Digital Humanities:
Envisioning a Collaborative Tool for Mapping, Evaluating, and Sharing Reconstructed Colonial American Parcel Maps

Mary Brent Ruvane

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

2011

Approved by:
Dr. Deborah Barreau
Dr. G. Rebecca Dobbs
Dr. Ian Gregory
Dr. Richard Marciano
Dr. Paul Solomon
ABSTRACT

MARY BRENT RUVANE: Digital Humanities: Envisioning a Collaborative Tool for Mapping, Evaluating, and Sharing Reconstructed Colonial American Parcel Maps (Under the direction of Paul Solomon and Deborah Barreau)

The use of GIS technology for the humanities has opened up new avenues for visually exploring and asking questions of our nation’s historical record. The potential to harness new knowledge with tools designed to capture and preserve geographic links to the artifacts of our past is within our grasp. This research explores the common information needs of a community of interest to achieve their goal of reconstructing the spatial circumstance of America’s Colonial era and the information barriers they encounter. It envisions a suite of digital tools to help confidently join together the solitary efforts of dissimilar investigators to facilitate sharing, debate and long term preservation of their painstaking research.

The study described in this work examined the common information use behaviors of researchers whose goal is to reconstruct the missing geographic picture of British Colonial settlement in America. The scope of the analysis focuses on the work of two investigators who had constructed historic neighborhood maps of North Carolina’s early backcountry settlements. Using a domain analytic approach it examines the resources and different tools employed by each investigator in comparison to the cataloged data provided by the North Carolina State Archives’ online Manuscript and Archives Reference System (MARS). The results point to numerous barriers to information that could be overcome with the aid of standardized tools and a shared cyberinfrastructure. Envisioned is a suite of applications to enable a collaborative and authoritative reconstruction, over time, of the missing geographic picture of colonial America.
In memory of Donna Lynn Gilbert

January 28, 1960 – December 13, 2010
ACKNOWLEDGEMENTS

The pages of this dissertation hold far more than the culmination of years of study. They reflect the relationships with many generous and inspiring people I have met and worked with since beginning my graduate work. I am especially grateful to my long time advisor and co-committee chair, Paul Solomon, for his patience and guidance during this oftentimes interrupted endeavor. I have never met a kinder or more dedicated champion.

The members of my dissertation committee, Deborah Barreau (co-chair), G. Rebecca Dobbs, Ian Gregory, and Richard Marciano, have generously given their time and expertise to better my work. I thank them for their contributions and unwavering support.

To G. Rebecca Dobbs a special thanks for sharing her research and trusting in my abilities to support her research. A braver person I could not imagine. She is a dedicated scholar, inspired educator, extraordinary historical geographer, and a true friend.

To my mother, Joan S. Ruvane, there are no words to explain how much her editing efforts meant to me. My sincere thanks for getting me through this last step in the journey.
# TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................... vii
LIST OF FIGURES ................................................................................................................................ viii

Chapter

I. HUMANITIES AND COMPUTING .......................................................................................... 1
   Definitions................................................................................................................................... 1
   Research in digital humanities ................................................................................................. 3
   A community need: Colonial parcel maps ........................................................................... 12

II. LITERATURE REVIEW ............................................................................................................ 27
   Human information behavior ................................................................................................. 27
   The historical circumstance ................................................................................................. 39
   Geographic knowledge: Maps for historical research ........................................................ 56

III. METHODOLOGY & RESEARCH QUESTION ...................................................................... 62
   Objective .............................................................................................................................. 62
   Work centered research: Domain analytic approaches ........................................................ 62
   Research questions & methods ........................................................................................... 67

IV. RESULTS ................................................................................................................................. 75
   The Domain of Resources ................................................................................................. 75
   The Domain of Evidence .................................................................................................... 93
   The Domain of Uncertain Evidence ............................................................................... 107
   The Domain of Tasks ....................................................................................................... 113

V. DISCUSSION ............................................................................................................................. 118
   Summary ............................................................................................................................ 118

APPENDICES .................................................................................................................................... 123
REFERENCES .................................................................................................................................... 127
# LIST OF TABLES

Table

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Example of the hierarchal numbering scheme employed in MARS</td>
<td>81</td>
</tr>
<tr>
<td>4.2</td>
<td>Four Record Series considered for inclusion in the Domain of Resources</td>
<td>85</td>
</tr>
<tr>
<td>4.3</td>
<td>Example of content and format of consolidated inventory</td>
<td>86</td>
</tr>
<tr>
<td>4.4</td>
<td>MARS Record Series 12.12</td>
<td>88</td>
</tr>
<tr>
<td>4.5</td>
<td>MARS Record Series 12.13</td>
<td>89</td>
</tr>
<tr>
<td>4.6</td>
<td>MARS Record Series 12.14</td>
<td>91</td>
</tr>
<tr>
<td>4.7</td>
<td>The Domain of Resources</td>
<td>93</td>
</tr>
<tr>
<td>4.8</td>
<td>Deed references in MARS</td>
<td>110</td>
</tr>
<tr>
<td>4.9</td>
<td>Deed references in Dobbs’ database</td>
<td>111</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure

1.1 Facsimile of NW portion of Sharpe’s 1773 map ......................................................... 16
1.2 Historic parcel map of early settlement in Rowan County, NC ................................. 17
1.3 Historic settlements in Orange County, NC ................................................................. 18
1.4 Portion of historic Orange County located in present day Guilford County, NC ......... 19
1.5 Portion of McNeely map. Historic Burke and Lincoln counties, North Carolina ....... 20
1.6 Portion of a map created by a genealogist/family historian ........................................ 21
2.1 Variations in early land division practices .................................................................... 42
2.2 1755 land survey, Granville District, Orange County, NC ......................................... 43
2.3 The boundaries of the land granted the [eight] Lords Proprietors ............................... 46
2.4 The Granville District included nearly the northern half of North Carolina ............... 47
2.5 Dobbs’ overall study area covering the Piedmont region of NC ................................. 48
2.6 Dobbs’ Phase I (Dissertation) Study Area ................................................................ 49
2.7 Transcribed Granville District land record evidence by type from 1748-1763 ............. 52
2.8 Thematic map depicting parcel density ..................................................................... 59
3.1 Five stages in a common domain analysis process ..................................................... 66
4.1 Search screen options available in MARS: Basic, Advanced, and by Call Number ...... 77
4.2 Browse button displays a list of record types and collection categories .................... 78
4.3 View of the State Records collections ...................................................................... 79
4.4 Granville Proprietary Office: Land Entries, Warrants, Plats of survey Record Group ... 79
4.5 Item level description of a folder within the Anson County, A-E Record Series .......... 80
4.6 Search by MARS ID 12.12 ......................................................................................... 82
4.7 Finding aid for MARS ID 12.12: Granville Proprietary Land Office ......................... 82
4.8 Child Records of MARS ID 12.12 ............................................................................ 83
4.9 Child Records of MARS ID 12.12.1 ......................................................................... 83
4.10 MARS ID 12.12.1.1: Example of folder level details .............................................. 84
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.11</td>
<td>Series 12.14 finding aid variation</td>
<td>87</td>
</tr>
<tr>
<td>4.12</td>
<td>Dobbs’ Microsoft Access Database: Main data entry form</td>
<td>94</td>
</tr>
<tr>
<td>4.13</td>
<td>McNeely’s DeedMapper® Database: Main data entry form</td>
<td>95</td>
</tr>
<tr>
<td>4.14</td>
<td>Dobbs Database: evidence extracted from 97 sampled deed records</td>
<td>98</td>
</tr>
<tr>
<td>4.15</td>
<td>McNeely’s Database: evidence extracted from 100 sampled deed records</td>
<td>100</td>
</tr>
<tr>
<td>4.16</td>
<td>MARS catalog equivalent of Dobbs’ 97 sampled deed records</td>
<td>103</td>
</tr>
<tr>
<td>4.17</td>
<td>MARS catalog equivalent of McNeely’s 100 sampled deed records</td>
<td>106</td>
</tr>
</tbody>
</table>
CHAPTER 1: HUMANITIES AND COMPUTING

The humanities are the stories, the ideas, and the words that help us make sense of our lives and our world. The humanities introduce us to people we have never met, places we have never visited, and ideas that may have never crossed our minds. By showing how others have lived and thought about life, the humanities help us decide what is important in our own lives and what we can do to make them better. By connecting us with other people, they point the way to answers about what is right or wrong, or what is true to our heritage and our history. The humanities help us address the challenges we face together in our families, our communities, and as a nation.

-- Ohio Humanities Council

Definitions

In the National Foundation of the Arts and the Humanities Act of 1965 Congress defined humanities as including, “but...not limited to, the study of...language, both modern and classical; linguistics; literature; history; jurisprudence; philosophy; archaeology; comparative religion; ethics; [and] the history, criticism and theory of the arts” (NEH, 2010, p. 2). The Act further defines humanities to include “those aspects of the social sciences which have humanistic content and employ humanistic methods; and the study and application of the humanities to the human environment with particular attention to reflecting our diverse heritage, traditions, and history” by examining conditions of the past in relation to today’s circumstance--aspects of research equally germane in the study of human historical geography. Within the institutional halls of each of these disciplines individual methods for approaching problems and understanding information have evolved.

Since the advent of the computer, adapting its technology to support various humanities research problems has been an important focus of study (Jessop, 2004a; McCarty, 1998). Over time these studies have become associated with the field of humanities computing, and more recently under the new moniker of digital humanities (Svensson, 2009); a term intended to be more encompassing. At present both terms appear to be relatively interchangeable. Digital humanities continues to be a young but growing discipline, and in the past few years it has started to garner broader recognition and increased support for cyberinfrastructure (Tooby, 2007a,
Two definitions cited by Borgman (Frischer, 2009; Digital Humanities Manifesto, 2009) describe “digital humanities” as, firstly, “the application of information technology as an aid to fulfill the humanities’ basic tasks of preserving, reconstructing, transmitting, and interpreting the human record” and, secondly, “not a unified field but an array of convergent practices that explore a universe in which print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated.” Borgman blends these definitions into one, defining digital humanities as “a new set of practices, using new sets of technologies, to address research problems of the discipline.”

**Text, visualization, and computer technology**

The application of computer technology to humanities studies is not new. In the late 1940s a Jesuit scholar, Father Roberto Busa, first applied IBM computing technology to humanities scholarship. His work involved creating “the Index Thomisticus, a concordance with critical edition to the works of St. Thomas Aquinas” (McCarty, 1998). From the start the value of computers for organizing, analyzing, and codifying large volumes of text, such as the works of St. Thomas Aquinas (Busa, 1980), has helped to advance humanities research and enabled more in-depth studies into humanity’s cultural heritage.

While certain humanities disciplines were early adopters of computer technology, especially those dealing with large volumes of text such as in literature and linguistics (Svensson, 2009), the use of complementary visualization techniques is also a growing trend across disciplines (Friedlander, 2009; Frischer et al., 2006; Jessop, 2004b). As pointed out by CLIR (2009), citing Fisch (2007), “[t]here is substantial evidence that the next generation will be graphical learners and communicators” with the implication that “visualization will become increasingly important as a means of analysis as well as a mode of presentation and communication.” Essential forms of visualization are maps, which today are primarily constructed

---

1 The first definition is attributed to Frischer (2009, 15); the second attributed to the Digital Humanities Manifesto (2009).
with the aid of geographic information systems (GISs). Maps allow investigators to examine spatial patterns and analyze inconsistencies in the data they investigate.

The use of GIS technology for the humanities has opened up new avenues for visually exploring and asking questions of our nation’s historical record. The potential to harness new knowledge with tools designed to capture and preserve geographic links to the artifacts of our past is within our grasp. Online initiatives have been experimenting with the possibilities for years, yet few have addressed the complex issue of uncertain, fuzzy, incomplete, or ambiguous data and most fail to provide suitable distribution tools (i.e., to facilitate alternative methods for viewing, downloading, transforming, and printing of selected results), which limits their value as an aid for research. Further complicating matters are the growing number which, developed to facilitate educational research, lack essential search mechanisms to enable concurrent searching; for instance, limited to a particular place or feature, time, people, and/or event -- search features often cited as essential for facilitating humanities studies (Duff & Johnson, 2002; Tibbo, 1994).

**Research in digital humanities**

Humanities computing is committed to exploring both “information technology as a tool and written texts as a primary object of study” (Svensson, 2009). Two dominant areas of research are studies that investigate issues related to cyberinfrastructure and the tools necessary for facilitating humanities studies. Investigations into these concerns are often interrelated, particularly when the tools required and a community’s information needs necessitate high-powered computing (HPC) across a distributed heterogeneous network.

**Cyberinfrastructure**

Computer technology is rapidly changing the way scientific research and education is being carried out. A report by the National Science Foundation (NSF) states that “today’s scientists and engineers need access to new information technology capabilities, such as distributed wired and wireless observing network complexes², and sophisticated simulation tools until recently…

² Wireless sensor networks (WSNs) have been gaining popularity in a wide range of military and civilian applications such as environmental monitoring, health care, and control. A typical WSN consists of a number of geographically distributed sensors and a fusion center (FC) (Yi, S. & Schober, R., 2008). Further, as mobility demands increase
that permit exploration of phenomena that can never be observed or replicated by experiment” (NSF Cyberinfrastructure Council, 2007, p. 5). Cyberinfrastructure, or e-Science as it is more commonly known in the UK, joins together distributed computing systems, data storage systems, archival repositories, computational and visualization tools, and researchers who utilize software applications and high performance networks to enable discoveries never before imaginable, such as the Human Genome Project (Human Genome Project Information, 2010). The goal of such coordinated infrastructures is to facilitate collaboration in the ‘creation, dissemination, and preservation of scientific data, information and knowledge’ (NSF Cyberinfrastructure Council, 2007).

High performance computing (HPC) allows large data sets to be explored and analyzed in fresh new ways. Unfortunately, while “[i]nformation - and data-intensive, distributed, collaborative, and multi-disciplinary research is now the norm in the sciences, [it remains] experimental in the humanities”, even with the “many investments and years of development, basic infrastructure for the digital humanities is still lacking” (Borgman, 2009). Fortunately, headway is being made in this area according to a report cosponsored by the Council on Library and Information Resources and The National Endowment for the Humanities (CLIR, 2009), which points to the growth in funded initiatives and calls for humanities scholars to better define their vision and digital information needs for the 21st century in order to reap the full benefits of distributed technology. A “shared infrastructure of tools, services, and collections...would reduce unnecessary redundancy, allocate human and information resources efficiently, and, most interestingly, enable a different kind of scholarship” (CLIR, 2009). This sentiment is echoed by leaders in the field, including Besser (2004), Thomas (2004) and others (Schreibman, Siemens, & Unsworth, 2004; Unsworth, 2009), many who have championed the need for greater communication between information scientists, humanities scholars, libraries, and archives to better articulate and synchronize computing needs in the humanities.

more computers are "interact[ing] with more kinds of networks. Paths combining wired links and wireless links are likely to be as common tomorrow as cellular telephony for audio conversation is today" (Yarvis, M., Wang, A., Rudenko, A., Reiher, P., & Popek, G., 1999).
While progress is being made through sponsored workshops (CLIR, 2009; Cohen, Fraistat, Kirschenbaum, & Scheinfeldt, 2009; Frischer et al., 2006) and initiatives tasked with developing cyberinfrastructure for the humanities (Elliott & Gillies, 2009; Tooby, 2007a; Unsworth, 2009), Amy Friedlander (2009) explains that more research is needed to understand the common issues of this diverse community and its long-term information goals. She describes four overlapping areas of research that call for further study to identify obstacles and opportunities in order to advance digital humanities scholarship. These include scale, language and communication, space and time, and social networking, concerns shared across humanities disciplines. It is important to note that “the boundaries between [these research areas] are indistinct, and techniques that are developed in one may apply to problems in another.”

The first key area in need of more research as it relates to humanities computing is that of scale, which involves the challenge of providing “access to heterogeneous digital information of varying quality and in quantities that were unimaginable in prior generations.” How do you deal with burgeoning collections as new material (e.g., images, video, audio) is added and computationally complex cross-language requirements transpire? Additionally, while some information may “have been formally ingested into well-managed archives; other sources will have been captured on the fly and deposited into repositories with minimal attention”, thereby necessitating automated methods to authenticate various information resources. This is especially important in collaborative environments where “malicious or untrustworthy” contributions from users could become a problem (Ruth, Xu, Bhargava, & Regnierz, 2010). The issue of scale speaks to the need for understanding the size, scope, and future demands that an infrastructure might be pressed to support, suggesting that the domain (Talja, 2005) of a particular community, and its information needs, should be thoroughly understood.

The second key research area surrounds language and communication issues, which often rely upon both linguistic and geospatial techniques. Freidlander (2009) argues that “[l]anguage is central to much of humanities scholarship” and, citing Duguid (2007) in regards to quality, that “mass digitization projects are yielding extremely large digital corpora that are both problematic from the perspective of quality…and fascinating from the perspective of their
content”. Unfortunately, the ability to search across large heterogeneous digital collections is frequently impeded by a lack of overall quality, which includes but is not limited to barriers such as incomplete digitized collections, poor reproductions (for example, low resolution or cropped images), inadequate retrieval and image viewing tools, partial transcriptions that omit key elements of value to ongoing research, and transcriptions (or index terms) that fail to characterize uncertainties inherently found in historical data. These impediments serve to further exacerbate investigations that rely on geographic visualization (i.e., GIS) for analyzing and communicating their findings, considering that uncertain data carried over into the reconstructed geography of an historical circumstance introduces yet another layer of doubt in terms of spatial precision (Mueller, 2008).

Many humanities investigations fail to take into account the considerable amount of historical evidence that contains information of a geographic nature, which if parsed properly could enable in-depth spatial analysis, including one-of-a-kind manuscript material and government records collections. Investigations dealing with government records are particularly well-matched for employing cyberinfrastructure because such collections tend to be large, held by multiple institutions, and are of a fairly uniform nature. One obvious example is census records, a popular source material that has been the focus of numerous spatially focused initiatives (Fitch & Ruggles, 2003; Ruggles, 2003), but effectively using this growing body of digital material across distributed networks is tricky (Terras, 2009), especially when surrogates and transcriptions are imperfect (or inaccessible) and thereby difficult to validate. The issues surrounding language and communication suggest the need to better understand terminologically related barriers (i.e., indexing/metadata) in order to improve access and reuse of such frequently referenced baseline historical material.

The third area of investigation into the needs of humanities computing relates to space and time, which typically entails the use of visualization tools to facilitate temporal comparisons involving the use of new research methods made possible with GIS technology (Borgman, 2009). While “geographers …have made considerable headway with space … time is still a problem” (Friedlander, 2009). Despite this deficiency GIS permits researchers to reconstruct the historical
circumstance under study and, as a byproduct, considerable amounts of new knowledge connecting historical evidence to its geographic context and time period is generated. These new links enable historical collections for the first time to be organized by spatial and chronological characteristics; unfortunately, this new evidence is rarely recognized for its potential for incorporation into an institution's retrieval system. Geographic and time related search terms are highly desirable (Duff & Johnson, 2002; Tibbo, 1994) yet are typically absent from most archival and library catalog systems as well as from many of the emerging digital humanities initiatives. Yet, there have been some intriguing solutions proposed (Buckland, Gey, & Larson, 2004, 2007).

Friedlander cites two projects that from the onset set out to organize their collections based on space and time, the Our Americas Archive Project (OAAP) and the Persepolis Fortification Archive Project\(^3\), but neither appears to offer adequate search features to take full advantage of these new entry points. Instead, they both appear to rely upon a subject access model that is known to be problematic and oftentimes limiting (Tibbo, 1994; Wiberley, 1983). Because humanities investigations often need to analyze change over time, such as place names, political boundaries, and shifts in language terminology, along with other variables, subject access is rarely sufficient. Incorporating more precise elements, related to space and time as well as appropriate thesauri and gazetteers, appears vital for reducing these persistent barriers to humanities information and its reuse. This suggests the need to better understand how heretofore unknown spatial and time-related evidence, collected and documented during the course of an investigation, could be used to improve humanities computing initiatives.

The fourth area of common interest to humanities research is social networking. Friedlander points to the growing “proliferation of social networks, such as Facebook, Myspace, Linkedin, and H-Net, that connect very large and geographically extended social groups…that fosters the exchange of information on myriad topics”. Additionally, she sees the potential of “social networking algorithms …[as] a set of analytics that could be used to characterize text

---

\(^3\) The project that hosts this initiative, the Online Cultural Heritage Research Environment (OCHRE), is very interesting. It is an Internet database system for cultural heritage information. It is intended for researchers engaged in arifactual and textual studies of various kinds. It is especially suitable for(1) organizing and publishing the results of archaeological excavations and surveys and (2) preparing and disseminating philological (relating to the scientific analysis of written records and literary texts) text editions and dictionaries. See http://ochre.lib.uchicago.edu/index.htm
corpora, [thereby] enabling researchers to identify patterns and detect anomalies more generally”.
I believe what may be even more important to humanities computing initiatives are social networking technologies that harness user contributions, such as blogs, wikis, and social tagging sites (e.g., delicious.com, flickr.com) because of their potential for collaboratively improving access to information and facility to incorporate new knowledge. Perhaps with the aid of controlled vocabularies these tools can be improved or, better yet, incorporated into a hybrid system to supplement established retrieval mechanisms (Lemieux, 2009; Tennis, 2006).

Audience

To succeed, the cyberinfrastructure and tools called for in the humanities cannot be built in an ivory tower, designed only to fulfill a handful of scholars’ research and information needs. Unsworth (2009) believes “the most important audience for the humanities…is the general public”. Yet humanities scholars find it difficult to understand “why the work they do should matter” to the average citizen. Unlike the hard sciences, where investigations and distributed tools tend to facilitate critical but narrowly focused research efforts -- often beyond the comprehension of most lay people -- humanities studies do have a general audience. There is a "relatively large but widely dispersed and non-professional [audience] for almost any humanities topic."

In a speech given at the Second Biennial Wisbey Lecture Series in Humanities Computing, on the topic of public networks and vernacular computing, Unsworth (2006) noted that

Fifteen years ago, the challenge before us was to imagine how new technology might provide a new platform for the practice of scholarship in the humanities, but today our challenge is the reverse. It is no longer about opening the university and inviting the public in: it's about getting out where they already live, and meeting the public in the information commons, on the same terms that everyone else does.

While there are a growing number of K-12 distributed educational tools that have been, or are being, developed to encourage exploration into various humanities issues, such as the Virtual Burnham Initiative (LFC, 2008), there is also a mature and ever growing population of life-long
learners that could equally benefit from tools designed for fostering and simplifying their own humanities investigations.

**Tools**

To take full advantage of a cyberinfrastructure designed to facilitate humanities research requires not only “a critical mass of digital resources” but also “an integrated set of tools to work with them” (Borgman, 2009); tools flexible enough to accommodate the research needs of various domains, or communities of interest. As Matei (2008) explains:

>[A] future digital environment for humanities should meet a number of requirements… most importantly, it should mesh up with the most fundamental activities of…identifying sources, selecting them, learning and producing learning from them via a process of annotating, writing and rewriting text and/or other types of content (images, digital artifacts, numeric representations of social realities). In a word, it should center around an application that facilitates the creation process that is native to the humanities scholar…[as well as the general public, as this author believes.]

Yet, while a suite of tools is needed to accommodate the natural beginning-to-end process of knowledge discovery, the real challenge lies in designing tools in such a way that not only will the “new knowledge created from them…remain available for future generations” (Tooby, 2007a), but also the raw evidence collected be accessible to be reused, built upon, repurposed or challenged by later investigators as additional findings are uncovered.

The problem for the general public, as well as “[r]esearchers in the arts and humanities who attempt to use technology in their work [is that they] encounter significant barriers to access to digital content, work in isolation from systems of technological support, and spend far too much…time …constructing tools” to accomplish their investigative tasks (Broughton & Jackson, 2008). “The notion of sharing tools and/or data with other unrelated projects or systems is often foreign”, and as a result, the tools employed and data collected for one project are often “not reusable as components [in] other projects that might want to build on previous work instead of reinventing it” (Broughton & Jackson, 2008). In many instances, the tools utilized by solitary investigators who initiate the bulk of humanities research (Boonstra, Breure, & Doorn, 2004, p. 88; Tibbo, 1994, p. 607) are often limited in scope and represent a hodgepodge of general use

---

4 “The basic concept behind MeshUp technologies is to create Open-InfoSpace where individuals and organisations gain free and simplified access to an integrated open base of information services, contents, platforms and infrastructures and are free to integrate and modify them, the whole resource environment evolves. By satisfying their needs, users and resource providers enhance and upgrade the system” (Cosen.org, 2010).
applications such as spreadsheet, word processing, database, and other non-integrated software products. While humanities researchers may know the type of evidence they seek, most are unfamiliar with the complexities of computer technology and are unsure how best to structure their data to take full advantage of high power computing. As a result, essential data elements collected tend to be poorly defined and at times contain conflated information, which can be difficult to share, repurpose, and build upon.

One of the most important questions for the humanities is “What are data?” because “[t]he answer will determine what data are produced, how they are captured, and how they are curated for reuse. Data sharing in the humanities is a complex set of issues…that must be addressed” (Borgman, 2009). “If this new scholarship is to be more than a series of boutique projects” (Friedlander, 2009) the tools developed must enable the capture of incomplete, fuzzy, and ambiguous data elements to truly facilitate long term use and the cumulative incorporation of new knowledge (Fisher, 1999; Frischer et al., 2006). Central to the design of an effective tool would be the provision of comprehensive digital access to the scattered evidentiary material, along with the researchers’ individual transcriptions and assumptions, in order for others to substantiate the results, foster debate, and enable reuse in future projects. Above all is the need for “[c]ollaboration across traditional boundaries … particularly … for ambitious projects that require years of research and cannot be summed up in a single dissertation or monograph” (Friedlander, 2009).

Numerous tools developed early on have failed or been abandoned (Shilton, 2009), some as a result of insufficient long term funding, others appear to have been inadequate for the needs of their intended audience, or languished because they were of limited utility to a wider community of interest. Borgman (2009) points out that unlike the plethora of research exploring scientific practice, as represented in “social studies of science and technology” (Van House, 2004), there is “no comparable body of research on scholarly practices in the humanities … with the exception of research on information-seeking behavior”. Studies cited by Borgman (Anderson, 2004; Bates, 1996a, 1996b; Bates, Wilde, & Siegfried, 1993, 1995; Case, 2007; Siegfried, Bates, & Wilde, 1993; Stone, 1982; Tibbo, 2003; Wiberley, 2003; Wiberley & Jones,
1994) were not limited to information seeking, but also touched upon aspects under the broader realm of information behavior. Unfortunately, none of these studies examined the specific information sought in relation to the researchers’ consequent information use behaviors, defined by Wilson (2000, p. 50) as consisting of “…the physical and mental acts involved in incorporating the information found into the person's existing knowledge base. It may involve, therefore, physical acts such as marking sections in a text to note their importance or significance, as well as mental acts that involve, for example, comparison of new information with existing knowledge.”

Today data collection, note taking, and the analysis associated with information use behavior is often performed with the aid of computer technology (Schreibman et al., 2004), yet “humanities’ tools have not been systematically studied” (Friedlander, 2009, p. 4 n. 5), especially for their utility in support of a community's information needs. The “actual processes of knowledge construction becomes important for understanding the role of artifacts such as documents and information systems, and for designing systems and services to facilitate this work” (Van House, 2004, p. 11). Further substantiating this statement are the findings of Duff et al (2004), which imply that the historical documents themselves may play a significant role in determining the appropriate structure for the tools desired, yet most information behavior studies provide limited insight into the types of source material sought, let alone the format in which evidence is culled and organized.

A great deal can be learned by examining the tools used and data collected by a specific community of interest to better appreciate unresolved barriers and opportunities for advancing humanities computing. Therefore, this study focuses on the particular information requirements of a diverse but overlooked community that relies upon reconstructed colonial American parcel maps to make possible their varied historical investigations. It examines the tools, analytical requirements, and types of information collected for facilitating their work. This persistent but little understood need affects both academic scholars and the general public in their quest for collecting, organizing, and transforming historical information into geographic knowledge. While this study examines two similar North Carolina land-based research endeavors (Dobbs, 2006; McNeely, 2010), the scope of the problem crosses state and local boundaries and should be of
interest to every institution tasked with preserving and disseminating 18th century and earlier colonial land-related evidence. Friedlander’s (2009) universal concerns, described earlier, surrounding scale, language and communication, space and time, and social networking provided a framework for examining the results.

The following section introduces the complex nature of the information need and describes some of the varied motivations, objectives, and results to provide insight into this unique group of humanities investigators.

**A community need: Colonial parcel maps**

Because few, if any, parcel maps were drawn depicting the progression of backcountry settlement in America’s thirteen original colonies, humanities studies exploring such geographic circumstances repeatedly turn to evidence contained in colonial land records to guide in the reconstruction of the spatial relationships they seek to explore. This is a multifaceted yet fundamental activity that must be completed before an actual investigation can proceed. The desire to visualize a prior community’s dynamics is not an isolated information need. Various individuals including cartographers, geographers, environmentalists, social scientists, archeologists, family and local historians, students, and other laypeople have engaged in the practice of reconstructing past geographies.

The process of reconstructing early American property maps to visualize settlement patterns is undertaken by both academics and, in even larger numbers, the general public, in order to support their diverse historical investigations. This longstanding activity represents a little-studied aspect of human information use behavior, which involves the transformation of colonial land record data and supporting evidence into geographic knowledge. The task is time-consuming and often entails years of painstaking archival research and interpretation to piece together just one small geographic region of interest. Without these detailed maps many new insights into historical phenomena and their relationship to today’s environment would remain

---

5 Specifically State-Lands initially distributed by British colonial agents, or other state governments, including the present day states of Connecticut, Delaware, Georgia, Hawaii, Kentucky, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, North Carolina, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, and West Virginia (Hone, 1997).
undiscovered; not surprisingly, researchers lacking the skills, time, or tools necessary for reconstructing long-extinct neighborhoods often seek out such maps made by others.

Colonial era parcel maps not only support individual researchers’ investigative activities, they also provide an invaluable geographic key for locating spatially related archival records; a retrieval method absent from most institutional discovery tools (i.e., online or card based catalog systems). For example, historical cadastral maps prepared by various investigators provide the only visual method for identifying which colonial land parcels would fall within a state’s present-day county or township boundaries. They also facilitate locating tracts of land settled within proximity to physical features, such as which are within five miles of prominent trading routes or how many were settled along the south side of a particular river\(^6\). Spatial queries such as these are not possible without the aid of a map, which has prompted various archival patrons over the years to construct their own—an investigative activity that establishes them as the de facto experts on the neighborhoods they have chosen to reconstruct.

Regrettably, the underlying cadastral maps reconstructed by scholars are frequently omitted from their final publication(s); particularly when they are not the main focus of the investigation. For example, the foundation of this research draws upon insights gleaned over the course of a nearly 4-year collaboration with G. Rebecca Dobbs, an historical geographer, who at the time was working on her dissertation at UNC. This author’s role was to design and continually modify a relational database to support Dobbs’ evolving needs for collecting and organizing land record evidence essential for guiding the reconstruction of 18\(^{th}\) century North Carolina property maps. Neither Dobbs’ final publication, nor the two ad hoc research websites she maintained during her investigation\(^7\), provided access to her land record transcriptions, raw data, or the completed cadastral map(s) essential for informing future research.

In regards to parcel maps prepared by members of the general public, a different problem exists. Although various individuals have taken on the challenge of reconstructing colonial

\(^6\) For example, an historical map of North Carolina’s piedmont region would enable identifying which properties were located within five miles of the Indian Trading Path or how many were settled along the south side of the Eno River.

\(^7\) Dobbs maintained two personal websites during her investigation. Both were initially password protected. One served as a working reference to relevant literature and contacts, the other for presenting a selection of GIS maps prepared by Dobbs, which were ultimately incorporated into her dissertation.
backcountry neighborhoods over the years, many fail to consider the benefit of donating their completed maps to a local archive or historical society, although there are exceptions (Kluttz, 1995; McNeely, 2009). The reconstructed maps that have been donated and accessioned by various library or archival institutions typically lack the vital research notes and citations necessary for substantiating their accuracy (Kluttz, 1995; Markham, 1955, 1973a, 1973b).

Nevertheless, this type of fundamental historical land research, engaged in by both scholars and life long learners, deserves to be documented, preserved and shared. Historical cadastral maps provide an essential foundation for investigating a wide range of phenomena, such as migration motives (Ramsey, 1964), economic conditions (Earle, 1975), influence of transportation networks on settlement (Dobbs, 2006), landscape ecology and vegetation (Bain & Brush, 2004; Schulte & Mladenoff, 2001; Schwartz, 2007), genealogy and local history (Kluttz, 1995; Markham, 1973a; McNeely, 2010), and their role in various administrations (Kain, 1993). More importantly, they alone hold the key for enabling access to early land records and their related evidentiary material, as originally intended in a mapped format essential for visualizing individual land records in the geographic context in which they are meant to be displayed.

The benefits of a map interface for retrieving archival and library information based upon its spatial relationship to the earth is well documented (Hill, 2006), and the utility of such interfaces has been clearly demonstrated in the public sector as evidenced by the growing popularity of applications such as MapQuest, Google Maps, and Google Earth. Several libraries and archival institutions have been experimenting with methods to incorporate this missing but invaluable map-based search feature (ADL, 2004; Binkley, 2009; Earhart, 2009; Johnston & Jensen, 2009). To date, none have explored the possibility of employing a cadastral map interface to improve access to colonial land records. The major hurdle appears to be a lack of standardized tools and infrastructure for facilitating the exchange of existing and new information between the key stakeholders. These stakeholders include the patrons who need or reconstruct historical cadastral maps, the institutions charged with preserving land record evidence (e.g., archives, special collections, historical societies, libraries, museums, etc.), and the academic institutions and individuals who could build upon and benefit from such a shared resource.
Who constructs them, who uses them?

Past examples point to the diverse uses that such previously un-rendered maps have been put to. An early example is shown in a facsimile (Figure 1.1) of a sketch drawn by a private individual, William Sharpe (1773), to support his contention that a second church was warranted to support his community’s growing population. His map encompasses land once part of historic Rowan County, North Carolina, that today falls within present day Iredell County. Property outlines are not defined, yet the map illustrates the value of geographic visualization to understand a neighborhood’s physical distribution. Large-scale maps such as Sharpe’s, drawn in or near the time period under study, are a rare and coveted resource to the historical researcher. Although additional work is required to verify its authenticity, such as locating citations to and examining facsimiles of the supporting land records in order to verify the geographic relationships depicted, the settlers’ names provide a starting point for tracking down the corroborating evidence. Once located, the shape of each tract can easily be reconstructed (assuming a survey record exists) and situated appropriately onto a map with the aid of GIS technology.
More common are retrospective studies where the manual reconstruction of a parcel map is based entirely on land record evidence, pieced together without the guidance of an historical map such as Sharpe’s. For example, geographers have routinely mapped prior communities to explore a range of historical circumstances. Ramsey (1964) rebuilt selected neighborhoods in historic Rowan County, North Carolina, settled between 1747 and 1762 to study the migration motivations and ethnic diversity of settlers who moved from Pennsylvania, Delaware, Maryland, and Virginia to resettle on land in North Carolina (Figure 1.2). Earle (1975) prepared historic maps to study eighteenth century economic conditions in All Hallow’s Parish, located in present day Anne Arundel County, Maryland. Dobbs (2006) reconstructed the location of properties occupied between 1748 and 1763 within North Carolina’s present day Orange, Durham, and Wake Counties, including portions of Chatham, Alamance and Person, to understand the Indian
Trading Path’s influence on early settlement patterns in relation to today’s towns (Figure 1.3).

Figure 1.2 Historic parcel map of early settlement in Rowan County, NC (Ramsey, 1964, p. 108).
In a joint project an environmental geographer and a biologist reconstructed the location of land settled between 1664 and 1793 in the Gwynns Falls watershed area in Baltimore City and County, Maryland (Bain & Brush, 2004). Their work presents a scientific method for reconstructing early property mosaics as an aid for analyzing historical land use patterns, considered invaluable for studying aspects of landscape ecology. A local historian and mapmaker rendered several maps to document where land owners settled in sections of North Carolina prior to 1800 (Hughes, 1976a, 1976b, 1977, 1979, 1980) (Figure 1.4).
Some unexpected people have engaged in the practice, including a University of Tennessee chemistry professor (emeritus), Robert McNeely, and a United States Attorney (Alderman, c1985). McNeely’s unpublished work represents nearly 20 years of research, motivated by a desire to map the lands settled by his ancestors (McNeely, 2009, 2010). His maps depict conditions of historic Rowan, Lincoln and Burke counties, which encompass present day Lincoln and Burke counties, as well as portions of Catawba and Caldwell counties in North Carolina (Figure 1.5). The bulk of the parcels mapped by McNeely represent grants reissued by the State between 1778-1799, but it is likely many of these tracts had initially been granted to the same owner (or their relations) under a former colonial administration; some were perhaps granted more than 30 years earlier. Alderman, near the end or shortly after his term serving in the Regan administration, published a book containing parcel maps he had reconstructed to
accompany his historical analysis of present day Carroll County, Virginia, a region settled by his
father’s colonial ancestors.

The largest community engaged in mapping past neighborhoods appears to be the
general public, represented by family and local historians whose backgrounds, research skills,
and presentation styles vary considerably. Examples include land ownership maps drawn by
Markham (1955; 1973a; 1973b), Rendleman (1953), Holcomb (1974; 1975; 1980), Horvath and
Wilkinson (1980; 1981a; 1981b; 1982; 1985), Stanberry (1993) and Kluttz (1995) (Figure 1.6).
These particular maps serve to identify early settlers and the general vicinity of land they
occupied in North Carolina and Maryland; numerous other regions both within and beyond these
states’ boundaries have been mapped as well. The motivation for these mapmakers appears to
be personal, as many of the neighborhoods they have mapped seem to include ancestral
homesteads. While one might be tempted to categorize this group as genealogists, they have
gone far beyond the construction of a family tree to establish past familial relationships. In fact,
verifying ancestral relationships prior to the 19th century often depends upon land-based research
such as this in order to uncover the missing connections.

Figure 1.5 Portion of McNeely map. Historic Burke and Lincoln counties, North Carolina.
Representing State issued land grants after 1777. (NCGen Web Project, 2008).
Figure 1.6 Portion of a map created by a genealogist/family historian (Fitts, 2007, p. 27)

Of critical value to humanities studies are existing reconstructed parcel maps, which enable subsequent researchers to expand upon their predecessor’s investigations. Such an endeavor could add years to a project if one needed to first build the geographic circumstance under study themselves. For instance, an historian (Thorp, 1991) drew upon Ramsey’s (1964) maps to evaluate retail trade activity between 1755 and 1775 in the backcountry region near present day Salisbury, NC. Kluttz (1995) based portions of his landownership maps of Rowan County, NC, on Rendleman’s (1953) earlier work. Schwartz (2007) depended upon Dobbs’ (2006) parcel map reconstructions and survey transcriptions to analyze presettlement forest communities in the Piedmont region of North Carolina to understand vegetation change over time. Bain and Brush (2004) drew upon the maps prepared by Horvath and Wilkinson (1980; 1981a; 1981b; 1982; 1985). David Southern (nd), a North Carolina native and local historian

---

8 Eighteenth century landholdings in the area north of Chapel Hill. Section of parcel map prepared by Markham (1973a).
points to the value of Markham’s Durham County maps (1973a; 1973b) for informing Fish Dam Road research. Archeologists frequently rely upon reconstructed parcel maps to inform their research. For example, Simpkins and Pethernick (1986) cite Hughes’ map (1979) of Rockingham County, NC, and Markham’s map (1973a) were central to an assessment of the Carolina Commons tract located in Orange County, NC (Fitts, 2007).

These examples provide a glimpse into the myriad uses for historical cadastral maps and point to the diversity of contributors sharing this basic information need. Like William Sharpe many individuals over the centuries have created or utilized neighborhood maps to visualize or analyze the geographic circumstances they seek to study. While their motivations, methods, and output styles have varied, they each relied upon common source material to guide in the process of reconstructing their areas of interest or to substantiate the content of an existing reconstruction: extant land records and related evidentiary material in combination with historical maps of various scales. Their information need is fundamentally geographic yet existing reconstructed historical parcel maps are difficult to find as many are ‘hidden’ in files across scattered archives, libraries, historical societies, online genealogical sites, and private collections. Making these maps more accessible through a shared cyberinfrastructure, along with providing links and authenticated images of source material to which they correspond to would be of great benefit to a wide audience.

**The problem**

Reconstructed historical parcel maps fill an important information void, yet the archival institutions that maintain the corresponding land records appear to be overlooking their value as an integral aid for land research. Land records are intrinsically geographic information that only a map can adequately portray. Without a map to visualize the spatial relationships between land records their true context will continue to remain obscure. Facilitating access to reconstructed maps, both existing or under development, could resolve this gap in geographic information. Regrettably, few repositories tasked with maintaining early land records related to the colonial era have explored the potential of working together to construct collaborative tools for use by
individual mapmakers, which would help to preserve and consolidate their complementary and often hidden body of work.

Currently few one-of-a-kind reconstructed historical parcel maps, considered manuscript material by most archives, are listed in institutional catalog systems or even cross-referenced with the land records they depict. As for scholarly publications, wholly or partially reliant on such reconstructions, the catalog records that describe them tend to exclude the subject term ‘map’, making it difficult to locate these important contributions. For example, the term ‘map’ is not included as a searchable index term for locating Ramsey’s (1964) predominantly cartographic work. Other historical cadastral maps, prepared to facilitate scholarly investigations, have simply been omitted from the final publication, as mentioned earlier. More importantly the new knowledge these maps offer remains uncataloged; knowledge that could include each parcels’ latitude and longitude, vernacular and variant name spellings of individuals and features identified in the records, or the discovery of sequentially related land transactions and related supplemental evidence.

Complicating matters is the fact that existing reconstructed parcel maps are geographically scattered, maintained by unrelated public and private entities, and are characteristically produced in non-standard formats. More importantly, the majority of completed maps lack adequate citations to the underlying records employed in their construction, a significant oversight that limits their value in support of future research. These interrelated problems perpetuate the repetitive and unnecessary duplication of efforts as researchers, unaware of existing maps or unable to verify the evidence represented in those they find, are compelled to retrace the steps of their predecessors by reconstructing or reconfirming previously mapped geographies of interest.

**Opportunities, solutions & visionary statement**

To eliminate redundant efforts and facilitate the sharing, building upon, and preservation of reconstructed historical parcel maps requires not only a suitable infrastructure, but also a set of standardized tools to capture individual user contributions, such as transcriptions, links to document facsimiles, and authoritative citations that in turn could be coupled with the maps they
represent. It appears that aspects of social-networking technology may hold the key for fostering such contributions, and corresponds to the type of “vernacular-computing” environment digital humanities initiatives should strive to emulate, according to Unsworth (2006). Clearly, archives alone lack the manpower and funding necessary to solve this persistently overlooked barrier to information, but the benefit to them of participating in a collaboration to encourage this type of activity appears evident.

Picture having online access to a vetted interactive historical map\(^9\) of the United States that employs a geographic information system (GIS) capable of chronologically displaying each tract of land settled in America from the colonial period forward, ultimately linking to present day digital county cadastral maps. This multimedia tool would provide images of an institution’s primary source material and supporting data that could be viewed, annotated, discussed, and built upon by contributors from various communities of interest. Such an historical map could enable academic researchers and the general public to visualize and study changes in property ownership, jurisdictional boundaries, and related cultural features over time.

Imagine this map as part of a distributed network hosted by a consortium of academic institutions, government entities, libraries, archives, museums, and historical societies who, acting in partnership as the community’s hub, could reap the greatest benefit. As patrons identified new relationships between consortium members’ holdings, access to their scattered but related documents would improve and searching for material based on geographic content or a resource’s physical location could be made possible. While no such tool currently exists, research studies and ongoing initiatives point to viable aspects for consideration in the development of such a comprehensive historical GIS (Bearman & Trant, 2005; Buckland & Lancaster, 2004; Cappon, 1971; Chen, Smith, Larsgaard, Hill, & Ramsey, 1997; Crane & Wulfman, 2003; Ell, 2003; Hoelscher, 2001; Holdsworth, 2003; Jessop, 2005; Keene, 2004; MacEachren & Brewer, 2004; Merrens, 1965; Sager, 2001).

---

\(^9\)Google Earth represents one type of mapping application that may be worth emulating, but its proprietary format, inability to support analysis, and lack of geographic precision suggests a more robust environment may be required.
The potential number of patrons appears limitless. Generations of individuals and groups routinely investigate historical phenomena from a geographic viewpoint, including educators, students, government agencies, businesses, and the general public. Each could benefit from access to a map-based historical database of land transitions, which would improve over time as evidence became joined together in virtual space. Anyone fascinated with American history and geography, interested in visually exploring the places, people, and events that shaped our nation, would be prime candidates for adopting such a tool. To achieve this vision requires a set of common tools that could bring together the various communities of interest involved with cadastral map reconstructions, ranging from social and physical science scholars to those engaged in serious leisure activities (Hartel, 2003), such as genealogists and local historians. While these diverse groups’ motivations may differ their shared objective of visualizing land ownership inextricably links them.

This author’s long-standing dream (Ruvane, 2007) envisions a solution capable of providing access to early land record data via a map interface. Existing parcel map reconstructions could serve as the foundation and be dynamically grown as academic researchers and the general public added new maps to the interface, both fully functional GIS renditions and simple image files that could be later enhanced by others (e.g., converted into GIS format), along with corroborating content. By implementing an integrated suite of stand-alone and peer-reviewed social-networking tools, for collecting, analyzing, and sharing reconstructed maps and related evidence concerning historical neighborhoods -- linked to digital facsimiles of the evidentiary material -- the process of authenticating existing or generating new parcel maps could be streamlined. The ultimate goal would be to enable building upon past and future mapping endeavors to facilitate dissemination, intellectual debate, and preservation of this significant and enduring historical geographic knowledge. This would not only help to preserve and support ongoing original research, but also enhance access to institutional collections while providing a life-long learning tool for patrons of the humanities.

Dobbs' work (2006) and others', along with this author's prototype database, offer a starting point for the development of such a comprehensive tool. Over the course of three plus
years using the prototype database, Dobbs fully transcribed close to 6,000 North Carolina colonial land records\textsuperscript{10} and prepared digital copies of each survey document (albeit not of the highest quality), while annotating her observations (Ruvane, 2006), recording uncertainties inherent to the land record evidence, and later mapping a substantial portion of her findings using GIS technology. Her well-documented early American parcel map reconstructions represent untapped information with tremendous potential to advance historical research in North Carolina, as well as digital humanities tools (Frischer et al., 2006). Dobbs’ research could serve as an inspiration for various repositories charged with maintaining similar materials within and beyond the State of North Carolina. The thousands of early colonial land records as they exist today represent scattered pieces of a very large and important puzzle, one that should be permanently reassembled for use by future generations to enable the visualization and study of settled land in a mapped format as originally intended.

\textsuperscript{10} There are hundreds of thousands of colonial land record folders listed in the North Carolina State Archives MARS catalog, which will be examined more closely in the research phase. Most folders contain at least one record, but many contain two or more related land documents.
CHAPTER 2: LITERATURE REVIEW

The theoretical underpinning of the research described in this dissertation is based upon human information behavior, with particular emphasis on a little studied aspect of how information is actually used in humanities investigations to create new knowledge (e.g., information use behavior). The literature reviewed is drawn from both within and outside of the field of library and information science. It serves to highlight what is known, or not known, about diverse groups involved in historical research, their common information needs, motivations, challenges, and some of the tools they work with to facilitate their goal of constructing new knowledge. Although there is scant literature that explains in detail how geographic information culled from historical evidence is employed in the reconstruction of maps, the general importance of such maps as an aid in the historical research process is pointed out in various publications (Gregory, 2005; Knowles, 2002).

Human information behavior

Information behavior (IB) deals with a wide spectrum of theories surrounding a human’s information needs, exploratory processes, sources consulted, and results, as exemplified by Wilson (2000), Pettigrew et al (2001), Case (2002), and Fisher et al (2005). Issues investigated within this framework are typically considered a part of the broader field of communication science, which studies the conveyance of information and the role of human communication in life and work. Fisher (p. xix) builds upon the IB definitions of Pettigrew et al (p. 44) and Wilson (p. 50) by “…conceptualizing information behavior as including how people need, seek, manage, give, and use information in different contexts.”

Seminal studies have focused on modeling two discrete IB facets, information seeking (Dervin, 1983, 1997, n.d.; Ellis, 1989; 1993; Kuhlthau, 1988) and information retrieval (Belkin, 1980; Ingwersen, 1996), particularly in relation to human interaction with associated channels of communication (such as library catalog systems, web browsers, and face-to-face interactions). More specific studies have focused on information seeking in context, where a concrete information need is observed and outcomes analyzed. For example, papers presented at the 5th Information Seeking in 11

Communication studies: study of the activity of conveying information - by or to or between people or groups.
Context Conference held in Dublin, Ireland, in 2004, covered research on user needs in public libraries, health information-seeking among Latino newcomers, search behavior of engineering and science students, and information literacy of medical students (Information Research, 2004). In an Annual Review of Information Science and Technology, an entire chapter is dedicated to the topic of information seeking in context (Solomon, 2002).

One area underexplored within the framework of IB is information use behavior, which goes beyond information seeking and retrieval issues to examine an actual information need and the “…physical and mental acts involved in incorporating the information found into the person's existing knowledge base” (Wilson, 2000, p. 50). In addition, few studies situated their findings holistically within the framework of information use behavior in context to understand a particular community’s explicit information needs and how they transform original data to support their varied investigations.

**Historical researchers**

Certain information behaviors of historical researchers have been studied but cannot be easily categorized. The researchers observed represent a body of humanities scholars whose interests, motivations, and evidence requirements are diverse. People engaged in studying the past are not limited to historians, but also include genealogists, geographers, social scientists, anthropologists, and numerous others interested in our nation’s cultural heritage. Studies focused on the information behaviors of historians (Andersen, 1998; Case, 1991a, 1991b; Dalton & Charnigo, 2004; Duff & Johnson, 2002; Gilmore & Case, 1992; Jones, Chapman, & Woods, 1972; Tibbo, 2003; Yakel & Torres, 2003), and genealogists (Duff & Johnson, 2003; Friday, 2006; Ruvane, 2005; Ruvane & Dobbs, 2008; Yakel, 2004), have pointed to three characteristics that appear universal: 1) the need for access to historical documents and supporting material to answer the questions being explored, 2) the need for evidence related to a specific time, place, person(s), and event (or topic), and 3) the need for assorted tools to help locate, gather, transcribe, extract, record, organize, analyze, and at times share the results of their findings.

Many researchers that study the past, such as historians and historical geographers (Baker, 1997), initiate an investigation with the goal of sharing their results, while others simply seek to satisfy a personal information need. Historical researchers tend to work independently in exploration of their
specific topic, particularly humanities scholars and historians (Boonstra et al., 2004, p. 88; Tibbo, 1994, p. 607), but collaborative projects illustrate how certain communities of interest have joined forces in their efforts (About the Atlas of Historical County Boundaries, 2006; Electronic Cultural Atlas Initiative, 2009; Goeken, Nguyen, Ruggles, & Sargent, 2003). Despite the diversity of subjects explored, and varied research objectives, access to primary source material maintained in archives, museums, special collections, or private collections is critical to the success of an historical investigation (Andersen, 1998; Bidlack, 1983; Case, 1991a; Dalton & Charnigo, 2004; Dobbs, 2006; Duff et al., 2004). Equally important are secondary materials, such as books, maps, and journals found in libraries for supplying context and background information related to the subject under exploration.

Studies indicate that historical researchers prefer to examine original documents (Andersen, 1998; Case, 1991a; Duff et al., 2004), yet many papers housed in archives have been removed from circulation due to their fragile condition or for protection. In these instances reproductions such as Xerox copies, microforms (e.g., microfiche, microfilm), or digital copies are normally provided. Unfortunately, these substitutes are often of poor quality and difficult to work with; particularly microforms which are the least preferred format (Duff & Cherry, 2000; Duff et al., 2004). Several studies imply that a web-based delivery system would be a welcome replacement in situations where originals are unavailable; but only if the system is easy to use and offers high quality, persistent, and well-documented surrogates (Andersen, 1998; Duff & Cherry, 2000).

Aside from the constraint of working with inferior reproductions there are additional barriers related to locating and accessing suitable archival material. The geographic distance of records (involving travel time and expense), lack of finding aids, uncomfortable work environments, and limitations with current archival record discovery tools have all been mentioned (Arnold, 2004; Case, 1991a; Duff et al., 2004). A predominant concern is the need for retrieval systems to support searches beyond the limitations of a personal name search, such as entry points for locating content based upon its physical location (place), date, and subject context (Arnold, 2004; Case, 1991a; Galvez & Moya-Anegón, 2007; Hill, 2006). Unfortunately, many systems are inadequately designed to facilitate concurrent searching on such central terms. In addition, to do so effectively requires
dealing with the fuzzy (i.e., uncertain or outmoded) nature of historical evidence (Plewe, 2003), which calls for the incorporation of gazetteers (Buckland et al., 2004; Chappell, 1999; Goodchild & Hill, 2008; Mostern & Johnson, 2008), personal name matching programs (Borgman & Siegfried, 1992), and often a method for dealing with incomplete dates and vague time periods (Feinberg, 2003).

Further compounding the process of locating suitable primary source material relates to archival practices in general, which accepts that item level indexing is prohibitive and therefore not practicable for most collections; as a result, investigators are routinely hampered by the absence of vital search terms and adequate context. Much of this could be resolved by “…allowing researchers to contribute descriptive notes and other information to [online] archival collections” thereby increasing the available entry points and thus making “these materials…more…accessible to a wider variety of users” (Krause & Yakel, 2007), but archives have been slow to adopt social networking tools. Samouelian (2008) and Yakel (2006) studied the use of Web 2.0 technologies and described several features that offered promising solutions, including blogs, wikis, bookmarking, and tagging features. In a separate study Sedgewick (2008) found that patron-contributed annotations improved user satisfaction and accessibility in two out of three online collections. A number of articles point out additional Web 2.0 features worth consideration (CiteULike, (n.d.); Daines & Nimer, (n.d.)). The study of historical events depends upon efficient access to relevant information, and Web 2.0 tools appear to offer avenues for facilitating the process.

A common requirement of researchers studying historical phenomena is the ability to obtain copies of the documents pertinent to their investigation. Case (1991a, p. 76) observed that “…historians…tend to photocopy all but the briefest portions of primary material in archives.” In a separate study one respondent explained that copying “enable[s] better usage of documents” and “allows for closer reading [of the source material] over a long period of time” (Duff et al, 2004). While Duff and colleagues suggest copies (i.e., photo or micro) are a cost-effective method that enable researchers to reexamine sections of material as questions arise, this can be an exorbitant expense for those examining large collections. For example, Dobbs’ (2006) research involved the transcription

---

12 The third collection, Beyond Brown Paper, employed a system solution that did not make their annotations searchable. The user-supplied annotations were only viewable upon discovery of a particular item.
of nearly 6,000\textsuperscript{13} eighteenth century handwritten land records, which were only available for inspection on microfilm. The North Carolina State Archives at the time only provided four readers (one frequently out of service) capable of making substandard photocopies at a cost of twenty-five cents each. Because the machines were in high demand, patron use was limited to 1-hour blocks of time. For most students, spending over $1,500.00 on photocopies would be considered a prohibitive expense in addition to an inefficient use of time waiting for a machine, related parking expenses, and frustration associated with poor copies.

As an alternative, Dobbs purchased copies of the 33 reels of film required for her investigation\textsuperscript{14} and a small microfilm reader for use at home. This did not eliminate the need to make printable copies of selected land records, particularly the survey documents. For this she hired an assistant to make digital copies using state-of-the-art microfilm readers located at UNC’s Davis Library. From the digital copies Dobbs was able to print paper copies, which were also of much better quality, at a fraction of the cost charged by the archives. In reality, neither of the two choices available to Dobbs for working with this material was cost effective. In this author’s opinion, the most efficient method and one likely to reach a wider audience, would have been the provision to access high-quality facsimiles of the original document(s) directly from the online catalog while working in the comfort of one’s own home or office.

Some of the top-ranking materials cited as important for historical research include manuscripts\textsuperscript{15}, government records, newspapers, and maps (Arnold, 2004; Dalton & Charnigo, 2004; Duff & Johnson, 2003). Unfortunately, these studies failed to explain the type of data sought and how it was employed, making it difficult to fully appreciate the information behaviors of the respondents in context. In none of these studies were the respondents asked about the time period, place, person(s)...

\textsuperscript{13}Currently 5,951 land record transcriptions are in Dobbs’ database. A few represent records not available in the microfilm collections purchased. Seven of these records were added after the completion of her dissertation.

\textsuperscript{14}The initial 33 rolls of film employed by Dobbs related to the Granville District’s land records, 27 of which have been transcribed. This does not include the additional reels of Crown and State land records yet to be transcribed.

\textsuperscript{15}The specific types of records identified as important to historical researchers were not always clarified (i.e., did they rely upon mostly archives or manuscript evidence?). “The term ‘manuscript’ covers any type of work, either hand written or typed, which is not published. The terms archives and manuscripts are often used interchangeably to refer to unpublished records. However, a distinction is often made based on the source or creator of the records. Most personal records are considered manuscript material, and organizational or corporate records are considered archives.” Unfortunately, the original source for this quote, http://www.lib.unc.edu/instruct/manuscripts/glossary/, is no longer available.
or subject of their investigations. For example, Dalton et al’s study points to the importance of newspapers, but this media only began to flourish after the mid-18th century and primarily in populated areas. “History actually dictated the types of records kept, together with the format and content of those records” (Greenwood, 2000, p. 8). In other words, to truly appreciate the information needs of humanities-related investigations requires an understanding of the circumstance being explored in relation to the evidence available.

The findings reported by Duff et al (2004, p. 14) were the first to imply that the historical document(s) themselves play a central role in providing the information sought in support of an investigation (i.e., ranked by 93% as *very important*). It would be interesting to repeat these previously mentioned studies with a few additional qualifiers, such as asking each interviewee about the time, place, person(s), and topic under investigation, along with the specific type of records they deemed critical for satisfying their information needs. For example, locating personal records about a well-known individual or family is less challenging than finding similar manuscript material related to lesser-known personalities. Government records could mean anything from colonial land records, to Congressional meeting minutes, or judicial proceedings. This type of additional detail might prove insightful for identifying improvements to the search and retrieval functions of existing archival or library catalog systems.

The need for more in-depth analysis into historical researchers’ information behaviors has not gone unnoticed. For example, Case (1991a, p. 80) suggests narrowing the focus to a more homogenous cross section of historians, such as “social historians of the United States”. Yet the infinite variety of topics investigated by this group, compounded by the wide range of time periods explored, might preclude adequately identifying common behaviors due to the inconsistent nature of the evidence available. This suggests that limiting studies to a particular topic and time period, or conversely, to a specific set of resources may prove more useful to the process of improving access to our nation's historical records and shaping the design of suitable collaborative digital humanities tools. In regard to collaborative projects, aimed at facilitating on-line access to primary source materials, "...it is essential that archivists assess what users want and need and how they go about locating information before repositories spend precious resources on technology projects and digital
library design...” (Tibbo, 2002). Others concur that innovative solutions to web based delivery systems, based on user-oriented designs, is essential for insuring the success of archives in the future (Andersen, 1998; Duff, 2002; Sexton, Turner, Yeo, & Hockey, 2004).

As Case (1991a, p. 80) points out, “…the key to understanding information needs…lies in further probing of [a particular group’s] specific activities.” Studies exploring information behaviors “…should be designed and evaluated for their utility in the work task context” (Järvelin & Ingwersen, 2004). To date there appear to be no studies that have started with the archival resource itself to establish who is using it, for what purpose, and how the information culled from the records is organized and used to create new knowledge. For example, colonial land records have been and continue to be an essential resource in support of a wide variety of historical research, as discussed earlier. In fact, early land records along with land-ownership maps are one of the more heavily used resources in state and national archives, based on personal observations at the NC State Archives and as noted on the Library of Congress’ map division website16 (Geography and Maps: General Collections, 2000).

While the studies cited in this section provide glimpses into selected core information behaviors of historical researchers, such as the need for better retrieval tools and personal copies of primary source material, systems cannot be designed on these generalities alone. It is through an understanding of the historical circumstance itself, the evidence available, and the information gaps that remain, that the shortcomings and opportunities for improving access to archival collections are likely to be uncovered. It is especially important to recognize that the majority of archival patrons are not academic scholars, but instead tend to be pursuing an avocational interest or hobby (Ailes & Watt, 1999; Linder, 1983; Turnbaugh, 1983). Therefore, improving access to historical documents for all potential users should be the ultimate goal.

**Historical research & computers**

The use of computer tools to aid in locating, examining, organizing, analyzing and documenting historical evidence continues to evolve in stages. Many archives have replaced, or are

---

16 The Geography & Map division of the Library of Congress notes that county land ownership maps are among the most heavily used materials in their collection and are particularly valuable in support of genealogical studies.
in the process of replacing, their paper-based card catalogs and finding aids with online equivalents that offer enhanced search and retrieval features. For example, some provide links from their online catalog records to digital facsimiles of select documents. Yet, “…[w]hile more and more archival sources have become available in digital form, there has not been enough attention to the development of computational methods to process and analyze them” (Boonstra, Breure, & Doorn, 2004, abstract)\(^{17}\). The use of stand-alone applications, such as word processing, spreadsheet, citation, and other specialty programs to separately manage the various tasks involved in an historical investigation is often inefficient, especially in light of the variety of complementary research that could be shared, enhanced, and built upon if an integrated suite of standardized tools were made available.

A growing number of humanities-related investigations rely upon computer technologies to enable: (1) the evaluation of historical sources and conversion of the data into machine-readable formats, (2) some method of data quantification and aggregation, and (3) the processing, analysis, and communication of the resulting historical discoveries (McCrank, 2001, p. 599). Regrettably, according to Boonstra et al (2004, abstract) “[m]any historians … seem satisfied with standard office tools, which do not always suit their complex sources and research questions.” For example, database applications with user-friendly interfaces may provide false expectations that anyone can design a system for managing multifaceted historical information, a fallacy commonly discovered far too late after the database fails to meet the investigator’s objectives and preservation requirements. These breakdowns are often introduced during the database modeling stage (p. 32), where novices fail to consider the complete data structure of their source material and query expectations. “As a result, the level of technical sophistication in many historical studies has remained rather low” (Boonstra et al., 2004, abstract).

Boonstra et al and McCrank explore the issues central to the life cycle of historical research under the rubric of historical information science.

Historical information science is neither ‘history’ nor ‘computing’. It is a science of its own, with its own methodological framework. The object of historical information science is historical information, and the various ways to create, design, enrich, edit,
retrieve, analyse and present historical information with [the] help of information technology (Boonstra et al., p. 10).

The life cycle of historical information does not end with locating the historical evidence. It is an iterative and cumulative process that entails creation, enrichment, editing, retrieval, analysis and presentation of new findings (Boonstra et al, 2004, p. 22-23). Key concerns are the durability, usability, and effective project modeling of the tools employed for information gathering and project dissemination. While McCrank points out the evolution of numerous practical tools that have been applied in support of historical research, Boonstra et al suggest that these tools need further development and “[i]t is the duty of [historical information scientists] to adapt the information technology framework in such a way that historians can do their work” (p. 91).

Historians do not want to be burdened with the details of information technology, but they will adapt and are adapting individual computer tools to aid in their research process, whether suitable or not, that likely could “…be translated into more widely applicable and usable tools. Better infrastructure is needed in order to guarantee a transfer of results from the methodological and technical level to the daily practice of historical research” (Boonstra et al., 2004, p. 92). McCrank envisions a future of multiple historical Intranet research collaborations reliant upon Internet navigation in an “archival cyberspace” that could be managed collaboratively. “Currently, historical research overall is unmanaged, and one must wonder also about project management…of the means, opportunity, and cooperation so important to large-scale collaborative History” (p. 600). He suggests that the solution lies in the development of networked registries of historical databases constructed using a framework of shared templates to insure usability “…that would serve like metadata for reference, benchmarking, and specific kinds of historical analyses and methodologies.” Such a network would in turn enable persistent access and the sharing and building upon of this growing body of knowledge by future investigators.

Boonstra et al suggest that future research related to historical information could entail developing “generic” data models to accommodate variations of the same primary source type incorporating a basic functionality that characterizes the source in question. He proposes that “these
models be defined at a high conceptual level (with UML\textsuperscript{18} class diagrams, for example), not directly related to a particular technology...If such generic data models were available...data models for specific research exercises would be quicker to produce, and the resulting data collections easier to share” (p. 101-102). Software engineers employ a variety of diagrammatic methods (such as UML) to model the conceptual relationships between data elements and system interactions. A commonly employed model in the design of relational databases is the entity relationship (ER) diagram, a method which has advanced over the years (Chen, 2002) and which was used by this author in the design of Dobbs’ database (Ruvane & Dobbs, 2005). Paired with a functional specification to clarify the system requirements in terms of its input and output functionality, a generic model could provide the blueprint necessary for proceeding with a suitable design.

It seems possible that the research process for reconstructing historical parcel maps could be standardized by modifying a variety of existing computer tools (many identified by McCrank) to characterize the source material in relation to the stages of the historical research life cycle. For example, during the data gathering process existing methods for uploading digital images of individual land records (if not provided by the holding institution) would be useful for facilitating the process of full text transcription and extraction of key terms necessary to enable complex searches. Text encoding standards (TEI-c.org, 2009) might be appropriate for this task and existing editing tools potentially could be adapted for this purpose (Editing Tools: ImageMarkupTool, 2009; Editing Tools: OXygen, 2009). A newer tool, developed for diplomatic\textsuperscript{19} transcription of historical texts also appears promising (Dipper & Schnurrenberger, 2009), especially for version control to monitor multiple handwritten copies of the same transaction. As for the time-consuming process of citing each individual record, this could be automated by enabling investigators to link directly to individual catalog records themselves, thus insuring proper documentation, a persistent link to the source, and reliable access for future investigators.

There are numerous stand-alone applications available today that have been, or could be, used to organize pieces of the historical mapping process, but they need to be incorporated into a

\textsuperscript{18} Unified Modeling Language (UML). A general purpose diagrammatic modeling language used in software engineering.

\textsuperscript{19} “The study of records that identifies their generic characteristics, independent of their manner of creation, the nature of the individual or organization that created them, or the social system in which they were created” (Pearce-Moses, 2005).
comprehensive and integrated solution. For example, there are assorted programs for creating and
commercial applications (e.g., ESRI ArcGIS, AutoCad Map) and open source ‘build-your-own’ alternatives (FreeGIS; OpenSource GIS). Online collaborative mapping programs are also available (Google; OpenStreetMap; Wikimapia), although the accuracy of their user-provided content is often questionable (Goodchild, 2007) and these tools lack the precision and analysis tools required to support critical research. Several commercial programs designed specifically to facilitate historical parcel mapping also exist and are marketed for use by family historians (e.g, Deedmapper, Deed Platter, others). Additionally, development surrounding the use of Web 2.0 technologies for facilitating GIS information sharing is growing (Oxley, 2009). But in most cases, the mapping programs currently available are incapable of storing full transcriptions, capturing authenticated and properly formatted citations, dealing with uncertain (or incomplete) data, or facilitating collaborative projects where version control is critical.

**Map usage & availability**

There are few studies related to information behavior that explain the different types of maps, how they are employed, and the spatial information deemed valuable for resolving an unanswered question. Gluck (1996) reports the findings from five exploratory studies that highlight various behaviors related to the geospatial information needs of the general public. In one study, based on a log of map use at a public library in Florida, 25 patrons were referred to either the State archives, for historical maps, or to the county appraiser for parcel ownership information. Youngblood (2006), a university map librarian, observed researchers from 23 academic disciplines to understand map use across various fields of study. She provided brief examples of use under each category. Of note was the observation that social science researchers often seek historical geographical information for use as a basis in creating their own maps, a capability of tremendous academic value. Youngblood’s overall findings indicate that ‘map use plays a focal role not just across disciplines…but also in the areas of avocational, educational, and professional interests and work.’

It is from outside the field of Information and Library Science that the vital role that maps play in historical research can be gleaned. The need for maps and the types most beneficial in genealogical research are discussed in Greenwood’s guide (2000, pp. 46-50) as well as in a
pamphlet produced by the United States Geological Society (2002). They confirm that large-scale neighborhood maps are often vital in the study of family history. In a special issue on GIS as a tool for exploring social science history, several articles demonstrate how digitally enhanced reconstructions of historical maps have enabled the discovery of previously hidden knowledge related to past societies (Knowles, 2000). Historians and students of history have also adopted GIS technology in support of their various historical inquiries (Gregory, Kemp, & Mostern, 2003; Knowles, 2002).

The advantage of today’s GIS technology is that individual map themes, or layers, can be created digitally and stored, modified, and combined as needed to facilitate map customization for presenting the selected features of interest. “The users of geographic and land information systems increasingly agree] that common sets of geographic data on which users could build their own particular [maps] would promote greater sharing among various players” and “present a significant cost savings” (Frank, Goodchild, Onsrud, & Pinto, 1995). In an attempt to realize this objective, a survey, funded by the Federal Geographic Data Committee (FGDC) was conducted by the National Center for Geographic Information and Analysis (NCGIA) to identify the technical criteria and prioritize common data sets (referred to as “core” or “framework” data) targeted for inclusion in the National Spatial Data Infrastructure (NSDI) (Frank et al., 1995). Respondents were asked about their use and information needs related to six common classes of feature data including 1) transportation, 2) water, 3) cultural, 4) elevation, 5) land parcel, and 6) boundary data. Although this study was somewhat controversial for its sampling methodology (or lack thereof), the findings remain useful for understanding commonly employed layers within a GIS.

The FGDC’s vision is to provide universal access to digital representations of these core data layers via the NSDI to simplify the process of map construction, which in turn would reduce errors, duplications of effort, and lower production costs for government agencies and their private-sector partners. This same approach, providing central access to similarly themed historical map data, would be equally beneficial to those involved with investigating the past. At present only two

---

20 The FGDC’s core data sets were later expanded to include geodetic feature data and orthoimagery data; cultural feature data no longer is listed as a priority framework data set on their website (Framework Overview, n.d.).
organizations are focused on providing and preserving such “core” historical geographic information related to the United States. The National Historical Geographic Information System (NHGIS) provides downloadable files representing decennial census area boundaries, dating back to 1910, and state and county boundary files dating back to 1790, and The Newberry Library’s ongoing project, under John Long’s guidance, has documented the shifts in county boundaries for each state since their inception, with the most recent being constructed in digital formats (AHCB, 2006).

While there is a plethora of digital map data available for exploring present day conditions in the United States, historical map representations are just beginning to make their way onto the Internet (Rumsey & Williams, 2002). Many have been presented in constricted and often unusable electronic formats, impeding their utility and lacking the features users prefer (Peck, 1999). Even more challenging is locating functional digital spatial data – a topic not addressed in this paper but warranting study in its own right – yet those with patience and computer skills can find basic general-purpose or reference maps useful for exploring their themes of interest. To create customized historical map representations requires a higher level of skills, as few historical maps have been broken down into their individual component layers (i.e., vector or raster). Over time this problem could be reduced if the FGDC’s concept were adopted in the creation of a cyberinfrastructure to preserve and share commonly used historical framework data sets. The foundation of such a repository could start with thematic layers that have been digitized over the years in support of a variety of research, such as Dobbs’ base level parcel map reconstructions.

The historical circumstance

As Duff et al (2004, p. 14) and Boonstra et al (2004) have suggested, the historical records themselves play an important role in informing a researcher’s investigation into the past; consequently, it is important to understand the circumstance behind the records generated, or not, during the time period under study. The intent of this section is to provide a brief overview of the approaches employed in dividing the land across America, the various administrations involved, the use of land records in historical research, and the origination and characteristics of various North Carolina land records that were examined in this study. By understanding the land distribution practices employed by the British during their settlement of the American Colonies, the existence and
type of land records available to guide in the reconstruction of colonial era cadastral maps are easier to ascertain. This corresponds with Friedlander’s (2009) call for more research surrounding the scale, or subject domain, of a particular humanities investigation. This is especially important considering that standard archival practice calls for organizing collections by their producer (Duff & Johnson, 2002), and as a result often serves to disaggregate large bodies of related evidence that can be challenging to comprehend without suitable explanation.

For those interested in learning more about land distribution employed across the United States from the colonial era onwards, works by Kain (2002), Price (1995), and Hone (1997) provide good overviews. Hone also provides a starting point for locating records available by state. For more specific details regarding the different approaches employed during the formation of a particular colony a local library, state archive, or historical society will have additional resources. In North Carolina, good summaries of the shifting circumstances and various administrations tasked with distributing the land are available online in the form of a finding aid prepared by the North Carolina State Archives (Record Group 12, n.d.) and in Mitchell’s (1993) work related to land distribution in the northern half of the state.

**Dividing the land**

In order for the reader to understand why some historical researchers are compelled to reconstruct maps depicting past neighborhoods warrants an explanation. The scarcity of early property maps is not universal; the dearth is most noticeable in rural regions settled between the early 17th century and late 18th century21 within the original Colonies, particularly those in the South. The circumstance that led to this gap in geographic information can be directly related to the dissimilar approaches employed during the colonial era for dividing the land. While the distribution of land varied from colony to colony, in general there were two methods: one orderly and planned, the other irregular and haphazard. In the former approach land was surveyed first into large blocks, or townships, which was then divided into small tracts by its administrator(s) prior to being distributed to individual settlers and their dependents. Alternatively, in the irregular system individuals were free to

---

21 From the founding of Virginia in 1607, ending sometime after the US declared its independence from Britain (1776). By the mid-1800s commercially produced rural maps and atlases started to fill in the gaps, and today nearly every county maintains cadastral maps of the land under their jurisdiction.
choose land that appealed to them, irregardless of its proximity to previously occupied tracts, and then proceed with the survey and deed process to establish ownership\(^\text{22}\). The amount of land one could purchase was dictated by the presiding administration at the time.

Unlike the organized system, which relied upon the preparation of detailed property maps to efficiently distribute the land, agents in colonies utilizing the haphazard approach rarely prepared cadastral maps, much to the consternation of their overseas proprietors (Wacker, 1975, p. 223). For instance, in North Carolina, the heirs to the Granville District instructed their representatives to prepare a complete survey of the district and a map of the ungranted land; a request that apparently was never acted upon (Mitchell, 1993). Over time the lack of comprehensive property maps made it difficult to effectively monitor available land versus established claims, which often led to boundary disputes for the courts to resolve in later years.

In general, land in the Northern colonies was divided using an organized system, such as a township or rectangular scheme, with the exception of the eastern portion of New York (Figure 2.1). In the South, below Pennsylvania, the irregular system was more common and typically involved the distribution of larger tracts of land to individual households, organizations, and at times speculators. Throughout the colonies there were exceptions, areas where both the haphazard and organized distribution methods coexisted, but clearly large areas of settled rural land remained unmapped. There are numerous influences that fostered these different approaches to land distribution during the formative years of America’s colonization, including religious, social, and political concerns, which are best described by Kain (2002) and Price (1995) from whose work much of this section has been drawn.

\(^{22}\) Although land ownership was established with the issuance of a grant of deed (or patent) an annual quitrent was often a condition of the transaction. For example, in NC “Lord Granville rented his lands upon payment of three shillings sterling followed by an annual rent of three shillings sterling or four shillings proclamation money for each hundred acres” (Ramsey, 1964, p. 18).
Figure 2.1 Variations in early land division practices (Price, 1995).

Common to both land distribution methods was the metes-and-bounds survey method used for delineating property boundaries. The system had been employed for centuries in England and unsurprisingly was adopted for defining land in the Colonies. Both town boundaries and individual parcel boundaries were described using this “indiscriminate” measurement system, which did not always result in a rectangular plot. For example, over 1,000 surveys transcribed by Dobbs defined parcels containing eight or more sides, forming non-rectangular polygons; in an extreme example one parcel had twenty-four sides.

Starting at one corner of the land a surveyor would record the length and compass direction of each boundary line segment and continue sequentially around until returning to the beginning point. Landmarks such as trees, watercourses, rock outcroppings, or neighboring property lines served as references between directional shifts in each line. Most of these handwritten metes-and-bounds surveys were accompanied by a drawing (plat) of the land in question (Figure 2.2). A typical description for an individual parcel surveyed in 1755, located in Orange County, North Carolina, reads as follows (Dobbs db ID5414; MarsID 12.13.97.14):

---

23 The land surveying method in State-Land States, such as North Carolina, is often referred to as "the indiscriminate metes and bounds system" because plot descriptions were based upon measurements between imprecise land features and made reference to impermanent neighboring land owners.

24 William Eaton survey, Granville County, dated April 2, 1750. NC State Archives microfilm reel S.108.274; Dobbs db ID3761.
This plan represents a tract of land surveyd for Thos Lindley lying on the south side of Cain Creek & on the north side of Haw River about a mile above the Piney Mountain. Beginning at a white oak by the side of the sd creek running thence East 42 1/2 chains to a black oak then South 60 chains to a black oak then West 20 chains to the center of 4 hickorys then South 24 1/2 chains then N 44 W 67 1/2 chains to the sd creek so up the creek to the first station. Containing three hundred and fifty six acres. Surveyd the 26th of June 1755. Sw chain carrs Jas Lindley Jno Woody W Churton S.

Figure 2.2 1755 land survey, Granville District, Orange County, NC. Included with 1756 deed assigned to Alexander Mebane [Maborne].

The predominant land division system in the United States today is the Public Land Survey System (PLSS), or rectangular survey system, which was adopted by the newly formed federal government of the United States. Instituted in 1785, the PLSS was a systematic method that referenced a baseline or standard parallel (i.e., running east/west) and prime meridian (i.e., running north/south) for dividing newly acquired land in the MidWest, portions of the South, and west of the Mississippi (NRCCILDM, 1982). Land in these regions was first divided into townships, sections, and quarter-sections prior to distribution (Appendix A). This resolved many problems associated with the
less accurate metes-and-bounds system employed in the East and for the first time generated a wealth of contiguous historical cadastral map evidence; at present there is no comparable consolidated map for those interested in studying the Eastern states’ historical geography, thus the motivation for many a map reconstruction.

While the regions utilizing an organized survey system fall outside of the purview of this paper, particularly the post-colonial PLSS lands, it is interesting to note that the need for access to comprehensive historical neighborhood maps to support various investigations was recognized in 1982 by our federal government. The modernization (computerization) of the Public Land Survey System records was identified as a high priority for the development of a national multipurpose cadastre by the National Research Council Committee on Integrated Land Data Mapping (Dahlberg, 1984; NRCCILDM, 1982).

PLSS data is used by government employees, lawyers, educators, historical researchers, the general public, and many others. In academia, a growing number of ecological research studies have relied on PLSS data to aid in mapping presettlement vegetation (Bolliger, Schulte, Burrows, Sickley, & Mladenoff, 2004; Hotchkiss, Calcote, & Lynch, 2007; Schulte & Mladenoff, 2001). Family historians, perhaps the largest users of PLSS property maps, depend upon such visual information to establish links to their early ancestors. This author’s vision of an interactive historical cadastral map interface of the United States would encompass colonial era geography and ultimately incorporate access to the PLSS cadastral maps (BLM Land Records, n.d.), all linked to the archival documentation and GIS maps created to illustrate the unfolding land transitions.

**Land records: a universal but dubious resource**

Genealogists have long recognized the importance of land records in family history research. An authority on New England genealogy Donald Lines Jacobus (quoted in Luebking, 2006, p. 431) argued that “[t]he most important town records, genealogically, are the land records” especially “[i]n the South, which has far fewer vital records than New England, the land records are even more crucial to genealogical success” as they may be the only surviving evidence available for verifying the existence of long ago relations and the regions they once occupied. As previously discussed, land
records are equally invaluable to academic researchers in support of a variety of historical investigations.

The majority of our Nation’s earliest land records are available on microfilm (Luebking, 2006), indicating their longstanding value to decision makers and consideration of high priority for preserving. Their persistent and ongoing use by diverse researchers is evident. While the Genealogical Society of Utah\textsuperscript{25} may be best known for microfilming historical records their inventory is far from complete. For colonial land records the most comprehensive collections are likely to be found in state archives or libraries. Leading the way to improving access to original land records is the Bureau of Land Management (BLM) whose federal-land records, which number in the millions, are being systematically digitized and made available online (BLM Land Records, n.d.). The BLM also provides access to many of the corresponding PLSS township survey plats and field notes and continues to add new images as time and resources allow.

The archives charged with maintaining colonial land records have not been sitting idle. Several institutions have taken on the challenge of digitizing their historical land record collections (MD State Archives; PA State Archives; KY Land Office), but the different approaches, varying image quality, and lack of standard search mechanisms may limit their future ability to provide comprehensive access akin to the BLMs multi-state searchable land record database. Additionally, without access to corresponding maps the true context of unmapped colonial land records will continue to remain unclear, as the colonial metes and bounds surveys lack definitive reference points to determine the location of an individual tract of land. Conversely, each tract of land marked-off utilizing the PLSS was coded to a corresponding township survey map that readily established its precise geographic position on earth; thus, in order to provide a unified and comparable resource for the eastern states reconstructed cadastral maps are required.

While the actual process of obtaining vacant land varied between colonies and within each colony, it typically entailed a series of sequential transactions that culminated with the issuance of a first-title deed, referred to as a grant or patent. Despite the various administrations involved, the chronological records generated by each colony’s agents were comparatively uniform in format and

\textsuperscript{25} The Genealogical Society of Utah is affiliated with The Church of Jesus Christ of Latter-day Saints, Salt Lake City, Utah.
content, making them ideally suited for transcribing into a standardized machine-readable format.

During North Carolina’s Colonial period several different administrations were responsible for distributing the land prior to independence in 1776\(^{26}\); including the earliest administered jointly by eight Lord Proprietors [1660-1729] and later the Crown [1729-1775] and Lord Granville [1746\(^{27}\)-1763] (Figures 2.3 & 2.4). The coastal regions were the first to be settled, beginning in the 1650s with Virginian squatters moving into the Albemarle Sound Region (Walbert, 2008, p. 1.17) onto land granted in 1629 to a single Proprietor, Sir Robert Heath. By the 1740s settlers were arriving in large numbers in the central Piedmont region, which is the focus of Dobbs’ research and representative of the land records this study examined to identify opportunities for improved access (Figure 2.5).  

![Figure 2.3](image)

Figure 2.3 The (highly theoretical) boundaries of the land granted the [eight] Lords Proprietors under the charters of 1663 and 1665. Since no one at the time knew how wide North America actually was, no one knew just how much land was included under the grants (Walbert, figure 11).

Beginning with the earliest organized European government in North Carolina in 1663, the Lords Proprietors established the Office of Secretary, an administrative branch typically established in most British colonies. In North Carolina one of the secretary’s primary responsibilities was the recording and filing of land records (Record Group 12, n.d.):

\[\text{26}\] The State land office was opened in 1778 per NC State Archives finding aid (Record Group 12, n.d.).

\[\text{27}\] While Lord Granville’s grant of the northern half of NC was approved in 1744, his land office was not in operation until 1746, per NC State Archives finding aid (Record Group 12, n.d.). While some surveys were conducted as early as March, 1746/7, the first grants were not made until March 25, 1748 (Mitchell, 1993).
Of the functional records, the longest-lived, and largest, series is made up of the records of the land office. The secretary played a central and integral role in the granting of vacant public land to individuals and corporations from 1663 to 1959…[and]…was involved in every step of the granting process from the beginning entry to the recorded patent.

The one exception was vacant land distributed in the Granville District, which encompassed the northern third of the State (Figure 2.4). Here the distribution of land and filing of land records was under the purview of Lord Granville’s agents, who were not obligated to share their records with the secretary. Fortunately for the historical researcher, the Granville District’s paperwork became the responsibility of the secretary’s office in 1783 and these records are now part of the North Carolina State Archives, Secretary of State records collection (GPLO grants, 1748-1763; GPLO entries, 1748-1763).

![Figure 2.4 The Granville District included nearly the northern half of North Carolina. Land distribution was under the control of John Carteret, Earl Granville. The Crown controlled the southern portion, once part of the eight Lord proprietors grant. (Walbert, p. 261, figure 82).](image)

Dobbs’ long-term research goal entails reconstructing neighborhood-scale maps of the entire central Piedmont region (Figure 2.5), which encompasses land distributed by both the Crown and Lord Granville’s agents between 1746\(^28\) and 1763. To date Dobbs has fully transcribed the bulk of the microfilmed land records in the Granville District related to her study area. From these she has reconstructed a portion of the neighborhoods they represent, located within today’s counties of

---

\(^{28}\) While the Granville Record Group is described as containing land records between 1748-1763 (and Dobbs herself describes her study period as beginning in 1748), Dobbs transcriptions uncovered references to earlier record dates: nine entry records recorded between 1746 and 1747. (e.g., 1376, 3646#, 3652, 3677, 3678, 3722, 3907#, 4560, 4561), one warrant in 1747 (e.g., 3757#), seven Surveys in 1747 (e.g., 1122, 4178, 4333, 4223, 4309, 4317, 4324) and six grants recorded between 1728 and 1745/6 (e.g. 6322*, 6323**, 6324**, 6325**, 6326**, 6327**) *McCulloch lands; **crown lands; #date may be inaccurate (see Dobbs’ database record for details)
Orange, Durham, Wake, Chatham, Alamance and Person Counties (Figure 2.6). In the Granville District the process of land acquisition varied from time to time but generally began with the filing of an application, or entry, followed by a warrant (authorizing a survey), the actual survey, and finally culminated with the issuance of a grant of deed (Mitchell, 1993). A fifth document, an assignment of the land rights to another, at times may be found within the sequence of records.

Figure 2.5 Dobbs' overall study area covering the Piedmont region of NC (2006, p. 15).
The Granville land office and its records were not particularly well managed, unlike other colonial land offices that appeared to have been more diligent with their filing and recording practices. According to Mitchell (1993) after the death of John Carteret, the Earl of Granville, in 1763, the Granville land office closed and essentially remained inactive. Attempts to reopen it by agents of Lord Granville’s son, Robert, between 1773 and 1776 were ultimately unsuccessful. The Granville land office records changed hands several times between 1763 and 1783, at which time the secretary of state was authorized to take control of the scattered papers. By 1788 additional records were discovered to be in the custody of others and ordered to be returned. Between 1800 and 1922 the secretary’s office transcribed and made copies of old record books, prepared indexes to the grants, and continued to seek out missing records. At some time during this period the land grants were rearranged by county and more or less alphabetized under the grantees name.\footnote{It is unclear from Mitchell’s article whether related transactions (i.e., the warrant, entry, survey, and deed) had originally been filed together under the grantees name, which to this author would have been the most logical method for preserving the context of the records and for facilitating retrieval in a working office environment.}

In 1909 a card index system was instituted by the secretary of state to improve access to the
Granville land office records that entailed separating and placing the entries, warrants, and surveys into individual envelopes (i.e., shucks). There were more than 5,500 records of which approximately 1,500 were indexed and processed accordingly. In 1922 it appears the original deed records were transferred to the North Carolina State Archives, followed by the transfer of the un-indexed entries, warrants, and surveys in 1927. In 1965 the deeds were arranged and a card index prepared. In 1985 the state archives took receipt of the 1,500 previously indexed land records and together with the un-indexed records sorted and rearranged them by county and grantee name. Starting in 1986 the State Archives began to microfilm the entire collection of Granville land records. Deeds were filmed separately, while the entry, warrant, and survey records were combined with no guarantee that adjacently filmed records were related (Mitchell, 1993).

Mitchell explains, “There is no certainty about the number of grants made in the Granville District between 1748 and 1763.”

The evidence indicates that substantial numbers of the records of the Granville land office have been lost. Granville consistently referred to his records as books and papers, and no books are known to have survived. He prescribed a careful records system for the district, but none of his agents followed his instructions.

While the Granville heirs believed that 4,363 grants had been issued, Margaret Hofmann (1986) lists 4,979 deeds in her published abstracts of the Granville land grants. To date, Dobbs (2006) has transcribed 2,321 of the Granville grants likely to fall within her study region, in addition to 586 entries, 2,278 warrants, 1,044 surveys, and a handful of miscellaneous records. She has been able to reestablish numerous relationships between the separated land records, linking at least two and at times the complete sequence of records related to the acquisition of a particular tract of land. Some of the relationships remain tenuous, but the hope is through additional research, and perhaps new collaborative tools, in time these can be confirmed and additional relationships identified.

Unlike the approach taken by Dobbs (2006) most investigators who reconstruct early cadastral maps cull evidence only from the deed or patent records to guide them in building their maps, as these represent the culmination of the land transaction process and establish legal ownership. Deeds typically are recorded with a copy of the original metes-and-bounds survey, which

---

30 Surveys were also typically filed along with a grant of deed. Of the 1,044 surveys Dobbs transcribed many appear to have no matching deed record.
provide the measurements essential for recreating the shape of each tract of land prior to placement on a map. With this in mind Dobbs had initially planned to transcribe only the original survey records, assuming they would provide the most complete documentation, but realized early on that transcribing the entire collection (e.g., entries, warrants, surveys, deeds) would provide a more comprehensive picture of settlement in her study area in light of the imperfect records from the Granville District (Figure 2.7).

This decision turned out to be beneficial in a number of ways. For instance, at times only a deed might be available for a particular tract of land; in other instances only the earlier paperwork appears to have survived. Often times the entry and warrant records, which essentially contain the same information, include directional information not found in the survey or deed, which can be helpful in reestablishing the location of a particular tract of land. Likewise, land described in an entry or warrant that was initially associated with Anson County might later be identified in a survey as falling within the newly formed Rowan County, serving to reduce the area in which the tract might be located upon a map. Also beneficial was the ability to analyze the length of time between each transaction in the land acquisition process.
The fact that land records, particularly survey documents, are the only source material that contain the information necessary for reconstructing parcel maps further corroborates Duff et al’s (2004) premise that the type of records employed are an important factor to consider in the study of human information behaviors. This is particularly evident in the proliferation of published land record abstracts, as they tend to be a more efficient and comprehensive resource for locating specific records and in many cases have usurped the role of various archives’ catalog systems. Unlike catalog records, a good abstract will include an index to every person and place (or feature) named

---

31 Patron-produced abstracts, a common practice engaged in over the centuries (Duff, 2004), seems to be an activity that could be incorporated into archival catalog systems if such efforts were encouraged to utilize computer technology to enable linking to the records they describe.
within the record (Meyerink, 1998). Not only are grantees names important to the researcher, but the names of neighbors, former owners, chain carriers, witnesses, as well as surveyors. The names of places and features such as ‘old fields’, meeting locations, churches, watercourses, transportation routes, and paths further help to streamline the search process. Although rare, an index to land records by date issued is sometimes provided but oddly survey measurements and plat images almost always are excluded. In North Carolina two of the most prolific publishers of early land record abstracts include Margaret Hofmann (1982a; 1982b; 1986) and A.B. Pruitt (c1987; c1994; c1998).

A promising initiative, although much less comprehensive or authoritative than the published abstracts, are the efforts of USGenWeb’s genealogy social networking website. From their North Carolina site (USGenWeb/NCGenWeb deeds) one can find partial to full transcripts of selected early land records contributed by various volunteers, albeit the majority are improperly categorized and frequently not accompanied with the actual image transcribed. For example, one transcription submitted by Guy Potts was listed under the category of deed, but the linked image was of a survey document dated June 13, 1753 (see: Appendix B). It is possible this survey copy was an attachment to the deed, but if so, Potts failed to transcribe the entire document as evident after comparing Potts’ transcription to Dobbs’ equivalent record (db ID6284). Potts omitted the deed date, which was noted in the verso as December 29, 1760, as well as the names of two witnesses32. Their transcriptions also disagree as to the north/south boundary line measurements, which Dobbs recorded as 236, Potts 230. Another purported deed transcription, representing land located in historic Orange County, contains a link not to the images of the conflated records transcribed but to a topographic map depicting the parcel’s present day location in Guilford County; this is a very useful piece of visual information if accurate and the original land records or facsimiles can be located for comparison. In sum, while the work contributed by volunteers involved with USGenWeb’s initiative is promising, it is very difficult to validate.

What perplexes this author is that the majority of early land records from both the Crown and Granville land offices have been partially abstracted and indexed (in machine readable format) at the

---

32 William Churton and Robert Adams
item level by the North Carolina State Archives\textsuperscript{33}; presumably back in 1985 during the microfilming process described by Mitchell (1993). The Archives' online catalog system, referred to as the Manuscript and Archives System (MARS), includes index terms for selected personal names and key geographic features associated with each land record (MARS catalog). This is a commendable level of cataloging not typically employed by archival institutions. In fact, MARS contains much of the same information found in the commercially published abstracts mentioned previously, but unfortunately it is far less efficient to work with\textsuperscript{34}. It lacks adequate search features, organizes its collections in an unwieldy and non-intuitive hierarchy, and its help pages are incomplete. Also, there is apparently no provision for concurrently searching across specified record collections, making it difficult to browse across multiple colonial land record collections to generate a list of all records containing the same last name and/or geographic features. As for sorting by date, this does not appear to be an option, which prevents focusing one's search on a specific range of years, ideally by document type.

While it is laudable that a substantial amount of the NC State Archives land records have been indexed at near item level, because of the above mentioned limitations it is likely that very few people are able to navigate the MARS successfully to find the specific land records they seek, with the exception of the archivists intimately familiar with the system themselves. Additionally, the exclusion of other important index terms, such as the names of surveyors and witnesses, the role a person played in the transaction, and similar qualifiers related to important geographic features, prevents further refining of the search results. This problem illustrates one of the many barriers to access that could be resolved with the adoption of digital tools to facilitate the incorporation of user contributed index terms, perhaps via some form of social-networking technology or provision for sharing and combining the findings of individual researchers.

\textsuperscript{33} It is unclear who created the machine searchable index terms, although it is presumed that the process was overseen by archival staff familiar with 18th century handwriting techniques.

\textsuperscript{34} Therefore, in this author's opinion, the reason for the continuing success of published abstracts (i.e., abstracts serve as a substitute because MARS does not adequately address the patrons' search needs in relation to land record research).
In fact today's version of the MARS is relatively new, as indicated by its help page that states it is still a work in progress. Hopefully the search options will be expanded over time, but there remains yet another barrier to retrieving land records as they are currently indexed. Historical records contain inherently uncertain evidence and numerous contradictions can be found within the same document, as well as across related documents. For example, when performing an initial comparison of Dobbs' transcriptions to the subject terms assigned in the MARS for the same document they do not always agree. Dobbs routinely encountered spelling variations for an individual's name within the same record and did her best to indicate the discrepancy, yet the comparable MARS record may list the same name under a completely different spelling. While these differences may be the result of human error (on the part of Dobbs or the MARS indexer), they could also be viable interpretations of difficult to read text, or represent one of several alternative spellings for the same person as recorded in the original document.

Additionally, the MARS index terms do not clarify the role an individual played in the land transaction process, such as whether they were a grantee, co-grantee, chain carrier, assignee, neighbor, surveyor, witness, or played some other part; in contrast, roles associated with names are routinely specified in Dobbs' database. To improve access to historical land records anomalies and omissions such as these need to be addressed. While a phonetic name-matching algorithm (Borgman & Siegfried, 1992) may be able to alleviate some of the retrieval issues related to variant name spellings, considering the hundreds of thousands of records involved, the names that are missing and the roles played by each individual are equally important for narrowing a search.

These are not the only challenges involved with searching historical land records. Some barriers are related to the nature of space and time (Friedlander, 2009); such as shifts in county boundaries resulting in an initial record being filed under one county with later transactions being filed under a newly formed jurisdiction's. There were also surveyor errors, including inaccurate parcel boundary measurements and tracts erroneously recorded as residing in one county when in reality they fell within the boundaries of a neighboring administration. Use of vernacular names is also a

---

35 Regrettably the old MARS system, available via a telnet connection, has been retired. While it had similar search inadequacies this author believed it was considerably faster and easier to work with.
problem, where the same river or mountain may be referred to by two different names. Incorporating an historical gazetteer, such as Powell’s (1968) or the Getty Thesaurus of Geographic Names On Line (2003), could alleviate some of these problems although existing thesauri are not necessarily all-inclusive. For example, several historical features recorded in Dobbs’ database are not mentioned in either of these resources, pointing to the benefit of allowing patrons to contribute variant terms as they discover them during the course of an investigation.

**Geographic knowledge: Maps for historical research**

While this study’s focus concerns the content extracted and tools employed for collecting and converting evidence from colonial land records into historical cadastral maps, a variety of supplemental content is required to facilitate the procedure. Much of the auxiliary information is culled from existing maps, but there are a variety of items to consider in determining each map’s suitability for the task at hand. Different maps serve different purposes; each designed to highlight a selected combination of features depending upon the map’s intended function. The following section provides an overview of some of the considerations researchers contend with in the selection of suitable map resources.

**Revisions: a common practice**

For comprehending geographic space there is no equivalent to a map. Maps provide visual representations of physical and cultural features, spatial relationships, and the effects of human activity that no other type of document can capture or express as well. To understand a particular region and time period under investigation historical researchers often rely on a variety of maps to become acclimated to the conditions of interest. Both present day and earlier period maps are invaluable for such orientation. In most cases more than one map must be consulted to fully comprehend the circumstance under investigation, explaining the impetus behind many a map reconstruction. The ability to consolidate selected features onto one map is powerful, especially when diverse content requiring complex analysis can be more efficiently examined within a GIS (Youngblood, 2006).

Improving upon existing maps is a longstanding practice among mapmakers (Wiberley, 1980). In fact, with the exception of some early explorers’ maps most cartographic representations,
including today's, are theoretically secondary works. Meaning they are reinterpretations of previously published work with much of their content extracted from prior cartographic and related textual material. Map revisions are often motivated by the need to rectify earlier geographic inaccuracies, consolidate physical and cultural features illustrated separately across multiple sources, or to incorporate new knowledge such as changes in feature names or political jurisdictions.

Reconstructed colonial parcel maps are a rare exception as they represent original cartographic works of previously unmapped historical geographic conditions. There were few pre-1800s maps, if any, that illustrated backcountry settlements formed prior to the mid-18th century, leaving the unique task of piecing together the missing historical picture to later investigators. Antique maps, from the period, offer little guidance in the process as early mapmakers routinely misrepresented these less populated interior regions. Few watercourses had been charted, state and county boundaries were often not based on surveys, their scale precluded local analysis, and the great northwest was still unexplored. Also common to early period maps were inaccurate (but delightful) embellishments, including beautifully illustrated mountains and watercourses in locations where none ever existed (Sellers, Van Ee, & Library of Congress., 1981).

Not until the 1750s did the accuracy of colonial maps begin to improve as surveyed land formed the basis of new publications. Examples include Fry and Jefferson’s 1751 map of the middle and southern colonies, Evans’ 1755 map of the northeast region, and Mitchell’s 1755 map (Evans, 1755; Fry & Jefferson, 1755; Mitchell, 1755). But as was typical of the maps constructed during this time period, their small scale is of little value for informing property map reconstructions. For example, the aforementioned maps were inconsequential to Dobbs’ research, although a later work by Collet (1770), depicting only North Carolina, was somewhat helpful.

Scale and type

Despite improvements in geographic accuracy early maps of the British American colonies for the most part were small-scale, depicting broad regions such as the eastern shoreline or multiple adjoining colonies. While small-scale maps can be instrumental in orienting oneself to an historical setting they lack the detail necessary for studying regional or local area phenomena. Local-scale geographic representations, in contrast, have value for researchers based on their provision of long
forgotten names and pointing out the relative locations of major towns, ports, rivers, forts, trails, and Indian hunting grounds -- yet often are deficient in terms of wider context. Thus researchers must move between scales when drawing on historical maps as sources.

While certain layers of reconstructed historical maps utilize present-day digital content, such as watercourse data, modifications of contemporary content is often required and informed by evidence from historical maps. For example, areas obscured by present-day manmade lakes need to reflect the original circumstance, such as the position of prior watercourses and arable lands once occupied by settlers. Historical maps may include rare maps drawn during the time period illustrated or shortly thereafter, as well as recently drafted or digitized maps of past environments (Wiberley, 1980). This definition differs from that of the many who reserve the term ‘historical map’ for characterizing mid-nineteenth century or earlier publications. It recognizes that both antique maps and recently rendered maps of past geographies (e.g., published, unpublished, or digital) may contain invaluable information for the historical researcher. Unfortunately, little attention has been paid to the latter two categories, many of which have not been accessioned by an archive or are inadequately described in their catalog.

Cartographic literature (Clarke, 2003, p. 185) suggests that there are two basic map types employed by mapmakers: thematic maps and general reference maps. Thematic maps tend to focus on a limited number of features often employing a color code system to display aggregated data, such as variations in soil, land use, population density, or settlement density, as represented by Dobbs’ kernel density map (Figure 2.8). Reference maps typically incorporate a composite of themes, intended to display the precise positions of surface features in relation to their surrounding components. There is some debate as to whether maps can be so neatly categorized. Many fall into a gray area defying categorization. For example, while the primary focus of a parcel map is to depict land ownership (thematic) most are superimposed with one or more themes to clarify each tract’s relative position (general reference) to interrelated features, such as county boundaries, watercourses, transportation routes, or prominent cultural features (as depicted earlier in Figure 1.3).
Fitness for use

Not every map is equally useful for exploring and guiding in the reconstruction of historical conditions. To determine a map’s suitability an investigator must assess multiple map criteria. Of primary interest are its geographic scale, area covered, date created, time period represented, and the thematic features depicted. Another important factor is a map’s provenance, which is often murky (Ford, 1926). Ideally a researcher needs to know whether the physical and cultural features illustrated are based on recent survey data, a prior map’s content, textual evidence, the mapmaker’s imagination, or a combination of these. Metadata standards in place today, such as the FGDC’s content standards for geographical information provide this information for contemporary maps.
Although not always consistently. The same is rarely true when dealing with historical maps.

Assessing the accuracy of the subject matter depicted in an historical map can be challenging. For instance supporting documents for antique maps, such as the surveyor’s field notes, correspondence, measurement techniques, or notes about the motivation and resources that influenced its content are often difficult to locate. Supporting documents that do exist are often filed separately from the map itself, with no reference to this fact in the curator’s catalog record. It is also possible that some notes of this nature remain a part of a family’s archive waiting to be discovered. As for reconstructed maps, which would be considered manuscript material by archives, the few that have been accessioned typically lack similar supporting documentation such as the author’s source citations and corroborating research notes, which are essential for confirming a map’s provenance and how missing or uncertain data was handled. Regardless of whether an historical map is antique or reconstructed, published or unpublished, the supporting documents and final map together form the complement of the work.

The same is true in the evaluation of present-day digital maps whose content should be accompanied by descriptive metadata, though oftentimes it is missing or inadequate. One without the other is of little value to the serious researcher. In other words, if the evidence is doubtful and the process unclear the only recourse is to start over. By providing the infrastructure and tools to marry patron-contributed digital maps to the evidence, transcriptions, and inferences used in their construction, redundant efforts could be reduced and the new direction envisioned in digital humanities scholarship realized (Friedlander, 2009; Frischer et al., 2006; Unsworth, 2006).

Dobbs’ and McNeely’s land record databases and reconstructed maps, as well as earlier analog work prepared by others (Hartsfield, n.d.; Hughes, 1976a, 1976b, 1977, 1979, 1980; Lagle, 1976; Markham, 1955, 1973a, 1973b; Ramsey, 1870; Ramsey, 1964; Sharpe, 1773; Stanberry, 1993), offer a starting point for the creation of an historical land parcel data set; a core data set that would be of benefit to a large and diverse population interested in our nation’s early formation. To achieve this goal requires understanding the resources available, common information needs, investigation procedure, and system characteristics best suited for supporting a collaborative
environment for linking user contributions and cataloged land records in order to, over time, complete the missing geographic story the evidence reveals. Therefore, a multi-pronged domain analysis was selected for the purpose of examining this unique group’s geographically related information needs, the results are intended to serve as a preliminary blueprint for guiding in the development of a suitable suite of digital humanities tools and infrastructure to support this goal. A discussion of these methodologies, and their application for comparing Dobbs’ work with related historical land based content, is the subject of the next chapter.
CHAPTER 3: METHODOLOGY & RESEARCH QUESTION

Objective

The purpose of this research was to characterize the data elements, tasks, and system features essential for inclusion in an integrated “digital humanities” tool to support the process of constructing, sharing, analyzing, discussing, and preserving early American parcel maps and to offer an initial blueprint, or general specification, for guidance in its construction. As a byproduct the findings substantiated major shortcomings in an existing archival retrieval system that could be eliminated with the adoption of this author’s envisioned collaborative tool, thereby providing the impetus for garnering institutional sponsorship in its development. To this end, a work centered approach involving a multi-part domain analysis was employed to characterize the essential components.

Work centered research: Domain analytic approaches

Numerous conceptual frameworks have been employed by researchers to assess individuals’ information behaviors (Pettigrew, Fidel, & Bruce, 2001), although the majority of studies cited tend to focus on analyzing behaviors related to information seeking or retrieval and fail to examine the associated information use behaviors, as defined previously (Wilson, 2000, p. 50). By utilizing a domain analytic approach, considered the cornerstone of many work-centered methodologies, the context of these interrelated activities can be more readily understood. Friedlander’s (2009) four proposed areas of research, central for progressing the field of digital humanities, served as a framework for guiding the perspective of the questions to follow. They are geared to assess common problems and opportunities related to scale, language and communication, space and time, and social networking. But before examining the questions, the following discussion provides background on selected work-centered methodologies and their various approaches to domain analysis.

One of the more complex work centered frameworks is Vicente's (1999) multi-stage cognitive work analysis (CWA) methodology, in which a domain analytic approach is central to the process. Pettigrew et al (2001) note that work-centered methodologies, such as CWA, serve as a general framework "...to help information system designers analyze and understand complex interaction
between (1) the activities, organizational relationships, and constraints of work domains and (2) users’ cognitive and social activities and [...] preferences during task performance,” and observe further that “unlike ... most conceptual frameworks and models of information behavior” they can “lead directly to [system] design recommendations and specifications.”

Vicente’s (1999) framework focuses on five dimensions of information interaction, including 1) the work domain, 2) control tasks, 3) strategies, 4) social-organizational considerations, and 5) worker competencies. It is considered “a method for study rather than variables and other predictors that affect information behavior” (Pettigrew et al., 2001). The first step involves defining the work domain, or “lay of the land” in terms of what is being controlled, irrespective of the activities or tasks to be performed. The second step involves identifying the control tasks, or objectives that need to be accomplished, regardless of the how and by whom they are to be achieved -- while keeping in mind the domain constraints uncovered in step one. The third step, strategy analysis, involves defining how a task or process can be accomplished, whether by humans or a computer process. The fourth step entails understanding the social organization and cooperation required in relation to how the domain, control tasks, and strategies can be allocated. And the last step, worker competencies, serves to define the capabilities and limitations of the workers themselves.

The crux of a work analysis (WA), whether incorporating cognitive dimensions or not, entails first understanding the domain and its constraints in order to identify problems and potential solutions in relation to the tasks which are interrelated. But Kang, Cohen, Hess, Novak, & Peterson (1990) explain that “[w]hile much work has been done on methods for domain analysis there is no single definitive domain analysis method” nor a de facto list of dimensions that should be included in the analysis. There also appear to be differences in the goals and processes involved in a domain analysis. For example, in software engineering “domain analysis is the process of analyzing related software systems in a domain to find their common and variable parts” (Ma & Wo, 2007, p. 254). A frequent goal is to enable the reuse of design knowledge (Arango, 1994; Prieto-Diaz, 1990), but a

---

36 CWA is a conceptual framework developed by Rasmussen, Peijersen, and Goodstein (1994) that was presented in their work titled Cognitive Systems Engineering.
domain analysis (or broader work-centered analysis) is also used as a guide by computer scientists for designing system interfaces (Eggleston, Young, & Whitaker, 2000).

Akin to Vincente’s objective to analyze control tasks and strategies is task analysis, which shares a relationship with domain analysis (Hajdukiewicz & Vicente, 2004) and is frequently associated with studies related to Human Computer Interaction (HCI). HCI focuses on interface design, engaging practitioners and researchers from both computer science and information science, and “routinely advocate[s] building user-centered systems which enable people to reach their goal” (Preece et al, 2002, quoted in Crystal & Ellington, 2004). Crystal et al further point out that “central to the design of such systems is a clear understanding of what users actually want to do: What are their tasks? What is the nature of those tasks?” They explore the evolution of task analysis and various techniques employed and conclude that current approaches are “too difficult to perform” and the results “too difficult to understand and use”, yet they agree that the insight a task analysis provides can be invaluable. They suggest further work be aimed at developing simpler techniques that are “both easy to use and informative” by combining different models that enable an integrated view of the multiple levels associated with task analysis (e.g., technical, conceptual, work-process).

The same problem has been noted in relation to Vincente’s (1999) CWA approach. While the domain dimensions she describes embody aspects of information behavior that this study analyzed, the methodologies she recommended for capturing the characteristics are considered complex. While others have adapted the techniques to suit their needs (Lintern, 2006b; Naikar, Hopcroft, & Moylan, 2005; Peters, 2005; Tan & Helander, 2006), these remain equally difficult to apply unless one is intimately familiar with the process, such as the designers of the methodology themselves. For example, core to a CWA is the process of constructing an intricate “abstraction-decomposition space” to portray the work environment, or domain, yet subsequent articles aimed at clarifying the vaguely defined procedures (Simons, Dainoff, & Mark, 2006; Tan & Helander, 2006) have failed in that aim according to some (Lintern, 2006a). Lind (2003) believes that the CWA approach presents “problems of vagueness” and thereby an “openness to interpretation” that “defies falsification” rendering the process “untenable” without further clarification of the concepts.
In the field of information science, Hjørland & Albrechtsen (1995) describe the domain-analytic paradigm as being closely related to the “study [of] knowledge-domains as thought or discourse communities” explaining that “[k]nowledge organization, structure, co-operation patterns, language and communication forms, information systems, and relevance criteria are reflections of the objects of the work of these communities”. Bates (1998, p. 1185) draws attention to “research results associated with indexing and access to information … that seem … to have great potential and/or importance in information system design…” For example, to help inform improvements in system interface design or the shape of future tools for aiding “the searcher [to] find his or her way to the desired information,” she clearly sees value in “paying attention to domain as a way to improve retrieval”, as her examples confirm (p. 1200), giving added credence to the multifaceted domain analytic approach.

In light of the promise a multifaceted work-centered analysis affords, a choice of appropriate methodologies to accomplish the task remains. The domain-analytic approach is not really new. The desire to holistically understand a user’s information behavior in context has spurred the development of various methods, which have been applied and evolved over decades by IS researchers (Hjørland & Albrechtsen, 1995). Hjørland (2002) believes that current methods employed for analyzing a domain, and the tasks involved within that domain, are far more complex than necessary and therefore difficult to implement. Instead, he suggests eleven alternative methods that are based on long-standing approaches employed in IS, both traditional and innovative methods that are equally suited for the task and considerably easier to perform. These include preparing literature guides, classifications and thesauri, assessing indexing and retrieval specialties, empirical user studies, bibliometrical studies, historical studies, document and genre studies, epistemological and critical studies, terminological studies, studies of structures and institutions, and exploration of scientific cognition, expert knowledge and artificial intelligence.

While there are clearly numerous approaches to conducting a domain analysis (Burns, Bisantz, & Roth, 2004), one of the more articulate explanations is provided by Arango (1994), a systems engineer, and described below. To characterize a domain “require[s] a fairly good understanding of what is the domain, what we consider to be in the domain, and what we consider to
be outside of the domain… as well as our ability to access the knowledge needed to model the domain…” He describes the overall process (Figure 3.1) using a series of Hierarchical Task Analysis (HTA) graphical models to elucidate each step, which serve as the iterative guide for answering the questions in this study. The initial process (Step 0.1) entails selecting and describing the domain characteristic under study, identifying relevant data, creating an inventory of the data, and general project planning. As a preparation phase step one helps to establish the domain boundaries.

Figure 3.1 Five stages in a common domain analysis process

The second step in a common domain analysis involves data collection (Step 0.2), which is inextricably connected to the objectives of the data analysis and classification (Steps 0.3 and 0.4). Although interrelated, data collection “uses techniques that are different from those of data analysis. [T]he four most commonly mentioned approaches…[include] interviewing, protocol analysis of problem solving, questionnaires, and thought experiments” (Arango, 1994). Equally acceptable is the use of “[a]bstraction recovery from existing systems”, particularly when the researcher conducting the analysis is familiar with “the structure and behavior of [the] implemented applications” under study, as well as their shortcomings. This was the case for this study, where a reverse engineering approach was utilized to examine three existing systems. The intention of this step is to “collect raw data that then can be filtered, clarified, abstracted, and organized.”

The third step entails data analysis (Step 0.3) to identify basic elements such as entities, events, operations, and relationships, followed by modularizing the information and analyzing the similarities, variations, combinations, and trade-offs. “Data analysis and classification are the two
central steps in the domain analysis process...”. There are various approaches to modularizing data, including “object-oriented analysis, or function and data decomposition” but the method employed should be driven by the goal of the analysis, and the determination of its intent to evaluate “procedures, functions, design schemas, and so on.” The purpose of analyzing similarities is to identify equivalent concepts or elements that may be referred to “by different names by different people” thereby reducing the number of definitions required. Identifying variations enables the discovery of distinguishing characteristics or features that may be important for inclusion in subsequent system improvements or designs. “The concurrent analysis of similarities, variations, and relations is an effective means to disambiguate definitions” and will often “trigger new data collection activities.” As a by-product of these activities “trade-offs and constraints are documented.”

The fourth step is classification (Step 0.4), which typically involves clustering, abstracting, classifying, and generalizing descriptions in order to construct a common vocabulary or taxonomic classification. “Descriptions with similar meanings are clustered together” and “[a]s new descriptions are introduced as a [result] of the concurrent data collection and analysis activities, they are” either “classified into existing clusters or the clusters are reorganized” to accommodate the new information. These “[a]bstractions are [then] organized into generalization hierarchies” in order to group entities or features that share common attributes. This step enables the analyst to generate various models (e.g., UML, entity relationship, functional specifications, etc.), as needed, to communicate and facilitate evaluation of “the information uncovered through data analysis.”

The fifth step, evaluation (Step 0.5) is “the most underdeveloped step in published methods.” The objective is to confirm that the models generated capture the essence of the domain and therefore are deemed suitable as a guide for accomplishing the goals and objectives of the project under consideration. While Dobbs is an authority on the elements of the domain with which she has interacted, it is likely this step will require input from a wider audience before proceeding with the development of a prototype system for testing.

**Research questions & methods**

1. *What are the primary domain characteristics of universal interest to this community?*
The overarching question for this study encompassed analyzing four interconnected domain characteristics in order to understand the fundamental information needs and use behaviors associated with historical parcel map reconstruction and validation. Understanding these key characteristics is essential for shaping the design of an authoritative collaborative system for creating, sharing, and building upon the missing geographic picture of colonial settlement in America. Central to the domain are the information resources necessary for guiding in the reconstruction of (or analysis of existing) colonial era cadastral maps, as well as the new knowledge created as a result of such activity. The four questions were designed to evaluate a number of wide-ranging factors unique to the domain; embodying Friedlander’s (2009) suggested areas of humanities computing research (noted in italics) to identify the resources employed (scale), evidence culled (scale; language & communication; time & space), characteristics of data uncertainty (language & communication; time & space), and the tasks involved (language & communication).

The intent was to assess an investigator’s information needs in relation to the process of map-based, geographically dependent, research. Provided below is a brief explanation of each questions objective, a description of its scope and unit of analysis, and the method selected for analyzing and presenting the results. Like Vicente’s (1999) CWA process, these individual analyses were conducted in a sequential order, as successive iterations relied upon information garnered from the analysis of the previous question. In certain instances variations to the methods described below were employed. These variations are explained in the results section.

A. What resources are essential for facilitating parcel map investigations and reconstruction activity? The Domain of Resources

For investigations dependent upon colonial era property maps to resolve an information need one of the most important characteristics to comprehend is the domain of resources, namely the types and extent of documents available for supporting such map-based research activities. Not surprisingly, original land records or authentic facsimiles thereof are the primary evidence consulted. Of particular importance are the survey documents that contain the measurements needed for generating scaled versions of each tract of land, which serve as the puzzle pieces required before reassembling the geographic circumstance can proceed.
Scope/Unit of Analysis

While Dobbs’ and McNeely’s investigations were limited to specific regions in North Carolina, the system envisioned is intended to support researchers exploring the progression of rural settlement across the entire state. As such, the Domain of Resources incorporated all of the identifiable land records cataloged by and stored at The North Carolina State Archives (NCSA). The NCSA, located in Raleigh, is tasked with preserving original as well as facsimile or derivative documents of both legal and cultural importance to the State. The NCSA are believed to curate the most comprehensive and authoritative collections of land records sought by this study’s population.

Because the intended collaborative environment is envisaged to facilitate geographic exploration of colonial era neighborhoods, the scope was limited to land records containing dates of 1790 or earlier, and limited to land presumably within North Carolina’s present day geographic boundaries. While this date extends beyond the colonial era time frame, it seemed prudent to expand the scope in this first analysis to insure that later transactions, potentially related to earlier colonial era activities, were not overlooked in the initial analysis.

Methodology

The objective of this analysis was to quantify the size and characteristics of the various cataloged collections available at the NCSA, as well as to understand the base line data provided (e.g., index terms) and utility of the MARS ID for use as a permanent link to the source material routinely consulted. This is in keeping with the need to understand the scale of the computing environment envisioned, which would require access to the cataloged land records as well as the ability to incorporate authoritative user contributions to enhance the current search and retrieval mechanisms.

To establish the parameters, Dobbs’ (2006) and McNeely’s (2010) “parcel map” databases were examined to verify the resources they had employed for use in guiding their respective investigations. Both databases contain extracted content from individual land transactions, although neither included a field for capturing the ‘MARS ID’ associated with each record; MARS IDs are ideally suited for corroborating the precise collections consulted. Fortunately, having worked with the cataloged land records over the course of several years, it was relatively easy to ascertain which they
had employed and link them accordingly. To identify additional collections the MARS catalog was searched to locate land records that fell outside the scope of Dobbs' and McNeely’s investigations.

Initially, the intent had been to model the findings using a literature guide\textsuperscript{37}, a method Hjørland (2002) believes has been overlooked for its value in understanding the “structures of information sources from a systems point of view.” User studies depend upon knowing “the most important information sources … at a rather detailed level.” In conclusion, after assessing the complex hierarchy of the MARS catalog system and the different indexing approaches employed between collections, it was determined that a descriptive analysis of the current search tool (MARS) and quantitative assessment of each collection’s content would provide similar context and be better suited for classifying the findings.

To this end a ‘first tier’\textsuperscript{38} inventory of the cataloged land records, akin to the published abstracts discussed earlier, was prepared for each collection that in turn was consolidated into a statewide index of early land transactions in an Excel Workbook format. Two versions were prepared: one arranged by county, then sorted by grantee last name, the other was sorted by grantee surname. The inventory enabled taking stratified-samples, as necessary, of each county’s records, sorted by date and employing a sample size formula\textsuperscript{39} to insure 90% reliability. The item level descriptions provided by the catalog were then examined to facilitate estimating the overall size, by county, of the domain of resources.

B. What types of evidence are investigators extracting from the domain? (The Domain of Evidence)

A second key facet of the work domain, the \textit{domain of evidence}, consists of the information extracted by investigators from the domain of resources for incorporation into their own knowledge base (i.e., the database) to support their investigations. Assuming the data has been suitably structured and its source material verified, the new content produced from a manually intensive process of parsing and transcribing data from 18\textsuperscript{th} century hand written documents into a digital

\begin{itemize}
\item\textsuperscript{37} Typically a single page guide describing essential reference material germane to a specific topic.
\item\textsuperscript{38} The last view provided in MARS containing a list of individual land record folders in each box, as shown in figure 4.9.
\item\textsuperscript{39} The formula employed was: \(n=N\frac{1}{1+N(e)^2}\) where \(n=\)sample size, \(N=\)population size, and \(e=\)precision. Sample size precision was set at 90\% (.10) or greater.
\end{itemize}
format enables complex computational analysis as well as the facility for sharing across a broader computing environment.

Scope/Unit of Analysis

The individual records contained in Dobbs’ and McNeely’s databases along with their corresponding MARS catalog records served as the unit of analysis for comparison. While each record in the investigators’ databases represents an individual land record, the equivalent MARS catalog record describes the content of a folder that may contain one or more related documents generated during the process of acquiring a tract of land. Because McNeely’s data appeared to correspond to deed records only, unlike Dobbs’ database that included entry, warrant, and survey records along with a few miscellaneous transactions, the scope was limited to an examination of deed records to insure similar data was employed in the comparison. In addition, while Dobbs’ data was drawn from colonial era transactions recorded between 1748 and 1764, McNeely’s records included transactions processed up to the mid-1800s, which exceeded the time frame under study. Therefore, the examination of McNeely’s records for this analysis was limited to documents containing dates of 1780 or earlier.

Methodology

The purpose of this inquiry was to quantify the central terms investigators had extracted from individual land records to assess the types of information they deemed essential for supporting their map-based investigations. Both the similarities and variations between McNeely’s and Dobbs’s database records were examined and compared to the index terms employed in the corresponding MARS records. The objective was to establish how the data needs of an academic researcher (Dobbs) might differ from those of a family historian and chemistry professor emeritus (McNeely), and whether the MARS catalog (i.e., its index terms and search features) supports or thwarts the tasks associated with such land-based research. This is in keeping with Hjørland’s (2002, p. 429) belief that by studying aspects of “indexing, document representation, and retrieval [one] should be able to evaluate bad practices and improve them.”

40 During the domain analysis it was discovered that, while the bulk of McNeely’s parsed database records represented information culled from deed records, several of his records appeared to have been drawn from survey documents which lacked information pertaining to the deed date.
Because Dobbs’ and McNeely’s databases contained thousands of individual records a stratified sample from each was selected for use in the comparisons. This involved sorting the individual records of each database into a logical format (Babbie, 2001, pp. 201-204). Dobbs’ data was first sorted by county then by the earliest date associated with each document. McNeely’s data falling within the study’s time period were sorted by date since the records he had transcribed were associated with one County: Burke. Using the same formula employed in the previous question as a guideline a representative sample from each county was then selected to insure 90% reliability. The respective samples were then matched to their equivalent MARS record by searching the master index prepared for use in the preceding analysis of the domain of resources.

As “indexing is open to study by different methods”, a well suited but less frequently employed quantitative content analysis (Babbie, 2001, pp. 304-322) model was employed to present the findings. This technique was similar to the approach used by Yakel & Kim (2003/2004) in their study of state archive content on the Web. It also builds upon the Ruvane & Dobbs (2008) book chapter, which was part of an edited publication exploring the synergies between geography and genealogy.

C. What types of uncertainty are associated with the domain of evidence collected? (The Domain of Uncertain Evidence)

The domain of uncertain evidence is an important characteristic to understand; aspects of uncertainty are of primary concern in humanities related research and require special scrutiny in order to inform the design of digital humanities tools worth sharing across distributed networks. Both the types of uncertainty as well as their origins are often unclear as many of the tools adapted for use by individual researchers, or teams, are often not designed for capturing such nuances in a format suitable for efficiently searching across these important domain characteristics. As a result, the ability to reuse existing digital content is oftentimes limited; this omission tends to mask debatable circumstances, considered the crux and objective of many humanities investigations to foster discussion and enable further scrutiny of doubtful content.

While some uncertainty can be attributed to human transcription errors (transposed letters, mistyped words), a common problem involves accurately deciphering 18th century handwritten
documents rife with poor penmanship and archaic terminology or notations. There are problems associated with missing content as a result of document damage or loss as well as unfaithful transcriptions where self-defined standardized names have been substituted for the originals; substitution is considered a detrimental recording practice not recommended in historical research (Boonstra et al., 2004, p. 46). There are instances where only part of the content from an original document is made available. As corroborated in findings from the Census Holdings (ReACH) Project (Terras, 2009) this is not only misleading but limits such a resource’s usefulness in support of future research. Allen Renear’s remark, cited by Borgman (2009), sums up the problem by pointing out that “[i]n the humanities, one person’s data is another’s theory.” Users and archival institutions both play a role in introducing uncertainties of this nature when failing to appreciate that inclusive information, authentic facsimiles, and compete transcriptions are demanded.

One of the more complicated issues of uncertainty relates to characteristics inherently associated with time and space, for example how does one deal with alternate names and variant spellings that define the same place or person, gradual shifts in county, township, and related boundary lines, and well as other ambiguities typically encountered within this domain of evidence? Another type of uncertainty relates to the relative geographic accuracy of land record evidence, as transformed into new geographic knowledge in the form of a reconstructed parcel map. While an original rendition of a tract of land can be derived from a well-preserved survey record (i.e., transformed into a digitally precise polygon or ESRI shapefile), its original location in space may not be as easily deduced, thereby leaving doubt if not properly called out in the map itself. There are also inherited uncertainties carried over from the textual transcription process, which may further affect the accuracy of the visual representation.

Issues surrounding the characteristics of data uncertainty directly influence decisions in humanities computing projects and have been broadly discussed in several works (Boonstra et al., 2004; McCrank, 2001). Research surrounding more specific aspects of uncertainty cover a wide gamut including name-matching (Borgman & Siegfried, 1992; Galvez & Moya-Anegón, 2007; Wacholder, Ravin, & Choi, 1997), date and time disambiguation (LOC, 2010; Mostern & Johnson, 2008), indexing (Bates, 1998; Tibbo, 1994), thesauri (Janowicz & Kessler, 2008; Lopez-Huertas,
1997), gazetteers (Buckland et al., 2007; Hill, 2000; Janowicz & Kessler, 2008; Smith & Crane, 2001), and geographic data (Fisher, 1999; Jessop, 2008; Kemp & Mostern, 2001; Plewe, 2002, 2003; Stefanakis, 2003; Zhang & Goodchild, 2002). These and similar studies provide a base line for comparison for examining analogous characteristics of fuzzy, uncertain, or incomplete data.

**Scope/Unit of Analysis**

To understand the domain of uncertain evidence the previous question’s stratified samples of Dobbs’ and McNeely’s database records, and corresponding MARS records, were employed. Initially intended to compare commonly recognized types of uncertainty encountered between the evidence collected by each investigator a discovery during the analysis process changed the focus to an examination of references to related documents within each system (e.g., Dobbs’, McNeely’s, and MARS) to ascertain uncertainties associated with catalog deficiencies. Both the change in approach and methodology employed are described within the results section.

**D. What are the tasks involved in working with, extracting, or transforming evidence into geographic knowledge? (The Domain of Tasks)**

The domain of tasks is an important characteristic to understand about this community’s information use behaviors as it has strong implications for system design; particularly in relation to interface considerations and automated processes. By recognizing the requisite activities associated with reconstructing historical parcel maps and their evaluation, patterns are expected to emerge pointing to problems and opportunities in these areas. To examine the domain of tasks the process employed by Dobbs for gathering evidence and the utility of the database employed were described using a historical case study descriptive analysis that drew from my past observations and communications with Dobbs during the course of our three plus year collaboration to develop the database prototype.
CHAPTER 4: RESULTS

The Domain of Resources

Characteristics

The NCSA Manuscript and Archives Reference System (MARS) provides descriptions of the various collections preserved and administered by the Archives. The first rendition of this electronic catalog system was introduced to the public in April of 1990 and appears to have been a legacy system developed in-house. When first launched MARS provided searchable access via a telnet connection to a mere fraction of the Archives’ cataloged holdings; now these holdings encompass over 50,000 lineal feet of documents representing millions of individual items. Today, patrons access the catalog through a web-based Internet application and, in most cases, can examine the physical objects described, one box or folder at a time, in a secured research area provided on the fourth floor of the NCSA; to date, only a handful of original documents have been scanned to allow for online viewing.

Regrettably, while the colonial land records of North Carolina physically exist they have been removed from circulation for preservation purposes; in their place, microfilmed copies are made available on reels for use onsite, or for purchase for use elsewhere. The filmed documents are arranged on each reel by county, then by grantee surname, although at times they may be out of sequence. One of the many problems associated with this surrogate format, aside from the poor quality of the microfilmed images, is that the reels do not reference a document’s corresponding MARS ID, an essential call number for use in future retrieval and source authentication. Having omitted the ID during the filming process leaves doubt as to the record being viewed as well as uncertainty as to its relationship with the documents filmed alongside it; this is a barrier to information that will be discussed further in the domain of uncertain evidence results section.

From a laymen’s viewpoint the NCSA holdings are arranged in what one might consider a non-intuitive hierarchal structure (i.e., by Classes, Record Groups, Series, and ‘sub-Series’ or child records). However, this method conforms to archival practice, which dictates that where possible and reasonable accessioned record collections be preserved and organized as originally intended by its
creator(s) to maintain the provenance, or history of ownership. As was described in chapter 2, multiple administrations were involved with the distribution of land in North Carolina during its colonial settlement. Lord Granville’s agents handled the bulk of transactions in the northern half of the state while Crown agents handled the southern portion, although some overlap occurred during the years and other administrations were involved. Unsurprisingly, each administration employed similar application procedures and recording practices.

Unlike libraries, archives do not catalog, arrange and shelve their materials according to subject categories. At the NCSA related materials are organized, filed, labeled, and boxed in sequence under a creator’s name, which are assigned a class type, record group number and, when a collection is extensive multiple series numbers. Individual documents within a collection are rarely defined, such as a book might be in a library’s catalog (i.e., by author, date, title, etc…). Instead, finding aids are made available that broadly describe a collection and its assorted content, although not every collection is afforded such an aid. Fortunately for this study, three of the four critical collections containing colonial era land records were found to be described by both a collection overview as well as item level descriptions of each folder’s content; a level of cataloging uncharacteristically afforded in archival practice.

The current web-based version of MARS, introduced within the last few years, offers various methods for searching the NCSA collections (Figure 4.1). Regrettably, based on firsthand experience with the previous system, the new application is a disappointing replacement as its navigational system is both cumbersome and time-consuming. Its one advantage is that it is far easier to copy and paste a collection’s individual child records into tables, providing a method not offered in MARS for sorting and analyzing large and scattered datasets, such as North Carolina’s early land records; tools for consolidating and examining related records are necessary for enabling humanities computing investigations.
The MARS system was built around a hierarchal text-based numbering scheme, the MARS ID, which facilitated browsing, retrieval, sorting, and display of the cataloged records as arranged sequentially by collection. The MARS ID is less prominently displayed in today’s system, despite its importance as a reference number for authenticating original source material. Today, to surf through the catalog requires clicking on the browse button located on the basic search screen to retrieve a list of high level subject categories (Figure 4.2). Navigation is then accomplished by selecting a topic to expand using the plus buttons and wandering through the underlying hierarchy until reaching a collection title of interest which may or may not be linked to a finding aid.

To appreciate the browse sequence see figures 4.3 through 4.5 which illustrate the various expanded views encountered following a direct route to one of the land record collections transcribed by Dobbs: the Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey. Keep in mind that locating this collection required prior knowledge of the catalog system’s hierarchy and naming conventions, a serendipitous route would likely have taken considerably longer.
Figure 4.2 Browse button displays a list of general collection categories. The plus buttons activate expanded lists under each category. NC Land records are part of the State Records collection.

Navigating to the collection entailed expanding six titles, starting with the State Records category (Figure 4.2) before reaching the lowest level within the record series, an item level description of the content of one folder (Figure 4.5). Along this route separate finding aids become available, beginning with the Secretary of State Record Group heading (Figure 4.3), which can be launched by clicking on the title's name. Each finding aid provides a unique overview of the ever-narrowing scope of the collection's underlying record series.

---

41 Sequence of titles expanded to reach the lowest record in a Record Group Series included: 1) State Records, 2) Governor's Papers -- William Woods Holden (24 November 1818 -- 1 March 1892) ... Youth Services, 3) Secretary of State Record Group, 4) Accounts ... Voter Registration: Voter Registration Records, 5) Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey, and 6) Anson County, A-E.
Figure 4.3 View of the State Records collections. Expanding the lists under the topics titled “Governor’s Papers...”, Secretary of State Record Group, and “Accounts...” leads to one of the collections used by Dobbs, the Granville Proprietary Land Office: Land Entries, Warrants, and Plats of survey. This Series can be expanded for further examination.

Figure 4.4 Expanded view of the Granville Proprietary Land Office: Land Entries, Warrants, and Plats of survey Record Group Series. To the right is an expanded view of one of the Record Group Series within this collection, titled Anson County, A-E.
Figure 4.5 Item level description of the content of an individual folder within the Anson County, A-E Record Series, this one titled Aaron, Peter. Anson Co.

The browse feature is not the only method provided for searching the catalog. Another option available is by keywords using the Search Text box located on the basic search screen. There is also an advanced search screen that offers additional browse and filtering features, including the facility to display various index terms employed within the system (Figure 4.1; right). Many of the search alternatives are described in the online MARS help pages. An alternate method, not discussed in the help pages, is the search by MARS ID option, located in the pull down menu on the search by call number screen (Figure 4.1; lower left). This alternative offers the most efficient method for facilitating direct access to specific collections or individual finding aids of interest and is particularly useful for
dealing with large homogeneous record sets cataloged under multiple series numbers such as the NCSA land records.

For example, the same Granville land record collections, navigated to by means of the browse feature, can be readily accessed with a basic understanding of the hierarchal numbering scheme employed in MARS (Table 4.1). Entering any of the four MARS IDs presented below into the search field will return a view of the record title requested, along with general details about its scope (Figure 4.6). By clicking within the title row an associated finding aid will open in a new window providing more detailed information about the documents it encompasses (Figure 4.7). Navigation to successively lower levels within the series hierarchy, referred to as child records, can be accomplished by activating the "show list of child records" link located on the bottom left corner of each finding aid.

Despite the simplicity of the MARS ID for effecting a search there is no facility for examining the various land record collections in their entirety, let alone together, from within the catalog. For example, Series 12.12 encompasses 5,629 folders, stored in 82 boxes, but its cataloged content can only be viewed one 'box' at a time. Therefore, to facilitate further analysis an inventory was required.

<table>
<thead>
<tr>
<th>MAR ID (type)</th>
<th>TITLE</th>
<th>Child Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 (Record Group)</td>
<td>Secretary of State Record Group</td>
<td>124</td>
</tr>
<tr>
<td>12.12 (Series)</td>
<td>Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey</td>
<td>82</td>
</tr>
<tr>
<td>12.12.1 (Box)</td>
<td>Anson County, A-E</td>
<td>83</td>
</tr>
<tr>
<td>12.12.1.1 (Folder)</td>
<td>Aaron, Peter. Anson Co.</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.1 Example of the hierarchal numbering scheme employed in MARS
Figure 4.6 Search by MARS ID 12.12: First screen displays the cataloged record’s title and general details about its scope. Shown: *Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey*.

Figure 4.7 Finding aid for MARS ID 12.12: *Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey*. Note the Show List of Child Records link on the lower left of the screen.
Figure 4.8 Child Records of MARS ID 12.12: Activating the ‘Show List of Child Records’, located on the finding aid, returns the list of boxed items associated with this Series, arranged by county. The child count represents the number of folders in each box.

Figure 4.9 Child Records of MARS ID 12.12.1: Activating the ‘Show List of Child Records’ located on the finding aid returns a list of folder titles filed within one Box of this Series, arranged by grantee surname. This represents the lowest level of child records in the Series.
Figure 4.10 MARS ID 12.12.1.1: Example of folder level details provided in the catalog. Clicking on the individual title listed on the previous screen activates this screen.

**Inventory**

The MARS catalog was thoroughly examined using a range of search techniques. From this search four series were identified as containing relevant material for potential inclusion in the Domain of Resources and are presented in table 4.2. These cataloged collections, housed at the NCSA, represent what remains of the earliest land records generated by various administrations and serve to document the process of land distribution in North Carolina. The first two series (12.12, 12.13) were generated by agents of the Granville Proprietary Land Office; the remaining two (12.14, 12.15),
hearafter referred to as the State Land Office Series, comprise paperwork prepared by various Crown and State agents up through and including the year 1959.

<table>
<thead>
<tr>
<th>MARS ID</th>
<th>TITLE</th>
<th>TOTAL FILM + CHILD RECORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Reels</td>
</tr>
<tr>
<td>12.12</td>
<td>Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey (Yrs: 1748-1763)</td>
<td>14</td>
</tr>
<tr>
<td>12.13</td>
<td>Granville Proprietary Land Office: Granville Grants of Deed (Yrs: 1748-1763)</td>
<td>19</td>
</tr>
<tr>
<td>12.15</td>
<td>Land Office: Patent Books (Land Grant Record Books) (Yrs: 1683-1959)</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 4.2 Four Record Series considered for inclusion in the Domain of Resources

To determine which records from each collection to include or exclude in the domain of resources, a consolidated inventory of the relevant folders was prepared. Series 12.15 was excluded as it was not described to the item level and as such will be discussed separately along with the method used for estimating its folder count shown in table 4.2. The master index or final inventory (Table 4.3) included a list of every folder that contained at least one document dated 1790 or earlier and represented land transactions within the bounds of present day North Carolina.

Preparing the inventory entailed cutting and pasting folder level records associated with 220 different boxes or record series from three different series and parsing conflated data into individual fields to facilitate sorting. This was followed by removing any folders that represented transactions initiated in or after 1791, resulting in a master index containing 70,888 cataloged folder records. Because the final inventory encompassed well over 1000 letter-sized pages, it has not been included as an Appendix item but is available upon request for non-commercial use.

Stratified samples (as described in Chapter 3) were employed for two of the four series analyzed, Series 12.13 and 12.14. A different approach was employed for analyzing series 12.12 and 12.15, as defined in the results section.
<table>
<thead>
<tr>
<th>MARS ID</th>
<th>PRFX</th>
<th>LAST NAME</th>
<th>FIRSTNAME</th>
<th>MID_NAME</th>
<th>SUFX</th>
<th>OTHER NAME</th>
<th>COUNTY</th>
<th>YR1</th>
<th>YR2</th>
<th>YR3</th>
<th>YR4</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.12.25.5</td>
<td>Addaman</td>
<td>Thomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Granville</td>
<td>1753</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.13.31.28</td>
<td>Addaman</td>
<td>Thomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Granville</td>
<td>1760</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.66.24</td>
<td>Addamans</td>
<td>Thomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Granville</td>
<td>1760</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.27.4364</td>
<td>Addams</td>
<td>John</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Anson</td>
<td>1779</td>
<td>1783</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.44.182</td>
<td>Addams</td>
<td>Thomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chatham</td>
<td>1779</td>
<td>1780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.46.314</td>
<td>Adderly</td>
<td>John</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chowan</td>
<td>1720</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.57.1830</td>
<td>Adkinson</td>
<td>Will</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Duplin</td>
<td>1779</td>
<td>1780</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.57.1734</td>
<td>Adkinson</td>
<td>William</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Duplin</td>
<td>1778</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.107.1191</td>
<td>Admire</td>
<td>Elizabeth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rowan</td>
<td>1778</td>
<td>1783</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.50.2249</td>
<td>Aeroman</td>
<td>Frederick</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Craven</td>
<td>1757</td>
<td>1759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.56.13</td>
<td>Aery</td>
<td>Isaac</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Granville</td>
<td>1756</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.41.958</td>
<td>Aglisby</td>
<td>Thomas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Carteret</td>
<td>1788</td>
<td>1790</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.88.122</td>
<td>Agnew</td>
<td>Robert</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Guilford</td>
<td>1778</td>
<td>1779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.88.153</td>
<td>Agnew</td>
<td>Robert</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Guilford</td>
<td>1779</td>
<td>1779</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.88.456</td>
<td>Agnew</td>
<td>Robert</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Guilford</td>
<td>1779</td>
<td>1782</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.82.658</td>
<td>Akerman</td>
<td>Daniel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lincoln</td>
<td>1786</td>
<td>1791</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.12.55.4</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1757</td>
<td>1760</td>
<td>1761</td>
<td></td>
</tr>
<tr>
<td>12.12.55.5</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1761</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.12.55.6</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1761</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.12.69.2</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1757</td>
<td>1760</td>
<td>1761</td>
<td></td>
</tr>
<tr>
<td>12.12.69.3</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1761</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.12.83.3</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1762</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.14.95.308</td>
<td>Akin</td>
<td>Joseph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Orange</td>
<td>1762</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3 Example of content and format of consolidated inventory -- Alphabetically Sorted. (Contains over 1,000 8.5” x 11” pages that identify 70,888 individual MARS catalog folders.)
Results

Understanding the scale, or size, of the domain of resources is central to the design of a collaborative system; it is important to ascertain the computing resources necessary for interlinking catalog records, document images, and reconstructed neighborhood maps in a virtual environment. The results presented encompass a major segment of this universe: the land records required to guide in the reconstruction or validation of early colonial neighborhood maps. While not included in the analysis, a variety of digital maps and related supporting materials are integral to the domain of resources. Their role will be discussed in the domain of tasks.

While all three of the series analyzed were found to be cataloged using a standardized finding aid (Figure 4.10), the fields populated and data entry conventions varied within and across collections. A variation of this finding aid was employed in the State Land Office’s Series 12.14, which included a separate summary section and identified the patent book and page number where a reproduced copy of the deed could be found (Figure 4.11). Because of the variations between finding aid content, individual tables were employed to present the results. Regardless of the collection analyzed, each cataloged folder turned out to minimally hold one item; some contained just an empty shuck but most held two or more documents. Intermittently throughout the analysis typographical errors were encountered within the catalog records and where obvious were corrected.

![Land Grant Information Table]

**Figure 4.11 Series 12.14 finding aid variation: Land grant details incorporated below the standardized form**

The first two series analyzed, shown in Tables 4.4 and 4.5, encompassed records produced by the Granville Proprietary Land Office between 1748 and 1763: the *Land Entries, Warrants, and Plats of Survey* (Series 12.12) and their corresponding *Grants of Deed* (Series 12.13). As discussed
in Chapter 2, while sequentially related records were originally filed together prior to 1909, the Secretary of State Office chose to reorganize the paperwork by separating the deeds from their supporting documents (e.g., entry, warrants, surveys) and placing them into shucks, or envelopes, which were then placed into folders. External notations were recorded on each shuck indicating their content and relationship to other documents. When the Granville records were transferred to the NC State Archives, they continued the practice of disassembling related land records that had not yet been processed and, presumably sometime after their receipt, began putting the shucks into separate folders and assigned each a MARS ID.

![Table 4.4 MARS Record Series 12.12: Estimated number of documents related to the Domain of Resources]

<table>
<thead>
<tr>
<th>County</th>
<th>Year Formed</th>
<th>Series &amp; Sample Size</th>
<th>Years referenced per folder</th>
<th>Extended Test Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 yr</td>
<td>2 yrs</td>
<td>3+ yrs</td>
</tr>
<tr>
<td>Anson</td>
<td>1750</td>
<td>264</td>
<td>156</td>
<td>93</td>
</tr>
<tr>
<td>Beaufort</td>
<td>1705</td>
<td>50</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>Bertie</td>
<td>1722</td>
<td>106</td>
<td>56</td>
<td>45</td>
</tr>
<tr>
<td>Bladen</td>
<td>1734</td>
<td>56</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>Bute</td>
<td>1764</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Chowan</td>
<td>1668</td>
<td>17</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Cumberland</td>
<td>1754</td>
<td>7</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Currituck</td>
<td>1668</td>
<td>107</td>
<td>70</td>
<td>31</td>
</tr>
<tr>
<td>Dobbs</td>
<td>1758</td>
<td>53</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>Edgecombe</td>
<td>1741</td>
<td>884</td>
<td>394</td>
<td>433</td>
</tr>
<tr>
<td>Granville</td>
<td>1752</td>
<td>1,152</td>
<td>551</td>
<td>525</td>
</tr>
<tr>
<td>Hallifax</td>
<td>1758</td>
<td>246</td>
<td>173</td>
<td>69</td>
</tr>
<tr>
<td>Hertford</td>
<td>1759</td>
<td>7</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Hyde</td>
<td>1705</td>
<td>7</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>Johnston</td>
<td>1746</td>
<td>440</td>
<td>194</td>
<td>214</td>
</tr>
<tr>
<td>Northampton</td>
<td>1741</td>
<td>238</td>
<td>93</td>
<td>117</td>
</tr>
<tr>
<td>Orange</td>
<td>1752</td>
<td>1,117</td>
<td>608</td>
<td>397</td>
</tr>
<tr>
<td>Pasquotank</td>
<td>1668</td>
<td>151</td>
<td>93</td>
<td>48</td>
</tr>
<tr>
<td>Perquimans</td>
<td>1679</td>
<td>19</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Pitt</td>
<td>1760</td>
<td>56</td>
<td>53</td>
<td>3</td>
</tr>
<tr>
<td>Roxan</td>
<td>1753</td>
<td>587</td>
<td>384</td>
<td>170</td>
</tr>
<tr>
<td>Tyrrell</td>
<td>1729</td>
<td>64</td>
<td>43</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4.4 MARS Record Series 12.12: Estimated number of documents related to the Domain of Resources

The first collection analyzed was Series 12.12, which contained 5,629 folders. It was expected that each folder held up to three, at the most four, documents representing the complete set of paperwork generated prior to the issuance of a grant of deed. This series was evaluated in its entirety without the use of a stratified sample as it was believed that the cataloged dates provided sufficient information for estimating the number of documents within each folder; the assumption being that the different years noted represented a single document (see figure 4.9). The resulting
estimate (Table 4.4) suggests that 8,631 individual documents are contained within Series 12.12, an average of 1.5 documents per folder. The bulk of the folders (56%) were associated with Edgecombe (16%), Granville (20%), and Orange (20%) Counties. In hindsight, the use of a stratified sample may have been more appropriate as it was later discovered that two or more documents could have been recorded in the same year. This concern is examined further in the discussion section.

### Table 4.5 MARS Record Series 12.13: Estimated number of documents related to the Domain of Resources

<table>
<thead>
<tr>
<th>Year Formed</th>
<th>Ttl Folders In:</th>
<th># Deed copies per folder</th>
<th>Extended Ttl est items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Series</td>
<td>Sample</td>
<td>1</td>
</tr>
<tr>
<td>Anson</td>
<td>1750</td>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>Beaufort</td>
<td>1705</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>Bertie</td>
<td>1722</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Bladen</td>
<td>1734</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Chowan</td>
<td>1668</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Cumberland</td>
<td>1754</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Currituck</td>
<td>1668</td>
<td>86</td>
<td>71</td>
</tr>
<tr>
<td>Dobbs</td>
<td>1758</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Edgecombe</td>
<td>1741</td>
<td>819</td>
<td>274</td>
</tr>
<tr>
<td>Granville</td>
<td>1752</td>
<td>1,085</td>
<td>292</td>
</tr>
<tr>
<td>Halifax</td>
<td>1758</td>
<td>199</td>
<td>131</td>
</tr>
<tr>
<td>Hyde</td>
<td>1705</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Johnston</td>
<td>1746</td>
<td>303</td>
<td>172</td>
</tr>
<tr>
<td>Northampton</td>
<td>1741</td>
<td>205</td>
<td>136</td>
</tr>
<tr>
<td>Orange</td>
<td>1752</td>
<td>811</td>
<td>269</td>
</tr>
<tr>
<td>Pasquotank</td>
<td>1668</td>
<td>176</td>
<td>122</td>
</tr>
<tr>
<td>Perquimans</td>
<td>1679</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Pitt</td>
<td>1760</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Rowan</td>
<td>1753</td>
<td>626</td>
<td>244</td>
</tr>
<tr>
<td>Tyrrell</td>
<td>1729</td>
<td>165</td>
<td>117</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4,945</td>
<td>2,293</td>
<td>1,396</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3,486</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8,556</td>
</tr>
</tbody>
</table>

The second collection analyzed was the *Granville Grants of Deed*, Series 12.13, which contained 4,945 folders (Table 4.5). These folders contain only deed documents that are typically accompanied by a survey addendum. Presumably if three copies of a deed are found within a folder the grant ‘was not ripened’, in other words the transaction was never completed and the claim abandoned for one reason or another; this ‘not ripened’ statement was found to be not always accurate. In this analysis a stratified sample was employed to enable a detailed analysis of the individual folder content. The sample size is noted in a separate column for each county. In certain cases where the sample size was small, or if a series had been previously examined, 100% of the

---

Deeds were prepared in triplicate, one each for distribution to the grantee and Lord Granville, and the third retained as a file copy in the Granville Proprietary Land Office.
item-level descriptions were examined. Based on the stratified sample an estimated 8,556 individual deeds are contained within the 12.13 series, an average of 1.7 deed copies per folder. The bulk of the folders (68%) were associated with Edgecombe (17%), Granville (22%), Orange (16%), and Rowan (13%) Counties.
<table>
<thead>
<tr>
<th>County</th>
<th>Year Formed</th>
<th>Series 1 Folders in:</th>
<th># yrs referenced per folder</th>
<th>Sample Size</th>
<th># of Documents per folder</th>
<th>Extended Ti est Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anson</td>
<td>1770</td>
<td>8,622</td>
<td>6,500</td>
<td>3,675</td>
<td>1,399</td>
<td>6</td>
</tr>
<tr>
<td>Ashe</td>
<td>1799</td>
<td>5,725</td>
<td>600</td>
<td>65</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Bath</td>
<td>1696</td>
<td>490</td>
<td>185</td>
<td>178</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>Beaufort</td>
<td>1705</td>
<td>2,317</td>
<td>1,062</td>
<td>310</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bertie</td>
<td>1722</td>
<td>1,609</td>
<td>793</td>
<td>211</td>
<td>5</td>
<td>236</td>
</tr>
<tr>
<td>Bladen</td>
<td>1739</td>
<td>9,333</td>
<td>4,636</td>
<td>3,211</td>
<td>1,425</td>
<td>368</td>
</tr>
<tr>
<td>Brunswick</td>
<td>1764</td>
<td>3,522</td>
<td>419</td>
<td>259</td>
<td>160</td>
<td>205</td>
</tr>
<tr>
<td>Buncombe</td>
<td>1791</td>
<td>6,532</td>
<td>25</td>
<td>25</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>Burke</td>
<td>1777</td>
<td>7,056</td>
<td>1,450</td>
<td>351</td>
<td>1,089</td>
<td>314</td>
</tr>
<tr>
<td>Burke</td>
<td>1777</td>
<td>7,056</td>
<td>1,450</td>
<td>351</td>
<td>1,089</td>
<td>314</td>
</tr>
<tr>
<td>Cabarrus</td>
<td>1702</td>
<td>807</td>
<td>12</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Camden</td>
<td>1777</td>
<td>235</td>
<td>49</td>
<td>34</td>
<td>-</td>
<td>39</td>
</tr>
<tr>
<td>Carteret</td>
<td>1722</td>
<td>2,217</td>
<td>919</td>
<td>728</td>
<td>190</td>
<td>1</td>
</tr>
<tr>
<td>Caswell</td>
<td>1777</td>
<td>1,328</td>
<td>952</td>
<td>966</td>
<td>-</td>
<td>294</td>
</tr>
<tr>
<td>Chatham</td>
<td>1771</td>
<td>2,412</td>
<td>936</td>
<td>61</td>
<td>975</td>
<td>99</td>
</tr>
<tr>
<td>Chowan</td>
<td>1668</td>
<td>1,172</td>
<td>1,035</td>
<td>803</td>
<td>230</td>
<td>2</td>
</tr>
<tr>
<td>Craven</td>
<td>1705</td>
<td>6,376</td>
<td>4,442</td>
<td>4,009</td>
<td>425</td>
<td>8</td>
</tr>
<tr>
<td>Cumberland</td>
<td>7,725</td>
<td>2,524</td>
<td>1,738</td>
<td>705</td>
<td>1</td>
<td>95</td>
</tr>
<tr>
<td>Currituck</td>
<td>1658</td>
<td>1,023</td>
<td>405</td>
<td>283</td>
<td>122</td>
<td>201</td>
</tr>
<tr>
<td>Dobbs</td>
<td>1758</td>
<td>2,455</td>
<td>2,152</td>
<td>1,101</td>
<td>1</td>
<td>96</td>
</tr>
<tr>
<td>Duplin</td>
<td>1750</td>
<td>4,506</td>
<td>2,863</td>
<td>1,847</td>
<td>545</td>
<td>1</td>
</tr>
<tr>
<td>Edgecombe</td>
<td>1741</td>
<td>2,790</td>
<td>2,319</td>
<td>1,397</td>
<td>447</td>
<td>2</td>
</tr>
<tr>
<td>Franklin</td>
<td>1746</td>
<td>2,388</td>
<td>125</td>
<td>0</td>
<td>118</td>
<td>96</td>
</tr>
<tr>
<td>Gaston</td>
<td>1779</td>
<td>184</td>
<td>37</td>
<td>6</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>Glasgow</td>
<td>1791</td>
<td>118</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Granville</td>
<td>1752</td>
<td>1,757</td>
<td>1,580</td>
<td>1,122</td>
<td>458</td>
<td>94</td>
</tr>
<tr>
<td>Guilford</td>
<td>1752</td>
<td>2,480</td>
<td>1,498</td>
<td>185</td>
<td>1,313</td>
<td>94</td>
</tr>
<tr>
<td>Halifax</td>
<td>1758</td>
<td>662</td>
<td>391</td>
<td>208</td>
<td>183</td>
<td>80</td>
</tr>
<tr>
<td>Hertford</td>
<td>1759</td>
<td>270</td>
<td>72</td>
<td>30</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>Hyde</td>
<td>1705</td>
<td>1,461</td>
<td>621</td>
<td>449</td>
<td>172</td>
<td>85</td>
</tr>
<tr>
<td>Iredell</td>
<td>1788</td>
<td>971</td>
<td>125</td>
<td>125</td>
<td>-</td>
<td>56</td>
</tr>
<tr>
<td>Johnston</td>
<td>1746</td>
<td>4,310</td>
<td>2,664</td>
<td>1,774</td>
<td>799</td>
<td>1</td>
</tr>
<tr>
<td>Jones</td>
<td>1779</td>
<td>1,161</td>
<td>209</td>
<td>16</td>
<td>193</td>
<td>58</td>
</tr>
<tr>
<td>Lenoir</td>
<td>1791</td>
<td>551</td>
<td>11</td>
<td>-</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>Lincoln</td>
<td>1779</td>
<td>4,307</td>
<td>242</td>
<td>50</td>
<td>192</td>
<td>71</td>
</tr>
<tr>
<td>Martin</td>
<td>1774</td>
<td>646</td>
<td>318</td>
<td>26</td>
<td>292</td>
<td>26</td>
</tr>
<tr>
<td>Montgomery</td>
<td>1798</td>
<td>4,839</td>
<td>386</td>
<td>40</td>
<td>345</td>
<td>79</td>
</tr>
<tr>
<td>Moore</td>
<td>1774</td>
<td>5,214</td>
<td>56</td>
<td>11</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>Nash</td>
<td>1774</td>
<td>761</td>
<td>303</td>
<td>401</td>
<td>254</td>
<td>75</td>
</tr>
<tr>
<td>Mecklenburg</td>
<td>1762</td>
<td>7,092</td>
<td>3,927</td>
<td>2,649</td>
<td>1,246</td>
<td>2</td>
</tr>
<tr>
<td>New Hanover</td>
<td>1759</td>
<td>5,295</td>
<td>2,558</td>
<td>1,976</td>
<td>581</td>
<td>1</td>
</tr>
<tr>
<td>Northampton</td>
<td>1724</td>
<td>4,688</td>
<td>339</td>
<td>262</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>Onslow</td>
<td>1734</td>
<td>4,243</td>
<td>1,755</td>
<td>1,130</td>
<td>624</td>
<td>1</td>
</tr>
<tr>
<td>Orange</td>
<td>1752</td>
<td>3,131</td>
<td>1,767</td>
<td>912</td>
<td>853</td>
<td>2</td>
</tr>
<tr>
<td>Pascocauk</td>
<td>1668</td>
<td>581</td>
<td>419</td>
<td>373</td>
<td>46</td>
<td>204</td>
</tr>
<tr>
<td>Pender</td>
<td>1875</td>
<td>172</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Perquimans</td>
<td>1679</td>
<td>647</td>
<td>417</td>
<td>301</td>
<td>196</td>
<td>204</td>
</tr>
<tr>
<td>Person</td>
<td>1791</td>
<td>106</td>
<td>5</td>
<td>5</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Pitt</td>
<td>1750</td>
<td>1,991</td>
<td>955</td>
<td>280</td>
<td>674</td>
<td>1</td>
</tr>
<tr>
<td>Randolph</td>
<td>1779</td>
<td>3,502</td>
<td>521</td>
<td>108</td>
<td>412</td>
<td>87</td>
</tr>
<tr>
<td>Richmond</td>
<td>1779</td>
<td>3,157</td>
<td>77</td>
<td>39</td>
<td>37</td>
<td>44</td>
</tr>
<tr>
<td>Robeson</td>
<td>1787</td>
<td>2,300</td>
<td>48</td>
<td>45</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>Rockingham</td>
<td>1785</td>
<td>963</td>
<td>72</td>
<td>42</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>Rowan</td>
<td>1755</td>
<td>4,061</td>
<td>2,761</td>
<td>1,011</td>
<td>1,743</td>
<td>7</td>
</tr>
<tr>
<td>Rutherford</td>
<td>1784</td>
<td>4,634</td>
<td>237</td>
<td>79</td>
<td>158</td>
<td>79</td>
</tr>
<tr>
<td>Sampson</td>
<td>1791</td>
<td>2,014</td>
<td>39</td>
<td>29</td>
<td>-</td>
<td>29</td>
</tr>
<tr>
<td>Stokes</td>
<td>1786</td>
<td>1,598</td>
<td>73</td>
<td>3</td>
<td>72</td>
<td>42</td>
</tr>
<tr>
<td>Surry</td>
<td>1770</td>
<td>4,245</td>
<td>1,136</td>
<td>269</td>
<td>862</td>
<td>5</td>
</tr>
<tr>
<td>Tryon</td>
<td>1788</td>
<td>1,968</td>
<td>1,943</td>
<td>997</td>
<td>947</td>
<td>98</td>
</tr>
</tbody>
</table>

Table 4.6 MARS Record Series 12.14: Estimated number of documents related to the Domain of Resources
The third collection analyzed was the *[State] Office Warrants, Plats, and Related Records*, Series 12.14, which in total contains 160,516 folders, but only the 60,314 that fell within this study’s time period and geographic bounds were examined (Table 4.6). These folders contain land-transactions prepared between 1679 and 1790 by the various Agents involved. Like the 12.12 Granville series, finding three sequentially related documents within a folder would be the ideal but was rarely the case. Unlike the previous two series analyzed whose finding aids did not indicate how many documents were in each folder, the quantity field was consistently populated. In cases where a folder only contained a shuck, the quantity field would indicate ‘0 Items’ and include a note explaining “there were no documents in the shuck at time of filming.”

The results of this analysis suggest that there are an estimated 68,968 individual documents within the 12.14 series, an average of 1.4 items per folder. The folders analyzed were associated with 59 different counties, 11 of which (highlighted in blue) were, surprisingly, formed after 1790 yet contained records related to this study’s time period. Even more interesting was the discovery that numerous folders only contain shucks as indicated in the *# of documents per folder* column under the heading of ‘0’. This led to the realization that many of the deed shucks, related to the Granville Proprietary Land Office, had been erroneously filed within this State Land Office record collection; a fault not noted in its associated finding aid.

The last collection analyzed was the *State Land Office: Patent Books*, Series 12.15, which unfortunately was not described at the item level. This collection encompasses 168 patent books that were employed for recording deed transactions as they occurred in sequential date order; a duplicate recording practice still in use today by most county administrations. Manually examining these books and their content, only available on microfilm, was deemed beyond the scope of this study, as the time required to quantify and characterize the documents contained in each volume would have been prohibitive. Instead, a search by text field was employed to effect an estimate.

The estimate was based on a wildcard search for ‘Book No*’ references within the State Land Office collection, series 12.14, which provided a count of the patent book numbers associated with each folder and presumably represented a fairly accurate count. The search identified 187,000 book number references within the entire series, an average of 1.165 deeds per folder. This average was
multiplied by the folders associated with the study (e.g., 60,314) to come up with an estimate of 70,265 deeds. This suggests that the patent books may be the more comprehensive source out of the four series analyzed.

A summary of these findings is presented in table 4.7, which represents the estimated total number of documents deemed relevant for incorporation into the collaborative environment envisioned for a total of 156,420 individual land records whose related sequential transactions contained at least one record with a date of 1790 or earlier.

<table>
<thead>
<tr>
<th>Series Title</th>
<th>Total # of Folders</th>
<th>Estimated Document count</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Granville Proprietary Land Office: Land Entries, Warrants, and Plats of Survey</em> (Yrs: 1748-1763)</td>
<td>5,629</td>
<td>8,631</td>
</tr>
<tr>
<td><em>Granville Proprietary Land Office: Granville Grants of Deed</em> (Yrs: 1748-1763)</td>
<td>4,945</td>
<td>8,556</td>
</tr>
<tr>
<td><em>Land Office: Patent Books (Land Grant Record Books)</em> (Yrs: 1693-1959)</td>
<td>198 volumes</td>
<td>70,265</td>
</tr>
<tr>
<td>TOTALS</td>
<td>171,090</td>
<td>156,420</td>
</tr>
</tbody>
</table>

Table 4.7 The Domain of Resources: Total estimated number of documents in the domain

The Domain of Evidence

*Characteristics*

The second domain characteristic analyzed represents the types of evidence commonly collected by investigators and the tools they employ. The information gathered is vital for guiding subsequent map reconstruction process. To effect this analysis the similarities and differences between Dobbs’ and McNeely’s sampled database content were examined and classified into categories, which in turn were compared to their equivalent item-level MARS records.

Dobbs’ land record evidence was recorded within a custom built Microsoft Access database (Figure 4.12) (Ruvane & Dobbs, 2005). Along with the database Dobbs employed two other programs: one a GIS application for placing, or platting, each tract of land (ArcGIS v9.x by ESRI) the other an editing tool (Lloyd & Kimball, 2000) for generating the shape of each tract for use in this two
dimensional environment. While a consolidated system would have been preferable, the database design offered flexibility not available in existing tools, particularly the ability to classify records, people, and features by common as well as customizable categories (Figure 4.12).

The application employed by McNeely, DeedMapper® by Direct Line Software, combines both a database and a mapping application designed for use by the general public. Although the mapping tool does not offer the same analytical capabilities as a fully functional GIS, its simplicity is appreciated and it fulfils the needs of family historians whose backgrounds and computer skills vary widely. In addition, its price is reasonable at $99.00 dollars in comparison to the cost of purchasing
Microsoft Access 2010 ($139.99)\textsuperscript{43} and ArcGIS for Home Use ($100.00 annual license fee)\textsuperscript{44}; the NWF/DEM Data Editing Extension application employed by Dobbs was free but is now outdated (Lloyd & Kimball, 2000).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{deedmapper.png}
\caption{McNeely’s DeedMapper® Database: Main data entry form}
\end{figure}

DeedMapper® was not designed as an academic research tool, and as such is not well suited for use in a rigorous collaborative environment. This is partly related to its disproportionate use of free text fields containing commingled data and limited number of controlled fields for entering categorized data, which greatly reduces its sorting capabilities and use for in-depth analysis. As for its mapping application, it can only draw straight lines and therefore boundaries that fall along a meandering watercourse are inaccurately rendered. On the positive side, the mapping application employs individual lines to generate each parcel’s shape, thus enabling end points to be linked with the text describing each surveyor’s corner marker for display on screen (e.g, white oak tree, rock outcropping, neighbor’s corner, etc.).

\textsuperscript{43} Source: http://www.microsoftstore.com/store retrieved 10/27/2011
\textsuperscript{44} Source: http://store.esri.com/esri retrieved 10/27/2011

95
**Inventory**

Dobbs’ and McNeely’s sampled database records along with their corresponding MARS records served as the inventory in this analysis. Dobbs’ database, as of July 17, 2010, contained a total of 5,951 transcribed land records of which 2,267 represented deeds. The stratified sample selected from these consisted of 99 deed records, but two were eliminated because a corresponding match to the Domain of Resources index could not be found. McNeely’s database, as of December 30, 2009, contained a total of 1,912 abstracted land records, of which all were identified as deeds but only 100 were germane to the time period under study. The stratified sample selected from these initially consisted of 116 deeds, but 16 were removed as no corresponding catalog number could be found. The sampled records excluded from these analyses are discussed further in the domain of uncertain evidence results.

**Results**

Understanding the types of information extracted by investigators for incorporation into their own knowledge base is central to shaping the design of an efficient tool for use in a collaborative mapping environment. While Dobbs’ and McNeely’s reasons for mapping North Carolina’s early backcountry neighborhoods were decidedly different, the information they gathered to support their endeavors was surprisingly similar, although both did collect types of information not recorded by the other and the level of detail between equivalent categories often varied. These differences are examined in the quantitative analyses presented in Figures 4.14 and 4.15.

The models represent the collective content recorded in the sampled deed records within each database, which were classified into five categories: dates, people, features, county(s), measures, and reference. It was presumed that the data analyzed represented facts found within either the deed body, its survey addendum, or an administrative note that had been recorded on the back (verso) of the deed or shuck that may have accompanied it. Boxes were used to represent the different categories of information collected. Those with dashed lines indicate that the evidence was found recorded within a free text field, meaning that multiple unique facts were typically found commingled together. Solid-lined boxes represent evidence that was recorded separately within a
dedicated field; where occasional commingling was found dashed underlines were used to note these exceptions.
Figure 4.14 Dobbs Database: evidence extracted from 97 sampled deed records. (Dashed underlines indicate commingled data).
The date category included deed dates as well as dates that referred to earlier associated dealings (e.g., entry, warrant, survey, or other related transactions). The people category represented the individuals involved with each land-transaction, as classified by the role they played (grantee, assignee, neighbor, surveyor, chain carrier, attester, or other/unclear). The feature category corresponds to physical elements that would have been visible in the landscape (broadly categorized as animal, cultural, transportation, water, topography, or trees), which are important for narrowing down the likely geographic placement of each tract of land. County(s), an artificial feature, were categorized separately to appreciate the overlapping historic county administrations involved with the distribution of a single parcel. The measures category includes total deed acreage, number of boundary line pairs (i.e., a single length and direction), and the unit of measure depicted in each survey (e.g., chains, poles).

The remaining category, reference, included both internal and external relationships to other records associated with the deeds examined. Internal references represented records within an investigator’s database identified as being related to the deed (i.e., such as its corresponding entry, warrant, or separate survey document). The external references identify the patent book and page number where a duplicate of each deed record can be found, along with its assigned deed and entry number where provided.

Dobbs and McNeely identified a similar number of dates; 195 were found within 97 deed records and 196 were found within 100 respectively, representing an average of two dates per deed. The preponderance represented deed dates: 97, or 100%, in Dobbs’ database and 97, or 97%, in McNeely’s. There were significant differences between the remaining date types recorded. Dobbs had two entry dates, 94 survey dates, and two other dates, compared to McNeely’s 99 entry, zero survey, and 12 other dates.
Figure 4.15 McNeely’s Database: evidence extracted from 100 sampled deed records. (Dashed boxes indicated commingled data.)
The lack of survey dates in McNeely’s database may be reflective of DeedMapper’s limitations, as it provides only two fields for recording date information. The “other” dates shown in McNeely’s model were found commingled within various separate note fields. The low number of entry dates recorded by Dobbs (two) may be related to the fact that most of the Granville Proprietary Land Office deeds were not accompanied by shucks which provide this type of reference information. Neither Dobbs’ nor McNeely’s sampled deed paperwork appeared to reference any warrant dates.

The people involved in a land transaction provide an important clue for piecing together the geographic relationship between adjoining or nearby tracks of land. Particularly the names of bounding neighbors and chain carriers who presumably lived within some proximity to the surveyed land; both categories were well represented in each investigator’s database as were grantee names. Attestor(s) may also have lived nearby and, based on an examination of surnames, some appear to have been relations of the grantee, a connection often tracked in family history investigations yet absent from McNeely’s data. Likewise missing from McNeely’s data were assignee names, which are essential for linking earlier transactions to a previous “owner’s” paperwork, as were the majority of surveyors’ names.

Features of the landscape, like the people involved with each land-transaction, supply a critical clue for determining the location of a tract of land, particularly the watercourse(s) upon which a parcel may have been located or near. While the investigator’s feature types overlapped in most areas, no animal habitats were noted in McNeely’s database and Dobbs’ sampled deed records lacked topographic references; although such references can be found in many of the records not sampled. Perhaps most interesting was the number of trees recorded in Dobbs’ and McNeely’s databases, 381 and 252 respectively. Trees were used to demarcate boundary line corners and as such were prominently distinguished within a survey’s descriptive text. Notably, this is the first instance where key information in Dobbs’ database was not recorded in a controlled field, as indicated by the dashed underline in the model. This was equally related to limitations associated with the database design and ArcGIS’s polygon shape feature, which has no facility for linking individual polygon end points to separate attribute fields.
Less comparable between investigator’s databases were the number of county(s) associated with each parcel. This may relate to the different time periods and regions under study. Dobbs’ study encompassed a longer time period and wider geographic area. Conversely, McNeely’s sampled deed records were related to transactions located within present day Burke County, which was formed in 1777 from a portion of Rowan County. Many of McNeely’s deeds may have been reissued State deeds of land previously granted by the Granville Proprietary Land Office shortly after America’s Independence. Both Dobbs and McNeely captured detailed land measurement information and the survey’s unit of measure, key information to enable the automatic generation of a scaled copy of each tract of land for use in the mapping process.

The dissimilarity in reference types (i.e., internal versus external) was perhaps the greatest difference observed between Dobbs’ and McNeely’s sampled deed records. Absent from Dobbs’ database were external references to associated patent book numbers and pages, although deed numbers were found commingled within the transcription section of her database often identified as “verso” text. In contrast, McNeely’s database employs controlled fields for recording these important external references numbers. Exclusive to Dobbs were internal references (i.e., parcel matches), which identified the related land records that she had transcribed.

The deficiencies associated with the MARS index for supporting various types of research -- including land-based and family history investigations -- becomes readily apparent when comparing its finding aids to the equivalent sampled deed content as transcribed in Dobbs’ and McNeely’s databases. The catalog records describing Granville Proprietary Land Office (i.e., series 12.13) documents only provide deed dates. In contrast, the State Land Office catalog records examined by McNeely included both deed dates and entry dates, likely because this series of deeds (i.e., series 12.14) was cataloged using the more detailed variant finding aid (Figure 4.11); none of the cataloged deed records provided warrant, survey or other related transaction dates.
Figure 4.16 MARS catalog equivalent of Dobbs’ 97 sampled deed records. (Dashed lines and underlines indicate commingled data.)

*Originally (41) folders, (13) pointed to decidedly unrelated land records and were excluded from the model*
The personal names indexed in MARS are far from comprehensive, identifying a mere 1.2 names per deed. For example, Dobbs identified 592 people representing an average of six names per deed, compared to the mere 125 indexed in MARS, and McNeely recorded 507 representing five names per deed compared to the catalog’s 109. The MARS index consisted primarily of grantee names with an occasional reference to an assignee or neighbor and excluded surveyors, chain carriers, and attestors; in all instances none were identified by the role they played. A similar dearth in index terms for feature types was equally noted, as the bulk of MARS terms related to watercourses and an occasional identification of a cultural feature or tree type. There were no topographic or transportation terms found in the comparable MARS records, although some were found in records not examined in this comparison. Animal habitats were a feature type exclusive to Dobbs. County affiliations were a close match, although Dobbs had identified four more (i.e., 104 vs. 100) than the catalog and McNeely seven less (i.e., 104 vs. 111). With the exception of acreage, survey measurements were not included in the MARS index.

Internal references to sequentially related land records were found in both Dobbs's database and the MARS catalog, but not in McNeely's. While McNeely occasionally made note of related land records within his database, these typically referred to a neighboring or family members’ tracts of land, which was an interesting difference not represented in the model. Dobbs identified 78 related land records within her database (14 entry, 44 warrant, and 20 survey documents) in comparison to the 41 folders referenced in MARS associated with 58 related documents (nine entry, 32 warrant, and 17 survey documents). But as will be seen in the uncertainty results section, 13 of the MARS folder references (i.e., contained in the note field as “see also…”) clearly pointed to inaccurate relationships unrelated to the deed. The MARS equivalents of McNeely’s deed records provided reference notes to five folders each containing one land record (three warrants, and two survey documents).

While deed numbers were recorded in both investigators’ databases and MARS, Dobbs’s database contained no patent book/page numbers or reference to entry numbers and their date of issue. The omission of these in Dobbs’ database is likely related to the fact that the shucks provide this information but they were not filmed alongside the deeds they once held. External references are the only instance where MARS was found to provide slightly more information, but only in relation to the State Land Office 12.14 series. For example McNeely’s database contained 84 entry numbers and 81 book/pg
numbers versus the 87 entry numbers and 95 book/pg numbers found in MARS. Dobbs’ had recorded 94 deed numbers, found commingled within the database’s verso text transcriptions, compared to 95 identified by MARS.

The results indicate a 90% correlation between Dobbs’ and McNeely’s database content in terms of the common types of information collected, despite their dissimilar goals and map-based investigation objectives. Certain categories types were less prominent in one database as compared to the other. For example, Dobbs transcribed and recorded related supporting documents, whereas McNeely did not and as a result provided no references to complementary records. Dobbs also recorded a wider variety of people involved in the sequence of each transaction, particularly noticeable were the number of attestors and surveyors. In contrast, unlike Dobbs, McNeely had recorded the patent book and page numbers associated with each deed record.
Figure 4.17 MARS catalog equivalent of McNeely’s 100 sampled deed records. (Dashed lines and underlines indicate commingled data.)
The Domain of Uncertain Evidence

*Characteristics*

The third domain analyzed was the domain of uncertain evidence. As had been expected there were universally recognized types of uncertainty found within the investigators’ databases and comparable MARS catalog content. These included conflicting interpretations of 18th century handwriting, such as the spelling of a personal name or landscape feature, as well as variations of the same name within a single document. The quality of the microfilmed images, particularly those with poor contrast and faded text, rendered portions of content indecipherable. Additional uncertainty can be attributed to damaged or illegible documents where evidence is missing or unclear, a common problem encountered with old records particularly those that have been torn, are missing sections, contain smudged writing or employed poor penmanship. These along with other types of uncertainty have been well documented as a common concern in historical investigations, as described earlier in the related methodology section.

A different type of uncertainty relates to unverifiable source material within an investigator’s database. For example the initial sample drawn from Dobbs’ database to analyze the domain of evidence included 100 deed records, but two could not be matched to a comparable MARS catalog record and thus were excluded from further analyses. The eliminated records concerned two deed facsimiles supplied to Dobbs by a friend; these lacked sufficient information to corroborate the source, although Dobbs indicated that they came from one of the 12.15 Series patent books and suggested two microfilm reel numbers where an authentic copy might be located. Similarly, out of McNeely’s 116 initial sampled deed records 16 were excluded; one source referenced was a neighbor and 15 were identified as part of a local library’s collection, but a corresponding MARS IDs could not be found.

In actuality neither investigator provided adequate citations for authenticating the source of an individual document. While Dobbs’s database typically referenced a microfilm reel number associated with each document, a commendable level of detail not provided in MARS, microfilm numbers are insufficient because they represent numerous commingled documents, not the individual document’s catalog record. The deeds consulted by McNeely came from various local libraries, many of which were identified by a document title. These were clearly copies provided to his library by the NCSA as the titles
he noted were equivalent to those in the MARS catalog. Those identified by title could easily be linked to a precise MARS ID, whereas those without were typically unmatchable and therefore excluded.

Another type of uncertainty discovered, which had not been considered previously, represented impediments related to the NCSA hierarchal arrangement, limited indexing, and the extent of separately filed but related documents within MARS. This unexpected type of uncertainty, related to catalog deficiencies, became the primary focus in this analysis. As such, it examines the effectiveness of consolidating information from related land records to produce a more complete description of each tract of land.

Dobbs believed, as I did, that it was critical to examine the complement of documents (entry, warrant, survey, and deed) to effectively facilitate a land-based investigation. To this effect she matched interconnected records within her database wherever possible using a unique identifying number, akin to a virtual “office staple”, to enable creating a consolidated view of individual related transactions. For Dobbs, the process entailed comparing the database records she had amassed and tagging those she believed to be related. Central to the task was the ability to assess overlapping information between documents to insure a confident match, achieved when the personal names, features, acreage, and deed numbers could be reasonably correlated. Uncertain matches were noted as tentative.

MARS provides similar references to folders containing correlated records, but the extent of their use was unclear and based on observations appeared irregular. Unlike Dobbs’ database, MARS offers no feature for viewing related records and their content together, only one finding aid at a time can be viewed. It was presumed that the Secretary of State’s Office and later the NCSA staff, during the process of rearranging the land records, had made notations on each shuck as to its content and relationship to related records filed separately. What was unclear was whether the shuck notations had been incorporated into the corresponding MARS records to enable recreating the original creator’s filing system.

**Inventory**

The inventory and results for this analysis drew from three sources. First, the column labeled “Deeds” represents category totals from the records examined and reported upon previously in the domain of evidence. The second column titled “Reference” includes the count of any new information
found within the referenced supporting documents of each system; Dobbs’ references as discussed above were identified by a unique ID whereas the MARS catalog provided “See also [MARS ID]” references commingled within a note field. The third column, titled “D of R Index”, short for the domain of resources master index, presents a tally of supplemental content uncovered by searching the index and represents related land records that had not been identified previously in either system.

**Results**

Regardless of the historical topic under study, if the evidence being collected is incomplete the level of uncertainty will remain a barrier to a successful investigation. But as in most historical research, if given time additional facts can be located to resolve contradictory or missing data and eventually reduce the level of doubtful content as confirmed by this analysis. Recognizing that uncertainty is inherent to historical research is key for shaping a collaborative environment that would enable debate and facilitate user contributions intended to promote ongoing improvements. There is a need, often overlooked, for tools designed for collecting historical evidence to accommodate various forms of this characteristic.

Of the 97 cataloged deed records in MARS, 42 were found to contain references to supporting documents, although 14 were later excluded as they pointed to decidedly unrelated land records. Of the remaining 28 valid folder references, a meager 29% of the sample, 44 additional documents were identified that contained 44 new transaction dates, 41 personal names, and nine additional features (Table 4.8; see SOURCE). In contrast, the equivalent 97 deed records in Dobbs’ database provided 50 references to various supporting documents, representing 52% of the sample (Table 4.9; see SOURCE). These pointed to 78 related documents containing new information not found within the deed records, an average of 1.56 related documents per deed. Dobbs’ referenced supporting documents established 60 new transaction dates, 92 personal names, 32 features, and eight additional county affiliations.
Table 4.8 Deed references in MARS point to new information contained in supporting documents. Many relationships are missing.
Table 4.9 Deed references in Dobbs’ database point to new information contained in supporting documents. Very few relationships are missing.

The last column in the comparisons, which relied on the domain of resources master index to locate additional matches, identified 111 more folders containing 128 new “items” corresponding to the
MARS sampled deeds; these evident relationships had been omitted from the catalog (Table 4.8; D of R Index column). Twenty of the folders contained 37 documents, which provided 37 more transaction dates, 50 new personal names, 15 features, and six additional county affiliations. Unexpectedly, 91 of the folders identified were from the State Land Office record series (12.14). Each of these folders contained an empty shuck with external notations describing its (missing) content; all related to an original deed. Each shuck could be readily matched to the deed it once held, which were filed appropriately within the Granville Proprietary Land Office record series (12.13), sans shuck.

Following the same procedure with Dobbs’ sampled deeds a search of the master index unsurprisingly uncovered little new information as her database content was exhaustive. Only one new date, county affiliation, and two deed numbers could be found. Like the MARS deed records, Dobbs’ database made no reference to the 91 folders containing empty shucks that related to her study. This was understandable as her research involved examining the Granville Proprietary Land Office records that presumably ruled out the need to examine the State Land Office records, an inaccurate assumption in light of the results.

By comparing the consolidated totals between tables several similarities were found. The number and types of source materials associated with each system’s content were nearly identical, although there were minor variations between document types. For example, the MARS table drew from 269 individual items as compared to Dobbs 266, a 1% variation. An equally small difference was found between external references, denoted as XREF in the table; Dobbs identified 188, MARS 186. The COUNTY(s) category was in relatively close agreement where Dobbs had 111 administrative affiliations compared to 106 in MARS, representing a 4.6% difference.

The largest disparities observed were found between the people and feature content. MARS only identifies 216 people as compared to Dobbs’ 684, a 104% discrepancy. As to features, only 149 compared to 187 were noted, respectively, representing a 22.6% variation. The last comparison, deed dates and the dates of their associated records, shows Dobbs’ database with 255 as compared to the 178 in MARS, a 35.6% difference. This last discrepancy appears related to missing survey dates, which are typically found within the deed’s survey addendum but for some reason were not identified in MARS.

The results confirm that a significant amount of information, essential for map-based
investigations, is located within a deed’s supporting documents and not in the deed itself. Most troubling was the paucity of internal references to related documents within the MARS catalog, information of major importance to land-based investigations as well as the general public and family historians seeking similar content. In addition 14 out of 42, or thirty-three (33%) of the MARS deed references examined were found to be incorrect, suggesting that a significant portion of the references provided within the Granville Land Office Record Series (12.12, 12.13) are wrong. It is unclear if a similar percentage of errors would be found within the State Land Office 12.14 series, which was not examined here. Of note was the utility of the domain of resources master index in this analysis for identifying previously unrecognized matching records, confirming the veracity and usefulness of a consolidated index as an aid for reassembling the NCSA’s disassociated land records; this task would have been difficult to accomplish without a master index.

The Domain of Tasks

In general there are two distinctly different tasks involved with map investigations: the first is related to data gathering and its associated tools, which was the primary focus of this study, the second is related to the mapping process and allied system characteristics. While a map representation of the evidence collected is the ultimate objective for each investigator, analyzing the tasks involved with positioning tracts of land into geographic space was beyond the scope of this investigation. Nevertheless, many of the steps involved with collecting data directly impact the mapping process and as such are pointed out in the following description of data gathering characteristics.

The steps described are representative of the tasks employed by Dobbs, as gleaned over our three-plus year collaboration. It draws from discussions as well as observations recorded concerning the evolution of the database design as it grew to accommodate unanticipated categories of new information. It was presumed that McNeely engaged in similar processes in the collection of evidence, understanding that limitations within DeedMapper® would have precluded the use of certain advanced features unique to Dobbs’ database.

Data gathering characteristics

Many of the tasks involved with collecting evidence from historical land records have been introduced throughout this paper. The primary activities mirror those employed for carrying out the three
domain analyses presented earlier, such as identifying resources, parsing and recording evidence needed to enable map reconstructions, and joining together information from related documents to fortify the geographic picture. In most cases, Dobbs made partial copies of each microfilmed survey record to assist in the process, first on paper and later in digital format. Perhaps the greatest challenge entails interpreting 18th century British-American handwriting, as this requires knowledge of various early penmanship styles such as Round Hand, German Hand, Italian Hand, and other calligraphic standards in use during the time period.\footnote{For example see: How to Read 18th Century British-American Writing at http://dohistory.org/on_your_own/toolkit/writing.html}

Dobbs consulted several sources to familiarize herself with the various penmanship standards and found they became easier to decipher over time. Unlike McNeely, Dobbs recorded a full transcription of each land record within her database, including the body of the text as well as administrative notes found on the verso and accompanying shuck, where provided. In addition, she parsed critical content from the text into separate dedicated fields, as defined earlier in the domain of evidence results, to facilitate later searching and complex analysis. Incorporating a transcription tool with suitable markup language, such as TEI, to enable programmatically populating the individual search fields was noted as a future goal to eliminate redundant data entry steps and reduce errors.

For the most part, Dobbs transcribed personal surnames as well as feature names as written, in keeping with recommended best practice in historical research. Look-up up tables, akin to an index, were built into the database allowing the name and feature fields to be filled in with content from a drop down list of previously recorded items, which grew as new ones were entered. This saved time and helped to reduce data errors and duplication. Frequently, personal and feature names were found to be spelled two or more ways within the same document, which were essentially phonetic equivalents that rendered the correct spelling unclear; in these instances both versions of a name's spelling would be recorded.

Unique to feature descriptions was the use of local vernacular names, many of which are no longer in use such as the Saxapahaw River known today as the Haw, or Mark's Creek now referred to as Morgan Creek, or the Saponia River or Atkin River, both historical names for the Yadkin (Powell, 1968). Dobbs, as a geographer and native North Carolinian, was familiar with many of these but consulted Powell’s gazetteer as needed for confirmation and noted the findings accordingly in her database.
Interestingly, Dobbs encountered several historical names associated with watercourses that had not been included in Powell’s highly cited work, pointing to an opportunity to build upon his efforts.

Many conversations and emails surrounded the need to incorporate advanced system features to enable performing searches that would retrieve not only exact spelling matches but also their phonetic and vernacular equivalents. An idea to alleviate some of this problem using an ‘alias table’ was never implemented due to time constraints and the programming commitment involved. Fortunately, there are numerous existing solutions to consider in future revisions of the database for improving the process, such as integrating a personal name authority tool, similar to a thesaurus, and one or more gazetteers; these approaches have proven beneficial in historical research tools.

The recording of dates required particular attention as these could not always be deciphered from the documents. To accommodate this problem separate text fields were employed for recording the month, date, and year to allow for partial transcriptions where one or more could not be determined. A comment field was included to describe outstanding concerns, or uncertainty, related to the entry. Where dates remained incomplete an inferred date field was provided to enable substituting an estimated date to facilitate analysis. While partial dates could not always be resolved relationships between sequential documents often supplied the missing information to enable narrowing down the timeline. For example, if a survey’s date was indecipherable but its corresponding entry date and deed date were known it could reasonably be assumed that the survey date would fall somewhere within this time frame.

The database was programmatically designed to perform two tasks related to dates behind the scenes to automatically populate new fields. The first involved concatenating each record’s separate date fields (i.e., from ‘month’, ‘day’, ‘year’ into ‘mm/dd/yyyy’) and converting them into a controlled date format to accommodate reporting and analysis; this was done where complete dates were available. Second, the controlled dates, although historically accurate, were of concern to Dobbs because they contained both Julian (OS) and Gregorian (NS) calendar dates that she believed would distort the results of a timeline analysis. To address this concern Julian dates were converted to correspond with the Gregorian calendar, which enabled mapping the relative progression of settlement over time.

The survey data is the cornerstone of cadastral-based investigations, particularly the length and direction of each boundary line to facilitate generating scaled versions of each land parcel; essentially the
scaled versions represent the puzzle pieces awaiting reassembly. Boundary line measurements and directions were entered in sequential order following the route taken by the surveyor from beginning to end. The type of survey measurement (or unit) employed, either chains or poles, was noted accordingly in Dobbs’ database. These were programmatically converted into a more universal scale, such as feet or meters, to facilitate creating a digital ArcGIS shapefile suited for use in the 2D environment. Dobbs added three related check-box fields to identify problems encountered with the measurements, including possible directional errors, those with obvious errors that had been corrected, or parcels that would need to be manually drawn during the mapping phase.

Unlike McNeely’s DeedMapper® program, Dobbs’ database did not have a provision for linking boundary lines and property corners to their defining attributes, such as each line’s relationship to neighboring land and corner marker descriptions. This facility had not been considered in the design of Dobbs’ database, although in retrospect such information appears essential for guiding in the map reconstruction process. Like a puzzle piece, whose edges and color variations serve as clues to its placement within the larger picture, neighboring names and distinctive markers serve the same purpose -- but only if provision has been made to visibly display these characteristics alongside the edges or corners they define within the 2D mapping environment. In addition, identifying boundary line features is often central to environmental land-based investigations, such as Schwartz’s (2007) research that relied on Dobbs’ transcribed data but required database modifications in order to identify the various tree species employed as survey corner markers.

The administrative data collected by Dobbs was limited to five fields: the consulted document’s media format, its archival institution, associated call number, date of transcription, and the unique ID assigned to each record in the database. As expected, the majority of records examined were identified as microfilm facsimiles available from the NC State Archives. The call number was intended to be associated with the material format, which in this case referred to the microfilm reel number containing a copy of the land record transcribed. The last two fields, the data entry date and unique ID were automatically filled in as each new record was created. In hindsight additional administrative fields should have been incorporated to better describe the source material and range of operator actions. For example, the MARS ID is essential for enabling direct linkage to each land record’s finding aid as well as
a method to discriminate between multiple copies of the same land record, such as deeds. Additionally, investigations that involve collaboration would necessitate the facility to identify individual data entry contributors as well as capture ongoing revision dates as they occur over time.

One of the most important and frequent tasks associated with data gathering involves the ability to search the database to locate various document relationships and geographic similarities. In this capacity the database I designed to assist Dobbs in her data gathering process was adequate but far from perfect. While common search queries had been built into the database, such as the ability to retrieve a document by its unique ID, examine a list of features and personal names for editing, search all documents by feature or personal name, and print reports that joined together related land records along with a digital image of the related survey, these did not address every question Dobbs wanted answered. For example, we discussed the desire to locate tracts of land based on boundary line directions or lengths, limit the results to a particular year or water basin, and perform filtered searches capable of excluding certain content such as a search by personal names not including attestors.

As Dobbs identified new querying objectives we would discuss them and if time permitted I would help to formulate a new query. Ideally, if the database had been normalized\(^46\), Dobbs would have been better able to form many of her own queries as she was somewhat familiar with basic query construction. Unfortunately, because the database had not been normalized many of her attempts returned incomplete results or redundant rows of similar data and my explanations of how to avoid this problem were not very clear. Nevertheless, the database served her purpose and the problems encountered would be easy to resolve in a future rendition of the database, including the provision for additional standardized queries and a help page to guide those interested in constructing their own.

\(^{46}\) A final step in the process of designing a database to disaggregate elements into a single instance, or table, to eliminate redundancy. The decision to not fully normalize the database was intentional to permit table views more familiar to Dobbs.
Chapter 5: Discussion

Summary

Digital humanities is about exploring, developing, and implementing tools to enable inquiries into complex phenomena that before the advent of computers would have been unfathomable. The common goal in this growing field is to “fulfill the humanities’ basic tasks of preserving, reconstructing, transmitting, and interpreting the human record” in order “to make cultural heritage more widely available for teaching, research, and outreach” (Borgman, 2009). To support these efforts requires cyberinfrastructures akin to those being developed in the scientific arena that facilitate collaboration in the “creation, dissemination, and preservation of scientific data, information and knowledge” (NSF Cyberinfrastructure Council, 2007) by joining together distributed computing systems, data storage systems, archival repositories, computational and visualization tools. Unfortunately, “basic infrastructure for the digital humanities is still lacking” (Borgman, 2009).

To realize similar technological benefits humanities scholars need to better define their vision and digital information needs for the 21st century. A “shared infrastructure of tools, services, and collections…would reduce unnecessary redundancy, allocate human and information resources efficiently, and, most interestingly, enable a different kind of scholarship” (CLIR, 2009). Many in the field have championed the need for greater communication between information scientists, humanities scholars, libraries, and archives to better articulate and synchronize computing needs in the humanities (Besser, 2004; Schreibman et al., 2004; Thomas, 2004; Unsworth, 2009). This study is perhaps the first to have examined in detail the information needs of a group of investigators whose common goal is to explore the evolution of geographic relationships during America's colonial period; this group includes both scholars and surprisingly a large section of the general public engaged in the pursuit of family history.

Land-based investigations encompass many of the common challenges faced by humanities scholars engaged in analyzing and reinterpreting history, which rely on evidence contained in primary source materials. It involves making sense out of large sets of records that are essentially geographic in nature, yet the historic circumstance they embody remains unclear without a corresponding map and the missing information only it can provide. The lack of essential tools and potential for
cyberinfrastructure to mitigate the problem provided the impetus for this study, which examined the information needs and barriers involved that remain to be addressed to effect a permanent change in how land records are studied and the collective results disseminated.

**Evaluation of results**

The results of this study pointed to numerous barriers that could be overcome with the aid of dedicated tools and a shared infrastructure. Envisioned is a suite of applications to enable a collaborative and authoritative reconstruction, over time, of the missing geographic picture. Central to the system would be the provision for comprehensive access to the domain of resources, comprised of authoritative references to archived material, catalog records, and images. Unlike MARS, access to the domain of resources should not be limited to confined views but should also permit customizable configurations to suit an investigators needs, such as the facility to search, filter, and generate lists sorted by surnames, the roles people played, county affiliations, transaction dates, features, acreage, measurements and angles or a combination thereof. Ultimately, a similar search facility could provide the option to explore land records in geographic context in a mapped format with links to each parcel’s underlying supporting documents and corresponding digital images.

To accomplish this vision several inherent barriers to accessing North Carolina’s land records must first be resolved. For example, information and images of related records should be accessible in a single view, something not possible in the current catalog and microfilm formats. These should be replaced with individual digital images and appropriately identified (i.e., with a MARS ID) to insure authenticity and a venue to incorporate user-contributed index terms alongside cataloged content implemented. Resolving the first two issues would eliminate the regrettable result of two fateful decisions that changed the way North Carolina’s historical land records were arranged and made available to the public that effectively introduced unnecessary uncertainty. First, the Secretary of State Office’s directive back in 1909 to separate the deeds from their supporting documents, a practice later continued by the NCSA, dismantled the original creators’ intended filing system; a system that recognized geographic relationships and offered a single access point to interrelated land records. The domain of uncertain evidence analysis indicated that reestablishing these lost
relationships would play a key role in solving this major obstacle and be fairly straightforward to accomplish.

Secondly, the ability to uniquely identify the source material consulted is critical in historical research as is the need for access to copies of the original documents to enable transcriptions. The current microfilm format offered to the general public not only lacks authoritative references to the documents they represent, it is also the least preferred method for viewing documents as the media is cumbersome to work with, image quality is often poor, and the contrived arrangement is especially limiting. Clearly, the first two barriers to overcome, as well as the third related to insufficient indexing, would entail efforts beyond the resources of most archives. NCSA has an opportunity to join forces with investigators to achieve the goal of improving access to its historically important and highly popular land record collections and in turn help in the discovery of new geographic knowledge.

The envisioned solution would be a form of collaborative hybrid system or virtual catalog, including links to related documents within the catalog as well as external citations to relevant material that joins together the existing MARS finding aid information supplemented by controlled user contributions. This would address the information deficiencies identified in the results of the domain of uncertain evidence. Index term contributions could initially be limited to those defined in the results of the domain of evidence, although the potential for future adaptations should be considered. To insure the reliability of user contributions digital images of each folder’s documents should be accessible from the catalog; high-resolution images provided by users, but preferably images prepared under the guidance of the NCSA staff, to insure quality control and long-term preservation. By providing Internet accessible digital images a wider audience would be reached and the unwieldy microfilmed facsimiles could be retired.

While the collaborative cyberinfrastructure imagined is intended to improve access to the NCSA land records by facilitating the sharing of land record transcriptions, supplementary index terms, and user contributed maps, this information is commonly collected by solitary investigators who have employed non-standard tools. To enable effective information sharing would warrant the development of a stand-alone application for use by individual researchers; a database program capable of connecting to the 'virtual' catalog would facilitate ongoing individual research and insure
that completed findings can be confidently shared and will contribute to the whole. Of particular concern in the shared environment would be the ability to identify individual author contributions as well as concurring or divergent results as may be encountered in multiple author submissions concerning the same land records or geographically mapped reconstructions.

The need for advance search features within the virtual environment would be particularly important to address. For instance, the lack of essential search mechanisms to enable concurrent searching limited to a particular place or feature, time, people, and/or event are routinely cited as essential for facilitating humanities studies (Duff & Johnson, 2002; Tibbo, 1994). As the MARS catalog is clearly deficient in this area and its land record indexing is inadequate the value of “controlled” social tagging to alleviate this problem is evident. As Tennis (2006) has posited, “…information systems would benefit from employing multiple types of information organization frameworks – like social tagging along with library classification.”

**Limitations and conclusions**

The scope of this study was limited to the examination of two sets of data collected using dissimilar tools for different purposes: one set collected by a geographer pursuing academic research, the other by a chemistry professor emeritus pursuing a personal hobby to locate ancestral habitats. Unknown is whether the results would vary if the analysis were expanded to assess content collected by investigators with different academic goals, or by family historians with less experience in rigorous research and computer technology. Additionally, while the domain characteristics associated with land record content and resources were analyzed, the same analyses should be conducted to identify the domain characteristics and tools connected with the allied task of mapping the results.

Regarding the domain analytic method employed to accomplish each analysis, the process was both constructive and at times problematic. The iterative steps were very helpful for appreciating each unique domain and its characteristics although determining when each step was complete was at times difficult. A problem related to its sequential nature involved concerns discovered in subsequent analyses that suggested returning to a previous analysis to incorporate new insights gleaned. In certain instances revisions to prior steps were incorporated, but most reevaluations
considered were ultimately abandoned as the time required would have been prohibitive and not likely to have made a substantial difference.

For example, in the first analysis completed for assessing the domain of resources which examined the Granville Land Office records (Series 12.12) the estimated document count was based on the dates associated with each catalog record. In hindsight, a stratified sample may have been more appropriate. This issue was not fully appreciated until starting the third analysis regarding the State Land Office’s similar collection of non-deed related documents (12.14), at which time it was decided that a stratified sample should be employed to advance the process. The issue related to the dissimilar methods used in MARS to identify the content of each collection’s various folders.

While document counts were provided in the Granville cataloged deeds (series 12.13) the quantity field was not populated in either of the supporting document series (12.12, 12.14). For these collections it had been assumed that counting the dates associated with each folder would be a reliable indicator for perfecting an estimated document count. It had not been considered that two documents within one folder may have been prepared in the same year and as a result be counted as one. While a stratified sample from series 12.12 for use in examining selected scope notes may have provided a more precise estimate, the initial analysis using dates was deemed acceptable as the resulting document counts per folder were comparable between the two similar series (12.12, 12.14) despite the different methods employed.

The results presented provide an initial blueprint for opening up dialog as to the cyberinfrastructure required and associated tools needed for supporting repetitive land-based investigations. The system envisioned would help both academic scholars and the general public to authoritatively contribute and expand upon each other’s research, ultimately completing the missing picture of our nation’s past geographic rural settlement patterns. By providing researchers a stable and standards-driven infrastructure for sharing and searching across a common domain of interest, it is expected that through patron contributions over time, access to colonial era land records could be vastly improved and the method employed would serve as an exemplar for how uniquely geographic information could be better accessed and understood when presented in context: in the form of a map.
Appendix A. Public Land Survey System: Principal meridians and baselines.
U.S. Department of the Interior, (BLM map, 1988)
Surveyed for James Potts a tract of land containing 531 acres lying on the South side of the South fork of the Yadkin River beginning at a black oak standing on the South side of sd fork running thence So 230 pole to a black oak standing on John Potts line thence Et. 360 pole to a black oak thence N 230 pole to a black oak thence to the first station
Surveyed June 13 1753
Jacob Thomas
Allen Alex’d Jas. Carter
Chain carriers

Source:
Secretary of State Granville Land Grants
S.108.256

File at: http://files.usgwarchives.net/nc/rowan/photos/jamespot1277gph.txt

This file has been created by a form at http://www.genrecords.org/ncfiles/

File size: 1.2 Kb

Appendix B. Potts NCGenWeb Submission. Sources
### ORANGE COUNTY, NC - DEEDS - George Finley Grant

**USGENWEB NOTICE:** In keeping with our policy of providing free information on the Internet, data may be used by non-commercial entities, as long as this message remains on all copied material. These electronic pages may NOT be reproduced in any format for profit or for presentation by other persons or organizations.

Persons or organizations desiring to use this material for purposes other than stated above must obtain the written consent of the file contributor.

This file was contributed for use in the USGenWeb Archives by: Sharon Snitily sharigs@crcwnet.com

---

**THE GRANVILLE GRANTS LATER NORTH CAROLINA STATE LAND GRANTS, PRESENT DAY GUILFORD COUNTY, NORTH CAROLINA**

©1999 Sharon Snitily

Granville Grant Genealogy Research Project by, Sharon Snitily, with special thanks to Dave Simmer and Mary June Foulk. For a color printout and more information about our project, email your name and address with subject line of GRANT PROJECT to Shari - sharigs@crcwnet.com.

**GEORGE FINLEY**

Orange County, North Carolina

Granville Grant 3870

Patent Book 14, Page 399

Earl of Granville's Agent: 1 December 1753

Surveyed: 5 April 1754

Granted: 6 March 1755

Chain Carriers: Othio Boshears, Cornelius McDede

Witness: Hugh Forster, James Carter

Driving Directions: (from 7½' series USGS Map, 1994)

Take State Route 29 North out of Greensboro. Take McLeasnville Road Exit. Go South on McLeansville Road and turn right at the fifth unimproved road past the exit. This road winds and interesects two other unimproved roads before it eventually travels the north part of Finley's grant

**ORIGINAL GRANT DESCRIPTION**

...all that Piece and Parcel of Land, situate, lying, and being in the Parish of and County of Orange in the Province of North-Carolina, in America. On the North Side the Reedy Fork, Beginning at a poplar, on the Bank of said Reedy Fork, in the Line that divides the Counties of Rowan and Orange, thence along said line North 320 poles to a post, in said line, thence East 240 poles to a Maple Standing by a small Branch, thence South 229 poles to a Hicory Standing by the Reedy fork, thence up the various courses of said fork, to the Beginning. Containing in Whole Four hundred and four Acres of Land...

>>>> To view a map of the grant, go to: <<<<
[http://www.usgwarchives.net/nc/orange/map.gif](http://www.usgwarchives.net/nc/orange/map.gif)

Deed Abstracts-Guilford County

Bk 3:140 - George Finley, Sr. to George Finley, Jr.
4 November 1784, 202 acres, 50£. Wit: John Cunningham, John Clintock, Sr.,

Proved Guilford County November Court 1784

Bk 4:15 - George Finley, Sr. to his son James Finley
20 November 1786, 202 acres, 50£ Wit: William Covey, Josiah Finley, Proved

Guilford County November Court 1786

---

Sharon Snitily posting to NCGenWeb: retrieved 9/29/09 from:

[http://files.usgwarchives.net/nc/orange/deeds/finley01.txt](http://files.usgwarchives.net/nc/orange/deeds/finley01.txt)

---

**Appendix C (Exhibit A):** Online transcription categorized under “deed” heading (see related image Appendix C, Exhibit B)
APPENDIX C: (continued)

Appendix C (Exhibit B): Online map of parcel described in previous transcription (see related text Appendix C, Exhibit A)
REFERENCES


Collet, C. J. (Cartographer). (1770). *A Compleat Map of North Carolina From an Actual Survey* [Engraved map made of two joined sheets].


Fry, J., & Jefferson, P. (Cartographer). (1755). *A map of the most inhabited part of Virginia containing the whole province of Maryland with part of Pensilvania, New Jersey and North Carolina. Drawn in 1751*


Hughes, F. (Cartographer). (1977). *Davie (the forks of the Yadkin) County, North Carolina: Historical Documentation No. 4*

Hughes, F. (Cartographer). (1979). *Rockingham County, North Carolina: Historical Documentation No. 7*

Hughes, F. (Cartographer). (1980). *Guilford County, North Carolina: Historical Documentation No. 9*


Markham, A. B. (Cartographer). (1973a). *Land Grants to Early Settlers in Old Orange County, North Carolina: Parts of Present Orange, Chatham, and Durham Counties, Period 1743-1810*

Markham, A. B. (Cartographer). (1973b). *Land Grants to Early Settlers in Present Wake County, Period 1739 to 1820*


Rendleman, D. A. (Cartographer). (1953). *German Pioneer Settlers, Lower Rowan County*


Tooby, P. (2007a). *Preserving History on a Humanities Grid for the University of California*.


