited alongside social media and attending conferences as a notable method for engagement with the diverse digital curation community. In the absence of professional associations and academic departments, collaborative research initiatives serve as channels for cultivating social networks in digital curation and related areas.

The findings suggest that not only is collaboration a frequent occurrence in digital curation practice and research, but it is highly valued by the community. Specifically, collaborative research initiatives and projects were mentioned as core building blocks for the development of the area, promoting sustainable practical outcomes, and cultivating knowledge and social networks. Although the structure of the social network was not analyzed in this research, patterns of interdisciplinary co-authorship and cited references suggest that information
exchange across disciplinary boundaries is restricted. Further research is warranted to better understand the role that collaboration has in cultivating knowledge and social networks across disciplines.

**Summary**

In this chapter, I explored the first two themes that emerged from data. The theme of terminology is aligned with the element “discursive community” from the conceptual framework. The findings suggest that although there is generally a shared understanding of most field-specific jargon in the area of digital curation, a common language has not fully coalesced. In terms of emerging disciplines, and for that matter even a cohesive field, an agreed upon set of conceptual terms is foundational. The theme of collaboration aligns with the “social networks” element of the conceptual framework. The findings suggest that there are well-established social networks in place within the digital curation and related communities.

In the next chapter, I present the second part of the findings and discussion. In Part II, the themes of multiple discipline engagement and education are explored in detail. Part III presents the final theme, areas of professional and scholarly focus, and provides a summary of the findings and discussion.
CHAPTER 7: FINDINGS AND DISCUSSION: PART II

This chapter is Part II of the findings and discussion. The chapter is organized in two main sections. In the first section, I explore the theme of multiple discipline engagement. In the second section, I convey in detail the findings and discussion of the theme of education.

Theme Three: Multiple Discipline Engagement

Multiple discipline engagement was the third theme that emerged from the data analysis. Multiple discipline engagement was defined as any instance or mention of two or more people or institutions from different disciplines or fields working together, or the combination or application of concepts, methods, practices, or theories from two or more distinct disciplines or fields. The theme was derived from codes for discipline-specific discourse, interdisciplinarity and multidisciplinarity, relationships between fields, shared solutions, and required education, skills, and training.

The multiple discipline engagement theme was mapped to the “styles of subjectivity” element of a discipline from the conceptual framework (During, 2006, p. 266; D’Agostino, 2012, p. 335). Subjectivity is the “expression of self and the representation of a speaker’s [disciplinary] perspective or point of view in discourse” (Stein & Wright, 1995, p. 1). Members of a discipline are indoctrinated into that discipline’s culture, which includes norms for behaviors and established research practices that identify individuals as members of the discipline. These accepted behaviors and research practices become characteristic of that discipline, and are maintained through discourse and sense of identity of members.

Styles of subjectivity, or unique disciplinary perspectives, are embodied in the written discourse of a discipline. Each discipline has accepted criteria for assessing the quality and value of work in terms of student academic success, professional performance, and faculty research.
and teaching. A discipline engenders a style of written communication “that identifies its user as a member of the disciplinary community and so makes their contributions visible to others in the community” (D’Agostino, 2012, p. 336). Disciplines also possess a concept of rigor as demonstrated by a shared understanding about establishing the reliability, validity, and trustworthiness of research.

Multiple Discipline Engagement: Content Analysis Subset

An in-depth probe into a subsample of the content analysis provides evidence of the multiple stakeholders engaged in this area. Looking at only the data from the subsample of the content analysis for the top 5% most frequently cited articles in Scopus and Web of Science (n=170; 2001-2012; search terms: data archiving, data curation, data preservation, data stewardship, digital archiving, digital curation, digital preservation, and digital stewardship) provides evidence of the multiple stakeholders engaged in this area: 31% of articles from the sample were published in journals from the information fields, 28% from the life sciences, 19% from technology fields, 12% from the physical sciences, and 7% from the social sciences. Figure 7.1 shows the breakdown of the articles by field of the journal in which they were published. This chart is not intended to be representative of all articles, but rather conveys a story about the competing and shared interests.
Content analysis subsample (top 5% most cited): Percentage of articles by the field of the journal in which they were published.

Table 7.1 focuses on the 53 articles (31% of the subsample) published in journals from the information fields. Within the information fields, the majority of articles were published in LIS journals. Table 7.1 displays the publication titles and the number of article from each publication in archival studies, LIS, or museum studies journals.

As previously addressed, establishing the discipline of authors was not always feasible based on the information provided with the articles; however, identifying the domain in which they held a job post was. For the articles presented in Table 7.1, first authors for all five articles published in archival studies journals held positions in the archives field, and first authors from both articles published in museum studies journals were from the museum field. For articles published in LIS journals, six first authors held positions in the archives, 15 first authors held positions in technology, and 25 first authors held positions in the library and information science.
Table 7.1

Content analysis subsample (top 5% most cited): Publication titles for articles in the information field (n=53)

<table>
<thead>
<tr>
<th>Field of Journal</th>
<th>Journal Title</th>
<th>Number of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival Studies (n = 5)</td>
<td>American Archivist</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Archival Science</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>D-Lib</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Library Trends</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>International Journal on Digital Libraries</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Proceedings of the Joint Conference on Digital Libraries</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Journal of Digital Information</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Library Hi Tech</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Program: Electronic Library and Information Systems</td>
<td>2</td>
</tr>
<tr>
<td>Library and Information Science (n = 46)</td>
<td>Issues in Science and Technology Librarianship</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Journal of Documentation</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Journal of Information Science</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Journal of Library Administration</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Journal of Library Metadata</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>JASIST</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Library Quarterly</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Libres</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>OCLC Systems and Services</td>
<td>1</td>
</tr>
<tr>
<td>Museum Studies (n = 2)</td>
<td>Museum Management and Curatorship</td>
<td>2</td>
</tr>
</tbody>
</table>

The data for the subsample of the top 5% most frequently cited articles also shows how often publications from archival studies, digital curation, LIS, and technology fields are cited by articles in journals from other fields. Table 7.2 shows the number of articles that cited references from publications of different fields of study. For example, of the five articles published in
archival studies publications, four included at least one reference to a digital curation publication, four included at least one reference to articles published in LIS journals, one included at least one reference to an article in a technology publication, and none of them referenced articles from museum studies. Of particular note, 60 of the 84 articles published in any of the domain sciences journals did not reference any articles from archival, digital curation, LIS, or technology publications.

Table 7.2

*Content analysis subsample (top 5% most cited): The number of articles citing references from publications of different fields of study*

<table>
<thead>
<tr>
<th>Field of Article</th>
<th>Archival Studies</th>
<th>Digital Curation</th>
<th>LIS</th>
<th>Museum</th>
<th>Technology</th>
<th>Within Area Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archival Studies (n=5)</td>
<td>-</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Domain Sciences (n=84)</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>1</td>
<td>16</td>
<td>60</td>
</tr>
<tr>
<td>Library and Information Science (n=46)</td>
<td>19</td>
<td>13</td>
<td>-</td>
<td>0</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>Museum (n=2)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Technology (n=33)</td>
<td>6</td>
<td>4</td>
<td>14</td>
<td>0</td>
<td>-</td>
<td>13</td>
</tr>
</tbody>
</table>

Digital curation literature most commonly cited included the OAIS Reference Model standard, followed by articles from the *International Journal of Digital Curation* and DCC publications. The most common LIS publications cited in articles outside of the LIS field included articles published in *D-Lib, Library Trends*, and the proceedings of the Joint
Conference on Digital Libraries. The most common technology publications cited in articles outside of technology fields included *Communications of ACM* and *Lectures in Computer Science*. There was little consistency regarding the citation of archival publications; however, recurrent citations included items from the IS&T Archiving conference proceedings and the 1996 *Preserving Digital Information, Report of the Task Force on Archiving of Digital Information*.

**Multiple Discipline Engagement: Interviews**

Similar to the theme of collaboration, interviewees view the involvement of multiple disciplines as a key to effective and sustainable digital curation. One interviewee noted:

*Interviewee 06:* I think that it's kind of a cross disciplinary collision, hopefully not a collision, it's much more than that. It's sort of a nice productive coming together of skills where preservation works best.

Interviewees discussed engagement by multiple disciplines in the area of digital curation in relation to two fundamental areas. First, interviewees commented on the diverse knowledge and skills base needed to do activities related to digital curation. For example:

*Interviewee 02:* So you need to have or understand computer skills, you also need to understand collections. And you need to understand both, because guys that just deal with the computer stuff don’t necessarily understand the collections side...I don’t necessarily see digital preservation as being particularly special in terms of the skills, but it is how you are applying them and how you are drawing all these other skills sets into one area, I think which is actually important.

*Interviewee 05:* So I think there is no question…it is of relevance to many of [the information] disciplines, not necessarily to one, and many of those disciplines have a view on curation….But I think you can't do preservation properly unless you pick up something from the areas of the organization and the management of knowledge. Where I think there is a risk, is that there are lots of other things that come to bear on this. So a great deal is about quite fundamental computer science, for instance…I think there are [many] different set of skills.
Interviewee 14: So, I think a basic technical knowledge is important, a knowledge of the relevant models and theories in preservation, a really solid understanding of risk, and an ability to think extremely logically for example when you want to do preservation planning or something like that, and an appreciation of …what kinds of preservation approaches you might use in different context. Those are extremely important. But then there are going to have to be people who are actually able on the ground to do the really hard core technical work.

Interviewees also discussed how the diversity of research challenges in the area requires approaches and perspectives from a range of disciplines. For example:

Interviewee 04: I think that digital curation research, it is basically interdisciplinary, and so we need people from a diversity of backgrounds. I would say there are definitely a number of questions and research challenges in digital preservation/curation that are core computer science challenges and that surely need to be solved by computer sciences. There are other questions and changes that are more related to archival and library science, information science. And there are lots of challenges that are completely domain specific, you know, specific data formats, usage of data.

Second, although interviewees discussed the role of the domain sciences in digital curation education, practice, and research, the two main fields predominantly discussed as having the most influence in the area were the information fields and the technology fields.

Interviewee 02: So you’re either coming from the collection side, or you are coming from the computer side. And logically the discipline fits somewhere in between… We are a hybrid. [We are] in between. So I think it is going to be watered down within any one of the camps. I think you really do have to sit in the middle, but acknowledge that you are in the middle and that you are taking skills from two different sides.

Interviewee 05: So it touches on lots and lots of different areas. But it is the nature of the universities that things have to have a home, they have to sit in a faculty or school or college or something like that… I think that is true, the discipline would happily sit within the information disciplines [or within computer science]…except that if you [see] it exclusively as one or the other you’re missing half of what it was all about.

Interviewee 08: I think we have to start working more closely with computer scientists as well. There are kind of two strands in the professional world [related to digital curation]: The ones who have done libraries and archives and the ones who have done computer science.
The data also show how the themes of terminology, collaboration, and multiple discipline engagement are interconnected. As one interviewee noted:

Interviewee 14: The one thing that is extremely important is that anyone working in digital curation has an appreciation of the languages of communication that are involved in preservation and curation, and the languages that are used in disciplines like information management, computer science, so that they can actually communicate with other practitioners, because preservation is very much a collaborative activity and you must be able to communicate across communities.

Again, taking a closer look at the content analysis subsample of the top 5% most frequently cited articles, there are disciplinary differences in the use of terminology. Table 7.3 shows the number of articles retrieved by search term used and the field of the journal in which the article was published. The table indicates that digital preservation shows up more frequently as a search term in the information and technology fields, and data archiving found more often in the domain sciences and the technology fields. Data curation mostly retrieved articles published in journals from the life sciences and the information fields. These disciplinary differences highlight a preference in terminology, and also highlight a distinction in conceptual and practical interests among the fields.
Table 7.3

Content analysis subsample (top 5% most cited): Articles (n=170) from Scopus and Web of Science (2001-2012) by search term and field of journal

<table>
<thead>
<tr>
<th></th>
<th>General Science</th>
<th>Information Fields</th>
<th>Life Sciences</th>
<th>Physical Sciences</th>
<th>Social Sciences</th>
<th>Technology Fields</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Archiving</td>
<td>2</td>
<td>1</td>
<td>17</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>48</td>
</tr>
<tr>
<td>Data Curation</td>
<td>0</td>
<td>10</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>Data Preservation</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Data Stewardship</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Digital Archiving</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>13</td>
</tr>
<tr>
<td>Digital Curation</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Digital Preservation</td>
<td>0</td>
<td>31</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>12</td>
<td>50</td>
</tr>
<tr>
<td>Digital Stewardship</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Multiple keywords</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>53</td>
<td>47</td>
<td>20</td>
<td>12</td>
<td>33</td>
<td>170</td>
</tr>
</tbody>
</table>

Multiple Discipline Engagement Explored

Multiple disciplines working in the same problem space is not uncommon. Some problems intersect the boundaries that separate domains or are too complex for a single discipline to address. In many cases, collaboration yields more productive and valuable outcomes. These shared problem spaces may continue to draw resources or occupy space in multiple disciplines, and necessitate long-term collaboration and facilitate the transfer of knowledge between fields. In some instances, these collaborations can result in the formation of a subdiscipline or new discipline that combines elements of the collaborating fields, or the problem space may be subsumed into one of the collaborating disciplines as a new specialization.
There are different forms and levels of multiple discipline engagement, and an extensive body of research that addresses the dynamics, function, and impact of multiple discipline engagement exists. *Multidisciplinary, interdisciplinary, and transdisciplinary* are just three of the more common terms applied when professionals and researchers work across the boundaries of their home discipline. To complicate the matter, the definitions of these terms are often debated and some are used interchangeably (Petrisor, 2013; Krishnan, 2009; Barry, Born, & Weszkalnys, 2008; Choi & Pak, 2006; Strober, 2006; Stember, 1991). Each term implies some extent of integration or supplementation of the concepts, methods, and theories from more than one field for achieving outcomes or addressing problems.

*Multidisciplinary* is used as a general description for multiple discipline engagement, and as a more precise term that differentiates it from *interdisciplinary*. More precisely, multidisciplinarity refers to two or more disciplines working in the same problem space but retaining unique disciplinary perspective and respecting the boundaries of each discipline. “Multidisciplinary research draws on knowledge from different disciplines, but stays within the boundary of one primary field” (Natural Sciences and Engineering Research Council of Canada, 2012). The term does not imply integration of the disciplinary knowledge, but rather suggests that the contributions from other domains are supplemental or additive (Stember, 1991; Choi & Pak, 2006; Strober 2006).

*Interdisciplinary* is also used as a general description for multiple discipline engagement but more specifically refers to the integration of two or more disciplines. Interdisciplinarity implies a synergy of disciplinary knowledge from multiple domains, blurring yet preserving the boundaries between disciplines (Stember, 1991; Choi & Pak, 2006; Strober 2006). This convergence and integration of disciplinary approaches and perspectives brings to the problem
space “defamiliarization, fresh insights, [and] skills from one area of expertise enriching another and making up for another's limitations” (Torgovnick, 1996, p. 282). Through the synthesis of content, methods, perspectives, or theories from multiple disciplines, interdisciplinary practice and research creates a “coordinated and coherent whole” (Choi & Park, 2006, p. 359).

Transdisciplinary, on the other hand, is not used as a general description of multiple discipline engagement but does have two distinct meanings in the literature. The first meaning of transdisciplinary suggests a higher level of integration than interdisciplinarity by way of a holistic view of disciplinary knowledge bases and a comprehensive approach to the problem space (Choi & Pak, 2006; Strober, 2006, Stember, 1991). Individuals from multiple disciplines “work closely together on a common problem over an extended period but also create a shared conceptual model of the problem that integrates and transcends each of their separate disciplinary perspectives” (Rosenfield, 1992, p.55). The second meaning of transdisciplinary suggests a shift toward problem-oriented research, rather than discipline-oriented research. In transdisciplinary research, a problem space occurs beyond disciplinary boundaries, and involves participation by relevant fields and stakeholders outside of the academy. It is less about the disciplines that are involved in addressing the problem and more about involving a combination of expertise that can effectively find a shared solution (Ylijoki, 2000;). In both instances, the outcomes transcend disciplinary boundaries, and function both across and within the disciplines.

When referring to multiple discipline engagement in the area of digital curation, it was common among interviewees and in the literature to apply the term interdisciplinary to describe it. Interest in practice and research related to digital curation cuts across disciplinary boundaries. The rapid growth and adoption of computer technology has had an impact on the communication and research practices in the academy. In the domain sciences, research data management,
digital preservation, and data sharing, as well as new ways of conducting research in the digital environment necessitate forward thinking about the creation and maintenance of digital data. In the information fields, digital cultural heritage, digital formats of information and publication, and new ways of searching and using data have changed traditional practices and services. All the while, the technology fields continue to advance more quickly than our understanding of how to secure the longevity of digital materials. This is an area of concern for anyone working with digital content.

However, a closer look at the findings suggests that the area of digital curation is not altogether interdisciplinary. Although multiple disciplines are engaged in the discourse, the data indicates that there is a lack of integration of disciplinary knowledge. Analysis of the references of the subsample of the content analysis for the top 5% most frequently cited articles show that 60 of the 84 articles published in the domain sciences journals did not reference any articles outside of their general area, and specifically not any from archival, digital curation, LIS, or technology publications. However, articles published in archival, LIS, and technology journals regularly cited journals from other fields.

The data of this subsample also shows that there are differences between terminology used and the field of journal in which an article is published. This supports findings from the terminology theme: the language in this area is not uniform, and speakers select terms that will be most meaningful to their audiences. It also highlights that the fields have fundamentally different interests that drive their practice and research in the area of digital curation. The findings show an emphasis on \textit{data archiving} and \textit{data curation} in the domain fields (life sciences, physical sciences, and social sciences), and \textit{digital preservation} in the information and technology fields.
The findings suggest that concepts related to digital curation viewed through the lens of the information fields are approached as an interdisciplinary endeavor. It is not surprising that the data shows that articles published in journals representing the information fields display the most diversity in their references. Library and information science and related areas are inherently interdisciplinary and multidisciplinary, drawing on a variety of fields (Prebor, 2010; Saracevic, 2009). The findings also suggest that the technology fields not only incorporate literature from the information fields, but also view the information fields as part of their audience for the dissemination of work.

From the viewpoint of the domain sciences the data show little indication of an interdisciplinary approach to concepts related to digital curation. However, these articles are multidisciplinary; the subject matter is intrinsically about technology and information management. Although technology publications were referenced in 16 out of the 24 articles that cited outside of their general domain, overall there was little diversity or apparent integration of disciplinary knowledge. The apparent lack of integration does not preclude the use of concepts from other disciplines to address problem spaces within the boundaries of the home discipline.

Taking this into consideration, the area of digital curation is to some degree both interdisciplinary and multidisciplinary; which, depends on the disciplinary perspective of the individual. Practices related to digital curation are widely applied across disciplinary boundaries and advancing this area is important to diverse stakeholders. Many of the domain fields provide additive disciplinary knowledge to address unique concerns specific to those disciplines. But the disciplinary knowledge that has been integrated for the development of concepts related to digital curation, including best practices and conceptual models, is limited. The bulk of interdisciplinary collaboration appears to come from the information and technology fields.
Theme Four: Education

Education was the fourth theme that emerged from the data analysis. The theme of education was defined as any commentary about educating or training professionals for activities and employment related to digital curation. The theme was derived from codes for career paths, curriculum development, educational programs, professional training, and required competencies.

The education theme aligned with two of the elements that comprise a discipline according to the conceptual framework: “an academic organizational unit” and a higher educational “curriculum.” These two elements account for two of the six institutional aspects of a discipline and are the elements that expressly embed the discipline into higher education. These institutionalization components are essential factors in disciplinarity because it is this embeddedness in higher education that differentiates the definition of academic disciplines from those of professions and communities of practice.

An academic organizational unit is frequently embodied in a department, research center, or some other kind of organizing and political unit within the university system in which teaching and research takes place. These organizing and political units are responsible for hiring and promoting faculty according to accepted criteria within the discipline; allocating administrative, research, and teaching responsibilities; and “the reproduction of the discipline via the training of postgraduate students” (D’Agostino, 2012, p. 335). “Disciplines are thus in part identified by the existence of relevant departments: but it does not follow that every department represents a discipline” (Becher & Trowler, 2001, p. 41). Although the academic department is the most visible organizational unit, this is only one potential organizing structure and within any organizing unit there may be more than one discipline represented.
An academic discipline is also comprised of a core curriculum. A curriculum provides a coherent and teachable body of disciplinary knowledge (D’Agostino, 2012). “Disciplines provide the comfort of some stability in curricula and provide some general structure for the organisation of teaching, especially at an undergraduate level” (Krishnan, 2009, p. 43). The curriculum supports and is supported by identifying and teaching graduate attributes consisting of disciplinary-specific learning outcomes and skills relevant to real world application (Barrie, 2004); threshold concepts which are prerequisite principles, and tenets that function as a gateway to the discipline (Meyer & Land, 2003); and signature pedagogies defined as “characteristic forms of teaching and learning” aimed at preparing professionals “to think, to perform, and to act with integrity” within their field (Shulman, 2005, p. 52).

Education: Interviews

Interviewees engaged with the theme of education in three key ways. First, interviewees commented on the kinds of skills that are needed for professionals in this area. Most interviewees viewed the area of digital curation as requiring a combination of knowledge and skills from the information fields and the technology fields. However, from their particular vantage point, interviewees often considered skills related to digital curation as add-ons to existing curricula. In reference to biocuration, for instance, an interviewee commented:

*Interviewee 03: For most groups that I am aware of you can just be a very good biologist and that is fine. It is nice to be able to have…a big picture view, and make sure you don’t get drawn on too many details when you have to process so much information…also having the mind to understand how a database is structured, even if you don’t need to design the database yourself.

Interviewees from the LIS and archival fields posited being information professionals first. For example:
Interviewee 01: I don't think that you need to make archivists or librarians also IT professionals. But I think there are IT skills that you have to have.

Interviewee 10: Skills, I think it is really important for people to understand how systems are built and how systems are evaluated and how requirements are developed for systems... So, I think those are some of the areas and skill sets that are sometimes not there in terms of beyond a traditional library or archive skill set knowledge base and that kind of goes without saying.

Interviewee 12: Okay, first and foremost, they need the archival theory. They need to know what we are doing and why we are doing it. Second to that is an understanding of the way technology works.

Some interviewees expressed concern that there was too much focus on technology which put traditional skills at jeopardy of being displaced. For example:

Interviewee 01: I think it has, for some time, that the whole area of digital curation/digital archiving is becoming much more focused on IT solutions and not information management solutions. And I think that's the wrong emphasis... Because if you don't have some understanding of what you are actually trying to do conceptually then you are not going to get the IT solution to do it.

Interviewee 13: I think there has been too much focus on technology for its own sake and not enough understanding of how technology has developed over time... It’s a danger I think for us to be just technology obsessed – all we need is more and better, just lots of technology. Libraries, archives, museums are always going to be content focused or collection focused, so managing a collection as opposed to managing a technology infrastructure, that's what's important.

Second, interviewees commented on how digital curation education is related to the academic disciplines. For the most part, interviewees identified the area of digital curation as existing on the boundary between the information fields and the technology fields. Interviewees also commented on educational opportunities needed across disciplines. For example:

Interviewee 04: Minding that the core focus definitely will be in the archival/library sciences and computer sciences as those are the two core disciplines that will have to tackle the problem, but I would think, yes that most other disciplines would highly benefit from at least an offering into those aspects.
Interviewee 11: Yeah, so I think curation really spans disciplines and domains… But I think curation should be built into all domain training in the future, and certainly we are seeing that in the data world…You have people from physics and astronomy and biology, and there are sciences really taking a lead on curation-type activities and curation advocacy. So it really fits within all disciplines, but I think the information discipline is taking the lead on training and teaching and growing young professionals.

Several interviewees questioned the potential interest of computer science departments wanting to take responsibility for education related to digital curation. They tended to focus their observations on the types of students attracted to computer science, who are not necessarily interested in careers related to digital curation. For example,

Interviewee 02: I think generally computer science courses would see this as being probably a bit boutique. I would have to say it probably has to come from the library, museums, archives and galleries kind of world.

Interviewee 04: I am not aware of any masters program that builds on top of computer science degree to teach digital preservation or digital curation as concepts, the reason probably being that candidates in, or people who have a degree in computer science have such a big selection of job offers on the market and preservation/curation, archives and libraries are not amongst the first target employers that come to mind for them.

Third, interviewees discussed the responsibility of digital curation as falling on multiple people with diverse skills sets, not a single professional who possessed all of the required skills. For example:

Interviewee 06: I guess one of the things I like about the area and one of the things that I like about the community is that it welcomes all sorts of people with all sorts of skills, including educators and trainers. You can work with somebody at different levels, and it actually needs people with so many different skill sets and the mindsets to make the community and this activity fully functional.

Interviewee 07: So there are two main blocks I think of doing data management. There is the technical side of things which involves managing servers and moving files around and checking file formats or building databases or building metadata models…and it's software developers and technical types you need to be able to do those sort of things…That's important, but I also think that it's really important that you have the, I
don't know the softer more social skills as well. That would include things like going out
and talking to the researchers and engaging with them so that they feel like they are being
valued but also engaging with them to get that information out of them for the metadata
and encourage them, reward them, make them feel like they are part of the team when it
comes to actually submitting data.

*Interviewee 14:* There is no single person who is going to have all the skills you need in a
digital preservation person.

The theme of education echoes many of the same observations as the theme of multiple
discipline engagement. In both themes, the data analysis indicates a diversity of stakeholders
invested in this area. The principal academic and research communities engaged in the area of
digital curation appear to be those from within the information fields and the technology fields.

Several perspectives on the transmission of the knowledge and skills base related to
digital curation became apparent when interviewees talked about education. For the most part,
the findings indicated that the knowledge and skills base for the area of digital curation come
from a combination of the information and technology fields. Some interviewees viewed the
area as an interdisciplinary domain that is related but distinct from its originating fields,
underscoring a nascent disciplinary perspective. Others view this domain as additional
competences required for traditional degree programs or professions. For example, data on this
surfaced in discussions with interviewees in the archives field and in biocuration. In both cases,
the primary objective was described as the education of archivists or biologists who also possess
other skills required to address activities related to digital curation. Still others perceive digital
curation as a set of practices that need to be widely adopted, embedded in research practice and
within the information fields, and valued as part of the institutional structure. The emphasis is
not to distinguish the practices as something different, but rather as accepted conventions.
Education Explored

Whether emerging as an academic field or a set of professional practices, the education and training of professionals equipped to address skill gaps has been acknowledged as a significant challenge in the area of digital curation. Digital curation remains broad in scope, and has widespread implications for application and implementation. General competencies have been identified, but specific skill sets reflect narrower professional roles. Because of the scope of the area and lack of core curriculum, courses related to digital curation “are somewhat of a moving target, not only in the ways and frequency in which they are offered, but in the ways they are named and presented in the curriculum” (Bastian, Cloonan, & Harvey, 2011, p. 614).

Table 7.4 offers a selection of program names for degree-granting programs, specializations, and post-graduate certificates. This selection highlights the variety of names used in this area. The choice of program name is indicative of the originating discipline, learning objectives, the intended audience, the date the program was developed, and the influence of the creators of the program. The program name and curriculum also needs to be relevant to students and future employers. Unless otherwise noted, these programs are offered through the library and information science department. Further analysis into ways in which digital curation and preservation concepts are presented or integrated across LIS, computer science, and within the domain sciences curricula is warranted.
Table 7.4

*Examples of terminology used for programs in higher educational institutions*

<table>
<thead>
<tr>
<th>University/College</th>
<th>Program</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aberystwyth University</td>
<td>Master’s Degree</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>Johns Hopkins University (Museum Studies)</td>
<td>Certificate</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>Kent State University</td>
<td>Certificate</td>
<td>Digital Preservation</td>
</tr>
<tr>
<td>Kings College London/ Humboldt University Berlin</td>
<td>Master’s Degree</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>Robert Gordon University</td>
<td>Master’s Degree</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>Rochester Institute of Technology (Computer Science)</td>
<td>Specialization</td>
<td>Data Management</td>
</tr>
<tr>
<td>Simmons College</td>
<td>Certificate</td>
<td>Digital Stewardship</td>
</tr>
<tr>
<td>Syracuse University</td>
<td>Certificate</td>
<td>Data Science</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>Certificate</td>
<td>Digital Information Management</td>
</tr>
<tr>
<td>University of Dundee</td>
<td>Master’s Degree</td>
<td>Records Management and Digital Preservation</td>
</tr>
<tr>
<td>University of Glasgow</td>
<td>Specialization</td>
<td>Management &amp; Preservation (Digital)</td>
</tr>
<tr>
<td>University of Illinois</td>
<td>Specialization</td>
<td>Data Curation</td>
</tr>
<tr>
<td>University of Maine</td>
<td>Certificate</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>University of Maryland</td>
<td>Specialization</td>
<td>Curation and Management of Digital Assets</td>
</tr>
<tr>
<td>University of Michigan</td>
<td>Specialization</td>
<td>Preservation of Information</td>
</tr>
<tr>
<td>University of North Carolina</td>
<td>Certificate</td>
<td>Digital Curation</td>
</tr>
<tr>
<td>University of North Texas</td>
<td>Certificate</td>
<td>Digital Curation and Data Management</td>
</tr>
<tr>
<td>University of Texas at Austin</td>
<td>Certificate</td>
<td>Digital Archiving</td>
</tr>
<tr>
<td>Wayne State University</td>
<td>Specialization</td>
<td>Digital Content Management</td>
</tr>
</tbody>
</table>
There has been considerable growth in educational and training opportunities in the area of digital curation, namely curriculum development, continuing education and professional training opportunities, and research data management instruction. Although core competencies and curriculum development have been addressed (see Tibbo, Hank, & Lee, 2008; Kim, Addom, & Stanton, 2011; Kim, Warga, & Moen, 2013), the administrative, organizational, and political unit of this area within higher education is yet undecided. For the most part, the themes of multiple discipline engagement and education suggest that digital curation intersects the boundaries of information fields and the technology fields. Currently, programs tend to originate in LIS or archival programs, but the interdisciplinary and multidisciplinary nature of the area makes it relevant for many disciplines. The theme of education also highlights the multiplicity of terminology used in this area, and in turn the multiplicity of the conceptual and practical domains which each of those terms represent.

**Summary**

In this chapter, I presented the findings and discussion for the themes multiple discipline engagement and education. The theme of multiple discipline engagement that emerged from data analysis corresponds with the disciplinary element “styles of subjectivity.” The findings suggest that practice and research related to digital curation involves individuals with different expertise, knowledge, and perspective. This translates to stakeholders from different disciplinary backgrounds merging diverse interests and perspectives. At present in the area of digital curation, each discipline appears to bring its own style of subjectivity. The theme of education directly aligns with the academic organizational unit and higher educational curriculum elements. The findings suggest that education and training for areas related to digital curation are still developing and a higher education core curriculum still evolving. Although the skill sets
come from both the information and technology fields, currently programs in this area are mostly offered through library and information science departments, or as individual courses in other domains.

In the next chapter, I present the third part of the findings and discussion. The theme of areas of professional and scholarly focus is explored in detail and I provide a summary of the findings and discussion chapters.
CHAPTER 8: FINDINGS AND DISCUSSION: PART III

This chapter is Part III of the findings and discussion. The chapter is organized in two main sections. In the first section I explore the findings and discussion for the final theme, areas of professional and scholarly focus. In the second section, I provide a comprehensive summary of the findings and discussions that have been explored in Chapters Six, Seven, and Eight.

Theme Five: Areas of Professional and Scholarly Focus

The final theme addressed in the findings and discussion was developed from a set of open codes from the content analysis. This theme, areas of professional and scholarly focus, is defined as a descriptive summary of the main subject matters that characterize the area of digital curation as represented in the sample of published literature in this study, including informative articles, commentary, reports, and research. Each of the articles coded for the content analysis can be categorized into one of these six areas. The six areas of professional and scholarly focus identified in this study are: Development of the Field, Implementation, Functions, Preservation, Research Data, and Infrastructure.

The areas are described as follows:

1. Development of the Field.

This area of focus includes commentary and research about education in the area of digital curation; roles and responsibilities of professionals in practice; and theory building and model development. It also contains calls to action; surveys of the area’s historical development including commentary on exemplar or pivotal projects; and research agenda setting. This area of focus is both reflective in nature and represents a relevant object of research, exemplified by this study.
2. *Implementation.*

This area of focus is wholly about digital curation in practice and best practices. It encompasses practical case studies and reports of activities; program descriptions and commentary on the implementation of digital curation/preservation programs; application of tools or workflows; and digitization efforts and technology. Also within this focus are ethical and legal concerns. This area of focus is indicative of the applied and practical orientation of the area reinforced by the influence of collaborative initiatives and exemplar implementation projects.

3. *Functions.*

This area of focus represents commentary and research on the major actions, functions, and processes of digital curation. In the sample, these tended to be clustered around the Ingest, Access, and Archival Storage functional entities of the Open Archival Information System Reference Model. However, this area of focus also includes, for example, acquisition, appraisal and selection, and instruction and outreach, as well as general repository administrative and management issues.


Although preservation could be grouped within the Functions category, it has its own area of focus because it was one of the more prevalent subject matters in the literature. This area of focus is concerned with such subjects as preservation strategies, risk assessment, threats to longevity, and web archiving. It contains four subareas: *Policy/Planning*, preservation of specific *Content Types*, preservation of specific *Media Types*, and *Technology Solutions*. 
5. **Research Data.**

Research data is another subject matter that was assigned its own area of focus because of its prevalence in the literature, especially from within the domain sciences. This area of focus encompasses commentary and research specific to any field of research data as a particular type of content. Topics such as curated databases, data annotation, data management, data publication, and data sharing are included in this focus.

6. **Infrastructure.**

This area of focus pertains to commentary and research about building sustainable infrastructure for activities related to digital curation. It contains two subareas. First, *Cyberinfrastructure*: topics related to infrastructure for research activities, which includes data archives, e-Research activities, and scholarly communication. Second, *Digital Infrastructure*: topics related to infrastructure for digital preservation, which includes digital archives and digital libraries, interoperability, and standards. The Infrastructure area of focus addresses big picture interoperability and sustainability issues, and new ways of conducting research in the digital environment.

For each of the 460 items in the content analysis, codes were recorded for the subject of the article (focus) and for the topic of the article (locus). The subject is the broad general category about which the article was written. The topic is the narrower issue of investigation which falls into the subject category. For example, the focus of an article was preservation and the locus was the challenge of technological obsolescence; the purposes of the articles and their supporting material within a single area of focus were quite variable. The focus and locus codes were explored, aggregated into categories, and then refined into six areas of professional and
scholarly focus. Figure 8.1 provides a visual representation of the six areas of professional and scholarly focus and ways in which they relate to one another. The figure includes the number of articles from the sample (n) that were categorized into each area.

**Figure 8.1.** Content analysis sample: The six areas of professional and scholarly focus and their relationships.

There are evident relationships between the areas of professional and scholarly focus, represented by dotted lines in Figure 8.1. For instance, Research Data is associated with the Preservation subarea of Content Type. Also, the Preservation subareas of Policy/Planning and Technology Solutions are associated with the development of a sustainable infrastructure for
digital preservation. Implementation, as noted by its multiple relationship lines, has far-reaching implications. Additionally, commentary and research on the topic of use and users show up in areas such as access (Functions) and data reuse (Research Data); and the topic of metadata is everywhere and categorized based on the core subject matter, such as metadata at ingest (Functions) or metadata standards (Infrastructure).

Two of the areas of focus deserve special attention, namely Preservation and Research Data. Articles in these two subject categories comprised over half of the sample. Because preservation is a central activity of digital curation, articles concerning this subject could be categorized in the Functions category. The Functions category represents literature about the key activities and processes of digital curation. This subject category can be broadly understood as including articles about topics related to the functions of the OAIS Reference Model or the actions of the DCC Curation Lifecycle model. However, the literature that focuses on preservation issues was more extensive than other subjects. This is in part because preservation has been at the center of research agendas for more than 20 years and is at the core of digital curation. Long-term preservation, which ensures accessibility and counters obsolescence, is the main response that drives the area. For this reason, reservation constitutes its own subject area unique from other functions of digital curation.

Research data is the only specific content type that warranted its own subject area category in this study. Research data is the critical junction between the domain sciences and the information and technology fields in discussions about concepts and practices related to digital curation. Although the underlying practice and concepts for the long-term preservation and stewardship of research data are the same as other types of digital objects, the issues that surround research data constitute its own problem space and stimulate research questions specific
to the management, reuse, and sharing of scientific data. As Figure 8.1 shows Research Data is associated with the subject category Implementation, the Preservation subarea of Content Type, and the Infrastructure subarea Cyberinfrastructure. These relationships are significant because they link the stewardship of research data with other activities related to digital curation rather than isolate this content type from ongoing research and expertise.

**Research Agendas**

The theme of areas of professional and scholarly focus that was derived from the content analysis is mapped to the disciplinary element of “body of accumulated knowledge and skills” from the conceptual framework. A discipline is founded on an object or phenomenon of study which constitutes the agreed domain of practice, research, and teaching. Underpinning this body of knowledge and skills are characteristic methodologies; conceptual frameworks and explanatory theories; and a critical mass of disciplinary assumptions, definitions, explanations, and facts about the domain. The discipline gains institutional and social control over the knowledge and research of that particular phenomenon of study and in turn establishes boundaries around the domain.

The body of accumulated knowledge embraces an agenda for research, which identifies and prioritizes research problems and gaps in our understanding. In the area of digital curation there have been several efforts to outline a research agenda and codify past initiatives in an attempt at a more systematic approach to research. Since the early 1990s these research agendas have articulated the imperative for improved understanding and implementation of digital preservation practices, the value of digital curation and preservation research, and key challenges and problem spaces to address (Hedstrom, 1991; Bennett, 1997; Hedstrom et al., 2003; Workshop on Research Challenges in Digital Archiving and Long-term Preservation, 2003;
Digital Preservation Europe, 2007; Chanod, Dobreva, Rauber, Ross, & Casarosa, 2010; PARSE.Insight Consortium, 2010; Faniel & Zimmerman, 2011). Additionally, they have emphasized the need for a concerted effort across disciplines through cooperation and coordination, and the essential role of collaboration.

Research agendas also serve to identify areas for disciplinary growth. They not only identify gaps in the research, but are also used as a gauge for measuring the progress of the field. For example, in Digital Preservation Research Roadmap (2007), a report prepared by Digital Preservation Europe (DPE), researchers conducted a both an online Delphi study and a crosswalk analysis of 12 published research agendas from the previous 16 years. The researchers concluded that “after 16 years of research and the continuous refinement of a research agenda, very little progress has actually been made during this time.” The result of this study was the identification of 10 core areas for research: restoration; conservation; management; risk; significant properties of digital objects; interoperability; automation; context; storage; and experimentation. The report further observed that there has yet to be many successful shared solutions to the challenges faced by digital preservation: “The theoretical basis of digital preservation research is sound. Now the shift in focus needs to return to resolving the fundamental issues and apply the results in practice” (Digital Preservation Europe, 2007).

This lack of progress in the area of digital curation has been echoed repeatedly. Seamus Ross concluded in his keynote address in 2007 at the 11th European Conference on Digital Libraries that “research in digital preservation must in general be more rigorous, methodologically founded, repeatable, verifiable, contextualised, and more effectively reported; that is, it could conform better to the ‘scientific paradigm.’ It needs to be more ‘experimental’
than it has been up to now” (p. 18). In a 2011 plenary address at a symposium on preservation education hosted by the University of Michigan, Anne Gilliland noted:

If high-impact research will be fostered and sustained, then preservation must become an established presence in the academy. At the same time, it must remain embedded in practice—in undertaking preservation-centric research, and in preparing future professionals and academics who can undertake rigorous research in the field and in the academy. There must be a holistic conceptualization of preservation as a field (Gilliland, 2014, p. 54).

Research is essential for the development of a discipline and the area of digital curation in general. Research builds the knowledge base and refines the domain, establishes societal importance of the area, provides for evidence-based practice, and encourages critical analysis of the extant practice and research. Most recently, the 2015 National Agenda for Digital Stewardship report called attention to “the limited amount of empirical evidence available.” The report further observed that “the digital preservation community is beginning to develop a shared evidence base; however, studies must be broadened and repeated over time to establish a robust evidence base from which generalizable guidance can be drawn” (National Digital Stewardship Alliance Coordinating Committee and Working Group, 2014, p. 7).

The six areas of professional and scholarly focus noted in this chapter are not defined as research areas, but rather suggest areas of community interest observed in the published literature. The extant body of accumulated knowledge and skills does include research, but the literature is underpinned by professional and scholarly commentary, reports of collaborative research initiatives, exemplars of implementations, and conceptual model development. The content analysis sample did not focus only on research, but on any form of published material as
It appeared in the content analysis sample. The area of digital curation and preservation as a research field is emergent. As it emerges and as the area differentiates and refines itself, the literature reflects priorities in the developing domain, as well as problem spaces, research questions, and topics of interest.

These areas of focus do align with some of the problem spaces and research questions advanced in the published research agendas. For example, DPE’s *Digital Preservation Research Roadmap* (2007) largely focused on preservation issues, but also highlighted interoperability and storage as two areas for further research activity. These two topics are included in the areas of professional and scholarly focus within the Infrastructure and Functions categories, respectively. Faniel and Zimmerman (2011) target the research data sharing and reuse problem space, which is contained in the Research Data category in this study.

The accumulated body of knowledge and skills, and more specifically research in the area of digital curation, is emerging and maturing. Attempts at a more systematic approach to growing the area have been made and continue to be made. It is clear that some progress has been made; we are more aware of pressing research questions, the need for more rigorous research, and evidenced-based practice. There is an audience for the dissemination of this research, and conferences and journals for presenting findings. However, the challenge of communicating across disciplinary boundaries in a common language and to a variety of stakeholders needs to be overcome so that the progress can augment, application of research can be comprehensive, and shared solutions do not develop in isolation.

**Summary of Findings and Discussion**

In Chapters Six, Seven, and Eight, I presented and discussed the key findings from the analysis of the data. Qualitative and quantitative data from scoping the literature, content
analysis, and interviews were explored together and provided corresponding findings. Five major themes emerged from the data analysis: terminology, collaboration, multiple discipline engagement, education, and areas of professional and scholarly focus.

The first theme that emerged from the data analysis was terminology which aligns with the “discursive community” element from the conceptual framework. The theme of terminology was defined as words or phrases used to identify the area of practice and research related to digital curation. The findings suggest that although there is generally a shared understanding of most field-specific jargon in the area of digital curation, a common language has not fully coalesced. Choice of terminology is descriptive and situational, largely determined by what is most meaningful to the audience, rather than prescriptive and harmonized. The findings also indicate that there are disciplinary differences in the choice of terms employed. The examination of terminology also arises recurrently in the literature. The purpose of these discussions are often to specify the evolution of how a term came into use, identify the distinction between terms, or address why one term is used by the authors rather than another (Yakel, 2007; Dallas, 2007; Baker & Yarmey, 2009; Bastian, Cloonan, & Harvey, 2011; Flanders & Muñoz, 2011; Palmer, Weber, Muñoz, & Renear, 2013).

The theme of collaboration aligns with the “social networks” disciplinary element. The theme of collaboration was defined as any instance or mention of two or more people or institutions cooperating, sharing resources, or working together for mutual benefit or to achieve shared goals. The findings suggest that there are well-established social networks in place within the digital curation and related communities. These networks are cultivated through collaboration, and a conviction within the community of the necessity and value of collaboration. Interviewees approached the theme of collaboration from two main perspectives. First,
collaboration was observed to be necessary for successful and sustainable digital curation in practice. Second, interviewees associated collaborative projects with building communication and social networks with others working in the area. The data provided nominal detail as to the degree to which these social networks extend across disciplinary boundaries; while it is evident that some disciplines interact more than others, it appears that there is a level of disciplinary isolation.

The theme of multiple discipline engagement that emerged from data analysis corresponds with the disciplinary element “styles of subjectivity.” This theme was defined as any instance or mention of two or more people or institutions from different disciplines or fields working together, or the combination and application of concepts, methods, practices, or theories from two or more distinct disciplines or fields. The findings suggest that practice and research related to digital curation involves individuals with different expertise, knowledge, and perspective. This translates to stakeholders from different disciplinary backgrounds merging diverse interests and perspectives. When working in multi-discipline settings, styles of subjectivity are placed in conflict with each other. For instance, collaboration between multiple disciplines “requires rigor, but…it also requires…willful suspension of [one’s] own concepts of rigor” (Strober, 2006, p. 322). At present in the area of digital curation, each discipline appears to bring its own style of subjectivity. However, a nascent digital curation perspective characteristic of the area was discernible in some of the data.

The theme of education directly aligns with the “academic organizational unit” and a higher educational “curriculum” elements of the conceptual framework. The theme of education was defined as any comment about educating or training of professionals for activities and employment related to digital curation. Gaps in skills have been identified in the area, and are
being addressed by an increase in opportunities for professional development training and through higher education curriculum development projects. The findings suggest that education and training for areas related to digital curation are still developing and a higher education core curriculum is still evolving. Although the knowledge and skills base comes from both the information and technology fields, currently programs in this area are mostly offered through library and information science departments or as individual courses across domains.

The theme of areas of professional and scholarly focus is mapped to the “body of accumulated knowledge and skills” disciplinary element. This theme is defined as a descriptive summary of the main subject matters that characterize the area of digital curation represented in the sample of published literature in this study, including informative articles, commentary, reports, and research. There were six areas of professional and scholarly focus identified in this study: Development of the Field, Implementation, Functions, Preservation, Research Data, and Infrastructure. Each of the articles coded for the content analysis were categorized into one of these six areas. The findings suggest that there exists a considerable body of professional and scholarly knowledge in the area of digital curation, which is increasing and still evolving. However, because of the area’s interdisciplinary and multidisciplinary nature the domain is large in scope and unsettled. The interface and transmission of knowledge between the realms of education, practice, research, and theory within the area of digital curation continues to mature.

In the next chapter, I attend directly to the research questions and offer conclusions to this study. I discuss implications of the study, review limitations, and provide direction for future research. The aim of the concluding chapter is to pull together the evidence that has been presented in the previous chapters to provide a cohesive summary.
CHAPTER 9: CONCLUSION AND FUTURE RESEARCH

In this chapter I discuss the research questions and offer conclusions to this study. The chapter is organized in four main sections. In the first section, I supply a summary of this research. In the second section, I consider the research questions in detail and discuss implications of the findings. I also review the limitations of the study. In the third section, I offer a conclusion to this dissertation. In the last section, I present areas of future research.

Summary of Research

This dissertation research was designed to investigate whether digital curation, defined broadly in this study as the active lifecycle management of data for current and future use, is emerging as a distinct discipline. The purpose of this study was to explore the character, development, and education of the area of digital curation by conducting an in-depth analysis of how this area is evolving. The conceptual framework employed in this research positions the findings relative to elements that characterize it as an academic discipline.

The research questions addressed in this study were:

RQ1: Is digital curation emerging as an autonomous discipline?
   a  In what ways do interviewees describe the relationship of digital curation to areas within the information disciplines?
   b  Are there indicators that suggest digital curation is emerging as a discipline in its own right?

RQ2: Where does digital curation fit within the educational landscape?
   a  To what extent and in what ways do interviewees express the role of education in the knowledge transfer of digital curation practices, skills, and theories?
This dissertation employed a qualitative multimethod research design. The methodology used was scoping the literature, content analysis of published literature in the area of digital curation (n=460), and interviews with people engaged in this area (n=14). This study focused on identifying themes that emerge from the academic discourse and scholarly communication, and understanding the level of consensus among educators, practitioners, and researchers about how they view this domain.

In Chapter Three, I provided a context for understanding the nature and development of academic disciplines. This chapter also provided support for the conceptual framework used in this study. In this dissertation, I used the conceptual model put forth by D’Agostino (2012) that views a discipline as a cultural system, which is the interaction of ten elements that characterize a discipline interpreted within a framework of shallow consensus. D’Agostino contends that although we can define a discipline in terms of it meeting certain criteria, we must also contextualize the development of a discipline historically and accept that members of a discipline may have varying interpretations of what each element embodies.

In Chapter Four, I presented an historical perspective of the development of the area of digital curation. The narrative outlines the contributions of the archival, LIS, and research communities in the evolution of a digital preservation community and the eventual emergence of digital curation in 2001. Adopting the phrase digital curation was an effort to recognize an area of shared concern across all domains, and that longevity and reuse of data is an ongoing process influenced by complex cultural, organizational, and technical issues. In Chapter Five, I extended this narrative to address the growth of higher education curriculum and professional training in the area of digital curation.
In Chapters Six, Seven, and Eight, I presented the key findings and discussion. Five major themes emerged from the data analysis: terminology, collaboration, multiple discipline engagement, education, and areas of professional and scholarly focus. The five themes were mapped to the elements of the conceptual framework, and were used to support the conclusions presented in this chapter.

**Research Questions**

In this section I address each of the research questions and subquestions drawing from data I collected, the framework I used, and the analysis I conducted. In short, digital curation has not emerged as an autonomous discipline, but does meet several of the criteria to indicate its potential for emergence. Currently, it intersects the boundary between the information fields and the technology fields. Importantly, a nascent perspective characteristic of the area of digital curation was discernible in the data that differentiated it from the disciplinary perspectives of its originating fields. Education for this area fits well in coordination with archive and LIS higher education; however, there are many interested stakeholders and skill development is important across the domain fields.

**RQ1: Is digital curation emerging as an autonomous discipline?**

Observed in the area of digital curation are the reducibility of concepts to their originating discipline, and a lack of cohesiveness across the broad social network and intellectual domain. Digital curation comprises archival theory, information management, and technological innovation, among other concepts and skills borrowed from extant disciplinary knowledge bases. At present, these aspects have not been integrated to a point where they form their own body of knowledge that is developing autonomously. As such, digital curation is not the jurisdiction of a single discipline, but the result of a shared need across disciplines.
The first subquestion asked: In what ways do interviewees describe the relationship of digital curation to areas within the information disciplines? The aim of this question was to understand how the digital curation community articulates the roles of the fields of archival studies, and library and information science in relation to the area of digital curation. This was ascertained to gain a better appreciation for which fields influence the area of digital curation and to counter any bias arising from my perspective from within the information fields.

For the most part, interview participants viewed the area of digital curation as existing on the boundary between the information fields and the technology fields. Digital curation is fundamentally about information management and preservation. It borrows concepts from archival studies such as authenticity and provenance. The area extends to research areas that investigate how people access, search, and use/reuse digital information. This all occurs intrinsically within the confines of technology.

However, the role of domain knowledge for areas related to digital curation cannot be underrepresented. For the curation of domain-specific data and databases, expertise in the domain is essential for providing accurate annotations and metadata. In fact, many within these areas see this as a technology problem not an information problem. This is clearly evident in the findings that show domain specialists routinely using information management practices to enhance research possibilities without consulting information management experts. Again, these new research possibilities exist bounded by technological innovations.

This incongruity between the information fields and the domain fields creates a dichotomy between the treatment of digital cultural heritage, including library materials, and research data. This is observed in the use of the term digital curation more frequently in publications relating to archival studies, LIS, and museum studies, and the more accepted use of
the term *data curation* in domain sciences publications. This dichotomy is not only perpetuated by differences in terminology and a distinction of type of digital asset, but also by the more systemic lack of cohesion and harmony across digital curation and related areas.

The second subquestion asked: Are there indicators that suggest digital curation is emerging as a discipline in its own right? This question gets to the core of the analysis of the data. Its objective was to identify what elements, as defined by the conceptual framework, were observed in the area of digital curation. Table 9.1 outlines which disciplinary elements are observed for the area of digital curation and provides markers that illustrate progress.

D’Agostino (2012) contends that although we can begin to define a discipline in terms of it meeting the certain criteria, we must also contextualize the development of the discipline historically and accept that members of the discipline may have varying interpretations of what each element embodies. By acknowledging the elements that comprise a discipline in the context of the development of the area we can better realize how the discipline came to be constituted by those elements. D’Agostino argued that the elements that constitute a discipline have to be understood within the framework of “shallow consensus” (p. 341). Shallow consensus is the varying degrees of agreement that members of the same discipline share about their discipline or aspects of their discipline. It is this shallow consensus that is necessary to feed the exploration of new ideas, innovation, and growth of the discipline, or conversely divide the discipline.
Table 9.1

Summary of disciplinary elements for digital curation and markers of progress in the field.

<table>
<thead>
<tr>
<th>Elements of a Discipline</th>
<th>Present for Digital Curation?</th>
<th>Markers of Progress</th>
</tr>
</thead>
</table>
| Academic Organizational Unit | No | A number of non-academic organizational units are associated with higher education institutions. These tend to focus on advocacy, research and development, and training. For example:  
- Digital Curation Centre, United Kingdom (2004)  
- Digital Curation Institute, University of Toronto (2010)  
- Digital Curation Unit, Athena Research Center, Greece (2007)  
Certificates, specializations, and the few United Kingdom degree-granting programs are mostly offered from library or information schools or the occasional computer science department. |
| Accumulated Knowledge | Active Development | Efforts to establish research agendas have set research priorities for areas related to digital curation, principally for digital preservation, and agendas are revisited to encourage and gauge progress. |
| Curriculum | Active Development | There have been several curriculum development projects as well as local efforts. The two foremost projects were:  
- Digital Curator Vocational Education Europe (DigCurV, 2011-2013)  
| Discursive Community | Active Development | A shared terminology for the area has not yet emerged, although conceptually there is consensus. Effective collaboration and work across disciplines compels terminology, as well as other communication practices, to be more harmonized. |
| Narrative | Incipient | Digital curation has evolved alongside computer technology, only at a slower pace, with concepts and practices that have emerged from within a variety of disciplines out of need. Digital curation has a complex history with many actors and requires a narrative that represents the involvement and progress of all stakeholders, which has not yet been written. |
| Professional Association | No | Several professional associations have special interest groups for digital curation and related areas. The International Digital Curation Conference, established by the Digital Curation Centre, convened for the first time in 2005. The Digital Preservation Coalition (DPC) sponsors the Digital Preservation Awards. |
| Publishers | Existent | There is a growing body of literature in the area of digital curation including books, monographs, and scholarly journals that include in their scope articles and research related to digital curation. The *International Journal of Digital Curation* was established by the Digital Curation Centre in 2006. |
| Recognition in Classificatory System | No | *Digital curation* is recognized as a variant of *digital libraries* and *digital preservation* in the Library of Congress Subject Headings. Dewey Decimal Classification has *digital preservation* and *digital libraries* as subject headings but not *digital curation* or *data curation*. |
| Social Networks | Existent | Digital curation has a dedicated community that sustains social networks through collaboration, and both formal and informal methods of communication. |
| Styles of Subjectivity | Incipient | The interdisciplinary and multidisciplinary qualities of digital curation impede a shared “styles of subjectivity” from maturing. However, a nascent digital curation-oriented perspective is discernible. |
Digital curation has a complex history with many actors and as such requires a narrative that represents the involvement and progress of all stakeholders. The area of digital curation has evolved alongside computer technology, only at a slower pace, with concepts and practices that have emerged from within a variety of disciplines. These disciplinary-specific concepts and practices, from the archival, LIS, and research communities, developed parallel to each other without much integration. The confluence began as a digital preservation community emerged and was reinforced with the emergence of digital curation, which conceptually captured the set of practices and accompanying models for the lifecycle management and preservation of all data.

There are clear markers that demonstrate progress has been achieved in the area of digital curation; however, there are several aspects that impede its emergence as an academic discipline. In particular, these aspects revolve around the absence of a digital curation identity. Identity is “a sense of shared understandings and skills, experiences, common way of perceiving problems and their possible solutions. This identity is formed and reproduced through a shared and common educational background and professional training, work practice and memberships in professional associations” (Kallberg, 2012, p. 99). A commonly held sense of identity helps unify a community of people with shared interests and goals.

First, there is not yet consensus on the name of the field or how the field is defined. As noted in the findings and discussion, academic programs exhibit a variety of names for this area of study and we see even more diversity of terminology in publications. There is consensus that the area is about the lifecycle management of digital assets, and involves both adding value and preservation. However, there are distinctions drawn between types of digital assets. These distinctions influence not only how the area is referenced and the conceptual terms employed, but also influence problem spaces for research and areas for practical implementation.
This leads to the second aspect, which is that the domain of digital curation is broad spanning across disciplinary, practical, and research boundaries and there is not yet consensus on the direction of the field. There are calls for more research and evidence-based practice, more implementation of shared solutions including establishing a robust infrastructure, and more educational and training opportunities. But digital curation faces similar obstacles as other applied and professional fields in higher education, such as library and information science, nursing, and information technology. These obstacles include proving their legitimacy as a research and theoretical field as well as gaining acceptance from the established disciplines. Consensus on the direction of the field would concentrate efforts and resources according to the goals of the community.

From the interviews I gleaned three primary underlying points of view about the direction of digital curation. From the first perspective, which included that of the archivists, the domain of digital curation was a set of competencies, practices, and skills that is additional to current roles and responsibilities but not fundamentally different. Another perspective also approached the domain as practical, but not from within a particular field, rather as an advocate for digital curation as an integrated set of practices. A primary goal from this perspective was to embed activities related to digital curation into the research practice for improving data management, replication of findings, and data sharing; and into the practice of information professionals who work with digital materials in archives, libraries, and museums. The third perspective viewed the domain as having the potential to become differentiated as an autonomous specialization or eventual discipline, although the disciplinary knowledge and skills base that support the area are currently embedded in existing disciplines. Most of the interview participants provided commentary that corresponded to one of these first two perspectives.
Lastly, there is not yet a professional organization that advocates for the community and the profession, and provides a venue for collaboration. In addressing archival identity and the role of forming professional associations to centralize efforts, Shepherd (2012) contends that “working together, forging our common identity and self-belief, has brought us a success which we would never have won as a scattered handful of individual scholars” (Shepherd, 2012, p. 176). Part of the problem is that currently there are professionals who take on roles and responsibilities associated with digital curation, but a position for a digital curator is uncommon. It is partly in the name. There are archivists who work with electronic records, librarians who manage institutional repositories, biologists who curate databases, and so on. So, the question remains: are there digital curation professionals or are there professionals who use digital curation practices?

**RQ2: Where does digital curation fit within the educational landscape?**

The education and training for skills related to digital curation occurs across disciplinary boundaries; however, current certificate programs and specialization tend to emerge within library and information science schools. This positioning is filled with both promise and uncertainty: promise that the new digital curation professionals will possess an education that supports interdisciplinarity and a holistic view of information; uncertainty about the level of technological skills these professionals will emerge with and the stigma of outmoded archive and library fields.

The subquestion asked: To what extent and in what ways do interviewees express the role of education in the knowledge transfer of digital curation practices, skills, and theories? This question directly addressed issues of education and training in the area of digital curation. Specifically, it attempted to differentiate higher education programs in digital curation and
practical training for professional development. Additionally, it was intended to get at the state of current educational offerings in this area and how interviewees envision education in this area in the future.

For the most part, courses for digital curation and related areas are additive rather than integrated into the core LIS or archival curriculum. This indicates that digital curation competencies are not, at present, considered core knowledge areas for information professionals. There are some specialization tracks within LIS programs that focus on the curation, management, and preservation of research data or digital assets. Curriculum expansion in this area began and persists as post-graduate or additional credit certificate programs, which illustrates the notion that these complementary skills sets supplement traditional disciplinary education.

The value of taking coursework in digital curation and related areas was also addressed. The two key concerns discussed were not about the value of the skills gained or the value of digital curation as a set of concepts and practices, but about the value of investing in this career path. The first concern addressed was about the accreditation of digital curation programs, which currently does not exist. When the coursework is supplementary to computer science, LIS, or museum degrees, then it is the degree-granting program that carries the accreditation. This is observed in archival studies programs offered through LIS schools, which fall under the American Library Association accreditation for LIS degree-granting programs. The concern is that if students are investing in an education, they are more inclined to invest in an accredited program. The second comment on value focused on the interest of computer science students pursuing this area. Several interviewees commented that digital curation and related areas are not prevalent career paths for computer science students. As such, computer science
departments are more likely to invest in other opportunities more aligned with future employer and student goals.

There is also the notion of the convergence of professional education for archives, libraries, and museums for working with materials in the digital environment. Each profession comes from different disciplinary traditions and works with different types of collections; however, the competencies and skills related to digital curation for the long-term care of these digital assets are not different. Jennifer Trant, a leading researcher in this area, wrote that “while the traditions and historical areas of expertise in archives, libraries, and museums may differ, the new challenges facing all collecting cultural institutions are best addressed in concert, in an interdisciplinary forum that explores multiple solutions and takes advantage of many skills” (Trant, 2009, p. 377). Digital curation reflects this convergence with its holistic view of digital assets, conceptual models, and technology skills. The divergence, once again, focuses on the collection or content type. The challenge is how to reconcile the way that individuals view their digital materials uniquely with the fact that the underlying practices and principles for the long-term care of all digital materials is the same.

Implications of Findings

This study has provided markers for gauging future progress of the area of digital curation. At present, digital curation has not emerged as an autonomous discipline, although there are observable indicators that the area is maturing. This study has outlined those indicators and delineated barriers for future progress. But this study does not necessarily chart a path for digital curation towards becoming an academic discipline: rather it guides the area to become a more cohesive and harmonized field of practice and research. Emerging academic disciplines require both institutionalization within higher education, and recognition of their autonomy and
legitimacy by other disciplines. The community must come to a consensus about the direction of digital curation.

Yet, a nascent digital curation-oriented perspective is now becoming visible. It is distinctive from the domains from which the area borrows and integrates. In fact, it emerges as not being within the boundaries of an identifiable discipline. The perspective assumes a holistic view of data that includes all forms and types of digital asset, regardless of content or context. The perspective is grounded in the lifecycle approach to digital information and is inclusive of access and use, preservation, and technology. The perspective advocates for competencies, knowledge, and skills related to digital curation to be embedded across disciplines in practice and research. This perspective has been perpetuated on the job, within the community, and is gradually being reproduced via the developing curriculum emerging in higher education.

The findings from this study confirm many of the commonly held assumptions about this area, namely that it is collaborative and engages multiple disciplines, and that the area of professional and scholarly interest is broad in scope. But the study also identifies lines of demarcation between types of collections and content that reinforces the parallel growth of similar areas, rather than encouraging shared solutions.

**Discussion of Limitations**

As discussed in Chapter One, there are four main limitations to this study. The limitations that have the greatest potential impact to this study are the purposive sampling method; the availability of material for the content analysis; the publishing trends of academics and practitioners; and language restrictions on the samples selection. In this section, I elaborate on the potential impact of each limitation.
The use of purposive sampling rather than a randomized sampling method is common in qualitative research designs. Non-probability sampling restricts generalization of the findings because the sample is not representative of the population. Purposive sampling was used in order to select a group of participants who would be able to speak to the overall objectives of this research. As such, the sample over-represented the fields of archival studies, and library and information science. Additionally, the interviewees came from within the digital curation or preservation community and therefore already possess some shared acceptance and understanding of concepts, roles, and terminology.

Additionally, the sample for the content analysis is limited by the materials available. Discourse in the area of digital curation includes community-based publications and grey literature. They are more difficult to locate via conventional searching methods and are not always accessible in citation databases. The availability of databases for searching and the selection of the sample, although deliberate and methodical, also apply limitations on scoping the literature and content analysis sample.

Publishing trends differ between scholars and practitioners in the LIS fields, and they differ between disciplines (Schlögl & Stock, 2008; Haddow & Klobas, 2004; Powell, Baker, & Mika, 2002). This difference is present in publications related to digital curation. These trends not only include frequency of publication and type of articles but also number of authors, citation practices, and how abstracts are structured. Because the area of digital curation engages multiple disciplines, these differences likely have a subtle impact on the content analysis findings.

Lastly, all interviews were conducted in English and samples for the content analysis were limited to articles or presentations written in English. Very few foreign language articles were retrieved in scoping the literature or the content analysis searches. The area of digital
curation is an international venture, and although publishing in English is common for European countries and China, this limitation at least minimally restricted the sample.

**Conclusion**

Perspective is important. From the perspective of archival studies, digital curation is about digital archiving. For library and information science, it is about digital information management and preservation. For the technology fields, it is about finding technical solutions. For the domain sciences, it is about research data management. However, all share a similar objective: to ensure the authenticity and longevity of digital assets that are of value to a designated community.

The phrase “digital curation” was originally coined as an umbrella term to represent the sameness of the underlying concepts and processes for the ongoing lifecycle care of all types of digital materials. The heterogeneity observed in the area of digital curation does not preclude it from being or becoming a discipline. All disciplines are in fact heterogeneous to some extent when considering the various specializations and shallow consensus. This heterogeneity encourages improved communication, healthy debate, and rigorous research, which in turn supports growth of the field intellectually, institutionally, and in size and consequence.

The interdisciplinary and multidisciplinary nature of digital curation will evolve into a more coordinated and cohesive field; however, I do not anticipate the area emerging as an autonomous academic discipline within the next decade. There has been considerable growth in educational and training opportunities in this area, namely curriculum development, continuing education and professional training opportunities, and research data management instruction. However, the community requires a single set of conceptual terms that is extensible and enables effective communication; an association or organization to advocate for the field, build social
networks, and unite the community; a better grasp on advancing implementation; and an understanding of the area’s development and scope across tangential areas. Or else we continue to reinforce silos that are no longer relevant today.

**Future Research Directions**

This study exposes several areas for future research, both to extend the present analysis and to explore new research directions suggested in the findings. In particular, to provide additional insights to the findings of this study I plan to conduct a social network analysis. This will involve a more intensive analysis of authorship trends and citations, and include a larger and more disciplinary diverse sample of community members. A better understanding of the collaborative networks in the area of digital curation will provide insight into the spread and use of terminology, the extent of interdisciplinary collaboration, and the extent to which different stakeholders share knowledge across disciplinary boundaries. Additionally, it will serve to map the area and provide a view of where there is convergence or divergence.

New directions for research suggested in the findings largely focus on improving current practice. Notably, there was a discernible gap in the data in activities and research related to digital curation in the business sector that I plan to examine more closely. An exception is the grant-funded European project Timeless Business Processes and Services (TUMBUS, 2011-2014), whose vision is “to bring digital preservation into the realm of Business Continuity Management” (Timeless Business Processes and Services, n.d.). I plan to conduct an initial exploratory study to get a better sense of needs and current practices in this arena, and determine whether there is indeed a gap in our current understanding and practice.
REFERENCES


Inter-university Consortium for Political and Social Research. (n.d). *ICPSR: The Founding and Early Years.* Retrieved from [http://www.icpsr.umich.edu/icpsrweb/content/membership/history/early-years.html](http://www.icpsr.umich.edu/icpsrweb/content/membership/history/early-years.html)


JISC. (n.d.). *History.* Retrieved from [http://jisc.ac.uk](http://jisc.ac.uk)


*Communication systems and resources in the behavioral sciences* (Publication 1575). 


Navale, V., & Cameron, R. (n.d.). *Four decades of preserving electronic records at NARA.* 


212


APPENDIX A: SCOPING THE LITERATURE

Search Terms

Below is the list of search terms used in scoping the literature. The bold items indicate the main search terms used in to identify the top 5% most frequently cited articles in Scopus and Web of Science. Bolded with @ indicate terms used to identify conference presentations.

<table>
<thead>
<tr>
<th>Search Term</th>
<th>Search Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>data archiving</td>
<td>data management and stewardship</td>
</tr>
<tr>
<td>data curation</td>
<td>digital document*</td>
</tr>
<tr>
<td>data preservation</td>
<td>digital document* and archiv*</td>
</tr>
<tr>
<td>data stewardship</td>
<td>digital document* and curation</td>
</tr>
<tr>
<td>digital archiving</td>
<td>digital document* and preservation</td>
</tr>
<tr>
<td>digital curation @</td>
<td>digital document* and stewardship</td>
</tr>
<tr>
<td>digital preservation</td>
<td>digital humanities</td>
</tr>
<tr>
<td>digital stewardship @</td>
<td>digital humanities and archiv*</td>
</tr>
<tr>
<td>archival repositor* and data</td>
<td>digital humanities and curation</td>
</tr>
<tr>
<td>archival repositor* and digital</td>
<td>digital humanities and preservation</td>
</tr>
<tr>
<td>big data</td>
<td>digital humanities and stewardship</td>
</tr>
<tr>
<td>big data and archiv*</td>
<td>digital librar*</td>
</tr>
<tr>
<td>big data and curation</td>
<td>digital librar* and archiv*</td>
</tr>
<tr>
<td>big data and preservation</td>
<td>digital librar* and curation</td>
</tr>
<tr>
<td>big data and stewardship</td>
<td>digital librar* and preservation</td>
</tr>
<tr>
<td>data archiv* or digital archiv*</td>
<td>digital librar* and stewardship</td>
</tr>
<tr>
<td>data asset management or digital asset management</td>
<td>digital media</td>
</tr>
<tr>
<td>data management</td>
<td>digital media and archiv*</td>
</tr>
<tr>
<td>data management and archiv*</td>
<td>digital media and curation</td>
</tr>
<tr>
<td>data management and curation</td>
<td>digital media and preservation</td>
</tr>
<tr>
<td>data management and preservation</td>
<td>digital media and stewardship</td>
</tr>
<tr>
<td></td>
<td>digital museum*</td>
</tr>
</tbody>
</table>
digital museum and archiv*
digital museum and curation
digital museum and preservation
digital museum and stewardship
digital record or digital records
digital record* and archiv*
digital record* and curation
digital record* and preservation
digital record* and stewardship
data repositor* or digital repositor*
digitization
digitization and archiv*
digitization and curation
digitization and preservation
digitization and stewardship
electronic document*
electronic document and archiv*
electronic document and curation
electronic document and preservation
electronic document and stewardship
electronic media
electronic media and archiv*
electronic media and curation
electronic media and preservation
electronic media and stewardship
electronic record*
electronic record* and archiv*
electronic record* and curation
electronic record* and preservation
electronic record* and stewardship
e-science
e-science and archiv*
e-science and curation
e-science and preservation
e-science and stewardship
machine-readable
machine-readable and archiv*
machine-readable and curation
machine-readable and preservation
machine-readable and stewardship
online exhibit*
online exhibition and archiv*
online exhibition and curation
online exhibition and preservation
online exhibition and stewardship
virtual museum*
virtual museum and archiv*
virtual museum and curation
virtual museum and preservation
virtual museum and stewardship
web archiv*
web archiv* and archiv*
web archiv* and curation
web archiv* and preservation
web archiv* and stewardship
Databases and Search Fields

Below is the list of databases and search fields used in scoping the literature. Databases were accessed through Simmons College Library, except Sage Publications and Scopus which were accessed through my personal account.

Table A.1

*Scoping the literature summary: Article databases used in scoping the literature*

<table>
<thead>
<tr>
<th>Database</th>
<th>Search Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directory of Open Access Journals</td>
<td>Advanced Search: All Fields</td>
</tr>
<tr>
<td><strong>EBSCOhost</strong>: Academic Search Complete; America: History &amp; Life; Art Full Text (H.W. Wilson); Audiobook Collection (EBSCOhost); Biography Reference Bank (H.W. Wilson); Book Review Digest Plus (H.W. Wilson); Business Source Complete; Children's Core Collection (H.W. Wilson); CINAHL Complete; Communication &amp; Mass Media Complete; eBook Collection (EBSCOhost); Environment Complete; ERIC; Fiction Core Collection (H.W. Wilson); GreenFILE; Health and Psychosocial Instruments; Historical Abstracts; Hobbies &amp; Crafts Reference Center; Library &amp; Information Science Source; Library Literature &amp; Information Science Retrospective: 1905-1983 (H.W. Wilson); Library, Information Science &amp; Technology Abstracts; MEDLINE; Middle and Junior High Core Collection (H.W. Wilson); MLA Directory of Periodicals; MLA International Bibliography; PsycARTICLES; PsycINFO; PsycTESTS; Public Library Core Collection: Nonfiction (H.W. Wilson); Sears List of Subject Headings (H.W. Wilson); Senior High Core Collection (H.W. Wilson); Social Work Reference Center; SocINDEX with Full Text</td>
<td>AB Abstract [or] TI Title [or] SU Subject Terms [or] SO Source [or] AU Author</td>
</tr>
<tr>
<td>Emerald Group Publishing</td>
<td>All except full text</td>
</tr>
<tr>
<td>JSTOR</td>
<td>Author [or] Item Title [or] Abstract [or] Caption</td>
</tr>
<tr>
<td><strong>ProQuest</strong>: ProQuest: ABI/INFORM Complete; Boston Globe; ebrary ebooks; Library and Information Science Abstracts (LISA); ProQuest Dissertation; ProQuest Historical Newspapers</td>
<td>Ab Abstract [or] TI Document Title [or] SU Subject Heading [or] PUB Publication Title [or] AU Author [or] TAG Tag</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>Sage Publications</strong></td>
<td>Title [or] Author [or] Journal Name [or] Keywords [or] Abstract [or] Affiliation</td>
</tr>
<tr>
<td><strong>Scopus</strong></td>
<td>Article Title, Abstract, Keywords, Authors [or] Source Title [or] Affiliation</td>
</tr>
<tr>
<td><strong>Web of Science</strong>: Web of Science: Science Citation Index Expanded (1988-present); Social Sciences Citation Index (1988-present); Arts &amp; Humanities Citation Index (1988-present); Conference Proceedings Citation Index- Science (1990-present); Conference Proceedings Citation Index- Social Science &amp; Humanities (1990-present); and Medline Plus</td>
<td>Title [or] Author [or] Publication Name [or] Topic</td>
</tr>
</tbody>
</table>
APPENDIX B: INTERVIEW GUIDE

1. Can you tell me a little about your current role?
2. Can you describe how you became interested/involved in the preservation of digital materials?
3. What term do you use when you are discussing this area?
   a. How do you define this term and how is it different or similar to other terms that people use?
4. What do you consider to be the key concepts or principles and practices that drive this area? Do you think these are relevant to analog materials – if yes in what ways?
5. What do you consider to be the salient models, perspectives, or theories that influence this area?
6. Would you consider the area practice-driven, theory-driven, or something else?
7. Do you think this area fits within the information disciplines? Why or why not?
8. Tell me about up to three publications in this area that you regard as significant.
9. Broadly speaking and based on your experience, what types of knowledge and skills do you think are important for professionals in this area to possess?
10. What are your thoughts about the educational opportunities available in this area?
    a. How do you feel they serve the needs of the area? How could they better serve the area?
11. In what ways do you envision the educational landscape of this area maturing? In particular, do you see the education of digital curation knowledge converging into a comprehensive field or specialization within the information disciplines or fragmenting?
12. How do you engage, communicate, and stay up-to-date with the community?
13. Is there anything that you would like to add?
14. Is there anything that we had discussed that you would like to add to or follow-up on?
APPENDIX C: CODING SCHEMES

Fixed Codes for Content Analysis

This codebook defines coding units for recording information about each sampling unit (article, conference presentation or poster, editorial, etc.) for the content analysis. The keyword for each variable is highlighted in bold. These pre-set codes were recorded using a standardized coding form created in Microsoft Excel.

1. **Date** coded [yymmdd (e.g., 120110 for January 10, 2012)]

2. Variables related to the **Source Publication or Conference** in which the item is located
   a. Type of **source** (Journal, Conference, Book, etc)
   b. **Discipline** of publication/conference
   c. Title of **publication/conference** from which item is drawn
   d. **Year** of publication (e.g. 1999)
   e. **Volume (Number)** of publication, if applicable

3. Variables related to the **Article**
   a. **Title** of the item
   b. What **Type of item** is this?
      i. Article
      ii. Commentary/Opinion
      iii. Conference poster
      iv. Conference presentation
      v. Editorial
      vi. Peer-reviewed article
      vii. Review Article
      viii. Other (explain in Comments)
   c. **Abstract**

4. Variables related to the **Author(s)** of the Article
   a. **First** Author
      i. **Last, First Name** of first author
      ii. Organizational **affiliation** (e.g. Simmons College)
      iii. Location: **State** (State, Province, where the author works)
      iv. Location: **Country**
   b. **Second** Author (enter zero if no second author)
      i. **Last, First Name** of first author
      ii. Organizational **affiliation** (e.g. Simmons College)
      iii. Location: **State** (State, Province, where the author works)
      iv. Location: **Country**
c. Third Author (enter zero if no third author)
   i. Last, First Name of first author
   ii. Organizational affiliation (e.g. Simmons College)
   iii. Location: State (State, Province, where the author works)
   iv. Location: Country

d. Additional Authors in listed in Author Comment sections

e. Total Number of Authors

5. Variables related to the Topic and Subject Area of the Article
   a. Focus (primary) Subject Area of Article
   b. Locus (secondary) Topic Area of Article
   c. Primary academic Field of article topic

6. Methodology Comments
   a. Open option coding for type of research, research design, methodology, and data collection techniques.

7. Comments
   a. Comments section for notes about any issues, questions, challenges, observations, deviations from standard coding, etc.
Final Emergent Codes from Interview Transcripts and Content Analysis

This codebook defines the final 35 emergent codes derived from the data. NVivo was used for this iterative portion of the coding process. Complete interview transcripts and abstracts of each content analysis sampling unit were coded. These finalized coding categories were mapped to the elements that constitute a discipline (a priori codes) from D’Agostino (2012).

**EC-Archival Studies**
Comments and statements related to archival studies. Suggests a connection or relationship between archival studies and activities related to digital curation. Examples of subcodes: archival theory; archival thinking.

**EC-Best Practices**
Any description or mention of current best practices. Examples of subcodes: duplicate copies; early intervention; workflows.

**EC-Challenges**
Comments and reflections about the challenges of digital preservation. Examples of subcodes: benefits/value; changes; obsolescence; risk; security; sustainability (organizational).

**EC-Collaboration**
Comments and reflections about collaboration and the value of collaboration. Examples of subcodes: collaborative approaches; community; shared solutions.

**EC-Cyberscholarship**
Instances of new ways of communicating about, conducting, and disseminating research in the digital environment. Examples of subcodes: big data/science; digital humanities; eResearch; open access; scholarly communication.

**EC-Data Management**
Specific to the management, publication, and sharing of research data. Examples of subcodes: annotation; curated database; data citation; archiving/deposit of data; funder’s policies; peer review of data; replication.

**EC-Development of Area**
Comments and reflections about the area, the history of the area, and how the area is evolving. Examples of subcodes: becoming a discipline; call-to-action; early efforts; what drives the area.

**EC-Digital Object**
Explanations, descriptions, and comments about concepts related to the digital object (generic). Examples of subcodes: authenticity; provenance; significant properties.

**EC-Domains**
Any mention or anything related to a specific field. Suggests a connection or relationship between domain sciences and activities related to digital curation. Examples of subcodes: archaeology; biomedical; earth sciences; humanities.

**EC-Economics**
Statements about economic or financial matters. Examples of subcodes: costs; funding for projects; sustainability (financial).
EC-Education and Training
Identifies competencies and skills related to digital curation. Describes education or training programs. Examples of subcodes: on the job training; educational program development; skills; training opportunities.

EC-Employment
Comments and statements about jobs in areas related to digital curation. Examples of subcodes: careers; new profession; roles and responsibilities; self-defined roles.

EC- Formats and Types
Anything related to specific digital formats or media types. Examples of subcodes: 3D models; blogs; GIS; social media; video games; web archiving.

EC-Function
Explanations, descriptions, and comments about digital archiving functions or lifecycle actions. Examples of subcodes: access; appraisal and selection; ingest; storage. These subcodes contain child codes.

EC-Information Management
Statements about the management of digital information in different contexts. Examples of subcodes: digital asset management; personal information management; product lifecycle management.

EC-Infrastructure
Explanations, descriptions, and comments about establishing the infrastructure to support digital curation, preservation, and research. Examples of subcodes: cyberinfrastructure; institutional framework; policy framework; technological infrastructure.

EC-Initiatives and Projects
Explanations, descriptions, and comments about collaborative initiatives and projects for advancing digital curation and related areas. Examples of subcodes: DCC; NDIIPP.

EC-Library and Information Science
Comments and statements related to LIS. Suggests a connection or relationship between LIS and activities related to digital curation. Examples of subcodes: digital library.

EC-Legal and Ethical Issues
Explanations, descriptions, and comments about legal issues or ethical considerations. Examples of subcodes: data confidentiality; data fraud; freedom of information act; intellectual property rights.

EC-Local Digital Programs
Descriptions and comments about digital programs at the local level. Examples of subcodes: exemplar project; moving to digital; organizational readiness.

EC-Metadata
Any description or mention of metadata, or related concepts and tasks. Examples of subcodes: D&RI; markup language; ontologies.

EC-Models
Explanations, descriptions, and comments about models that have been developed for digital curation and related areas. Examples of subcodes: digital curation models; preservation models.
EC-Museum Studies
Comments and statements related to museum studies. Suggests a connection or relationship between museum studies, LIS, and/or activities related to digital curation. Examples of subcodes: convergence; preservation of digital cultural heritage.

EC-Multiple Discipline Engagement
Comments and statements about multiple fields working together. Examples of subcodes: cross-disciplinary grand challenges; disciplinary differences; location among disciplines; multiple perspectives; relationship between areas.

EC-Preservation
Explanations, descriptions, and comments about preservation actions, services, or technology. Examples of subcodes: distributed digital preservation; microservices; preservation strategies; conceptually the same as analog; technological challenges.

EC-Related Areas
Any mention of tangential areas. Suggests a connection or relationship between activities/concepts in these areas and digital curation activities/concepts. Examples of subcodes: biocuration; digital forensics; informatics; records management.

EC-Repositories
Comments and statements about repositories and their role. Examples of subcodes: context/collections; data archive; eprints; institutional repository; interoperability.

EC-Research Areas/Activities
Suggests a connection or relationship between research areas/activities and the aims of digital curation. Examples of subcodes: data mining; information seeking behavior.

EC-Stakeholders
Statements and descriptions about categories of stakeholders. Examples of subcodes: audiences for dissemination; data creators; publishers; users.

EC-Standards
References or descriptions of standards and their function. Includes ISO standards. Examples of subcodes: CIDOC CRM; DOI; OAIS; TDR.

EC-Sociology of Digital Environment
Explanations, descriptions, and comments about social behavior and the digital environment. Examples of subcodes: incentivization; social aspects of data sharing.

EC-Technology
Descriptions of technologies and tools for activities related to digital curation. Comments and reflections about technology. Examples of subcodes: focus on technology solutions; software/systems; tools.

EC-Technology Fields
Comments and statements to the technology fields. Suggests a connection or relationship between the technology fields and activities related to digital curation. Examples of subcodes: information technology perspective.

EC-Terminology
Identification of the area by name; terminology used to describe the area. Definitions of field-specific terminology. Examples of subcodes: field of...; referred to area as...; definition.

EC-Types of Content/Collections
Reference to specific kinds of digital content or collections. Examples of subcodes: digital art; personal digital archives; research data; scientific workflows.
APPENDIX D: CONTENT ANALYSIS BIBLIOGRAPHIES

Conferences Presentations


Garrett, L., Gramstadt, M. T., & Silva, C. (2013). Here, KAPTUR this! Identifying and selecting the infrastructure required to support the curation and preservation of visual arts research data. *International Journal of Digital Curation, 8*(2), 68-88.


Top 5% Most Cited (Scopus and Web of Science)


Parry, O., & Mauthner, N. S. (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology, 38*(1), 139-152.


Wicherts, J. M., & Bakker, M. (2012). Publish (your data) or (let the data) perish! Why not publish your data too? *Intelligence, 40*(2), 73-76.


