MAINTAINING PERSISTENT SCHOLARSHIP:

THE CASE OF UNIVERSITY AT ALBANY DISSERTATIONS

by

Carol Anne Germain

A Dissertation

Submitted to the University at Albany, State University of New York

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

College of Computing and Information

Department of Informatics

2012

Committee:

Dr. Deborah Lines Andersen (chair)

Dr. Teresa Harrison

Dr. Abebe Rorissa

UMI Number: 3516764

UMI

Dissertation Publishing

UMI 3516764

ProQuest®

Maintaining persistent scholarship:

The case of University at Albany dissertations

by

Carol Anne Germain

Abstract

Citation is a highly valued practice in the academic community. This mechanism supports an author's ideas, theories, and research; it acknowledges the scholarly contributions of others; and integrates academic works to enrich scholarly communication. Well-constructed citations, in addition to providing the appropriate bibliographic information needed to retrieve a supporting document, acknowledge the contribution of previous literature to the topic and create important links between the citing and cited sources which often have similar contents. The development of the World Wide Web has generated the use of digitally-based citations. Web pages can be volatile and are often not static publications, so when an author cites a URL resource there is a risk that it may vanish and thus the cited work is lost.

Of particular interest are citation strategies of doctoral students who are developing their research protocols. This study analyzed the patterns of freestanding URL citations found in University at Albany dissertations between 1996 and 2007. Three distinct years of dissertation citations (2000, 2003, 2006) were tested for accessibility and content accuracy. Each URL was searched using a URL Web browser (Firefox), the search engine Google, and two Web archival tools, the Internet Archive and WebCite. In addition, members of the University's doctoral granting departments were interviewed to determine departmental dissertation citation policies and practices.

The findings of this study indicate that the use of URL citations in dissertations is increasing over time. After testing the URL citations using the four

information retrieval tools, the results showed that 42% of Humanities, 61% of Science, and 63% of Social Science citations could be located in at least one of the tools. Data from the departmental interviews showed that three departments (20%) had a policy regarding dissertation citations (defer to discipline style manuals) and no department had engaged in departmental discussions on the use of URL citations.

This research contributes data on the changing nature of scholarly attribution across disciplines, empirically documents the rising use of online citations in scholarly works, and highlights the need to increase discussions about citation policies across academic disciplines if citations are to continue over time as the means of acknowledging the work of others and providing intellectual context to new research.

To Chris, Dale, and Dan, with love.

ACKNOWLEDGEMENTS

Table of Contents

Chapter 1. Introduction

Chapter 2. Literature Review

Chapter 3. Research Methods

Chapter 4. Results

Chapter 5. Results Part II- Interviews

Chapter 6. Discussion and Conclusion

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

Overview

One of the primary motivations for creating the Internet was to facilitate communication among researchers, scholars, and academics (Abbate, 1999; Jonscher, 1999; Lehnert & Kopec, 2008; Whittaker, 2004). Over the years, this government-sponsored project made it possible to exchange research information through text-based protocols such as email, telnet, ftp, listservs, and usenet (Abbate, 1999; Doty & Bishop, 1994; McCain, 2000; Online searching with Internet, 1992). In 1990, Tim Berners-Lee developed an easy-to-use Web browser and the World Wide Web. This development simplified the online navigation process and opened the way for a more viable commercialized Internet (Abbate, 1999; Brody, 1996; Press, 1994). Over the next two decades, the World Wide Web grew at an exponential pace. It expanded as a commercial entity as well as for research resources (Farrell & Petersen, 2010; Meszaros, 2002).

The World Wide Web gained popularity among researchers, faculty, and students because it was easy to access and contained information on research topics of interest. Increased use prompted many academic users to start incorporating hyperlinked information into their publication materials. This included the practice of citing Internet sources to support ideas, models, and theories. While some of these hyperlinks were to formally published materials (e.g., journals articles within a database), others pointed to freestanding Web sites, lacking formal publication or archival structures.

Several studies measuring the increase in the number of URL citations in bibliographies have noted that URL citations are beginning to outnumber the traditional format counts in the reference sections of some journal articles (Gatten, 2010; Vaughan & Shaw, 2003). Problems may arise when citing digital information, however, since these information resources can be volatile and sometimes retrieve broken hyperlinks. When URL sources are unavailable the cited material, which supports an argument, idea, or theory, is corrupted because of inaccessibility. The citing source thus loses its credibility since its support documentation is no longer available.

A number of studies have alerted scholars and researchers to the issues related to the inaccessibility of electronic citations (Bugeja & Dimitrova, 2009; Dimitrova & Bugeja, 2006; Jasco, 2002; Johnson, Hester, Schilling, & Dellavalle, 2004; Wu, 2009). The focus of these research efforts, however, is largely on citation patterns within specific disciplines and do not include cross-discipline, electronic citation analysis. Moreover, these studies do not address the practices and policies associated with the use of URL citations within the educational community. In the present study, a multi-discipline, multi-year approach presented a mechanism for observing the current state of access rates to electronic citations across a variety of fields in one particular case at the University at Albany. Using a sample of the University's dissertation citations offered a viable metric for this type of research since they provided a comparable unit of analysis which could be tested across multiple disciplines over time.

Throughout this dissertation specific terms will be used to describe access to URL sources and its impact on citation. Table 1.1 presents the terms and definitions used for the purposes of this study:

Table 1.1 Definitions of Terms Used for Purposes of the Study

Term	Definition for this Study
Accessibility	The ability to hyperlink to an active Web page or site which contains no error messages. Note: Inaccessibility is the antonym.
Citation Erosion	This occurs when cited material cannot be retrieved either because a URL source is inaccessible or the bibliographic information is insufficient.
Content Accuracy	The content of an accessible Web page matches the content of the dissertation citation. Note: Content Inaccuracy is the antonym.
Effectiveness	The cited URL is accessible and the content matches the dissertation citation.
Freestanding Web Pages (Sites)	Web pages or sites containing features such as text, graphics, and statistics that are uploaded and maintained by individuals, government agencies, corporations, and educational institutions. Examples of freestanding URLs include links to government reports, association white papers, course materials, and organizational statistics. These do not include materials from publisher databases (e.g., Communication Abstracts), typically available by subscription, which have a formal publication and archiving process.
Web Browser Search	A search for a uniform resource locator (URL) source using a Web browser such as Firefox. The search is executed by entering the URL string into the browser location bar.

Research Questions and Contributions

The intent of this research was to analyze a dataset consisting of University at Albany dissertation citations. The dataset incorporated variables of time

(dissertations published between 1996 and 2007) and discipline variables (Humanities, Science and Social Science). Interviews with members of the University's doctoral granting departments addressing departmental policies and practices in the dissertation research process form the qualitative component of this dissertation.

This study aims to identify answers to the following questions:

- What general trends are present in Web citations referenced in doctoral dissertations at the University at Albany (e.g., proportion of URL citations/to total citations; URL citations per academic group (discipline))?

- What relationship, if any, exists between disciplines and the rate of accessibility for University at Albany URL dissertation citations using a Web browser access tools (e.g., Firefox)?

- Does the passage of time affect the accessibility rate of online citations, and does this vary across disciplines?

- What is the rate of accessibility for cited resources using the Internet search engine Google, and two Web archival tools, the Internet Archive and WebCite?

- Are there differences between the content accuracy of the accessed results of the four retrieval tools (Web browser search, Google, Internet Archive and WebCite) and the bibliographic information in the dissertation citation?

- Is there a relationship between departmental policy on academic citation and the use or accessibility rates of electronic citations? More specifically, do particular policies influence use and accessibility rates of URL citations?

Research on citation analysis in dissertations has focused chiefly on cited journals and monographs to inform collection development strategies in libraries (Cox, 2008; Gluibizzi, 2002; Pancheshnikov, 2007; Vallmitjana & Sabaté, 2008; Waugh, & Ruppel, 2004). This study, however, relies upon dissertation citation analysis to investigate the state of URL citations within doctoral theses. In doing so, this research provides timely empirical data on the accessibility and content accuracy rates of cited electronic sources across disciplines. These findings may better inform digital preservation efforts impacting the use of URLs in cited academic sources in general, and dissertations, in particular.

Doctoral Education in America

Because the focus of this study is citation practices in dissertations and doctoral programs, a certain amount of background information related to both follows. While doctoral education has a long history, this level of study did not emerge in America until the mid-nineteenth century. Early collegiate programs in the United States were rooted in the English tradition, which essentially consisted of acquiring a bachelor's degree and then engaging in three years of literary activities (Rudoph, 1962; Thelin, 2004; Thwing, 1928). Early colleges, most particularly Harvard, were influenced by their connection with the two English schools, Oxford and Cambridge, "where Puritan theology and Puritan aspiration had been especially nurtured" (Rudolph, 1962, p. 4). Once a student completed the bachelor's program, he could gain a master's degree by acquiring additional knowledge typically by

extending upon his field of interest in a professional venue (Rudolph, 1962; Thelin, 2004).

This limited form of graduate education continued into the nineteenth century; to attain a higher level of education, many American students chose to study abroad at German institutions. The German model provided rich experiences within which students gained advanced knowledge through research and the ideal of academic freedom (Association of American Colleges, 1960; Boyer, 1973; Buchanan & Hérubel, 1995; Hofstadter & Hardy, 1952; Nettles & Millett, 2006; Storr, 1953). German education spirited the pursuit of more intellectual engagement. While the British-style programs in America were acceptable, this European version of graduate education offered a much more rigorous educational experience and was respected for its higher work standards along with graduate professionalism (Buchanan & Hérubel, 1995; Hofstadter & Hardy, 1952; Rudolph, 1962; Thelin, 2004).

As the popularity of earning a doctoral degree increased, American colleges worked toward integrating the German model into advanced degree programs. Several institutions, such as Yale, Harvard, and the University of Michigan followed this doctoral template, with Yale issuing its first non-honorary PhDs in 1861 (Berelson, 1960; Rudolph, 1962; Thelin, 2004).

The last quarter of the nineteenth century was the formative period of doctoral education in the United States. During this time, major education efforts generated the creation of several scholarly societies which published scholarly journals (Berelson, 1960). From that time forward doctoral education placed a

6

heavy emphasis on research and scholarship (Gale & Golde, 2004; Thelin, 2004; Walker, 2008). These changes in university structure and doctoral education prompted far reaching change at the state level which resulted in the establishment of state systems of higher education in many states in the late 1800s (Rossi, 1966; Rudolph, 1962). New York was the last state in the contiguous United States to commit to the development of such an entity with the formation of the State University of New York (SUNY) in the mid-twentieth century (Clark, Leslie, & O'Brien, 2010).

State University of New York

The State University of New York was established under the direction of the State Legislature. The mission of the State University was to create a large educational system designed to meet the needs of New York citizens through two-year community colleges, four-year liberal arts colleges, and technical colleges as well as large university research centers and medical schools. A major goal with respect to the university centers was to offer comprehensive graduate programs at the doctorate level (Birr, State University of New York at Albany, 1961, 1977; Vogt, 1984). With the implementation of this large education system, the State Legislature designated the Teachers College at Albany as one of four university centers within the SUNY system (Birr & Reda, 1994; Carmichael, 1955; Clark, Leslie, & O'Brien, 2010; State University of New York, 1961; Vogt, 1984). According to the State University of New York revised plan of 1960 "the multi-purpose college in Albany [was to] begin graduate work in the humanities and social sciences. The

introduction of master's level work in these subject areas [was to] be effected

immediately…. Programs in the humanities and social sciences at the doctoral level

[were to] be developed as early as possible" (State University of New York, 1961, pp.

49-50). The college at Albany, with its vibrant history as a teachers' college, was

ready to take on the opportunity to offer doctoral education programs.

The University at Albany – Brief History

The University at Albany traces its roots to the State Normal School. Founded

in 1844, the Normal school had as its mission the training of teachers for the

common schools of New York State. As such, it ranks among the oldest teachers

colleges in the United States. Establishing pedagogy as its foundation the goal of

Normal School was to improve the quality of education by training teachers to

provide adequate instruction at a time when it was lacking. The college was

nationally recognized in 1921 by the Association of American Universities as a

college of liberal arts, and in 1931 it became the first teachers' college to be

approved by the American Association of University Women (French & French,

1944).

In 1948, the Teachers College at Albany became an entity of the State

University of New York (Birr & Reda, 1994; Clark, Leslie, & O'Brien, 2010; State

University of New York, 1961). Because it was one of the State's four university

centers, a large amount of energy was devoted to the development of doctoral

programs across the disciplines. In 1960, the college offered its first doctoral degree

program, an EdD in Educational Administration and Policy. Two years later, the

college's first PhD program was implemented in the College of Arts & Sciences'

Department of English (Birr & Reda, 1994; Select Committee on Academic Programs

and Priorities, 1975; State University of New York at Albany, 1962). Soon thereafter,

the college was admitted to the Council of Graduate Schools, an organization of

universities offering graduate programs (Vogt, 1984). Throughout the 1960s there

would be a number of doctoral programs that emerged, including French, Criminal

Justice, History, Psychology, and Public Administration; these and other doctoral

programs are listed in Table 1.2. Over the next fifty years, many additional doctoral

programs were implemented. A review of the implementation of doctoral programs

by decade shows there were 28 programs in the 1960s; six in the 1970s; ten in the

1980s; three in the 1990s and two in the 2000s (see Table 1.2).

Table 1.2 Doctoral Granting Departments at the University at Albany from 1960 to present

University at Albany Doctoral Programs from 1960 – Present	
Discipline	**Year**
American History	1963
Anthropology	1971
Astronomy and Space Science	1968
Biology/Biological Sciences	1965
Biochemistry	1971
Biomedical Science	1985
Biometry and Statistics	1988
Chemistry	1965
Classics	1967
Communications	2009
Computer Science	1982
Counseling Psychology	1981

Continued on next page

TABLE 1.2 Doctoral Granting Departments at the University at Albany from 1960 to present (Cont'd)*

Criminal Justice	1965
Curriculum Supervision	1963
Earth and Atmospheric Sciences	1967
Economics	1969
Educational Administration and Policy	1960
Educational Foundations	1967
Educational Philosophy & Measurement	1964
Educational Psychology & Methodology	1964
Educational Theory and Practice	1994
English	1963
Environmental Health and Toxicology	1985
Epidemiology & Biostatistics	1988
European History	1965
French	1964
Geology	1971
German	1971
Guidance and Personnel Services	1962
Health Policy, Management & Behavior	1998
Humanistic Studies	1984
Information Science	1990
Languages, Literatures and Cultures	1968
Latin American and Caribbean Studies	1999
Mathematics and Statistics	1968
Nanosciences and Materials	2001
Organizational Studies	1985
Philosophy	1970
Physics	1965
Political Economy	1966
Political Science	1966
Psychology	1963
Public Administration	1962

Continued on next page

TABLE 1.2 Doctoral Granting Departments at the University at Albany from 1960 to present (Cont'd)*

University at Albany Doctoral Programs from 1960 – Present (Continued)	
Reading	1963
Romance Languages (now Languages Literatures & Cultures	1968
School Administration	1960
School Psychology	1985
Social Welfare	1984
Sociology	1970
Spanish	1966

*Table years indicate either the year of the doctoral program's approval or the year in which the program first appeared in the University's graduate bulletin.

(Birr & Reda, 1994; Coughlin, 1971; Fiess, 2001; Select Committee on Academic Programs and Priorities, 1975, 1990; State University of New York at Albany, 1962, 1963, 1966, 1967,1968, 1969, 1970, 1971a, 1971b, 1990; University at Albany, 2004; Vogt, 1984).

During the 1970s, the direction of doctoral programs at the University would change for several reasons. These included changes affecting government funding as well as concerns on the part of the State regarding an overabundance of doctorates. In 1971, "the SUNY Central Administration placed a moratorium on all new graduate programs" (Vogt, 1984, p. 248). At that juncture there were 34 disciplines in which students could acquire a doctorate. As a result of the moratorium, many doctoral programs which had received program approval did not come to fruition. Among such programs was the doctoral program in Library Science (Coughlin, 1971; Select Committee on Academic Programs and Priorities, 1975; State University of New York at Albany, 1971a). Some doctoral programs were terminated, as in the case of

Astronomy and Space Science, and Political Economy; others were suspended and later reinstated, e.g., Chemistry, History, and Sociology (Florentino, 1979; Select Committee on Academic Programs and Priorities, 1975; Vogt, 1984).

Over the next several decades, new doctoral programs would be added, such as Social Welfare, Computer Science, Information Science/Informatics, and Nanotechnology (Fiess, 2001; Select Committee on Academic Programs and Priorities, 1990). As of this writing there are 34 different doctoral departments/divisions (hereafter referred to as "departments") at the University.

Current Demographic Information of Doctoral Departments

Demographic information was collected on each doctoral department and included the doctoral program's age, number of faculty (FTEs), and number of students enrolled in the fall 2011 semester. Since the approval of the University's first doctoral program in 1960, there have been many changes in the number of programs offered. Of the thirty-four doctoral programs reviewed in this study eighteen were introduced in the 1960s, three in the 1970s, eight in the 1980s, three in the 1990s, and two in the 2000s (see Table 1.3).

Table 1.3 Doctoral Departments, Faculty, and Students Enrolled (2011)

University at Albany Doctoral Programs			
Department	Faculty (FTEs)	Doctoral enrolled	Year*
Anthropology	19.5	68	1971
Atmospheric and Environmental Sciences	15.3	39	1967
Biology	33.5	39	1965
Biomedical Sciences	10.5	48	1985
Business Organizational Studies (suspended)	46.2	6	1985
Chemistry	22.9	44	1965
Computer Science	14.4	28	1982
Communications	16	14	2009

Continued on Next Page

12

Table 1.3 Doctoral Departments, Faculty, and Students Enrolled (2011) Cont'd

University at Albany Doctoral Programs			
Department	**Faculty (FTEs)**	**Doctoral enrolled**	**Year***
Criminal Justice	19.5	47	1965
Economics	23.3	52	1969
Educational Administration and Policy Studies	13	90	1960
Educational and Counseling Psychology (ECPY)	37.2	118	1962
Division - Curriculum and Instruction	with ECPY	with ECPY	1963
Division - Education Psychology and Methodology	with ECPY	with ECPY	1964
Division - School Psychology	with ECPY	with ECPY	1985
Educational Theory & Practice	15.7	82	1994
English	53	64	1963
Environmental Health & Toxicology	10.5	21	1985
Epidemiology & Biostatistics	13.9	50	1988
History	27.1	58	1963
Humanistic Studies (suspended)	0	0	1984
Informatics	7	33	1990
Languages, Literatures and Cultures	27.1	35	1968
Latin American, Caribbean, and US Latino Studies	8.3	20	1999
Mathematics	31.3	39	1968
Nanoscale Science and Engineering	N/A	N/A	2001
Philosophy	13.3	42	1970
Physics	17.8	36	1965
Political Science	22.1	51	1966
Psychology	32.4	89	1963
Public Administration and Policy	23	48	1962
Reading	8.4	30	1963
Social Welfare	28.7	67	1984
Sociology	36.4	108	1970
Total	647.3	1466	

*Table years indicate either the year of the doctoral program's approval or the year in which the program first appeared in the University's graduate bulletin.
(Birr & Reda, 1994; Coughlin, 1971; Select Committee on Academic Programs and

Priorities, 1975, 1990; State University of New York at Albany, 1962, 1963, 1966,

1967,1968, 1969, 1970 1971a, 1971b; University at Albany, 2004; Vogt, 1984).

Many of the earlier doctoral programs, such as Educational Administration and Psychology are still vibrant. A review of the number and percentage of students enrolled in doctoral programs in 2011, shows that there were 1466 students, with 287 (20%) in Humanities programs, 316 (21%) in Science, and 863 (59%) in Social Science. For all doctoral departments at the University there were 647.3 faculty (FTEs); with a distribution of 148.3 (23%) in Humanities, 155.7 (24%) in Science, and 343.3 (53%) in Social Science. These departmental statistics include the numbers for faculty who teach graduate as well as undergraduate instruction (see Table 1.4 and Figure 1.1).

Table 1.4 Number and Percentage of Faculty and Enrolled Doctoral Students In Doctoral Granting Departments

Number and Percentage of Faculty (FTEs) and Enrolled Students in Doctoral Programs							
	Humanities		Science		Social Science		Total
	Number	**Percent**	**Number**	**Percent**	**Number**	**Percent**	**Number**
Faculty (FTEs)	148.3	23%	155.7	24%	343.3	53%	647.3
Students Enrolled	287	20%	316	21%	863	59%	1466

Figure 1.1 Percentage of Faculty and Enrolled Doctoral Students in Doctoral Granting Departments
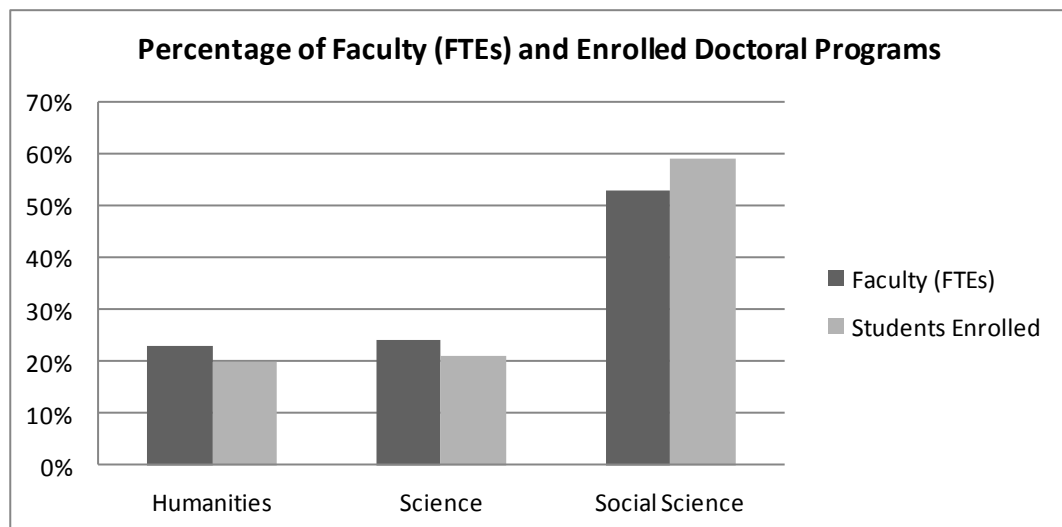
Dissertation Process

Similar to other doctoral programs in the United States, University at Albany students must fulfill several major program requirements: coursework (including core courses of the discipline), a comprehensive examination which assesses skills learned within the program, a dissertation, and an oral defense of the dissertation (Buchanan & Hérubel, 1995; Gardner, 2009). "An important goal of the doctoral program is to prepare students to meaningfully engage in the kind of scholarly writing that is expected of professionals in their field" (Kucan, 2011, p. 229).

As the capstone project of a doctoral program, the dissertation is a culmination of several years of learning, collaborating, researching, and writing. It provides the venue for the doctoral student to demonstrate her "thorough and sophisticated grasp" of a "field of study" (Boote & Beile, 2005, p. 11). Traditionally, the dissertation has been an original, often groundbreaking, research concept (Brause, 2000; Creedy, 2008; Elphinstone, & Schweitzer, 1999; Nettles, & Millett, 2006) or one which builds upon prior problems with a novel approach (Locke & Spirduso, 1976). The dissertation is "synonymous with intellectual enquiry and research activity" (O'Hara, Carte, Dewis, Kay, & Wainwright, 2011, p. 3). The dissertation not only demonstrates and communicates knowledge about a specialized topic but adds to the existing body of knowledge on that topic.

The doctoral thesis is a scholarly, peer-reviewed, and validated monograph (Boyer, 1973; Brause, 2000; Elphinstone, & Schweitzer, 1999; Garson, 2002; Locke & Spirduso, 1976; Lovitts, & Wert, 2008a; Lovitts, & Wert, 2008b). It often constitutes the "rite of passage" into the world of scholarship and may be the first

major work of an individual's lifetime of research (Hawley, 2003, p. 22). For the neophyte scholar, the dissertation may be one of the most exhaustive pieces of work she accomplishes (Lovitts, 2005; Lovitts, 2007; Rudestam & Newton, 2007). The dissertation is considered the major artifact produced in a doctoral program and the candidate is held to rigorous standards of research and methods (Mauch & Park; Pelikan, 1983). The dissertation corroborates the student's ability to use the tools and ideas needed for conducting research (Koefod, 1964; Madsen, 1983). Dissertations demonstrate fluency in a particular topic, and a deep understanding of key research and theoretical concepts, as well as knowledge of appropriate methods (Davis & Parker, 1997; Lovitts, 2005; Dunleavy, 2003; Swetnam, 2004; Ungar, 2001). "As a product, the dissertation is an essential representation that [a student has] achieved the expertise necessary to be a professional in [her] field" (Lovitts, & Wert, 2008, p. vii).

Clearly, the dissertation is one of the most important research initiatives conducted by new scholars in higher education and sources cited serve to support the scholarship of the research. Accuracy when citing the works of others is vital in order to "create reliable and valid knowledge based in a profession" (Heppner & Heppner, 2003, p. 346). As the culmination of scholarly engagement and a sustained research contribution (Lovitts, 2007; Riley, Wood, Clark, Wilkie, & Szivas, 2000), the dissertation lays the groundwork for professional endeavors. It is the "introduction to a career of research and scholarship" (Council of Graduate Schools, 1990, p.22). With changes in the access modes of information, from the traditional print medium

to digitally-based resources, the dissertation needs to continue to maintain a high level of scholarship.

Digital Scholarship

As publications move from print format to electronic versions, and digital scholarship efforts are accepted as publication criteria for faculty tenure and promotion (Andersen, 2004; Fountain, 2004; Kling & Callahan, 2003; Shatz, 2004; Vaughan & Shaw, 2003) it would be expected that dissertation advisors and mentors more readily use and cite electronic resources. Since students engage in similar research threads, they too may reference their mentor's URL citations and/or recommended digital readings.

With the availability of more resources via the World Wide Web, citations within dissertations are found in URL formats in growing numbers. This is particularly true in the case of freestanding websites which contain access to government-generated information, white papers, and professional organization reports. Because the dissertation is the central research endeavor of the doctoral program, its academic standing as a contribution to the body of knowledge may erode and come into question if its supporting documents are difficult or impossible to access.

Structure of Dissertation

This dissertation examines the importance of scholarship and citation within dissertations. More specifically, it addresses the state of accessibility rates of URL dissertation citations through Web browser searches, an Internet search engine, and

Web archival tools. It focuses on policies and practices which may affect the dissertation citation process.

Chapter 2 contains a literature review focusing on several aspects of this topic. It highlights the importance of scholarship and digital scholarship as well as the role and significance of citation. It includes a comprehensive assessment of studies addressing the accessibility rates of URL citations within academic resources and examines current Web archival initiatives and strategies which may help reduce citation erosion.

Chapter 3 introduces the multi-method approach undertaken in this research. It outlines the testing strategies for checking the accessibility rates of URL citations in University dissertations through a Web browser search (Firefox), a search engine (Google), and two Web archive searches (Internet Archive and WebCite). This chapter also discusses the research strategies for collecting information on departmental policies and practices on the use of electronic sources cited in department dissertations. Finally, it addresses the strengths and limitations of the study.

Chapter 4 provides an analysis of the University at Albany dissertations and dissertation citation patterns to determine the level of consistency in the output of these resources. It addresses trends in the use of URL citations within these dissertations, and tests for differences and similarities in the publication rates of dissertations, dissertation citations, and dissertation URL citations. This chapter also presents the results of the dissertation URL citations accessibility rates after each was tested using a Web browser, Google, and two Web archival tools. It

identifies the rate of accessibility for the digital citations with a comparative analysis of the impact of time and academic discipline had on these metrics.

Chapter 5 reviews the results of the qualitative component of this research which consisted of the doctoral departmental interviews. Interview questions were formulated to investigate scholarship, publication engagement, and the departmental policies and practices addressing the use of URL citations in dissertations.

The final chapter, Chapter 6 discusses the analyses of both the use patterns of URL citation in dissertations and doctoral program interview responses. An appraisal of the relationships drawn for the two sets of data -- the URL citation analysis and the departmental interviews -- is discussed. This chapter offers recommendations informed by the findings which may help limit citation erosion and includes suggestions for future research.

Chapter 2. Literature Review

Scholarship

Introduction

Scholarship is defined in various ways in the literature. This multidimensional concept has been interpreted as education (Boyer, 1990), discussion among intellects (Hall, 2007; Van de Ven, 2007), an altruistic desire for knowledge (Benjamin, 1960; Rice, 2005), and research (Matusov & Hampel, 2008), as well as the transmission of ideas and wisdom (Jones, 1979; Paulsen & Feldman, 1995). The *Oxford English dictionary* defines scholarship as "the attainments of a scholar; learning, erudition; esp. proficiency in the Greek and Latin languages and their literature" (Scholarship, 1989, p.630).

Certainly, ancient scholars strongly influenced the pursuit of knowledge and education, emphasizing it as the ideal perfection of citizenship (Kronman, 2007; Nicholls, 2005; Sandys, 1906). Early Greek philosophers/thinkers rejoiced in investigation, ideas, and discourse (Cherniss, 1968). With the Renaissance came the printing press, an invention offering a technology which could rapidly mass produce pamphlets, books, and other print media (Gorniak-Kocikowska, 2001). This provided the impetus for a more efficient and democratic means of transmitting information. It opened new mechanisms for governments to communicate, artists to share their creative endeavors, and scholars to exchange ideas and build upon their knowledge (Roberts, 1993). With the emergence of Enlightenment and the Age of Discovery came an intellectual community highlighting a more communal and

dynamic interest in sharing ideas (Grafton; 1997; Grafton & Jardine, 1986; Kronman, 2007). The publication of scholarly journals in the seventeenth century introduced a stronger means of academic communication through the dissemination of important research findings to the wider research community (Correia & Teixeira, 2005).

The eighteenth and nineteenth centuries witnessed an expansion of educational institutions (Hofstadter & Hardy, 1952) which established a mechanism to teach students from one generation to another (Boulding, 1983; Jencks & Riesman, 1968). The German universities' model encouraged teaching scholars to engage in research initiatives to broaden the scope of scholarship (Gumport, 2005; Herbst, 1965; Hofstadter & Hardy, 1952; Hutchins, 1956; Nettles & Millett, 2006). With research as the ideal, scholarship remains the key focus in higher education today (Budd, 2009; Kronman, 2007).

Boyer's 1990 seminal work, *Scholarship reconsidered,* highlighted four aspects of scholarship: the scholarship of discovery, the scholarship of integration, the scholarship of application, and the scholarship of teaching. He advocated for a more enriched and expanded discussion of intellectual life in academia. In this multi-faceted approach to scholarship, Boyer (1990) termed research as the "scholarship of discovery" and asserted that "no tenets in the academy are held in higher regard than the commitment to knowledge for its own sake, to freedom of inquiry and to following, in a disciplined fashion, an investigation wherever it may lead" (p.17). He supported the idea that "[r]esearch is central to the work of higher learning" (Boyer, 1990, p.17). Glassick, Huber, and Maeroff (1997) reiterated Boyer's ideas, arguing that

21

academic curiosity adds to human knowledge and enhances the "intellectual climate of a college or university" (p.9).

As an integral component of scholarship in higher education, research is best promoted and fostered through intellectual conversation. Van de Ven (2007) suggested that through engaged scholarship, researchers and practitioners can develop collaborative relationships which "produces knowledge that can both advance the scientific enterprise and enlighten a community of practitioners" (p. 7). Scholarly discussion provides a cross-fertilization of ideas (Stimson, 1968), and this type of conversation among intellectual peers and their predecessors encourages inquiry (Hall, 2007). Scholarly ideas that reinforce research efforts, in turn, generate publication, which serves as a legitimating device and a means by which research initiatives "can be further authenticated by the scholarly community" (Horowitz, 1986, p.35). Consequently, society reaps the benefits from these open exchanges through the development of new products, technologies, policies, and theories (Borgman, 2007; Hutchins, 1956; Partridge, 1912).

Many of those new ideas are the byproduct of work originated in doctoral dissertations. Doctoral students preparing dissertations find that this major research endeavor is indeed a learning process in scholarship. Provided they follow Boyer's (1990) elements of scholarship, these doctoral students will encounter inquiry and the desire to know more about their research topic, discovery in exploring new and seminal aspects of their research area, and synthesis of these events through the integration of knowledge gained through their research pursuit. Part of this knowledge building process is rooted in the exploration of the literature.

Citation

Traditionally, much of the literature in a discipline is connected through citation, a scholarly tool that promotes the dynamic interaction of ideas and research. The act of crediting others for their contributions to a work, citation is a fundamental aspect of good scholarship. It is an authoritative mark acknowledging the contributions of others. As a mechanism that links the ideas of one scholar with those of another, citation affords the new scholar, as well as the expert, the opportunity to access ideas, thoughts, and theories with the capacity to prompt new contributions. Through the use of citation, the intellectually curious expand upon the thoughts and ideas written in one text (whether digital or print) to cited work in another. Frequently, this type of inquiry prompts the reader to further explore the bibliography of a text in order to locate a cited book, article, or other resource.

When describing why an author or scientist cites another text, Garfield, Sher and Torpie (1964) observed that a citation reflects and symbolizes the subject of a cited work. Garfield (1979) also advocated that "citations are the formal, explicit linkages between papers that have particular points in common" (p. 1). Henry Small (1978) referred to the citation as a "symbol." These "symbols of concepts or methods" function as connections to earlier works that an author-researcher has embedded as a reference in his or her writings. Authors cite these works because they embody the ideas of the author's discussion and thus the cited reference becomes a "symbol" of those ideas (Small, 1978). Blaise Cronin (1984) eloquently promoted the need for the practice of citation as follows:

Metaphorically speaking, citations are frozen footprints on the landscape of

scholarly achievement; footprints that bear witness to the passage of ideas.

From footprints it is possible to deduce direction; from the configuration and

depth of the imprints it should be possible to construct a picture of those

who have passed by, whilst the distribution and variety furnish clues as to

whether the advance was orderly and purposive (p. 25).

One of the most important functions of the citation is that it links the written

work to a much larger community. When an informatics student cites Shannon and

Weaver's (1948) model of communication to uphold an argument, a connection is

established between valuable works of the past and those of the present. Citation

thus provides a way for the researcher to enhance the integrity of his or her own

work to the reader. This practice supports and adds credibility to an individual's

arguments by connecting them to existing works with recognized authority. In

scholarship, the researcher/writer strives to persuade and put forth new ideas,

models, and theories. Burke (1969) highlighted that the "goal of rhetoric is to

persuade," and one needs authoritative support from other scholars in order to

persuade. By means of citation, the writer calls upon a broader community to add to

his/her own voice to persuade the reader (Gilbert, 1977). "In every field of research,

one's own articulation must engage with previous articulations" (Hall, 2007, p. 73).

The academic community promotes intellectual growth that may perhaps stimulate

the advancement of new medicines and cures, novel writing techniques, or

breakthroughs in technology. The dialogue that is encouraged with the use of

citation supports a learned and dynamic fellowship. As previously stated, Boyer

(1990) promoted scholarship as a form of conversation. Because his thesis is one of

the seminal works in the field (Austin & McDaniels, 2006; Candy, 2000; Gordon, 2007; Gurm, 2009), Boyer's four ideal elements of scholarship are cited in most published intellectual discussions of scholarship and higher education (e.g., Austin & McDaniels, 2006 and Feezel & Welch, 2000).

Another important aspect of citation is that it keeps the writer honest (Abilock, 2009; Van Leunen, 1992; Steneck, 2006). That the academic community is able to rely on and use the studies, arguments, and findings of other scholars is vital. The reader then has the ability to verify the cited work and check the semantic connection, the accuracy of a statistic, or the intricate details of a model. It is imperative that the reader can go back to access the cited work (Axel-Lute, 1982). Citations are critical components of scholarship and "may be as important as the work itself" (Wilkerson, 2006 p. 333).

Citation also makes possible the acknowledgement of the works of others that support a piece of research. By citing the materials of others, a scholar recognizes the cited authors. The use of citations provides "pointers for the reader, and in some cases, endorses[s] or recommend[s] the work and contributions of significant others" (Cronin, Snyder, Rosenbaum, Martinson, & Callahan, 1998, p. 1319). A good scholar cites every publication to which he or she owes an intellectual debt (American Psychological Association, 2010; Modern Language Association of America, 2009; Kochen, 1987). Rudolph and Brackstone (1990) maintained that "scholars who are cavalier in reporting their sources should realize that their credibility could someday rest on the sources they cited" (p. A56). The erudite, by contrast, will aim to acknowledge scholars of the past for the benefit of

today and tomorrow (Sandys, 1906). "Scholarship does not grow in a vacuum. Rather, it is always related to previous work, if even just tangentially or in opposition to" (Bermejo, 2007, p. 7).

Scholars, researchers, and students alike have relied on the accuracy of citations in order to retrieve cited work for years. Citations are not always documented correctly, however, and this phenomenon is recognized as a widespread problem in the scholarly community (Awrey, Inaba, Barmparas, Recinos, Teixeira, Chan, & ... Demetriades, 2011; Benning & Speer, 1993; Boyce & Banning, 1979; Buchanan, 2007; Davids, Weigl, Edmonds, & Blackhurst, 2010; Morrisey, 2002; Oren & Watson, 2009; Pope, 1992; Rudolph & Brackstone, 1990; Sweetland, 1989; Wyles, 2004). In a recent study on citation accuracy, Raida Gatten (2010) reviewed reference lists from three fashion periodicals. The results of the study revealed a 49% error rate. Inaccurate citations not only make it difficult for the reader to locate a particular resource; they also reflect poorly on the author and "may raise questions about the validity of the rest of the work" (Kronick, 1985, p. 96). If cited resources are inaccessible due to poorly constructed references, then the integrity of the research as well as its scholarship comes into question. Since scholarship is based on previous works, it is vital that consulted resources be accurately cited. The use of inaccurate citation "brings the scientific validity of the research into question and may create difficulty when accessing the cited background data" (Awrey, et al., 2011, p. 475). This is no less true in the case of electronic sources used for the purpose of supporting one's scholarly ideas and arguments.

Digital Scholarship

With the development of the World Wide Web, many cited sources are

digitally-based. This is hardly surprising given that one of the primary reasons for

the creation of the Internet was the development of a venue to broaden scholarly

exchange using computers and networked systems (Abbate, 1999; Jonscher, 1999;

Lehnert & Kopec, 2008; Whittaker, 2004). The design of the World Wide Web was

also driven by the need to share scholarly endeavors and efficiently manage

research documents more readily (Borgman & Furner, 2002; Deegan & Taner, 2002;

Feather, 2008). With its expansion, a vast amount of information became readily

available to scholars, students, and the general public (Feather, 2008; Jonscher,

1999; Schmiede, 2009; Tonta, 2005). Over time, scholarly resources, such as

monographs, conference proceedings, and peer-reviewed articles—traditionally

print publications—became available in electronic format. Since the mid 1990s, the

migration from print to electronic sources of journals (Harter & Kim, 1996; Jansen,

2006; King, Boyce, Montgomery, & Tenopir 2003; Mogge, 1999), books and

reference materials (Bradford, Costello, & Lenholt, 2005; Robbins, McCain &

Scrivener, 2006) organizational papers (Griffin, 2003) and government publications

(Block, 2003; Brown, 2004; Brown, 2011; Griffin, 2002; Hernon & Dugan 2004;

Jorgensen, 2006; Kendall, Bakke & McDaniels, 2006; Singer, 2003) has grown

exponentially. Large quantities of electronic access links are continually added to

resources that index these materials, such as library catalogs and research

databases (Baruth, 2000; Dillon & Jul, 1996; Jul, 1997).

Despite the concerns that have been raised in connection with this transition

(Borgman, 2007; Eaton, 2005; Fialkoff, 2003; Waters, 2006; Weller, 2001), the

format preference of academic resources has steadily shifted from print to digital

(Lewis, 2007; Okerson, 2000; von Bubnoff, 2005). In 1991, Horowitz questioned

whether a publication with no hard copy version could be considered a publication.

Concerns with migrating to electronic counterparts from print collections surfaced

since "long-term access –10 years plus -" of electronic documents "is speculative"

(Kling & Spector, 2004, p. 87).

The benefits of digital scholarly publications, however, cannot be discounted.

One important consideration is that the time during which electronic resources are

readily available online is shorter than that of the availability of print versions

(Shatz, 2004). Given that "scholarship is a cumulative process," it is most efficient

when disseminated as rapidly as possible in order that bad ideas can be dismissed

and good ideas promoted and expanded upon (Borgman, 2007, p. 35). In many

cases, the speed of digital publication expedites the dissemination of scholarly

information. Additionally, networked communication fosters scholarly

collaborations within academe (Ross, 1999). As publications migrate from print to

electronic versions, digital scholarship efforts have gained acceptance in terms of

the scholarship criteria of tenure and promotion processes (Andersen, 2004;

Fountain, 2004; Kling & Callahan, 2003; Shatz, 2004; Vaughan & Shaw, 2003).

With the transition to digital publications, citation patterns have witnessed a dramatic change, as more scholars and researchers value the benefits of online resources (Campbell & Campbell, 1995; Herring, 2002). Because electronic resources are easier to access than more traditional formats (i.e., print), their usage has increased dramatically (Bugeja & Dimitrova, 2009; Davis, 2003; Fulda, 2000). Uniform Resource Locators (URLs) are regularly referenced in scholarly publications due to the growth of free and open access scientific literature (Goh & Ng, 2007; Kurtz et al., 2005; Wouters & de Vries, 2004), as well as government documents (Barnes, 2006). Journal articles (Denehy, 2005; Falagas, Karveli, & Tritsaroli, 2008), dissertations (Clarke & Oppenheim, 2006), and newspapers (Trumbo & Trumbo, 2002) have likewise experienced a marked increase in the use of referenced electronic resources. The percentage of articles containing at least one URL in dermatology journals, for example, increased from 2.3% in 1999 to 13.5% in 2004 (Wren, Johnson, Crockett, Heilig, Schilling, & Dellavalle, 2006). Research by Hester et al. (2004) found that the number of oncology articles containing URLs rose from 9% in 2001 to 11% in 2002, and subsequently to 16% in 2003. In a study tracking the citation patterns of two Chinese journals, the number of web references per article increased by 262%, while the number of total references per article, increased by 43.6% (Wu, 2009).

Persistence of URLs
Overview

The use of electronic resources is not without its drawbacks, particularly in the case of freestanding URLs. It is frequently difficult to retrieve linked information

at a later date because the anticipated page has experienced an error, such as "404 Not Found," making it difficult, if not impossible, to reach the expected electronic destination. This phenomenon, often referred to as link rot (Ashman, 2000; Benbow, 1998; Evans & Furnell, 2001; Feldman, 1997; Herring, 2007; Lawrence & Giles, 1999; Lynch & Horton, 2009; Garcia, Nagarajan, & Croll, 2003; Smith, 1995), renders the linked information worthless (Broughton, 2008). Due to the ephemeral nature of the World Wide Web, this wreaks havoc with accessibility (Casey, 1998; Dimitrova & Bugeja, 2006; Jasco, 2002; Jorgensen & Jorgensen, 1991; Landesman, 1997).

The inability to access online information presents problems for business (Griffin, 2003; McMillan & White, 2001; Potts, 2007; Thelwall, 2000; Weiss, 2001), law (Aldrich, 2008; Beazley, 2006 Lyons, 2005; Rhodes, 2010; Rumsey, 2002; Whisner, 2005; Wilkerson, 2006), government (Hernon & Dugan, 2004; Martin & Eubank, 2007; O'Mahony, 1998; Quint, 2008), and, most notably, the health fields (Johnson, Hester, Schilling, & Dellavalle, 2004; Schilling, Wren, & Dellavalle, 2004; Sotos, 2006). Estimates on the life span of a website range from 44 days (Kahle, 1997; Lyman, 2002) to 75-100 days (Whittaker, 2004). Martha Anderson (Preserving Bush Websites, 2008), Director of Program Management of the Library of Congress's National Digital Information Infrastructure and Preservation Program, expressed concern that government information in digital formats was at risk due to link rot. There are several estimates on the life span of a website ranging from 44 days (Kahle, 1997; Lyman, 2002) to 75-100 days (Whittaker, 2004).

This raises concerns with the ability to access electronic citations and the impact on scholarship. When a cited source becomes a broken link it reflects poorly

on the citing document (Hilse & Kothe, 2006).The disappearance of Web resources, in particular within cited references cannot be overlooked or ignored (Bugeja & Dimitrova , 2009; Bar-Ilan & Peritz, 2004; Carlson, 2005). One of the most important requirements of scholarly citation is that there is intent of indefinite availability (Davis, 2003; Germain, 2000; Russell & Kane, 2008). Without a viable link, whether in print or electronic format, it is difficult to preserve scholarly communication (Davis, 2003). Concern with the increased rate in the instability of electronic sources has come from researchers, editors, and professional associations (Bugeja, Dimitrova, & Hong, 2008; Dimitrova & Bugeja, 2007a; Lawrence, et al., 2001).

Researchers have attempted to quantify the reliability of Web sites. Wallace Koehler (1999) described the results of a four-year longitudinal study to investigate the stability of Web objects. In a follow-up study, he monitored changes to a fixed set of Web pages over time and found the average half-life of a Web page to be about two years (Koehler, 2002). In a separate longitudinal study, a set of 50 Web sources was tracked between 1996 and 2004, at the end of which only 22% of the original links were still accessible—revealing a half-life of less than four years (Smith, 2005). A 2005 study of pediatric Web sites found an 11% annual disappearance of sources (Hernandez-Borges, Jiménez-Sosa, de Arcaya, Macías-Cervi, Gaspar-Guardado & Ruíz-Rabaza, 2005). A longitudinal study focused on health-related Web sites documented a 73% survival rate (McMillan & White, 2001). Germain (2000) tracked a set of URLs cited in academic journal articles over time, and found that nearly half of them could no longer be accessed at the end of a three-year period.

31

Over the past fifteen years, there has been a strong push to add electronic materials to academic resources, such as online library catalogs and databases (Childress, Jul & Miller, 1998; Jul, 1997, 1998). Maintenance issues occur with uploaded URL sources in catalogs and databases (Beam & Copeland, 2001; Boydston & Leysen, 2002; Brown, 2004; Burke, Germain, & Van Ullen, 2003; Martin, 1996; Matthews & Owings, 1999; Strader & Hamill, 2007). A transition such as this requires that staff stay on top of changes in linked websites which may include a broken hyperlink or an alteration to content. Constant changes in the URL sources within bibliographic records translate into a labor intensive commitment for libraries (Beam & Copeland, 2001).

Disciplines

The issue of link decay is not new to scholars and researchers in varying academic fields. A number of studies have tested the rate of accessibility for URL sources cited in academic sources. The following review of these studies illustrates the magnitude of this issue within specific disciplines. It should be noted, however, that few studies address cross-discipline comparisons.

Reports from the health sciences reflect a significant concern about the practice of using URL sources in scholarly materials (Johnson, Hester, Schilling, & Dellavalle, 2004; Schilling, Wren, & Dellavalle, 2004; Sotos, 2006; Wagner, Gebremichael, Taylor, & Soltys, 2009). On average, there is at least one cited URL Web page in most major journal articles in this field (e.g., *Science, New England Journal of Medicine*). Because journals are the primary source for accessing professional information in the medical field, it is vital that the cited material in

these periodicals is maintained for accessibility and content accuracy (Ducet, Liu, & Fontelo, 2008; Veronin, 2005).

Studies analyzing the accessibility rate of URL references within medical journals have varying results. A review of electronic citations in the *New England Journal of Medicine*, the *Journal of the American Medical Association (JAMA)*, and *Science* found that 13% were inaccessible within 27 months of post-publication (Dellavalle, Hester, Heilig, Drake, Kuntzman, Graber, & Schilling, 2003). In a similar study, 4.4% of URL citations in five major medical journals had eroded after three months from date of publication (Crichlow, Davis, & Winbush, 2004). An analysis of URLs published in five biomedical journals between 1999 and 2005 found a 5.4% annual decay rate of cited URLs (Carnevale & Aronsky, 2007). Another study of random Web citations in the four top-tier journals reported that over 50% of the Web citations failed within 4.5 years (Snyder & Peterson, 2002). Yet another study reported an inaccessibility rate of 14.6% and 17.9% of URL citations in *Lancet* and the *New England Journal of Medicine*, respectively (Falagas, Karveli, & Tritsaroli, 2008).

Another study addressing Web citations in an open access journal found that *Public Library of Science Medicine* (PLoS Medicine) articles published between 2005 and 2007 had a substantial rate of defunct URL sources. Of the 28,177 references, 2,503 (8.9%) were identified as URLs. The researchers found that almost 17% of the URL references were no longer accessible and the rate of URL attrition increased over time (Nagaraja, Joseph, Polen, & Clauson, 2011).

Specialized areas of the medical field are also experiencing problems with Web citations. A citation analysis of the *Annals of Emergency Medicine,* found a 78% inaccessibility rate after five years of publication (Thorp & Brown, 2007). In a study of oncology journal citations, the proportion of inactive URLs was 33% 29 months after publication date (Hester, Helig, Drake, Johnson, Vu, Schilling, ... Dellavalle, 2004). A review of URL references in *AIDS Patient Care and STDs* found an increase in the inaccessibility rates from 21.3% to 41.7% over a four year span (Olfson & Laurence, 2005). In a study of dermatology journals the rate of URL citation access was 81.7% (Wren, Johnson, Crockett, Heilig, Schilling, & Dellavalle, 2006). Several studies in the field of nursing have checked the accessibility rates of Web citations in nursing journals. One such study (Oermann, Nordstrom, Ineson, & Wilmes, 2008) found a 28% rate of dead links, while a previous study (Denehy, 2005) discovered that only 21% of URL citations were still accessible five years post publication.

Several studies have tested the accessibility rates of hyperlinks in Medline records (a medical database). One study revealed that almost 19% of the Web sites in the abstracts were consistently inaccessible (Wren, 2004); a subsequent analysis indicated no change in the rate of URL decay (Wren, 2008). In 2006 Ducut, Liu, and Fontelo tested a set of URL sources from 1994-2006 database records; these were randomly accessed on a daily basis for one month and returned a 16% inaccessibility rate. Using PubMed, Cheung's (2001) attempt to access Web links in medical abstracts found that 57% were live links, 36% were broken links, and 7% were redirects. In another PubMed study, researchers reviewed citations from randomly selected forthcoming publications and reported that within two days of an

article's public release date, 11.9% of the references were expired (Aronsky, Madani, Carnevale, & Duda, 2007).

Other science-related fields suffer various rates of Web citation attrition. A study tracking URL-based citations in four Ecological Society of America journals, dated 1997 to 2005, found that although 19-30% of the electronic sources were inaccessible, Web searches were able to recover 72-84% of the lost URLs at different addresses (Duda & Camp, 2008). In an analysis of electronic citations within eight high-level chemistry journals for the years 1996, 2000, and 2004, the rates of accessibility were 34%, 73% , and 83%, respectively (Brown, 2007).

Similar research conducted in the communications/journalism field points to the same types of problems related to the persistence of URLs used both in article citations and footnotes. In two separate studies, Bugeja & Dimitrova, (2006; 2007a) tested URL-based citations of communication articles published between 2000 and 2003. The results of the initial study showed that by 2004, 37% of nearly 1600 URLs were inaccessible. The second study of 1126 URL citations revealed an inaccessibility rate of 39%. In a study of URLs referenced in footnotes between 2000 and 2003, Bugeja & Dimitrova (2005a; 2005b) found that only 61% of 416 online footnotes were still accessible.

One would expect that the business world would be very interested in the permanence of URLs since commercial entities rely heavily on the Internet and want customers to have consistent access to their online sources. Only a few articles highlight this issue (Controlling your Web site, 1998; Machrone, 1999), however, and the amount of in-depth research is modest. One study examining electronic

references in the publication, *Business Communication Quarterly (BCQ)* for the years 1998 to 2000. The results showed that one-third of the online references from 1998 were active; approximately 50% of the 1999 references were accessible and just above the 50% of the 2000 references were accessible (Griffin, 2003).

The use of Internet footnotes in the legal field, which relies heavily on citation, has increased dramatically since the Web's inception. According to a study conducted by Susan Lyons (2005), three law review articles published in 1994 contained a total of four Web URLs in footnotes, but by 2003 that figure grew to nearly 97,000. A 2001 study assessed randomly selected law review articles from Westlaw's Journals and Law Reviews database for the years 1997-2001. The results established high levels of link rot: 70% in 1997, 66% in 1998, 58% in 1999, 37% in 2000, and 38% in2001 (Rumsey, 2002). In a smaller sample, an examination of online sources in law review articles found that 60% of the articles contained footnotes that were no longer available (Neacsu, 2002). In another study, a review of Internet citations in federal judicial opinions also fared poorly, with 85% of its 1997 cases inaccessible (Barger, 2002). An appraisal of Internet citations referenced in U.S. Supreme Court Opinions dating from 1990 through mid-2005 found that 85.5% of the sources were still accessible (Wilkerson, 2006).

The library and information science (LIS) field has developed an interest in this issue, as access, preservation, and retrieval of scholarly materials are part and parcel of this discipline. In 2004, a report of a four-year review of informatics-related Web sites revealed that nearly 40% of the Web sources vanished and approximately half of the remaining pages were altered (Bar-Ilan & Peritz, 2004).

Casserly and Bird (2003; 2007) conducted two separate studies on the persistence of URLs. In their preliminary research, when the authors studied Internet citations published in library and information science journals between 1999 and 2000, they were able to retrieve more than half (56%) of the sources using the original hyperlink; a Web search increased the retrieval rate to 81%, and a search using the Internet Archive produced 89% of the sources (Casserly & Bird, 2003). Using identical criteria, the researchers searched for the same sample between August 2005 and June 2006, and found approximately 39% of the links were valid, 73% were accessible via a Web search, and 80% were obtainable through the Internet Archive (Casserly & Bird, 2008).

Carmine Sellitto conducted several important studies on URL citation persistence. In 2004, after examining a total of 2162 bibliographic references in scholarly papers, he reported that 45.8% of the 1041 electronic citations were missing (Sellitto). A second study involving 1068 Web citations in 123 academic conference articles published between 1995 and 2003 returned a 46% rate of inaccessibility; "collectively, the missing citations accounted for 22.0% of all citations, which represents a significant reduction in the theoretical knowledge base underpinning many scholarly articles" (Sellitto, 2005, p.695).

A study of Web bibliographies published in the 1995 to 2001 issues of *College & Research Libraries News* reported a half-life of between four and five years (Tyler & McNeil, 2003). A more recent study analyzed references in two Chinese LIS and computer science journals. According to the findings, 44% of electronic references published between 1999 and 2003 were accessible and had an annual

decay rate of 10%, with an access rate of only 10% by the sixth year of the study (Wu, 2009). A study reviewing top tier LIS journals published between 2006 and 2008 found that 95% of the cited Web sources were still available (Riahinia, Zandian, & Azimi, 2011). In a qualitative approach to the issue of URL persistence, results from a survey on the use of electronic publishing in LIS revealed that respondents expressed concern about the ephemeral nature of e-publishing and the lack of stability with citing electronic sources (Zhang, 2001).

A study of computer science literature found that within four years, nearly 50% of referenced URLs became inaccessible and had a four-year half-life (Spinellis, 2003). In 2001, a review of URL citations from computer science journals showed a varying degree of URL inaccessibility ranging from a high of 53% for 1994 to a low of 23% for 1999 (Lawrence, Pennock, Flake, Krovetz, Coetzee, Glover, & ... Giles, 2001).

Surprisingly, a review of the humanities literature produced little research examining the accessibility rate of URL sources. This may be due to the fact that most scholars in these fields rely heavily on books and archival materials in their research endeavors. Bugeja, Dimitrova, and Hong (2008) compared the stability of online sources in the two journals, *American Journalism* and *Journalism History.* According to their findings, URL sources in journalism history periodicals were fewer than those in communication and journalism journals. After testing the online footnotes of the two journals, the researchers found that 53.8% were inaccessible. A 2008 study analyzing the reliability of URL citations in the leading history journals, *Journal of American History* and *American Historical Review,* revealed that

38

articles published seven years earlier had a 38% rate of dead URL citations. The researchers were able to retrieve 57% of the missing resources, however, with the use of digital archives (Russell & Kane, 2008). Since the current study undertakes a cross-discipline analysis of URL citations trends, the results of the Humanities findings will contribute to the literature in this area.

It is quite evident that the rate of accessibility for URL sources is a common theme in the literature. The extent of the conversation varies across disciplines, with a stronger discussion in the medical sciences which has conducted over fifteen studies than in Humanities, which offers only two. The literature indicates that with more than thirteen research studies conducted between the two disciplines, LIS and Communications are dedicating more research efforts to this topic than the other Social Science fields. These studies signify the value of maintaining and preserving online citations, reinforcing the notion that citations are fundamental elements of scholarship and add value to the scholarly exchange of ideas.

Academic Issues

Many educators have found student and pedagogical use of URLs to be problematic (Gresham, 1996; Markwell & Brooks, 2002a; 2002b, 2003; Thompson, 2003; Tomaiuolo, 2006; Trotter, 2002). Markwell and Brooks (2002a; 2002b, 2003, 2008) have conducted extensive research on the implications of using Internet sources in their academic teaching initiatives, with a focus on the sciences. These instructors found that 12% of their online course hyperlinks were down and noted a half-life of 55 months for their course links (Markell & Brooks, 2002a; 2002b).

Issues of URL persistence in other types of academic resources include the use of inaccessible online citations within student papers. As the Internet has grown and information has become easier to access online, it has become the information source of choice (Davis, 2003; Hovde, 2007; Radia, & Stapleton, 2009; Thompson, 2003). Most students view the Internet as the best place to search, locate and access information. No doubt, the ability to easily access assignment information plays a key role in choosing this option (Thompson, 2003). At times, the World Wide Web may be the only place where certain information, such as pre-publications, is available (Barry, 1997). Rather than going to a physical library, students leisurely remain at their workstations, where they are able to retrieve information through the Web, as well as through research databases.

Several researchers have tracked students' work cited pages in order to analyze trends and policy. In a 2007 study, more than 1600 Internet citations from freshman English composition papers for the years 1999 and 2004 were checked for accessibility rates. The 1999 Web citations had a 38% access rate after the first seven months and were down to 9% after seven years. The 2004 electronic references had a 45% access rate after the first year of review (Hovde, 2007). An investigation of bibliographies in undergraduate economics research papers for the years 1996 and 1999 found that only 47% of the 1996 Internet citations and 55% of the 1999 references were accessible (Davis & Cohen, 2001).

Despite a recent citation analysis of dental theses, which reported minimal use of Websites (Cox, 2008), the literature suggests that graduate students are also avid users of electronic sources in their academic pursuits (Bomba, 2000;

Kushkowski, 2005; Thomas, 2004). A dissertation from 2000 (Bomba) investigated patterns of Internet citations in Oklahoma State University dissertations from 1993-1998. Although the results of this early initiative suggest an increase in the use of URLs in doctoral dissertations, the sample was too small to make generalizations. The results of a study examining citation patterns in print and electronic graduate theses published between 1997 and 2003 indicated only a slight difference in the number of electronic citations in each sample with a low accessibility rate of 55% for both results (Kushkowski, 2005). An investigation of electronic references in social work theses for the years 2000 and 2002 found that 77% were accessible after 22 months of submission (Thomas, 2004).

Access Resources

Uniform Resource Locators

In describing Uniform Resource Locators the URI Working Group of the Internet Engineering Task Force noted that this access tool was "the syntax and semantics for a compact string representation for a resource available via the Internet" (Berners-Lee, Masinter, & McCahill, 2000, pp. RFC 1738.1). Providing further detail, the group explains that:

URLs are used to 'locate' resources, by providing an abstract identification of the resource location. Having located a resource, a system may perform a variety of operations on the resource, as might be characterized by such words as 'access,' 'update,' 'replace,' 'find attributes.' In general, only the 'access' method needs to be specified for any URL scheme. (Berners-Lee, Masinter, & McCahill, 2000, pp. RFC 1738.2)

URL form and universal resource identifier syntax is widely accepted and has been used on the World Wide Web since 1990 (Berners-Lee, 2000). That notwithstanding, a serious flaw in electronic referencing accompanies the use of the volatile Uniform Resource Locator (URL). This locator, with its technical limitations, has proven to be a poor choice in the addressing scheme of the World Wide Web (Evans & Furnell, 2001; Evans, Phippen, Mueller, Furnell, Sanders, & Reynolds, 1999; Gresham, 1996; Langston & Tyler, 2004). Machines require that URLs are syntactically correct in order to properly resolve an address. Problems arise when humans err in constructing clear, accurate, and concise location information (Evans & Furnell, 2001; Evans, et al., 1999). Issues surface with site redesigns, cumbersome directory configurations, misspellings, typographical errors, and material removal (Emtage, 1994; Germain, 2000; Notess; 2003). That digital information would remain in one virtual location, with the oversight of one consistent system administrator in perpetuity is unrealistic (Sollins, 1994). Digital addressing strings are transient and may become invalid at any time (Emtage, 1994). "URLs were never intended to be long-lasting names for content; they were designed to be flexible, easily implemented and easily extensible ways to make reference to materials on the 'Net'" (Lynch, 1998 p. 33).

Among basic strategies for limiting link rot is the approach taken by Web developers who make access more reliable by simplifying URL structures (DiCarlo, Pastor, & Markovitz, 2000; Gresham, 1996; Leuf, 2006; McClure, 2006; Tomaiuolo, 2006). Using URL shortening services, such as TinyURL.com and Shorl.com, to create reasonably-sized addresses (Notess, 2003), as well as standardizing citations (Beck

& Beck, 2006), implementing better naming and robust URLs (Phelps & Wilensky, 2004), and commissioning link checking software to maintain electronic collections (Brown, 2004) go a long way toward improving accessibility reliability.

A more viable solution for the persistence of digital resources, however, calls for tool maturity (Leuf, 2006). Recognizing the limitations of URLs, the Internet Engineering Task Force has been working to develop a more dependable system, which incorporates the use of URNs. A Uniform Resource Name (URN) identifies a resource by name within a particular namespace. Unlike URLs, its construction is not dictated by the location of the source. Rather, it is a naming scheme with oversight at a central registry similar to a monograph's ISBN (Ashman, 2000; Lynch, 1998; Molholm, 2006; Shafer, Weibel & Jul, 200l; Schroeder, 2005). URLs are set up at the discretion of individuals, whereas URNs are created under the auspices of an organization (Coyle, 2006; Shafer, Weibel, & Jul, 2001). URNs are globally unique, which enhances persistence. Organizational hierarchies afford a more stable and consistent mechanism for managing Internet addressing as compared with individual Web developers (Coyle, 2006). Several projects addressing persistent locators are currently underway and include:

ARK (Archival Resource Key)

Ark is a naming scheme for networked information. It is a specially constructed identifier with a unique identity which is independent of the current host. Each ARK contains three elements: object access, object metadata, and a faceted commitment statement. These are mainly designed for use by digital object archivists (Kunze, 2004, Molholm, 2006).

43

- **Digital Object Identifiers (DOIs)**

  Created by the Association of American Publishers with assistance from the
  Corporation for National Research Initiatives, DOIs are widely used by
  publishers, commercial and academic, to ensure persistence with book,
  journal, and journal article access (Scharf, 2002). These have been widely
  used in journal article databases because DOIs can be reliably used as
  content hyperlinks from one article to another (Jasco, 2002; Kennedy, 2004;
  Langston & Tyler, 2004). Registration agencies assign DOIs, which is an
  identifier consisting of two parts: a prefix identifies the registering publisher,
  and the suffix which identifies the individual item; an ISBN or ISSN can be
  used as a part of the suffix (Arms, 2001; Chandrakar, 2006; Egghe, 2000;
  Paskin, 1999).

- **Persistent Uniform Resource Locators (PURLs)**

  PURLs were designed and implemented by OCLC (Online Computer Library
  Center) after this organization encountered extensive link rot while working
  on several Internet cataloging projects (Jul, 1997, 1998). This locator was
  originally intended to serve as an interim solution until URNs became more
  sophisticated and then implemented (Lynch, 1998). PURLs point to an
  intermediate resolution service rather than a direct Internet location. This
  resolution service maintains and updates actual URL locations and redirects
  the PURL to updated sites. The intent is that PURLs will act as a permanent

address for an item. Since PURLs are compatible with URNs, the solution will

provide a smooth and transparent transition from PURL to URN technology

(Arms, 2001; Bross, 2003; CENDI Persistent Identification Task Group, 2004;

Jul, 1997, 1998; Lynch, C, 1997; Marill, 1999; Shafer, Weibel, & Jul,2001).

United States government agencies use PURLs extensively. GPO Access uses

PURLs to meet its commitment to providing "permanent public access" to

government-supported Web sites (United States, Government Printing Office, p. 6).

OCLC worked with the government to implement this resource to address several

issues including permanency, authenticity, and the archiving of digital government

documents. These are crucial issues, then and now, to public accessibility of

government information (Hathaway, 2006). Some libraries have implemented PURL

servers for a similar outcome to ensure access to electronically-based resources

(Bracke, 2001).

Several researchers have expressed the following concern regarding the use

of PURLs. When a Web site owner no longer maintains a specific site, the PURL may

be deactivated; end users receive an error message highlighting the deactivated link.

Burke et al. (2003) found that while PURLs had a higher accessibility rate than

URLs, the resource locators still had a 7.5% rate of unavailability. Another study

noted instances of PURLs being redirected to useless information without a

verification process (Brown, 2004).

Digital Access

During the early stages of the World Wide Web, many digital efforts concentrated on indexing and cataloging Internet sources. OCLC worked on numerous projects involving the cataloging and archiving of resources found on the Internet. InterCAT, a project funded by the U.S. Department of Education, was one such endeavor (O'Leary, 2000; Oder, 2000; Oder, Hinman, & Leita, 1998). With the support of libraries and institutions of higher education, the creation, implementation, testing, and evaluation of a searchable database of USMarc records containing electronic location and access information was initiated (Childress, Jul, & Miller, 1998; Woodward, 1996). Many URL access problems surfaced with InterCat, however, including a 3% rate of inactive hyperlinks at any given time (Jul, 1997). The discovery of link rot promoted the creation of PURLs. InterCAT was replaced by NetFirst, although NetFirst, too, was discontinued in 2002, when its records were migrated to the WorldCat database (Dean, 2002; O'Leary, 2000; Oder, 2000; Oder, Hinman, & Leita, 1998; Schieber, 2009).

Digital Preservation

Overview

A review of the literature leaves no room for doubt that issues related to access to electronic sources abound. The World Wide Web offers the opportunity for anyone to self-publish in a world where "information consumers prize immediacy of access over provenance…[and] the emphasis is very much on information publication rather than retention" (Brown, 2006, p. 3). Digitally-born resources

enjoy a shorter life span and call for a different set of preservation requirements than traditionally preserved materials. "The short physical life and fragility of digital documents implies that we need constant backup in order to protect sudden loss of information. Being digital means fragile and ephemeral" (Liu, 2008, p. 11). Conversations regarding digital preservation efforts have been expressed in government agencies, higher education institutions, libraries, and other cultural entities. While the advantages of ready accessibility of information gained from the rapid evolution of technology are obvious, the challenges of maintaining and preserving electronic materials are great—an unforeseen drawback (Russell, 2002).

Lack of preservation efforts for Internet sources, particularly scholarly materials, can limit their longevity and value. In 2006, Donald J. Waters, Program Officer of Scholarly Communications at the Andrew W. Mellon Foundation, wrote that dissemination, preservation, and access embody the life cycle of the scholarly resources used and produced in higher education, and serve as the objects of scholarly communications. The migration from print to digital formats has affected traditional modes of dissemination and access, but most notably preservation (Brown, 2006; Hswe, Kaczmarek, Houset, & Eke, 2009; Liu, 2008; Weiss, 2001). With the development and use of Web archives, digital content can be preserved and may thus gain greater potential for persistence. These electronic repositories will "benefit researchers and practitioners in many areas, ranging from history and sociology to product development and marketing" (Jatowt & Tanaka, 2007, p. 77). Several Web archival projects, discussed below, have addressed the proliferation of Web-based documents (Waters, 2006).

Internet Archive

While there have been several international Web preservation projects (Masanès, 2005; Phillips, 2005) one major independent solution for maintaining the Web's collective knowledge is the Internet Archive (Brown, 2006; Lyman, 2002; Ó Dochartaigh, 2007; Panos, 2003). Brewster Kahle created the Internet Archive in 1996 with the intent of capturing screen shots of Web pages and storing them for future use (Kahle, 1997; Rosenzweig, 2003; Szydlowski, 2010; Veronin, 2003). The archive is thus distinguished by a feature which retains all retrieved copies of Web pages (indexed by their URL) so that changes made to a Web page can be tracked over time, in addition to providing the user with an accessible document. The Internet Archive is a massive endeavor which does not discriminate based on the perceived importance of a site or lack thereof (Thelwall & Vaughan, 2004; West, 2007). Using the Wayback Machine, users are able to retrieve Web pages that have been deleted. Since searches are URL-based, a correct locator is necessary for retrieval purposes. The Internet Archive has proved useful for several academic research projects (Hawking, Craswell, & Thistlewaite, 1999; Vaughan & Thelwall, 2003). It has a corresponding subscription service, known as Archive-It, which libraries can use for searching, cataloging, and preservation purposes (Schmidt, Shelburne, & Vess, 2008). Up until December 2010, authors were able to submit cited Web pages to the Internet Archive for archival purposes (Johnson, Hester, Schilling, & Dellavalle, 2004). This function has since been discontinued (see Appendix G)

Although the Internet Archive has become known as a prime resource for accessing inaccessible Web sources, the project has not been without its limitations. Among such limitations are problems with broken links and overlaying text of retrieved pages (Schmidt, Shelburne, & Vess, 2008). In some cases, hyperlinked pages are not captured, resulting in failed storage of images, supporting documents, and content information (Ryan, Field, & Olfman, 2003; West, 2007). In other cases, password protected sites are missing from the Archive due to lack of access (Panos, 2003). Additionally, because Web page content changes on a constant basis, neither the Internet Archive nor any other digital archives is able to capture every Web page update (Król & Litwin, 2010; Smith, 2003; Whittaker, 2004). According to Masanès (2005), the time lag for capturing pages—approximately two months—creates lapses in the dynamic Web environment. While its original goal was to capture the Web every few months, over the years the Internet Archive has lagged in its timeliness (Nelson, McCown, Smith, & Klein, 2007; What's up with the Wayback Machine?, 2010). Finally, the Internet Archive is not without copyright and intellectual property issues (Howell, 2006; Knutson, 2009), though these can be addressed by using the Internet Archive's opt-out option (Brown, 2006).

WebCite

Another archiving solution, comparable to the Internet Archive, is a utility called WebCite initiated by the International Internet Preservation Consortium. This archiving system permits authors, editors, and publishers to archive freestanding (non-periodical) resources. When citing Internet sources, an author will submit a hyperlink to a document to WebCite for preservation, including both the Web page

URL and the WebCite ID (Ducut, Liu, & Fontelo, 2008; Eysenbach, 2008). When a reader comes across a broken link, he or she can use the WebCite ID number to access a snapshot of the cited Web site as it was when originally cited. Since the Internet Archive no longer accepts Web page capture requests (see Appendix G), WebCite appears to be the best option for proactive storage and preservation of transitory Web sources. WebCite has been recommended as one of several solutions in medical publishing (Eysenbach, Diepgen, & Coiera, 1998; Wagner, Gebremichael, Taylor & Soltys, 2009).

Other Digital Archive Examples

Other Web archiving resources include country/region-specific initiatives, such as the Nordic Web Archive Access Project (Nordic countries), Pandora (Australia), Koninklijke Bibliotheek (Netherlands), and the UK Web Archiving Consortium (United Kingdom countries)(Brown, 2006; Glanville, 2010; Hakala, 2004). Additional projects include the Library of Congress's Minerva, DACHS (Digital Archive of Chinese Studies), and the Chesapeake Project, which collects digital law and government materials (Brown, 2006; Guenther & Myrick, 2006; Rhodes, 2010).

Many of these Web archives are proactive in collecting essential digital sources, aiming to bring collection development into the process as opposed to more passive, intermittent data collection with crawlers (Brown, 2006; Nelson et al., 2007). Problems persist, however, despite collection development strategies. When Rhodes (2010), for example, reviewed the sources archived in the law-related utility, Chesapeake Project, to verify URL stability and explored the accessibility rate

of sources from a three year span (2007-2010), he discovered an increase in URL instability.

LOCKSS

LOCKSS, which stands for "lots of copies keep stuff safe," allows institutions to capture subscribed journals and articles, and store them on a server for preservation purposes (Eaton, 2005; Jacobs & Reich, 2010; Maniatis, Roussopoulos, Giuli, Rosenthal, & Baker, 2005; Moghaddam, 2008; Pool, 2011; Schneider, 2007; Seadle, 2006; Tomaiuolo, 2006). LOCKSS promotes collaborative efforts among libraries to cache authorized electronic journals. The intention of the project is to allow libraries to permanently keep materials they have purchased through subscriptions and preserve them (Maniatis et al. 2005; Moghaddam, 2008). One of the strengths offered by LOCKSS is that materials are stored in their original configuration, ensuring the integrity of the data (Seadle, 2006). CLOCKSS (Controlled LOCKSS) builds on the LOCKSS principle while serving as a shadow archive (Reich, 2008; Tenopir, Baker, Robinson, & Grogg, 2006). Materials can only be used, however, when they are no longer available from the publisher (Horrell, 2008). Both systems work to ensure access to scholarly works over time (Maniatis et al. 2005; Reich, 2008; Tenopir et al., 2006). Unfortunately, the present focus of this solution is limited to journals and does not include other types of resources, such as freestanding Web pages.

Digital archiving and preservation efforts are important. Maniatis et al. aptly point out, however, that "[s]torage alone will not solve the problem of digital preservation. Academic materials have many enemies beyond natural bit rot:

ideologies, governments, corporations, and inadequate budgets. It is essential that sound storage and administration practices are complemented with the institution of communities acting together" (Maniatis, et al., 2005, p. 41).

POLICY
Overview

The literature contains little addressing the need, or call, for better policy to ensure digital preservation of Internet sources. Policies are essential to ensure the preservation of electronically-based information. "Basic to any successful undertaking are sound policies" (Alford, 1942, p. 1332). Policies help organizations, such as corporations, government agencies, or university departments, address important issues, "and are intended to have long life, i.e., to act as a stabilizer" (Juran, 1974, p. 3-1). Policy lays the groundwork for quality-related activity (Alford, 1942, Caplan, 1980, Juran, 1974). Good policies generate standards which are designed to reflect a positive and consistent outcome (Caplan, 1980; National Research Council (U.S.) International Standards, Conformity Assessment, and U.S. Trade Policy Project Committee, 1995).

Research efforts have established that the use of URL citations in academic resources is increasing and that linked information is regularly lost (Bugeja & Dimitrova , 2009; Bar-Ilan & Peritz, 2004; Carlson, 2005). Several researchers have argued that "loss [of academic citations] will continue until better preservation policies are adopted" (Wren, Johnson, Crockett, Heilig, Schilling, & Dellavalle, 2006 p. 1147). Galloway (2004) advanced the idea that in order for digital preservation to be actualized, the information science discipline would need "to be active in the

formulation of, and advocacy for, national information policies" (p. 549). In two separate studies, Cloonan and Sanett (2000; 2005) explored electronic resource preservation strategies and policies in national and university archives. According to their findings, few institutions had acted on or implemented policies for the preservation of digitally-based resources—indicating a large longevity gap for these collections.

Nick Szydlowski (2010) advocated that while the Internet Archive is a notable undertaking, "librarians and archivists should not be lulled into thinking that the job of archiving the Web content that is most important to our patrons will be done by someone else" (p. 38). Abby Smith (2007) noted a need for policy, but expressed concern with respect to financing access to digital information. She highlighted that few people are aware of the economic costs or the potential societal benefits of maintaining and preserving cultural resources for future access, and emphasized the need for funding and commitment to retool "preservation infrastructures … to assure long-term access to digital content" (Smith, 2007, p. 5). The financial burden associated with digital preservation may fall on information owners, such as organizations, who find it too difficult to undertake (Chen, 2001; Cloonan & Sanett, 2000; Cloonan & Sanett, 2005; McCown & Nelson, 2009; Pool, 2011).

In a 2006 letter to the editors of the *Annals of Internal Medicine,* Sotos questioned the journal's practice of accepting URL references when a print counterpart may have been available. The journal editors' response stated that

while they did not wish to be restrictive with the use of Web citations, they had several recommendations to offer to authors:

1.   Avoid citing a Web site with content that would be considered unscholarly if found in print

2.   Avoid using an online reference if a more scholarly print reference is available

3.   Archive (preferably in a durable governmental or institutional archive) cited Web pages, and, if possible, archive free material from e-journals and other Web-based scholarly materials that you might cite in the future if that material is not published in a journal with an established archiving process (Badgett, 2006, p. 77).

In a similar letter addressed to the editor of the journal *BURNS* (Alexander, Ghoneim & Fadhli, 2004), three physicians questioned whether "future researchers" will be able to "verify the claims, theories and findings" that current "scientists have put forth in digital format where archiving systems cannot be trusted." They encouraged the editor to adopt "strict guidelines" for cited Web resources, promoting the idea that print resources need to be used until digital access, preservation, and storage become as reliable as print (p. 251). Without preservation initiatives and/or policies, researchers of the future will not be able to access research of the past (Galloway, 2004; Kirchhoff, 2008).

Some have suggested that policies might encourage authors to submit cited Web pages to WebCite for archival purposes (Eysenbach, 2008; Johnson, Hester, Schilling, & Dellavalle, 2004). Other recommendations along this line advocate for

policy stipulating that a photocopy of each reference accompany manuscript submissions and that editors return manuscripts with extensive link rot (Wyles, 2004). Nonetheless, while the conversation in the literature concerning the lack of persistence of URL citations is extensive, discussion addressing the preservation of these academic materials is limited. This study undertakes to address both aspects by identifying trends in the use of URLs in academic sources, as well as trends in policies and practices of the overarching administration. Chapter 3 discusses the methods used for this study, provides specifics pertaining to data collection of the dissertation citations from the University at Albany, reviews the particulars of the techniques used to measure accessibility rates of the URL citations, highlights the details of the departmental interviews, and reviews the strategies used for data analysis.

Chapter 3. Methods

Introduction

As noted in the previous chapters, research findings confirm an increase in the use of URLs cited within academic resources. In addition, many of these studies have reported concerns regarding the rate of accessibility of cited materials. The literature is lacking, however, in terms of comparable measures of accessibility rates of URL citations across time and disciplines. A meta-analysis approach to these research findings is not sufficient to produce a useful summary due to the high level of variability in types of reviewed publications. Some researchers, for example have checked access rates of URL citations in journal articles (Duda & Camp, 2008; Hester et al. 2004; Lawrence et al., 2001; Olfson & Laurence, 2005), while others have tested federal judicial opinions (Barger, 2002), websites (Hernandez-Borges et al., 2005)or online course hyperlinks (Markell & Brooks, 2002a; 2002b). The research on citation erosion has also demonstrated inconsistencies across the studies' method choices. Some studies were longitudinal and measured the same URL dataset multiple times over an extended period of time (Germain, 2000; Koehler, 1999; Smith, 2005). Others have limited the testing of URL sources to a certain timeframe without repeating the testing process periodically over a longer span of time (Sellitto, 2004; Russell & Kane, 2008).

Two strategies were used in the present study to address the foregoing limitation. First, this study used units of analysis (i.e., dissertation bibliography citations) from the same institution spanning a variety of disciplines to determine patterns in rates of accessibility of cited URL materials. As noted in Chapter One,

dissertations have a high level of academic rigor and are regarded as scholarly

contributions across varying subject areas. The ability to review a consistent

publication format (dissertations citations) from the same institution (University at

Albany) provided a mechanism for comparing similar units of analysis. This is not

feasible using other venues, such as journal article citations which generally fall

within a single discipline having a variety of publication requirements.

Second, a review of the literature identified a gap in the research relevant to

how institutions of higher education, in general and academic doctoral granting

programs, in particular, are addressing the increasing practice of electronic citation

within student work and the resulting impact on scholarship as these resources

decay over time. It is important to take a closer look at these academic programs in

light of the reality that many of today's doctoral students will be the educators of

tomorrow's scholars.

Part I – Citation Analysis Data Collection
Unit of Analysis

This study focused on several issues relating to the persistence of Internet

sites cited in dissertations. It aimed to answer the primary research questions

highlighted in Chapter One. The quantitative units of analysis for this study were the

references cited in all dissertation (doctoral) bibliographies published at the

University at Albany between 1996 and 2007. The year 1996 was selected as the

starting date because of its close temporal proximity to the commercialization of the

World Wide Web circa1995. With the emergence of this information technology, it is

reasonable to expect that students would begin to use Web-based resources for research purposes. It is likewise reasonable to expect that earlier published dissertations (pre-1996) would contain few to no URL sources since access to online materials was limited. The end date of 2007 was selected in order to create a time lapse which allowed for measurability of link decay (i.e., a sufficient number of years elapsing between date of publication and measurement of accessibility).

A close analysis of the URL citations for the period in question was conducted to verify accessibility rates. This analysis revealed that the citations contained numerous freestanding URLs, which are hyperlinks to Web pages or sites that are not subscription-based databases. Subscription resources include such materials as newspapers, magazines, or journals (e.g., *New York Times, Scientific American, Journal of Psychology)*. Because subscription databases manage sources with a formal publication process, there would be a greater likelihood of retrieving these indexed materials than freestanding URLs. As pointed out in Chapter One, freestanding Web pages are online sources containing features such as text, graphics, and statistics that are uploaded and maintained by individuals, government agencies, corporations, and educational institutions. Examples of freestanding URLs include links to government reports, association white papers, course materials, and organizational statistics. These do not include materials from publisher databases (e.g., *Communication Abstracts*), typically available by subscription, which have formal publication and archiving utilities.

Data Collection

Stage I – Review of Dissertation Bibliographies

As a preliminary to this research, a review of the dissertation bibliographies was conducted to ensure an adequate dataset for analysis. An earlier dissertation (2000) using similar quantitative research methods found only a small number of URLs within the final dataset, thus limiting both the conclusions of the analysis and the ability to identify significant patterns (Bomba, 2000). A 2007 study of doctoral theses bibliographies found that only 18% of the cited references were web-based (Ansari and Ahmad, 2007). This preliminary step was thus necessary in order to ascertain that there were enough Web citations for an adequate sample.

The current study's dataset was acquired by conducting an initial search of the University Libraries' online catalog, Minerva, for retrieval of the University at Albany's dissertations' catalog records. This search was performed by limiting the catalog's "library/collection/format" option to "UA Dissertations/Theses," selecting a specific publication year in the date field, and entering "wcl=mcm or wsl=albr" into the catalog's expert search window, where wcl searches a specific collection, mcm limits to microfilm, wsl searches a specific sublibrary, and albr confines the results to online sources only. This search resulted in the removal all master's theses which are not available on microfilm or online. Individual searches were conducted for each of the twelve years in question (1996-2007).

The catalog searches retrieved the respective records for each dissertation. Each record contained bibliographic information (e.g., dissertation title, author, year of publication, doctoral granting department), call number, and if available, an

electronic link to the ProQuest database, *Dissertations & Theses*. All dissertations

available from the *Dissertations & Theses* database were downloaded. The

dissertations available in print or microform-only format were manually retrieved

and their bibliographies were scanned or photocopied.

Once retrieved, the bibliography of each dissertation was manually reviewed

and the following information was documented:

1. Dissertation title

2. Doctoral granting department

3. College (e.g., Rockefeller)

4. Year of publication

5. Total number of cited sources

6. Number of URL sources cited

    a. Top-level domain (e.g., .org, .gov)

    b. Reference to any archiving information (e.g.,
       submission to Internet Archive, print copy held
       by author, etc.)

7. PURL status

During the 12 year period in question, 1876 dissertations containing a total

of 263,727 cited references were published. Referenced materials include books,

journal articles, government documents, and freestanding URL sources (see Table

3.1). Each of the 263,727 citations was reviewed for its URL status. If the citation

contained a link with "http," "www," "ftp," "gopher," or other URL indicator, it was

marked for further investigation. This examination aimed to eliminate periodical

sources.

Table 3.1 Distribution of University at Albany Citations and Dissertations by Year

Number of Citations and Dissertations		
Year	Citations	Dissertations
1996	18027	140
1997	20117	156
1998	23226	167
1999	17790	142
2000	23073	164
2001	19998	151
2002	24155	170
2003	24232	172
2004	22712	158
2005	20722	136
2006	22052	147
2007	27623	173
Total	263727	1876

Once the review was completed, 603 dissertations emerged with at least one

freestanding URL in the bibliography section. In total, 4,258 URLs were cited, with a

low of six in 1996 and a high of 815 in 2006 (see Table 3.2). From this initial query it

was determined that the dataset contained a sufficient number of units of analysis to

identify quantifiable patterns across both variables of interest—years and subject

areas.

Table 3.2 Distribution of Freestanding URL Citations by Year

Number Freestanding URL Citations	
Year	URL Citations
1996	6
1997	5
1998	82
1999	97
2000	196
2001	275
2002	493
2003	413
2004	507
2005	558
2006	815
2007	811
Total	**4258**

Stage II – Accessibility Rate of URL Citations

Three years within the dissertation set—2000, 2003, and 2006—were selected for in-depth analysis in order to identify patterns in the accessibility rates of URL citations within the dissertations represented by these years. This subset contained 1424 URL citations. The year 2000 was selected as the initial dataset year based on the observation that there was minimal use of the World Wide Web as a medium for cited materials in the mid to late 1990s. Given that dissertation research and writing takes several years, it was expected that a stronger trend of citing URLs in dissertations would begin only after 2000. The latter data year, 2006, was incorporated into the analysis since it allowed for enough time between the date of publication and the years of analysis (2010-2011) to determine measurable accessibility rates. The year 2003 was selected as the midpoint between the other two dates.

To verify the persistence of the URL citations, each address (n=1424) was searched to determine whether the site was still accessible. Using the Firefox Web browser, each cited URL was entered into the Firefox search window. This procedure was conducted once a month in three separate months between May 2010 and January 2011. In all cases, at least 31 days transpired between each measurement. The three different access days were used to insure against temporary interruptions caused by downed servers, poor network connections, or issues of bandwidth. All searches were performed between 5 a.m. and 12 p.m. EST. Whereas, the methods in some studies use automated link checkers or other software applications for checking access (Goh & Ng, 2007; Strader & Hamill, 2007; Wren 2004, 2008), all searches in this study were conducted manually in order to verify not only search results, but actual retrieved page content as well. Any sources that were not freestanding URLs, and could be identified at this point as a periodical-type Web source were removed from the dataset. There were 137 URLs that fit that criteria and were non-freestanding hyperlinks. These were removed from the 1424 set which left 1287 remaining.

In instances where simple URL errors were detected during the manual searches, some effort was made to access a site by correcting for misspelling or syntax errors. This included, for example, the addition of periods where an author may have omitted the pre-html decimal, addition of a "w" to "ww," or correction for a more appropriate top-level domain, such as .edu rather than .com.

As specified in Chapter One, accessibility of a URL citation was understood to mean the ability to access an active Web page or cite which contained no error

messages. If an onsite index or search tool was retrieved, and could link to or

retrieve the cited work through a search, the URL citation was considered

accessible. Citations containing URLs which had been moved or redirected, or

which accessed a host or site other than that indicated in the bibliography URL were

also regarded as accessible, but were identified as retrieving content inaccuracies.

Error Messages

When a URL cannot be accessed, a variety of error messages may be elicited.

The error messages "File Not Found," "404 Not Found," and "Not Found" indicate

that the browser cannot locate the specified Web site, due either to relocation or

removal of the site. The message "Failed DNS lookup" appears when the Domain

Name System (DNS) cannot resolve the URL to a valid IP address. "Host Unknown,"

"Unable to Locate Server," "Socket Error," and "No Response" messages

communicate the inability to connect to the remote server. This may occur when the

remote server is either too busy or is no longer in existence. Generally, remote

computers send only error messages. Unless instructed, they provide no forwarding

address or any indication of a site's relocation.

Recorded Data

On each of the attempted access dates, the following coded notations were

recorded for each of the URL citations. These notations include both coding for error

messages for inaccessible links (indicating why the link was unavailable) and coded

messages for information pertinent to the accessible sources (indicating content

inaccuracies).

      1.   Accessible link

a. No

b. Yes

2. Type of error message for Non Accessible links

    a. Sorry Not Found/404 Not Found

    b. Article Not Found

    c. Does Not Exist

    d. Password Needed

    e. Cannot Find Server

    f. Domain Expired

    g. Google Redirect

3. Type of notations for Accessible links with Content Inaccuracies

    a. Main Website (redirected to host)

    b. Move Forward Given (user linked to relocated Web page of URL citation)

    c. Not Cited Content

4. Top level domain

5. PURL status (it is a persistent URL)

6. Citation access date(date indicated in the citation)

7. Access attempt date

8. Additional notes for persistent URLS not accessing the cited data

    a. Links but Not Content

    b. Redirect to Main Host Site

    c. Bibliography Entry (listed in a bibliography)

d. Via Page Search (searched through site index)

e. Move Forward Given (new URL provided)

Because these URL citations were tested on three separate occasions, each of which could return a variety of results, it was necessary to define criteria for whether the URL would be considered ultimately accessible upon completion of all three attempts. It was decided that if the URL citation was accessible on at least two of three attempts or upon the last testing date, the URL was to be considered accessible. A review of the error message logs from the three access attempts revealed that only ten URL citations did not have identical error messages for the different access attempts. In these ten cases, the error message that occurred most frequently among the three responses was recorded as the status error for any given URL. For example, if the first testing retrieved a "404 Not Found" response and the other two access attempts received "domain expired" responses, the error was recorded as "domain expired."

Google Searches

In addition to attempting to locate sources through the cited URL, each source (n=1287) was evaluated for accessibility by conducting a search using the search engine Google. This search engine was selected because of its large collection of indexed Web sites and its relevancy ranking. Pertinent information from each citation—author, title of source, and publisher—was entered and a search was executed. This method was employed in as many as six searches using various configurations of the citation search terms. The first thirty retrieved links were then reviewed to identify any matching sources. Here the trend should be noted that in

studies examining the accuracy of search engine retrieval results and ranking, "most investigators only consider the first ten to twenty results" (Lewandowski, 2008, p. 917). Moreover, in Casserly and Bird's (2003) URL accessibility study which used similar methods when searching for URL sources using Google, the researchers viewed only the first twenty-five of the search engine's results. On this basis it was determined that a review of the first thirty results was adequate for the purposes of this study. Searches that retrieved a matching source using the citation information provided in the dissertation were coded as accessible. If no appropriate match for the citation information was found in the displayed set, the URL source was considered inaccessible. The following data were recorded for further analysis: Accessible

1. Accessible – New Server

2. Accessible – Found via URL

3. Accessible – Host Page

4. Inaccessible – Not Cited Source (usually cited in another bibliography)

5. Inaccessible -   Not Enough Information in Citation

6. Inaccessible – Terminated (in results but no longer available)

Web Archival Tools

Two final attempts to retrieve each of these sources (n=1287) were made by using two Web archival utilities—the Internet Archive (www.archive.org) and WebCite (http://www.webcitation.org/)—in order to determine whether the electronic citation was archived and available using either of these tools. As with the browser searches, the URL was entered into each utility's search window and

similar efforts to correct for typographical or other errors were made to retrieve the cited work. Two common error messages occurred on the Internet Archive. The first of the two, "No Retrieval Fails," was a message appearing on the Archive's retrieval screen telling users that the requested source could not be located in the Archive. The other common error message, "No Robots," indicated that the requested URL was blocked from access and had not captured the content of the cited page (Robots exclusion standard, p. 417).

The following testing information was coded and recorded for future analysis:

1. Web archival tool (Internet Archive or WebCite)

2. Ability to Access Source

3. Error Messages

   a. No Retrieval Fails (Internet Archive)

   b. No Robots (Internet Archive)

   c. Not Cited Materials

   d. Host Site Not Cited Content

4. Date website was captured and archived in the Internet Archive or WebCite

As noted in Chapter 2, the Internet Archive and WebCite log the date that a webpage is captured or posted, respectively (see Appendix A and Appendix B for screen shots of these applications). When searching for the cited dissertation URLs, this date information was used to verify whether the corresponding date within the citation's recorded access date (if provided) was available on either Internet Archive or WebCite. Otherwise, the date closest to the year of dissertation publication was

selected as it could reasonably be expected that all cited freestanding URLs would have been checked for availability just prior to final submission. Within the Internet Archive searches, it became apparent that many of the results linked to error messages (e.g., "Index Metadata Error. Sorry, we are unable to locate this page in the archive due to an indexing error") or to sources which contained different content. Some of the content change may have resulted from page redirects (i.e., the URL worked but retrieved another page, with different content, because the page owner had redirected the link), as observed during the direct link checking process. Results of the Internet Archive searches were documented as accessible only if similar content to the cited freestanding URL was retrieved within three link tries of the listed sources, since this would be a standard (n= 3 tries) typical of a conscientious searcher.

Once the results for the accessibility of each URL citation were completed, they were added to the preliminary data. The resulting dataset was subjected to further analysis of the relationships between URL accessibility and dissertation disciplines across years of publication.

Interrater Reliability

To determine the search performance reliability of the main researcher, several information professionals (n=3) were engaged to conduct searches of a sample of the study's dataset. All three interraters hold Master's degrees in either Library or Information Science and have two or more years of employment experience in an academic or public library. The task here was to measure

agreement between the results obtained by the main researcher and the findings of the information professionals. Agreement in this instance was based on the notion that the task was to find the content cited in the URL rather than merely ascertain that the URL does or does not exist.

Each of the three interraters was provided with an instruction sheet (see Appendix C) outlining the same strategies taken by the main researcher, including a search for the cited URL, searching Google for the cited work, and an access attempt using the URL on the Internet Archive. The interraters were not asked to search the Web archiving resource, WebCite, since the main searcher's results in this search category were very low, with only 31 of the study's URL citations (n=1287) producing active and accurate links.

Each of the interraters received a flash drive containing an Access database utility with 54 URL citations from the main dataset. By selecting this subset, the main investigator took into account realistic expectations for the interraters. The estimated time for searching each URL citation from a Web browser, a Google search, and on the Internet Archive was approximately ten minutes, making possible five to six searches per hour. Each interrater was asked to provide ten hours, which, at the estimated search rate, would result in the search of fifty to sixty cited URLs.

The URLs for the subset were selected using a stratified sample with proportional units of analysis taking into account academic discipline of the dissertations, year, and accessibility status. The subset consisted of 54 URL citations, 22% of which were from the Humanities, 19% from Science, and 59% from Social

70

Science. Six URL citations were from 2000, seventeen from 2003, and 31 from 2006.

Accessibility rates were accounted for by using 25% inaccessible URLs and 75%

accessible URLs (see Table 3.3). The final subset was chosen using an Excel random

number generator.

Table 3.3 Interrater Reliability Stratified Sample

Interrater Reliability Stratified Sample					
	Humanities	**Science**	**Social Science**	**Total**	**Percent**
2000	17	31	132	180	14%
2003	135	44	202	381	30%
2006	130	172	424	726	56%
Total	282	247	758		
By Academic Group	22%	19%	59%		
25% Inaccessible	3	3	8	14	
75% Accessible	9	8	23	40	
Final Dataset	**12**	**11**	**31**	**54**	

The results were reviewed to ascertain access consistency (reliability). Two

popular measures of interrater agreement, percent agreement and Cohen's (1960)

Kappa, were calculated between the main researcher's results and the three

professional searchers. The computed percent agreement and Cohen's (1960) Kappa

values were 0.76; 0.81; and 0.72, for each of the three interraters respectively. These

measures exceed the frequently cited agreement thresholds of 0.70 (Neuendorf, 2002).

In addition, the main researcher summarized the searcher performance against

the standard that at least one of the four searchers found the sought after URL. That

is, if one of the searchers found the URL to be accessible then perfect performance

would be for all searchers to find that URL accessible. The possibility of an

erroneous "found" report was discounted given that the searchers are information

professionals with expertise in retrieval of information sources. The percentage

agreement results for each searcher were 0.94 agreement for the main researcher

(S1); 0. 94 for Interrater #1 (S2); 0.87 for Interrater #2 (S3) and 0.91 for Interrater

#3 (S4) (see Table 3.4 for sample results).

Table 3.4 Sample of the Interrater Search Performance

Searcher Performance Summary									
					S # Match	51	51	47	49
S=Searcher					S %	94%	94%	87%	91%
UrlId	S1	S2	S3	S4	UrlExists	S1	S2	S3	S4
U00-0122	No	No	No	No	No	1	1	1	1
U00-0152	Yes	Yes	Yes	Yes	Yes	1	1	1	1
U03-0026	No	No	No	No	No	1	1	1	1
U03-0159	Yes	Yes	No	No	Yes	1	1	0	0
U03-0173	No	No	No	No	No	1	1	1	1
U03-0177	No	No	No	No	No	1	1	1	1

Part II –Department Interviews – Data Collection

Doctoral - Granting Department Interviews

Another component of this research consisted interviews with members of

the University at Albany's doctoral granting departments' chairs, chair assistants,

and faculty. Prior to conducting the interviews, demographic data were collected on

each department including age of doctoral program, number of faculty, and student

enrollment. From the 12 years of dissertation data, the dissertation authors self-

identified that they were from 59 different departments. This figure represents a

discrepancy from the actual number of departments (34) and is explained by the

changes in department names (e.g., Information Science changed to Informatics). As of

this writing (2012), there are 34 doctoral granting entities (departments/divisions) at the University at Albany. To create a more manageable metric, the data were coded into three major academic groups: Humanities, Social Sciences, and Sciences. Dissertations from the Department of Psychology, for example, were coded as Social Science, while theses from the Department of Biology were coded as Science and those from the Department of English as Humanities.

A non-probability sample was used in the interviewee selection component of the study. The sample consisted of fifteen representatives (mainly department chairs) from the five departments with the highest number of published dissertations from each of the three major disciplines—Humanities, Social Sciences, and Sciences. McCracken (1988) emphasizes that "less is more" when selecting a sample. He notes that a representative group offers the "opportunity to glimpse the complicated character, organization, and logic" of a particular "culture" (p.17). The interview process began in November 2011. Each interview with the 15 departments was scheduled through an initial email query to the department chair. An in-person follow-up request was made if there was no response to interview requests. In some cases, the chair opted for a department representative to participate in the interview; frequently this was the point person for the department's graduate studies program.

In-person interviews were selected over surveys or phone interviews, as in-person interviews facilitate a natural rapport between the interviewer and the interviewee, enhancing the likelihood of more accurate answers and the potential for self-generated responses (Bryman, 2004; Shuy, 2003). Research indicates that

more thoughtful responses are given in face-to-face interviews (Sykes & Collins,

1988). These were structured interviews within which interviewees were asked

identical questions; the consistency of questions provided reliability when

aggregating the interviewees' responses (Bryman, 2004; Conrad & Schober, 2008).

Each interview was scheduled for one half hour to one hour. The interview

questionnaire (see Appendix D) contained eleven questions, the first ten of which

focused on the interviewees' research endeavors, as well as departmental practices

and policies concerning the dissertation process. These were a mix of open-ended

questions (e.g., When you write articles, books, other materials, do you use URL

citations? What are your thoughts on the use of electronic citations, mainly free Web

sites and pages?) and closed-ended questions (e.g., Have you used Web archival

utilities, such as the Internet Archive's Wayback Machine or WebCite, to archive

cited Web resources?). While the closed-ended questions afforded interviewees

ease of processing, the open-ended questions allowed for in-depth conversation and

the opportunity to reveal important details (Bryman, 2004). At the end of the

session, each interviewee was prompted for recommendations of individuals in

her/his department who had voiced an opinion on this topic (see question 11 of the

script, Appendix D). Finally, if the department had a policy addressing citation

usage, a request was made to see the requirements for writing citations,

instructional practices, and faculty advisement. Each session was recorded,

transcribed, coded, and analyzed. Because the study involved human subjects, the

research proposal was reviewed and approved by the IRB at the University's Office

of Research Compliance on September 22, 2011 (see Appendix E for the IRB

informed consent documents). The interviews began in November 2011 and the last interview was conducted on March 15, 2012.

Part III - Data Analysis

Stage I - Numeric

The literature points to numerous studies (Casserly & Bird, 2008; Olfson & Laurence, 2005; Rumsey, 2002; Wren, 2008; Wu, 2009) from a variety of disciplines, which confirm that URL citations decay over time and at varying rates within specific academic disciplines. This study seeks to expand upon those studies, investigating similarities and differences in the accessibility rates of URL dissertation citations across disciplines as well as determining the relationship between those rates and the departments' academic citation policies. The analysis in this study looked at the data from different perspectives, including descriptive and bivariate. First, descriptive statistics were used to discover patterns in the 1996-2007 dissertations, as well as the three year subset (2000, 2003, and 2006) of URL citations that were verified for accessibility. Second, bivariate analysis, which aims to determine relationships between two variables (Bryman, 2004), was used to confirm relationships between time, disciplines, and rate of accessibility.

Descriptive

A more detailed analysis of the twelve year comprehensive dissertation dataset (1996 through 2007) was performed using descriptive statistics. The analysis included a comparison of trends across years and disciplines, and reviewed

the number of citations, and number of URL/PURL citations for each discipline and year. An analysis of the top level domain for each URL/PURL citation and the corresponding discipline and year was calculated.

After the URL citations were tested to determine accessibility rates with Web browser searches, the results were analyzed to determine trends in URL access across years and disciplines. A summary of the accessibility rates included the number of URL dissertation citations that linked directly to the cited page, as well as the number of accessible URL citations that did not link to the same content as cited in the dissertation citation. A similar analysis was conducted for Google searches and the Web archival resources, the Internet Archive and WebCite.

Bivariate Analysis

Prior studies suggest that there may be real differences across disciplines (e.g., 38% loss rate/ 7 years in Humanities (Russell & Kane, 2008); 50% loss rate/4 years in Computer Science (Spinellis, 2003); and 78% loss rate/ 5 years in Medicine (Thorp & Brown, 2007)). This study used bivariate analysis to confirm whether there were differences across disciplines by using a dataset containing consistent units of analysis with varying subject areas and dates of publication. A multiple regression model (Mendenhall & Sinich, 2003) was used to determine whether there were differences in the output of (a) dissertations, (b) citations, and (c) citations with at least one URL, across years and by academic group for each of these categories.

The methods of this study called for a review of the accessibility of each URL citation using several online resources, including a Web browser search, a Google search, and an Internet Archive query. Additional t-tests were run to examine differences between accessibility rates by academic group and all years. In this study, t-tests were used to establish whether there were any significant differences in the effectiveness of the retrieval modes (Web browser search, Google search, and Internet Archive search) by academic group.

Stage II – Analysis of Interviews

The interviews with members of the doctoral departments/divisions included a questionnaire containing six closed-ended questions, four open-ended questions and a request for contact information for follow-up interviews. The interviews were conducted and transcribed. Descriptive statistics were used to analyze the closed-ended questions.

Each open-ended question was analyzed following Weber's (1990) standard content analysis procedures. The researcher identified key terms (relevant words or phrases) for each of the questions and used them as recording units. Concepts were drawn from the extracted terms and grouped according to similar terms to form categories of scholarship and citation practice.

Strengths of the Research

Review of Dissertation Citations

The quantitative component of this research consists of a large dataset with a similar unit of analysis (citations within published dissertations). Since all dissertations are peer-reviewed through the dissertation committee and a dissertation defense, expectations with respect to level of scholarship are steady across academic groups. The units of analysis also contain variables (e.g., field of study, year) that afford the ability to explore differences across the accessibility rates of URL citations.

The use of interrater agreement provides a mechanism to measure the reliability of the researcher's coding results by ascertaining whether they are consistent between different coders (Bryman, 2004). To verify the reliability of the study results, interraters checked for agreement of the URL accessibility rates against those of the main researcher. Because all results exceeded the generally accepted rate of 70%, the testing results are deemed reliable.

Department Interviews

Interviews with experts who oversee the dissertation process within departments provided insight into the practices and policies governing dissertation research and writing. These interviews provided the investigator with an "agile instrument" with which to capture how respondents view this aspect of the world (McCracken, 1988, p. 65).

The questionnaire contained both closed- and open-ended questions. One obvious advantage of closed-ended questions lies in their clarity for the interviewee.

Closed-ended questions also increase the capacity to compare answers and reduce variability. Open-ended questions, by contrast, afford interviewees the opportunity to respond on their own terms, to explore the study topic in their own words, and to provide the researcher with unusual responses (Bryman, 2004).

Multi-method Approach

As pointed out by McCracken (1988), quantitative and qualitative methods cannot be substituted one for the other because "they observe different realities, or different aspects of the same reality" (p. 18). A mixed method approach provides a complementary strategy by blending quantitative data (consisting of numbers and trends) with qualitative materials (interviews which deliver words, context and meaning) to produce richer conclusions (Brewer & Hunter, 1989; Bryman, 2004; Creswell, 2009). The multi-method approach provides added value to research, producing results that may be stronger than those produced by a single method approach, in that this strategy allows the researcher to observe varying aspects of an event including both process (practices within doctoral granting departments) and outcomes (use of URL citations within dissertations) (Plano Clark, Creswell, O'Neil Green, & Shope, 2008). In this study the quantitative data were the product of doctoral dissertations. By reviewing the dissertation citations, and the use of URLs within those citations, this study measures the practices students are following in the University's doctoral programs. The interviews provided an essential qualitative counterpart on addressing the state of departmental

instructional oversight of practices and policies governing doctoral students'

research publications.

Limitations of the Research

Review of Dissertation Citations

One of the limitations of the quantitative data in this study is directly related

to the organization of the doctoral departments into larger aggregated group units

(i.e., Humanities, Science, and Social Science). Although this step was taken for the

purpose of eliminating department identification, it resulted in the removal of

granularity and the ability to identify potential differences within the three specific

domains (e.g., one department may have few URL citations, while another has an

overabundance, thereby removing the extremes).

An additional limitation stems from the fact that when writing a dissertation,

doctoral students use the style guides required within their discipline field. Since

manual styles (e.g., *MLA handbook for writers of research papers, Publication manual*

of the American Psychological Association) differ across disciplines, each containing

varying techniques for citing URL sources, some dissertations may have contained

freestanding resources without citing a URL. In such cases, the citations would not

have been included within the dataset.

Lastly, the study uses only one search engine, Google, to verify whether the

cited URLs are indexed in a search engine. In the event that Google's bots have not

indexed the referenced sites, the cited source could not be retrieved during testing.

Bar-Ilan and Peritz (2004; 2009) advocate the use of multiple search engines when

testing for accessibility while emphasizing that there are still limitations since the whole Web is not completely indexed.

Department Interviews

The study had several limitations in the interview component, the first being that with a non-probability sample, there is a possibility that "some units of the population are more likely to be selected than others" (Bryman, 2004, p. 87). The studies' interviewees were major stakeholders in the educational process of doctoral students, thus introducing the potential for bias. Certainly, these interviewees wanted to present the best possible answers to create a positive impression of their respective departments.

Although the open-ended questions produced valuable insights into the doctoral dissertation process at the University, there is room for error and inconsistencies when coding the results to these types of questions (Bryman, 2004). When coding the study's responses, efforts were extended to ensure that categories did not overlap and were distinct, thus leading to stronger validity.

Summary

This chapter has outlined the methods used for both the quantitative and qualitative components of the present study. Chapter 4 presents an analysis of the quantitative data, which consists of doctoral dissertations and the accessibility rates of URL citations contained within them. In Chapter 5, a review and analysis of the responses from the departmental interviews on URL citation practices and policies

is provided. Chapter 6 offers a discussion of both the quantitative and qualitative

results of the study, as well as an analysis of how the two sets of results affect the

state of URL use in doctoral programs.

Chapter 4. Results I

Part I - Citation Analysis

As previously noted, the University at Albany dissertations published between 1996 and 2007 yielded a large dataset containing over a quarter of a million citations. This figure included close to five thousand referenced URL sources. To facilitate greater coherence in this section, the data are reviewed in several stages with a focus on academic groups across years. The data review incorporates the following elements:

- An analysis of the production of dissertations and dissertation citations to help confirm output patterns. Determining similarities and differences between these resources helps establish a baseline for comparison of URL citation use;

- A summary of the number and percentage of dissertations containing at least one URL citation within academic groups across study years. This analysis provides insight into the changing trends of URL citations in dissertations;

- A review of the dataset's URL citations within the dissertations for identification of trends in academic groups across years;

- An examination of a three year subset (2000, 2003, 2006) of URL citations to be tested for accessibility (Can the URL hyperlink access the cited resource?) as well as effectiveness of content (Does the material on the retrieved Web page match the cited source?). This testing was conducted using a Web browser search (Is the URL link still a viable access mechanism?); a Google

Search (Can the cited information be located in this Internet index?); and the
Internet Archive and Webcite (Has the cited Web page been captured and
stored in one of these two Web archival storage facilities?); and

- An assessment of the differences between the access utilities (URL Web
browser search; Google; and Web archival tools) to determine the role they
play with effectively retrieving URL citations by academic group.

Stage I - Review of Dissertation Bibliographies

When evaluating the trends of URL citation use, a review of the
demographics of the dissertations and citations under investigation is valuable in
order to establish context. The data of this study include 1876 doctoral dissertations
approved by 34 doctoral granting departments/divisions of the University at Albany
between 1996 and 2007, inclusive. The mean number of dissertations published per
year was 156, with a low of 136 in 2005 and a high of 173 in 2007 (see Table 4.1).
Although the dissertation authors identified some fifty-nine sponsoring
departments, this figure does not reflect the overlap between departments resulting
from name and/or department changes during the period in question. Several
dissertation authors, for example, listed French Studies as their department,
although this program presently falls under the auspices of Languages, Literatures
and Cultures. In actuality, all of the 34 current doctoral-granting departments are
represented (see Appendix F). This numerical discrepancy was of no consequence
since the data were aggregated into three broader academic groups: Humanities
(n=7), Science (n=9) and Social Science (n=18) in order to streamline analysis and
remove any identification to departments, divisions, or stakeholders.

It was clear from the data that the highest rate of dissertations across study years was produced by the Social Science doctoral programs, which together contributed more than 50% of the total number of dissertations for each year. The distribution of dissertations (n=1876) according to academic groups was as follows: 277 (15%) in Humanities, with a mean of 23.08 per year; 424(22%) in Science with a mean of 35.33 per year; and 1175(63%) in Social Science, with a mean of 97.91(see Table 4.1). The number of dissertations published per year ranged from 13(1996) to 31(2001) in Humanities, 25(2005) to 43 (1998) in Science, and 85(2006) to 114 (2007) in Social Science. The percentage of dissertations published per year ranged from 10% to 20% in Humanities, 18% to 26% in Sciences, and 58% to 69% in Social Sciences (see Table 4.1).

Table 4.1 – Number and Percentage of Dissertations by Academic Group across Study Years

Number and Percentage of Dissertations by Academic Group across Study Years							
	Humanities		Science		Social Science		Total
	Number	Percent	Number	Percent	Number	Percent	Number
1996	13	10%	30	21%	97	69%	140
1997	28	18%	38	24%	90	58%	156
1998	23	14%	43	26%	101	60%	167
1999	15	11%	34	24%	93	65%	142
2000	23	14%	42	26%	99	60%	164
2001	31	20%	33	22%	87	58%	151
2002	29	17%	31	18%	110	65%	170
2003	27	15%	32	19%	113	66%	172
2004	27	17%	36	23%	95	60%	158
2005	20	15%	25	18%	91	67%	136
2006	24	16%	38	26%	85	58%	147
2007	17	10%	42	24%	114	66%	173
Total	277	15%	424	22%	1175	63%	1876
Mean	23.08		35.33		97.91		156.33
SD	5.78		5.51		9.88		13.1

To test for differences in output over the twelve year span, a multiple

regression model (Mendenhall & Sinich, 2003) was run (y=number of dissertations;

x_1=Humanities, x_2=Social Science, x_3=years). The results determined that years

made no significant contribution to differences (R^2=.957, F (3,32)=235.39, p=.555)

in dissertation publication. Rather, the results indicate a significant difference

between the academic groups, Humanities (p=.0003), Science (p=.0000) and Social

Science (p=.0000). Figures 4.1 and 4.2 illustrate the disparity between the high

percentage of dissertations published in Social Science as compared to Humanities

and Science. They also reflect the continuity output of the three academic groups

across the twelve year set.

Figure 4.1. Number of Dissertations by Academic Group across Study Years
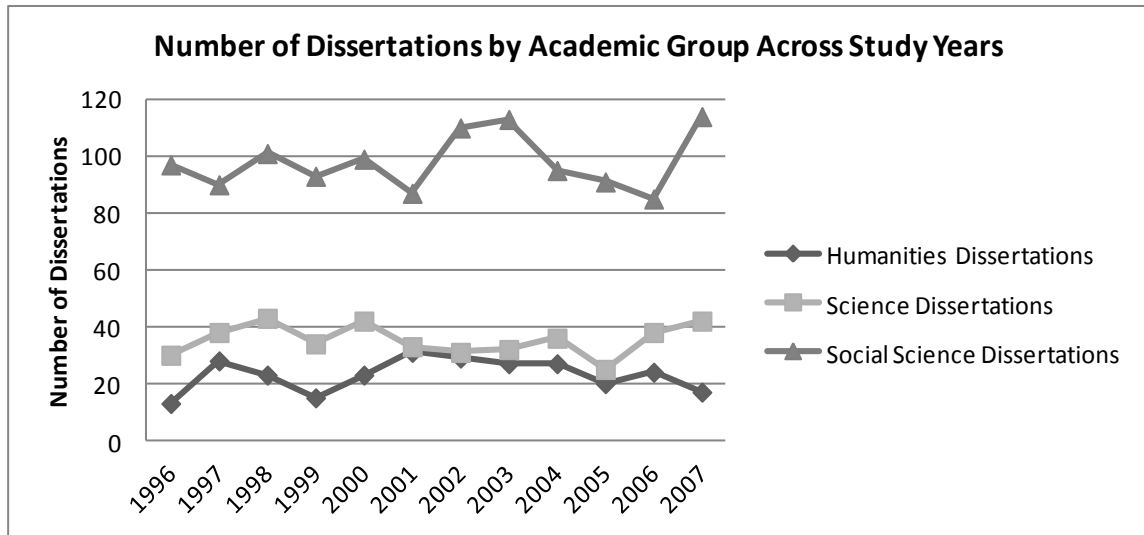
Figure 4.2. Percentage of Dissertations by Academic Group across Study Years
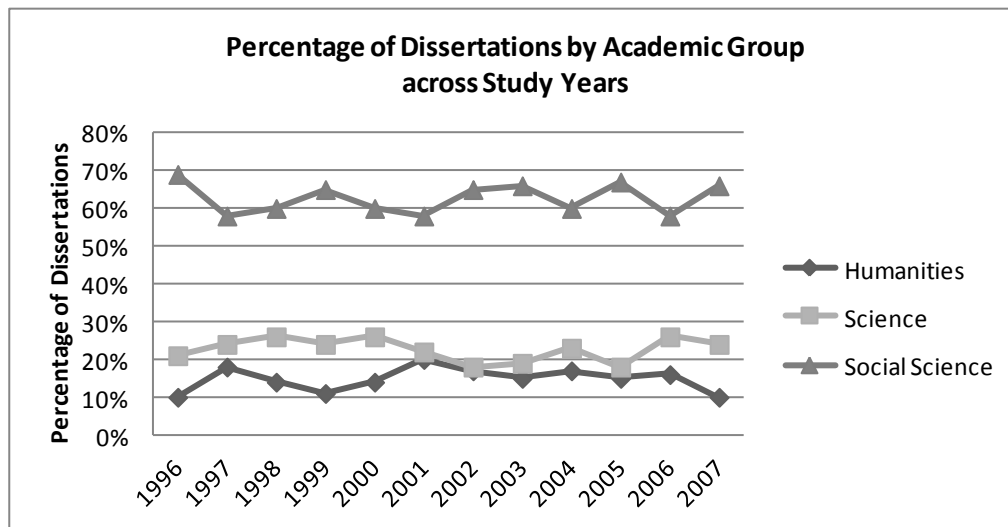
Stage II – Review of Dissertation Citations

As with the review of dissertations published over the study years, an analysis of the citations within the dissertations is helpful to establish differences and similarities in output according to year and academic group. By examining the number and percentage of dissertations and citations, and testing for significance, a base line was established for future comparisons.

The dissertation bibliographies (n=1876) generated a total of 263,727 citations over the twelve years, with a mean of 21977.2. The Humanities produced a total of 44,134 (17%) cited sources with a mean of 3677.8. Science dissertations contained 57,730 (22%) cited sources, with a mean of 4810.8. There were 161,863 (61%) citations in the Social Science theses, with a mean of 13488.5. The percentages of citations per year for the three academic groups fell within the following ranges: 11% (2007) to 26% (2000) in Humanities; 17% (2002) to 26%

(both 2000 and 2006) in Science; and 50% (2001) to 69% (1996) in Social Science

(see Table 4.2).

Table 4.2 Number and Percentage of Citations by Academic Group across Study Years

Number and Percentage of Citations by Academic Group across Study Years							
	Humanities		Science		Social Science		Total
	Number	Percent	Number	Percent	Number	Percent	Number
1996	2390	13%	3315	18%	12322	69%	18027
1997	3058	15%	3942	20%	13117	65%	20117
1998	3762	16%	5370	23%	14094	61%	23226
1999	2354	13%	3447	19%	11989	68%	17790
2000	3694	16%	6077	26%	13302	58%	23073
2001	5135	26%	4765	24%	10098	50%	19998
2002	4493	19%	4072	17%	15590	64%	24155
2003	4903	20%	5453	23%	13876	57%	24232
2004	4507	20%	4891	21%	13314	59%	22712
2005	3081	15%	3766	18%	13875	67%	20722
2006	3777	17%	5745	26%	12530	57%	22052
2007	2980	11%	6887	25%	17756	64%	27623
Total	44134	17%	57730	22%	161863	61%	263727
Mean	3677.8		4810.8		13488.5		21977.2
SD	935.8		1128.5		1895.4		2813.8

As with the analysis of dissertations, these data were tested using a multiple

regression model (Mendenhall & Sinich, 2003) (y=number of dissertations;

x_1=Humanities, x_2=Social Science, x_3=years) to identify differences in citation output

across the twelve year span and between academic groups. When testing across

years, a significant variation (R^2=.93, F (3, 32) =143.77, p=.0142) emerged. The

slope of 160 was very low since the mean number of citations was 21977.2. As with

the dissertations, the results determined that there were significant differences

between the dissertations in Humanities (p=.0372), Science (p=.0000), and Social

88

Science (p=.0000). Again, this is accounted for by the large number of publications in Social Science (see Figures 4.3 and 4.4).

Figure 4.3 Number of Citations by Academic Group across Study Years
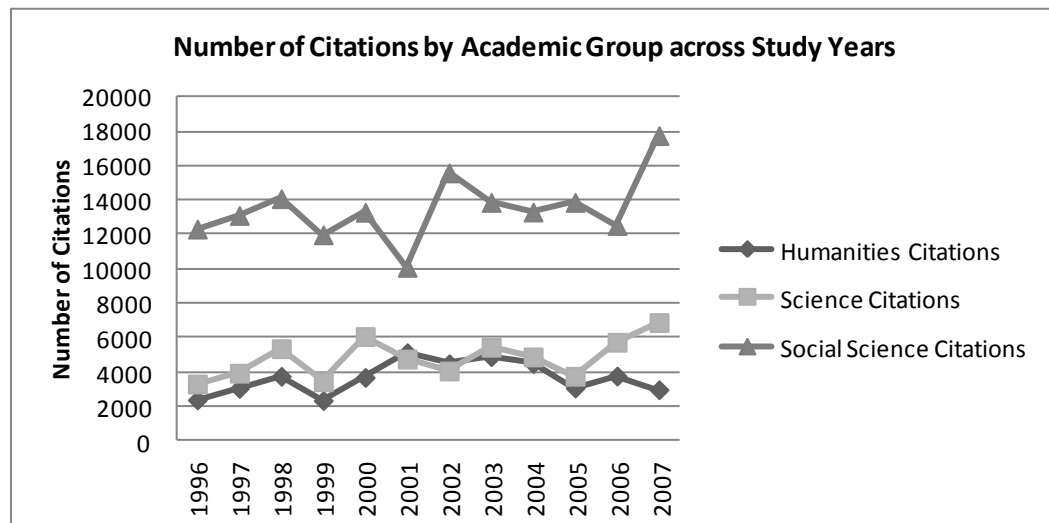
Figure 4.4 Percentage of Citations by Academic Group across Study Years
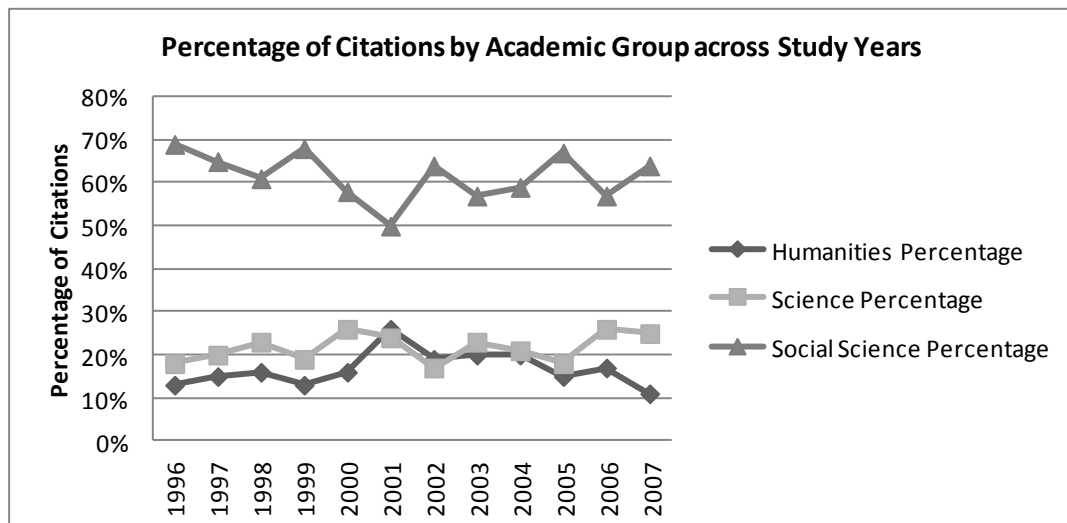
A bar graph comparing mean percentages of dissertations and citations shows that there is little difference between the output percentages within the three academic groups (see Tables 4.1, 4.2 and Figure 4.5).

Figure 4.5 Percentage of Dissertations and Citations by Academic Group

Percentage of Dissertations and Citations by Academic Group
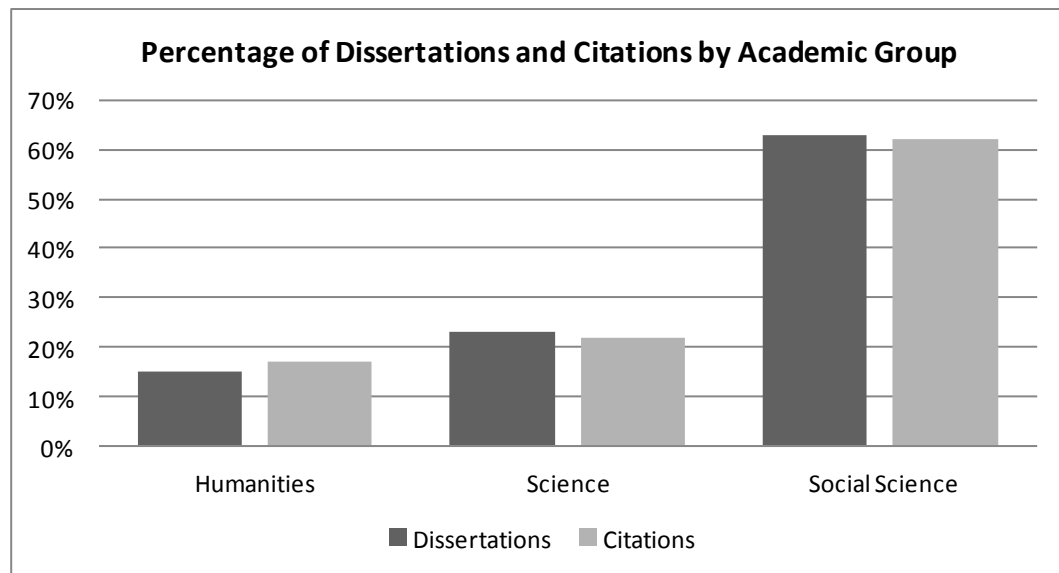
Dissertations with URL Citations by Year

As stated in the literature, the use of Web resources for scholarly citation has increased over time (Goh & Ng, 2007; Kurtz et al., 2005; Wouters & de Vries, 2004; Wren, et al., 2006). A review of the dissertations containing at least one URL citation was undertaken to establish changes in the proportion of URL citations in dissertations. Of the 263, 727cited references, 4,256 (16%) were freestanding URLs. Within all of the dissertations (n=1876), 603 (32%) had at least one freestanding URL. A review of total dissertations by academic group indicates that 98 (5%) Humanities dissertations; 99 (5%) Science dissertations and 406 (22%) Social Science dissertations contained at least one freestanding URL. From the data it can be determined that few URL citations were used in the first four years of the data set. Over the first eleven years, there was a continuous increase in the percentage of dissertations containing URL citations, from 0% in 1996 to 61% in 2006. A slight decrease occurred in the 2007 dissertations (n=92), dropping to 53%. From 1998

onward, Social Science had the highest percentage of dissertations with URL

citations (see Table 4.3).

Table 4.3 Number and Percentage of Dissertations with at Least One URL Citation by Academic Group across Study Years

	Humanities		Science		Social Science		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
1996	0	0%	1	1%	2	1%	3	2%
1997	2	1%	0	0%	1	1%	3	1%
1998	0	0%	2	1%	16	10%	18	11%
1999	2	1%	3	1%	14	10%	19	13%
2000	7	5%	9	5%	20	12%	36	22%
2001	8	6%	9	6%	20	13%	37	25%
2002	14	8%	10	6%	44	26%	68	40%
2003	12	7%	8	5%	55	32%	75	44%
2004	13	9%	13	8%	51	32%	77	49%
2005	13	10%	7	5%	60	44%	80	59%
2006	16	11%	22	15%	57	39%	95	65%
2007	11	6%	15	9%	66	38%	92	53%
Total	98	5%	99	5%	406	22%	603	32%
Mean	8.17		8.25		33.83		50.25	
St. Deviation	5.84		6.4		23.91		34.55	

(Table title row spanning: Dissertations with at Least One URL Citation by Academic Group across Study Years)

A multiple regression model (Mendenhall & Sinich, 2003) (y=number of

dissertations; x_1=Humanities, x_2=Social Science, x_3=year, x_4=Social Science

interaction with year, x_5=Humanities interaction with year) was used to test the

differences between dissertations with at least one URL citation by academic group,

year, and the academic group interaction with year. The results indicate a significant

difference (R^2=.949, $F(5, 30)$ =112.1, p=.0004) in the number of dissertations with

at least one URL citation across years. Though not steep, the slope was 2.

Humanities made no significant (p= .9805) contribution to the model. By contrast,

both Science and Social Science made significant (p=.0000 for both academic

groups) contributions to the changes in URL citation use over time. Figure 4.6 and

Figure 4.7 illustrate these test results.

Figure 4.6 Number of Dissertations with at Least One URL Citation by Academic Group across Study Years
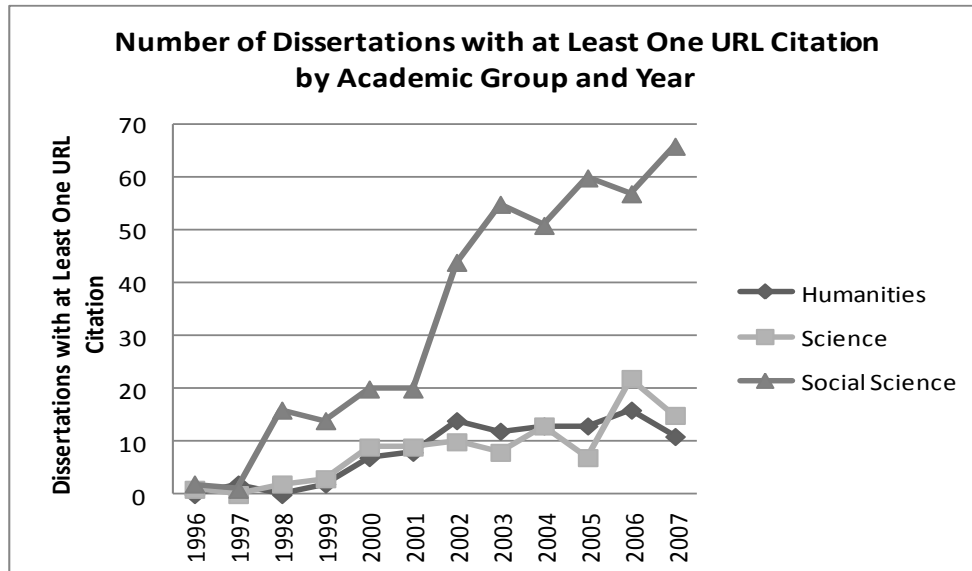
Figure 4.7 Percentage of Dissertations with at Least One URL Citation by Academic Group across Year
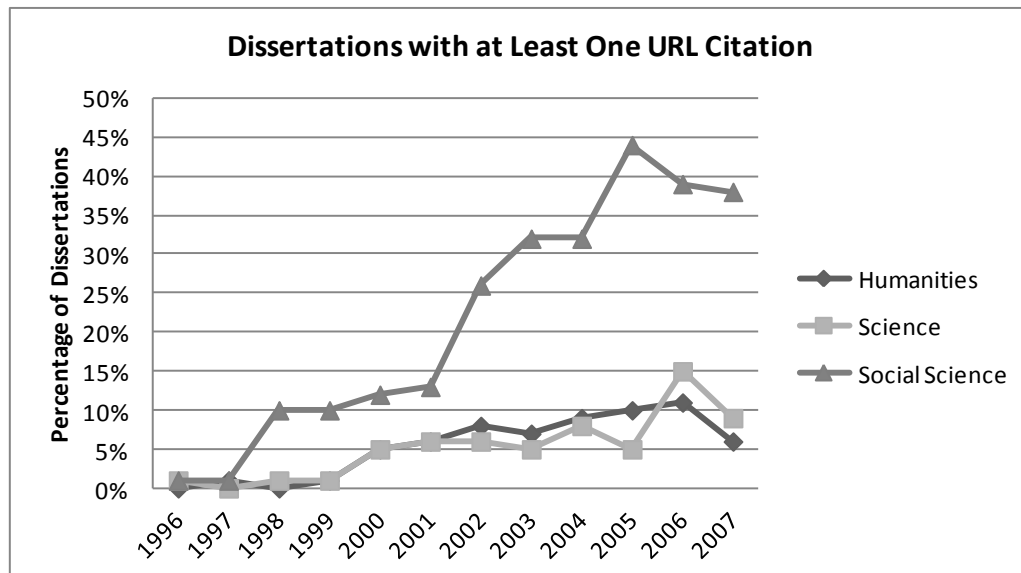
Dissertations with no URL Citations

The pattern of dissertations with no URL citations reveals a reverse trend. A total of 1273 (68%) dissertations contained no freestanding URLs. Further analysis of the no URL subset reveals that 439 (34%) of these dissertations were published within the first three years of the dataset (1996-1998). The first six years of the study included 804 (63%) dissertations with no URL citations. Table 4.4 divides the study years into quartiles. Across the quartile years, the percentage of dissertations with no URL citations dropped from 97% to 34% in Humanities, 97% to 48% in Science, and 93% to 37% in Social Science (see Table 4.4). This downward trend in the percentage of dissertations with no URLs is illustrated in Figure 4.8.

Table 4.4 Number and Percentage of Dissertations with No URLs by Academic Group across Quartile Years

Number and Percentage of Dissertations with No One URL Citations across Quartile Years									
	Humanities			Science			Social Science		
	No URLs	All	Percent	No URLs	All	Percent	No URLs	All	Percent
1996-1998	62	64	97%	108	111	97%	269	288	93%
1999-2001	52	69	75%	88	109	81%	225	279	81%
2002-2004	44	83	53%	68	99	69%	168	318	53%
2005-2007	21	61	34%	61	105	48%	107	290	37%

Figure 4.8 Percentage of Dissertations with No URLs by Academic Group across Quartile Years

Percentage of Dissertations with No URL Citations across Quartile Years
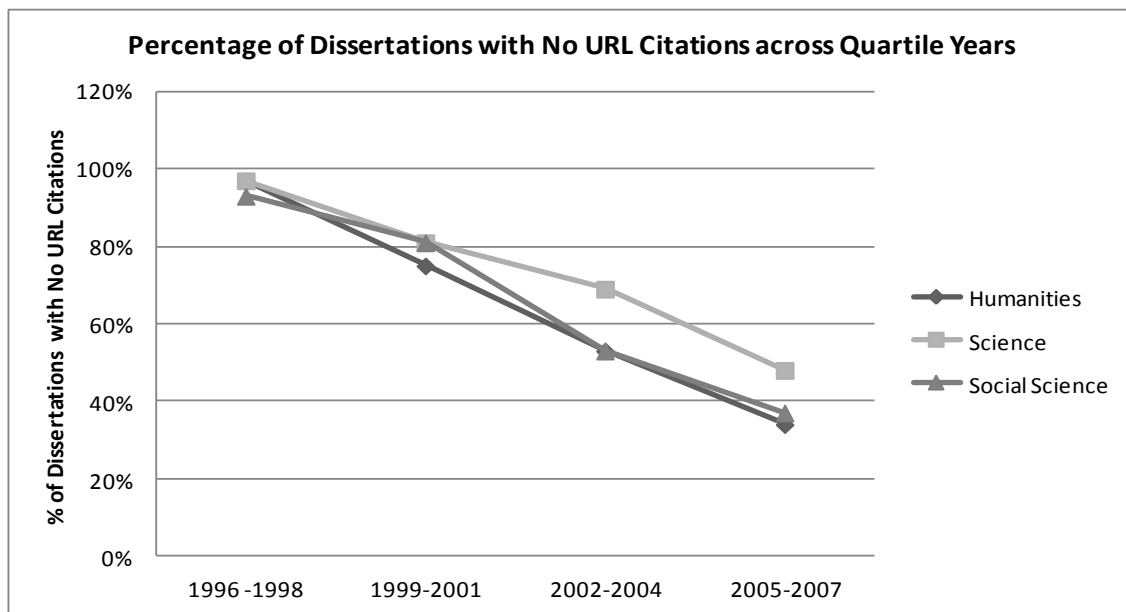
Dissertations with at Least One URL Citations by Academic Group

Humanities dissertations for the years 1996 and 1998 had no URL citations. A high level of URL citation use occurred in Humanities in 2006, accounting for 11% (n=16) of the total humanities dissertations (n=140) published that year. Science contributed at least one dissertation with URLs in all but one year, 1997. The year producing the most Science dissertations with at least one cited URL was 2006. There were 22 (15%) dissertations that met that contained at least one URL citation. In Social Sciences, all years contained dissertations with URL references. In 1997, only one dissertation contained a single URL citation (1%). The peak year for dissertations with URL citations in Social Science was 2005 when 60 (44%) dissertations (n=136) contained at least one URL citation (see Table 4.3, Figures 4.9,

94

4.10, and 4.11). The mean values of dissertations with at least one URL citation

were 8.17 for Humanities; 8.25 for Science; and 33.83 for Social Science.

Figure 4.9 Percentage of Humanities Dissertations with at Least One URL Citation and All Humanities Dissertations across Study Years
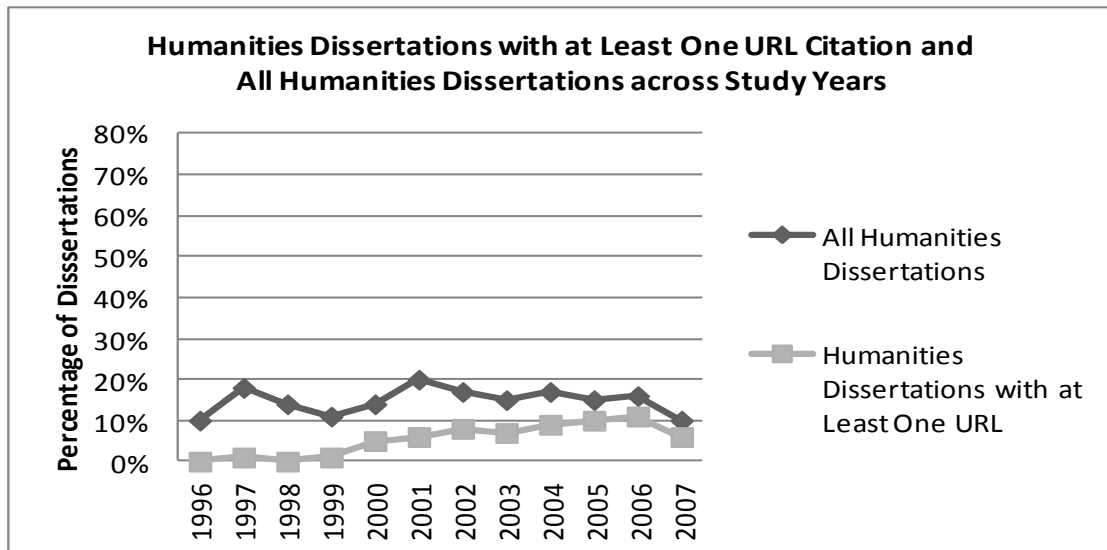
Figure 4.10 Percentage of Science Dissertations with at Least One URL Citation and All Science Dissertations across Study Years
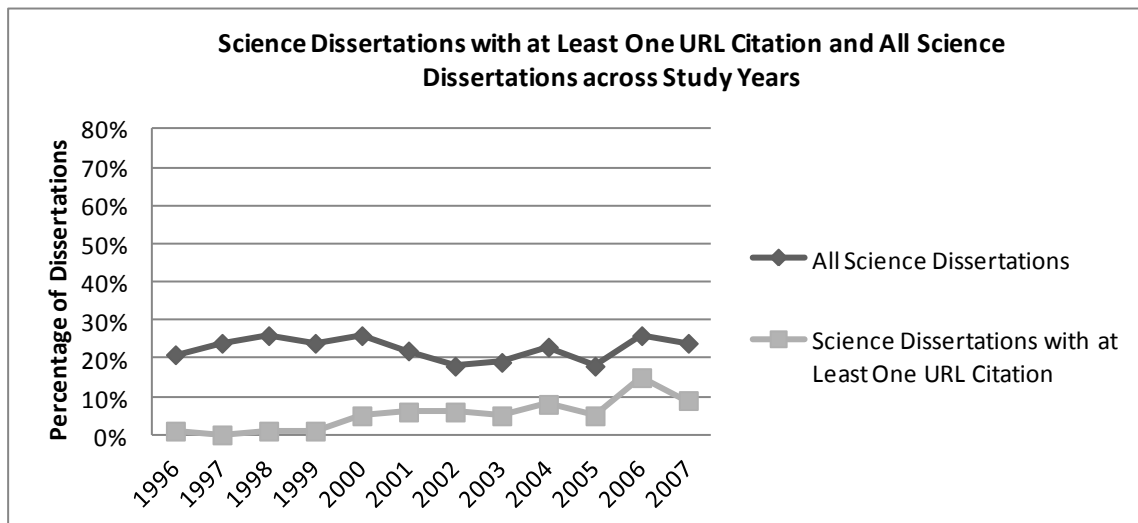
Figure 4.11 Percentage of Social Science Dissertations with at Least One URL Citation and All Social Science Dissertations across Study Years
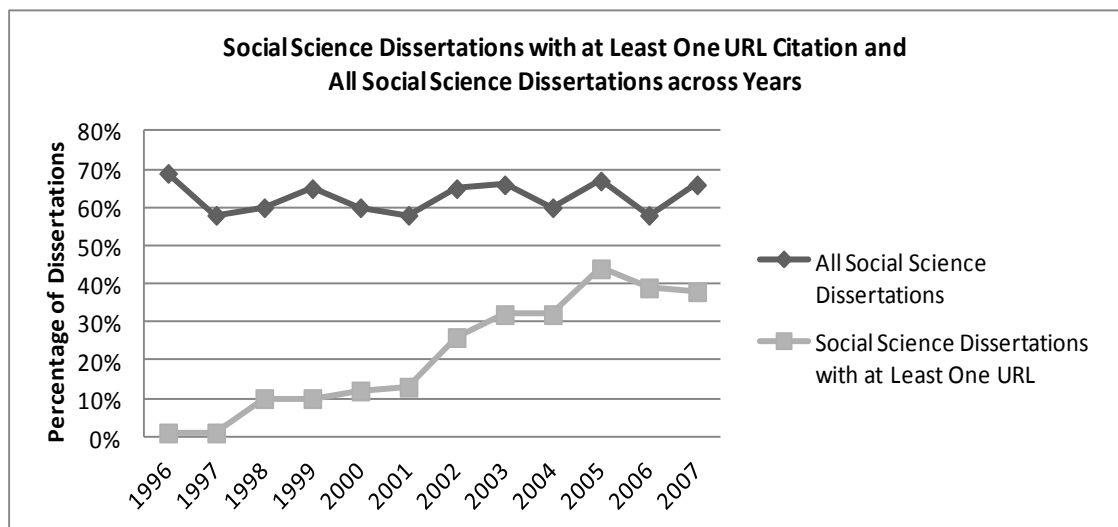
An examination of the subset of dissertations with at least one URL (n=603) by percentage and academic group for each year, indicates that Humanities (n=98) accounted for 16% of the total dissertations with at least one digital citation for the twelve year span while Science (n=99) contained 16% and Social Science (n=406) accounted for 68%. The distribution for the highest percentage of URLs by academic group was 67% for Humanities (n=2) in 1997, 33% for Science (n=1) in 1996, and 89% for Social Science (n=16) in 1998. The lowest percentages occurred in the same years, with 0% for Humanities in 1996 and 1998, 0% for Science in 1997, and 33% for Social Science in 1997 (see Table 4.5). As evident in Table 4.5, the data are skewed by the first four years of the study since the number of dissertations with URL citation for these years was very low (e.g., zero in several academic groups).

Table 4.5 Percentage of Dissertations with at Least One URL within the Academic Groups by Study Years

Percentage of Dissertations with at Least One URL by Academic Groups within Year			
	Humanities	Science	Social Science
1996	0%	33%	67%
1997	67%	0%	33%
1998	0%	11%	89%
1999	11%	16%	73%
2000	20%	25%	55%
2001	22%	24%	54%
2002	20%	15%	65%
2003	16%	11%	73%
2004	17%	17%	66%
2005	16%	9%	75%
2006	17%	23%	60%
2007	13%	16%	71%
Total	16%	16%	68%

The previous analysis affords a better understanding of URL citation use in all dissertations. Additional perspective can be gained by observing the patterns of these resources within each of the academic groups across the 12 year span. This analysis tracked the proportion of dissertations with at least one URL citation against the total number of dissertation within each academic group across study years. In both Humanities and Social Science, 35% of the total number of dissertations contributed by the discipline over the 12 year span contained at least one URL citation. Dissertations with at least one URL citation accounted for 23% of the Science dissertations from all years. Analyzed by year, the percentages ranged from a 0% low to a high of 67% in Humanities, compared with a 1% to 58% range in Science, and 1% to 67% in Social Science (see Table 4.6 and Figure 4.12).

Table 4.6 Percentage of Dissertations with at Least One URL Citation _within_ their respective Academic Group across Study Years

Dissertations with at Least One URL Citation within Academic Group				
	Humanities	**Science**	**Social Science**	**Total**
1996	0%	3%	2%	2%
1997	7%	0%	1%	2%
1998	0%	5%	16%	11%
1999	13%	9%	15%	13%
2000	30%	21%	20%	22%
2001	26%	27%	23%	25%
2002	48%	32%	40%	40%
2003	44%	25%	49%	44%
2004	48%	36%	54%	49%
2005	65%	28%	66%	59%
2006	67%	58%	67%	65%
2007	65%	36%	58%	53%
Total	35%	23%	35%	32%
Mean	35%	23%	34%	32%
SD	25%	17%	24%	22%

Figure 4.12 Percentage of Dissertations with at Least One URL Citation _within_ their Respective Academic Group across Study Years
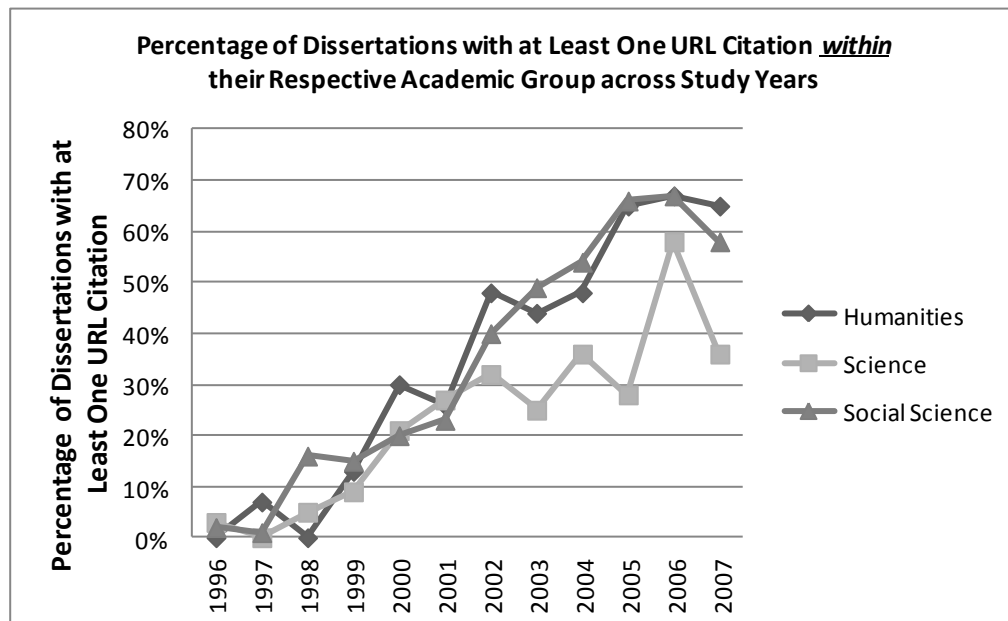
In light of the low returns of URL citations in the earlier years of the twelve year dataset, a determination was made to aggregate the data by time. For the purposes of analysis, the data were distributed into quartiles based on three year intervals. The results indicate an increase in the use of URL citations across the quartiles. The percentage of dissertations with at least one URL citation in Quartile One (1996-1999) ranged from 3%, in both Humanities and Science, to 7% in Social Science. The second quartile (1999-2001) shows an increase to 19% in Science and Social Science and to 25% in Humanities. In Quartile Three (2002-2004), dissertations with at least one URL ranged from 31% in Science to 47% in both Humanities and Social Science. The last quartile (2005-2007) shows the highest percentage of dissertations with at least one URL with a breakdown of 66% in Humanities, 42% in Science, and 63% in Social Science (see Table 4.7). This increase in the use of URL citations in dissertations is illustrated by the trajectory in Figure 4.13.

Table 4.7 Number and Percentage of Dissertations with at Least One URL Citation and All Dissertations by Academic Group across Quartile Years

	Number and Percentage of Dissertations with at Least One URL Citation								
	Humanities			Science			Social Science		
	One+ URL	All*	Percent	One+ URL	All *	Percent	One+ URL	All*	Percent
1996-1998	2	64	3%	3	111	3%	19	288	7%
1999-2001	17	69	25%	21	109	19%	54	279	19%
2002-2004	39	83	47%	31	99	31%	150	318	47%
2005-2007	40	61	66%	44	105	42%	183	290	63%

*ALL indicates all dissertations published in the 3 years (e.g., 83 dissertations were issued in Humanities between 2002 and 2004).

Figure 4.13 Percentage of Dissertations with at Least One URL Citation by Academic Group across Quartile Years

Percentage of Dissertations with at Least One URL Citation by Academic Group across Quartile Years

Y-axis: Disserations with at Least One URL Citation (0% to 70%)
X-axis: 1996-1998, 1999-2001, 2002-2004, 2005-2007
Legend: Humanities, Science, Social Science
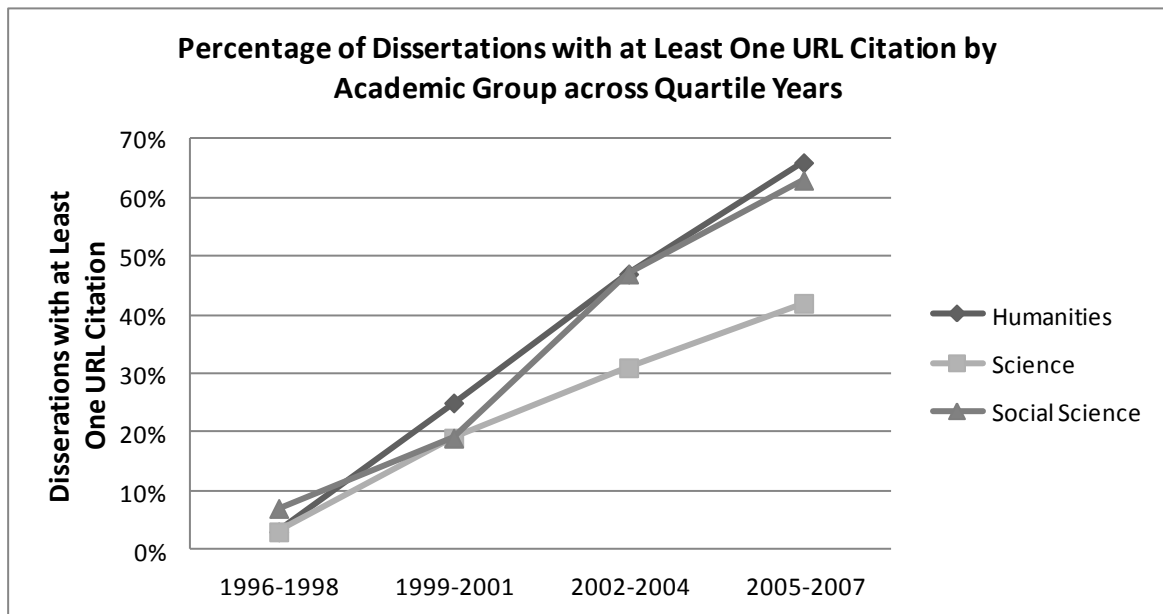
Dissertation URL Citations

Of the total 263,727 citations, 4,258 (1.6%) contained freestanding URLs. The distribution for the academic groups shows that there were 840 (0.3%) URL citations in the Humanities with a mean of 70 per year, Science contained 632(.2%) URL citations, with a mean of 52.7, and Social Science citations containing URL sources totaled 2786(1%), with a mean of 232.2. A review of the URL citation use within each academic group indicates that 1.9% of the 44,134 citations in Humanities were URLs, 1.1% of the 57,730 Science citations were URLs, and 1.7% of the 161,863 Social Science citations were URLs (see Table 4.8 and Figure 4.14).

100

Table 4.8 Number and Percentage of Citations and URL Citations by Academic Group across Study Years

Number and Percentage of URL Citations by Academic Group across Study Year	Humanities			Science			Social Science			All Citations		
	All Cit *	URLs	Percent	All Cit *	URLs	Percent	All Cit*	URLs	Percent	All Cit*	URLs	Percent
1996	2390	0	0.0%	3315	1	0.0%	12322	5	0.0%	18027	6	0.0%
1997	3058	4	0.1%	3942	0	0.0%	13117	1	0.0%	20117	5	0.0%
1998	3762	0	0.0%	5370	3	0.1%	14094	79	0.6%	23226	82	0.4%
1999	2354	4	0.2%	3447	8	0.2%	11989	85	0.7%	17790	97	0.5%
2000	3694	25	0.7%	6077	33	0.5%	13302	138	1.0%	23073	196	0.8%
2001	5135	31	0.6%	4765	38	0.8%	10098	206	2.0%	19998	275	1.4%
2002	4493	44	1.0%	4072	37	0.9%	15590	412	2.6%	24155	493	0.2%
2003	4903	144	2.9%	5453	44	0.8%	13876	225	1.6%	24232	413	1.7%
2004	4507	168	3.1%	4891	81	1.5%	13314	258	1.9%	22712	507	2.2%
2005	3081	95	3.0%	3766	54	1.3%	13875	409	2.9%	20722	558	2.7%
2006	3777	147	3.9%	5745	179	3.3%	12530	489	3.9%	22052	815	3.7%
2007	2980	178	6.0%	6887	154	2.2%	17756	479	2.7%	27623	811	2.9%
Total	**44134**	**840**	**1.9%**	**57730**	**632**	**1.1%**	**161863**	**2786**	**1.7%**	**263727**	**4258**	**1.6%**
Mean	3677.8	70.0		4810.8	52.7		13488.6	232.2		21977.2	354.8	
SD	935.8	71.4		1128.6	58.7		1895.5	178.6		2813.9	289.7	

*ALL CIT indicates all citations published by the academic group during the year specified (e.g., citations published in 2002 Humanities dissertations totaled 4493).

Figure 4.14 Percentage of URL Citations by Academic Group across Study Years

Percentage of URL Citations by Academic Group across Study Years
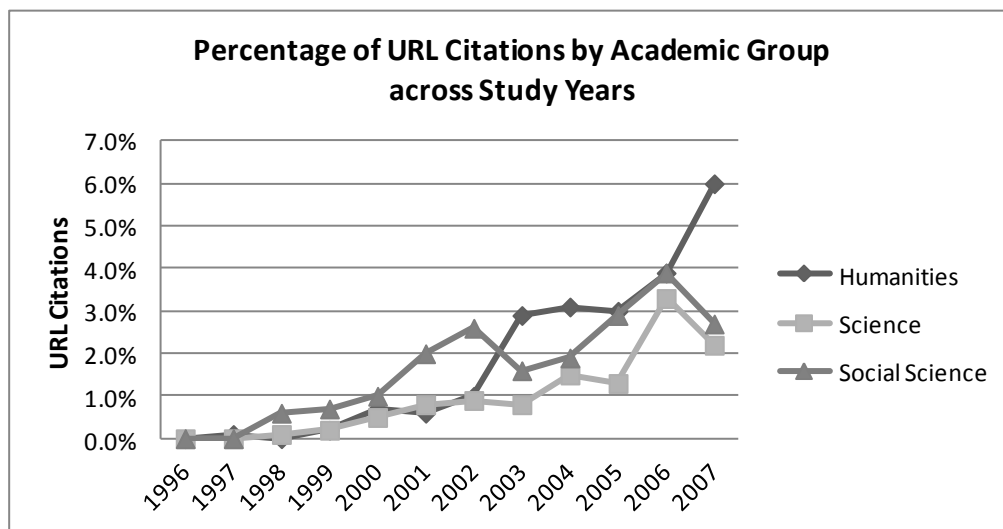
Stage III – Accessibility Rates of URL Citations

Overview of URL Citations

As laid out in Chapter 3, three distinct years of the dissertation set—2000, 2003, and 2006—were selected for an in-depth analysis of accessibility rates in order to identify patterns in the URL citations of University at Albany dissertations. The year 2000 was selected as the initial dataset year because use of the World Wide Web as a medium for cited materials prior to that date was minimal. Since dissertation research is a lengthy process, it was expected that a stronger trend of citing URLs in dissertations would not emerge until after 2000. The last year of the data subset, 2006, was selected in order to allow sufficient time between the date of dissertation publication and the years of analysis (2010-2011) to determine measurable accessibility rates. The year 2003 was selected as the median point between 2000 and 2006. This three year subset contained 1424 (33%) of the total URL citations in all dissertation (n=4258, 1996-2007) from the academic groups. After the initial search of these URL citations, revealed that many contained periodical sources, rather than freestanding URLs. The periodical sources, which had 138 hyperlinks, were removed from the dataset, leaving 1287 (30%) freestanding sources in the subset.

Like the larger dataset consisting of all dissertations having one or more URL citations for the 1996 to 2007 period, this subset shows a trend toward an increase in the percentage of URL use across years for all academic groups. In Humanities, for instance, 17 (.5%) URLs among total Humanities citations for the year 2000

(n=3694) increased to 135 (2.8%) in 2003 (n=4903) and leveled at 130 (3.5%) for

the year 2006 (n=3777). The three year subset for Science included 31 (.5%) URL

sources among total Science citations in the year 2000 (n=6077), which increased to

44 (.8%) URL citations in 2003 (n=5453) and 172 (3%) URLs in 2006 (n=5745).

Social Science URL citations amounted to 132 (1%) of total Social Science citations

for the year 2000 (n=13302), increased to 202 (1.5%) in 2003, and ultimately to

424 (3.4%) in 2006 (see Table 4.9 and Figure 4.15).

Table 4.9 Number and Percentage of Citations and URL Citations by Academic Group across Study Years (2000, 2003, 2006)

	Number and Percentage of Citations and URL Citations by Academic Group across 2000, 2003, 2006											
	Humanities			Science			Social Science			All Citations		
	All Cit	URLs	% URL	All Cit	URLs	% URL	All Cit	URLs	% URL	All Cit	URLs	% URL
2000	3694	17	0.5%	6077	31	0.5%	13302	132	1.0%	23073	180	0.8%
2003	4903	135	2.8%	5453	44	0.8%	13876	202	1.5%	24232	381	1.6%
2006	3777	130	3.5%	5745	172	3.0%	12530	424	3.4%	22052	726	3.7%
Total	12374	282	2.3%	17275	247	1.4%	39708	758	1.9%	69357	1287	1.9%

Figure 4.15 Percentage of URL Citations by Academic Group across Study Years (2000, 2003, 2006)
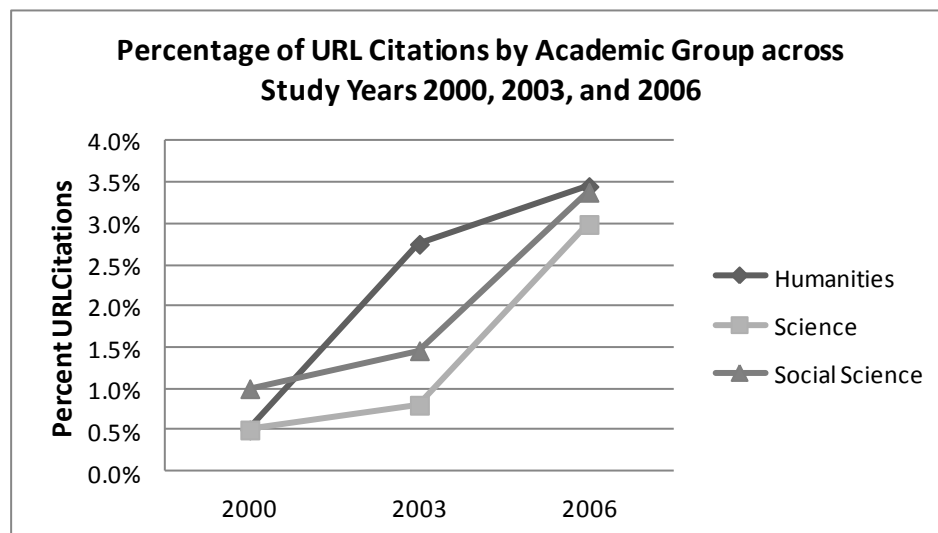
A closer review of the three year subset shows that in 2000, 9% of the URL citations (n=180) were in Humanities, 17% were in Science and 74% were in Social Science. For 2003, the breakdown of URL citations (n=381) was 35% in Humanities, 12% in Science and 53% in Social Science. Of a total 726 URL citations for the year 2006, 130 (18%) were in Humanities, 172 (24%) in Science, and 424 (58%) in Social Science. URL citations for all years (n=1287) comprised 282 (22%) in Humanities, 247(19%) in Science and 758 (59%) in Social Science (see Table 4.10).

Table 4.10 Number and Percentage of URL Citations by Academic Group within Study Years (2000, 2003, 2006)

Number and Percentage of URL Citations by Academic Group within Study Years							
	Humanities		Science		Social Science		Total
	Number	% by Year	Number	% by Year	Number	% by Year	Number
2000	17	9%	31	17%	132	74%	180
2003	135	35%	44	12%	202	53%	381
2006	130	18%	172	24%	424	58%	726
Total	282		247		758		1287

PURLs (Persistent Uniform Resource Locators)

As noted earlier, PURLs were created by OCLC (Online Computer Library Center) in order to deter link rot of URL sources (Jul, 1997, 1998). The United States Government Printing Office uses PURLs to maintain permit access to Web pages and sites (United States, Government Printing Office). Of the 1287 URL sources included in this subset and the 4258 URL citations in the larger dataset, only one was identified as a PURL. It was a .gov accessible via a Web browser search.

Top Level Domains

A review of top level domains (TLDs) in the subset revealed a pattern similar

to that of the full set of URL citations, with the distribution of the subset falling

largely in the .coms – commercial (22%), followed by .govs – government (15%),

.orgs – organization (26%), .edus – education (16%), and .trys – countries (16%).

The .com distribution (n=277) was 71 (26%) in Humanities, 71 (26%) in Science,

and 135 (48%) in Social Science. The .gov sources (n=199) appeared most

frequently in the Social Science citations, which totaled 172 (86%). Social Science

citations also accounted for a high percentage of the .orgs. The breakdown for the

.org TLD (n=336) was 64 (19%) in Humanities, 67 (20%) in Science, and 205 (61%)

in Social Science. The .trys, which accounted for .16 of the total TLDs (n=1287), were

distributed more or less evenly across Humanities (43%), with 87 sources, and

Social Science (45%), with 91 sources (see Table 4.11).

Table 4.11 Number and Percentage of Top Level Domains by Academic Group across Study Years (2000, 2003, 2006)

Number and Percentage of Top Level Domains by Academic Group and Study Years 2000, 2003, 2006								
	. com	.gov	.org	.edu	.try	.us	.net	.etc	Total
Humanities									
2000	4	0	5	7	1	0	0	0	17
2003	37	0	22	26	46	0	3	1	135
2006	30	5	37	12	40	2	4	0	130
Science									
2000	12	6	2	8	2	0	1	0	31
2003	20	3	1	9	7	1	3	0	44
2006	39	13	64	32	16	2	5	1	172
Social Science									
2000	21	38	38	20	6	5	1	3	132
2003	33	43	54	43	24	3	1	1	202
2006	81	91	113	52	61	13	9	4	424
Total	**277**	**199**	**336**	**209**	**203**	**26**	**27**	**10**	**1287**
Percent	22%	15%	26%	16%	16%	2%	2%	1%	

Frequency Distribution of URL Citations within Dissertations

An analysis of the distribution of URLS cited in the 2000, 2003, and 2006 dissertations revealed the following patterns in terms of frequency of cited URLs. The highest number of URL citations was found in a single dissertation containing 69 URL sources. At the lower end of the frequency distribution, each of 49 dissertations contained a single URL reference (see Table 4.12).

Table 4.12 Frequency Distribution of the Number of URLs Cited within Dissertations for the Study Years (2000, 2003, 2006)

Frequency Distribution for Number of URLs within Dissertations		
Number URLs	**Number Dissertations**	**Total URLs**
1	49	49
2	40	80
3	24	72
4	16	64
5	9	45
6	9	54
7	3	21
8	9	72
9	7	63
10	3	30
11	4	44
12	2	24
13	3	39
14	1	14
16	3	48
18	1	18
19	2	38
20	4	80
22	1	22
24	3	72
25	3	75
29	2	58
37	1	37
46	1	46
53	1	53
69	1	69
Total	202	1287
Mean=6.37	Median= 3	

The mean number of URLs citations contained within these dissertations was 6.37, the median was 3. Figure 4.16 illustrates the pattern whereby a much larger number of dissertations have a small number (one to four) of cited URLs as compared to a few dissertations which contain a larger number of URL citations.

Figure 4.16 Frequency Distribution of the Number of URL Citations within Dissertations
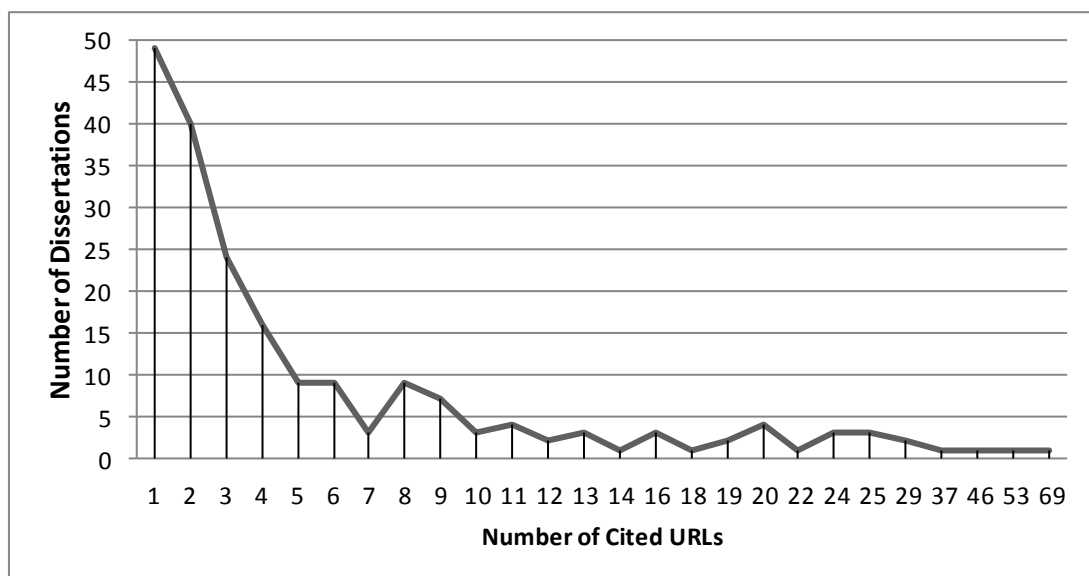
A frequency distribution according to academic groups revealed that the number of URLs used in Humanities dissertations ranged from one in the case of seven different dissertations to a high of 69 in a single dissertation. The mean for the use of URL citations in Humanities dissertations was 8.37; the median was 2.0 (see Figure 4.17).

Figure 4.17 Frequency Distribution of the Number of URLs within Humanities Dissertations for the Study Years (2000, 2003, 2006)
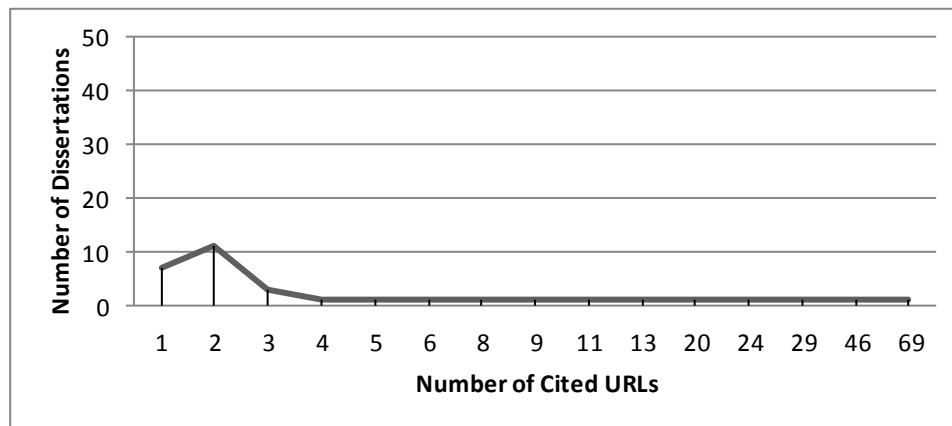
In the case of Science dissertations, the frequency distribution for the number of cited URLs ranged from one in seven dissertations to 46, again in a single dissertation. The mean for use of URL citations in Science was 6.28, with a median of 3.0 (see Figure 4.18).

Figure 4.18 Frequency Distribution of the Number of URLs within Science Dissertations for Study Years (2000, 2003, 2006)
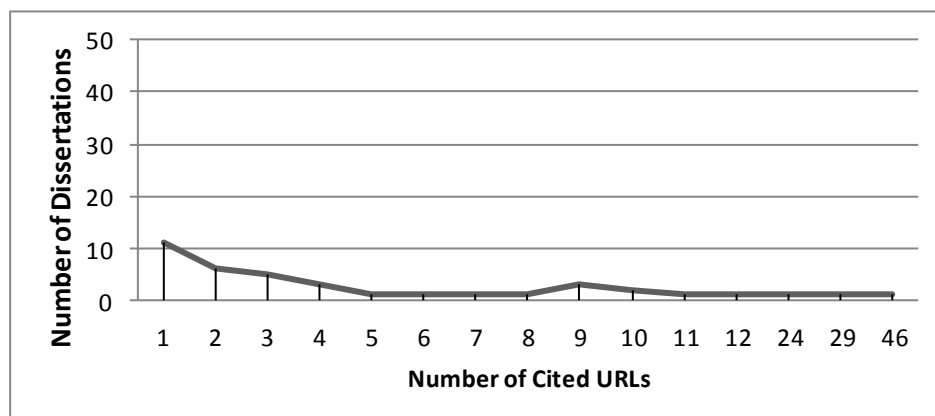
Of the 102 dissertations issued in Social Science, the frequency distribution of the number of cited URLs ranged from one reference in three dissertations to a single dissertation containing 37 URL citations, producing a mean of 7.16. This academic group had the highest median with 4.0 URL citations per dissertation (see Figure 4.19).

Figure 4.19 Frequency Distribution of the Number of URLs within Social Science Dissertations for Study Years (2000, 2003, 2006)
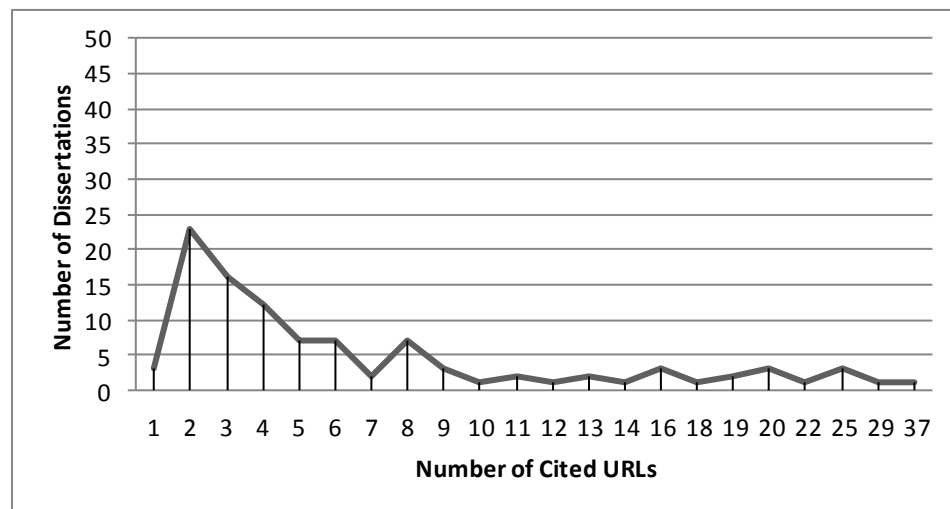
Rate of Accessibility via Web Browser

To verify the persistence of the URL citations, each digital address (n=1287) was searched to verify whether the citation URL was accessible. Using the Firefox Web browser each cited URL was entered into the search window. This was done on three separate dates, with at least a month between each, from May 2010 and January 2011. The results of the three searches were recorded for future analysis.

This research study considered persistence of a URL citation to be understood as the ability to access a cited URL. In the analysis of the three test

results, a source was considered accessible if it could be accessed at least twice or in the last testing. Error messages for the inaccessible sources were recorded and analyzed. Citations containing URLs that had been moved, redirected, or accessed a host or other site, not containing the cited material, were regarded as accessible but were identified as not retrieving the correct information.

Of the total URL citations (n=1287), 626 (49%) were identified as accessible. A distribution of accessible URLs by year showed that for the 2000 URL citations (n=180), 81 (45%) were accessible, for 2003 (n=381) 144 (38%) of the hyperlinked citations were available, and for 2006 (n=726), 401 (55%) URL citations were accessible. Taking academic group into account revealed 124 (44%) Humanities URL sources were accessible, 159 (64%) Science freestanding websites were retrieved, and 343 (45%) Social Science URL references were not dead links (see Table 4.13 and Figure 4.20).

Table 4.13 Number and Percentage of Accessible URL Citations via Web Browser Search by Study Years (2000, 2003, 2006)

Number and Percentage of Accessible URL Citations via Web Browser Search by (2000, 2003, 2006)							
	All URLs	2000	2003	2006	Humanities	Science	Social Science
Number URLs	1287	180	381	726	282	247	758
Total Accessible	626	81	144	401	124	159	343
% of URL Totals	49%	45%	38%	55%	44%	64%	45%

Figure 4.20 Percentage of Accessible URL Citations via Web Browser Search by Study Years (2000, 2003, 2006)
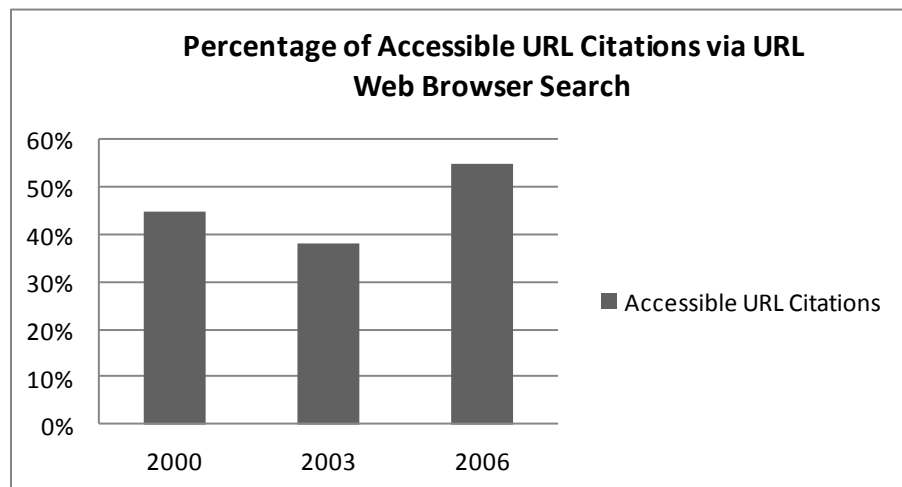
An examination of the accessibility rates of URL citations via a Web browser, indicated that in Humanities, nine (53%) URL sources were accessible in 2000 (n=17); 54 (40%) were available in 2003(n=135), and 61 (47%) were accessible in 2006 (n=130). For Science, of the 31 URL citations from 2000, 18 (58%) were accessible, in 2003, 17 (39%) URLs were accessible of the 44; and in 2006 there were 124 (72%) available links for the 172 digital citations. In Social Science, 54 (41%) of 132 year 2000 sources were active, for 2003, 73 (36%) of 202 were found, and of the 424 URL citations in 2006, 216 (51%) were retrieved (see Table 4.14 and Figures 4.21 and 4.22).

Table 4.14 Number and Percentage of Accessible URL Citations via Web Browser Search by Academic Group across Study Years (2000, 2003, 2006)

Number and Percentage of Accessible URLs via Web Browser				
		All URLs	Accessible	Percent
2000	Humanities	17	9	53%
	Science	31	18	58%
	Social Science	132	54	41%
	2000 Total	180	81	45%
2003	Humanities	135	54	40%
	Science	44	17	39%
	Social Science	202	73	36%
	2003 Total	381	144	38%
2006	Humanities	130	61	47%
	Science	172	124	72%
	Social Science	424	216	51%
	2006 Total	726	401	55%
	Total	1287	626	49%

Figure 4.21 Percentage of Accessible URL Citations via Web Browser Search by Academic Group across Study Years (2000, 2003, 2006)

Percentage of Accessibile URL Citations via Web Browser Search by Academic Group and Study Years (2000, 2003, 2006)
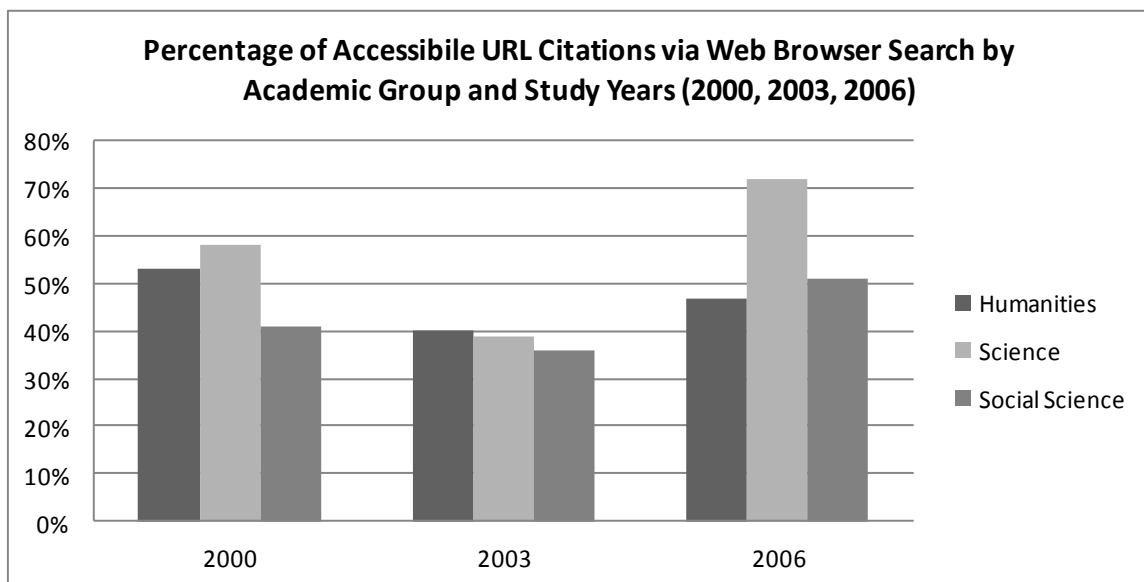
Figure 4.22 Percentage of Accessible URL Citations via Web Browser Search by Study Years (2000, 2003, 2006) and Academic Group

Percentage of Accessible URL Citations via Web Browser Search by Study Years (2000, 2003, 2006) and Academic Group
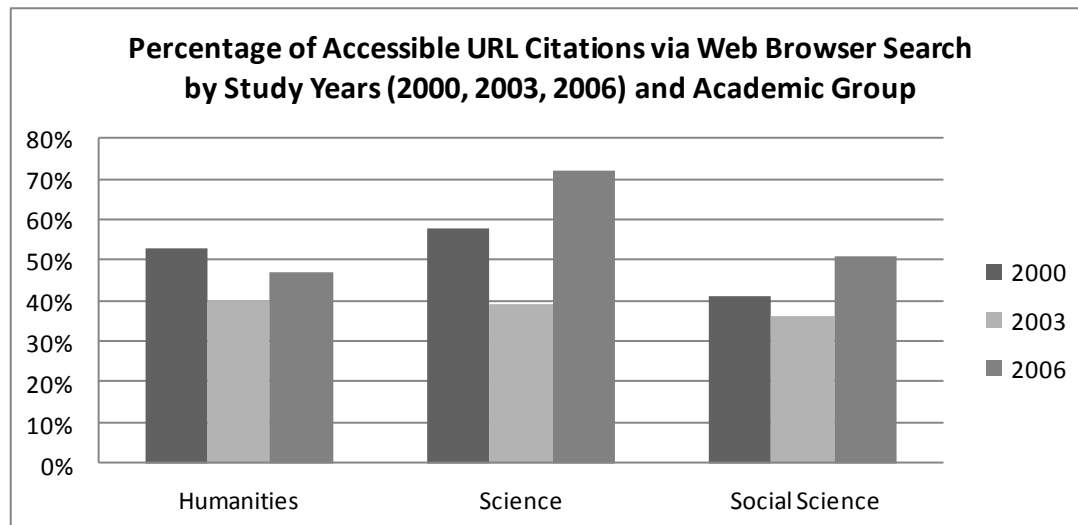
Error Messages Web Browser Search

Errors for inaccessible URLs were recorded and analyzed for patterns. In view of the fact that each URL was tested on three separate occasions, the possibility of different error messages arose. In reviewing the three sets of data, this occurred only ten times, all other access attempts consistently contained identical error messages for all inaccessible tries. For these ten occurrences the most common error message of the three was used. For example, if the first testing retrieved a "404 Not Found" response, and the other two access attempts received "domain expired" responses, the "domain expired" response was recorded as the error.

The most frequent error messages were the 404/Sorry Not Found notices, with 71.1% of the 661 errors. Server errors accounted for next highest warning and

accounted for 102 (15.5%) of the total errors. The distribution of errors by academic group did not indicate unusual trends (see Table 4.15).

Table 4.15 Number of Error Messages for Web Browser Searches by Academic Group

Number of Error Messages for URL Web Browser Searches by Academic Group								
	Humanities		Science		Social Science		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
404/Sorry Not Found	84	13.0%	77	12.0%	309	47.0%	470	71.1%
Cannot find Server	42	6.0%	10	2.0%	50	8.0%	102	15.5%
Directory Denied	7	1.0%	0	0.0%	2	0.3%	9	1.3%
Does Not Exist	0	0.0%	0	0.0%	6	0.9%	6	0.9%
Domain Expired	0	0.0%	0	0.0%	5	0.8%	5	0.8%
Google Redirect	0	0.0%	0	0.0%	4	0.6%	4	0.6%
Password	25	4.0%	1	0.2%	39	6.0%	65	9.8%
Total	158	24.0%	88	13.0%	415	63.0%	661	

Content Issues - Web Browser Searches

While many of the tested URL citations were labeled as accessible a number of these sources were identified as not containing the same content as within the dissertation citations. An analysis of those qualifiers indicates that the rate of citation content was much lower once these measures were taken into account. There were two major themes found in these errors, 'Main Page – No Content' and 'Not Cited Content.' 'Main Page – No Content' refers to the URL attempts redirected to a Web page's main site (e.g., http://www.albany.edu/its would redirect to http://www.albany.edu). These retrieved links did not contain the cited dissertation material. 'Not Cited Content' meant that the URL was accessible; however, the accessed page did not contain the cited works' content. The majority of the

114

inaccuracies were the 'Not Cited Content' with 66% (n=267) of the errors (see Table

4.16).

Table 4.16 Content Errors for Web Browser Searches by Academic Group

Content Error Messages for URL Web Browser Search by Academic Group								
	Humanities		Science		Social Science		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Main Page - Not content	32	12%	12	4%	47	18%	91	34%
Main - subscription	0	0%	0	0%	1	0%	1	0%
Not Cited Content	30	11%	14	5%	131	49%	175	66%
Total	62	23%	26	10%	179	67%	267	

After taking these errors into account, the number of accessible results that

could be identified as effective and containing the correct citation content dropped

to a retrieval rate of 28%. The rates of access were 21% for the 2000 URLs (n=180),

20% for 2003 URLs (n=381), and 34% for the 2006 URLs (n=726). In the

Humanities 22% (n=282) of the originally cited sources in that discipline were

available; 53% for Science (n=247) and 22% in Social Science (n=758). A *t*-test

indicated that Science performed significantly better (t=2.11, df=8, p=.0361) than

the other two academic groups in retrieving URL citations that were accessed

containing the cited information (see Tables 4.17 and Figure 4.23) .

Table 4.17 Rate of Accessibility and Content Accuracy via Web Browser Search by Study Years (2000, 2003, 2006) and Academic Group

Rate of Accessiblity and Content Accuracy via Web Browser Search			
	All URLs	Accessible	Percent
2000 Humanities	17	6	35%
Science	31	13	42%
Social Science	132	19	14%
2000 Total	180	38	21%
2003 Humanities	135	24	18%
Science	44	12	27%
Social Science	202	40	20%
2003 Total	381	76	20%
2006 Humanities	130	32	25%
Science	172	107	62%
Social Science	424	106	25%
2006 Total	726	245	34%
Total	1287	359	28%

Figure 4.23 Percentage of Accessible URL Citations with Content Accuracy via Web Browser Searches by Academic Group

Percentage of Accessible URLs with Content Accuracy via Web Browser Searches by Academic Group and Study Years

116
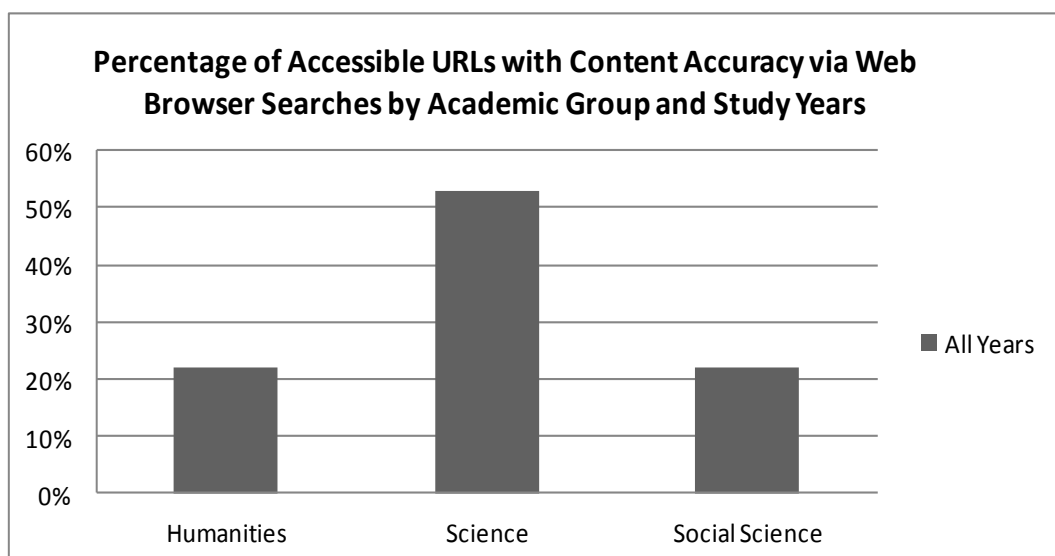
Rate of Accessibility via Google Searches

As noted in Chapter 2, URL sources are often problematic due to human err in URL construction, typographical mistakes (Evans & Furnell, 2001; Evans, et al., 1999), and failures caused by down servers and Web site redesigns, (Emtage, 1994; Germain, 2000; Notess; 2003). Sometimes, URL citations can be located through the Internet via Google keyword searches using a citation's bibliographic information. To test this option, Google was searched to verify if the URL sources (n=1287) could be retrieved through keyword searches.

The information from the URL citations, including author, title, and publication information were used to create searches on Google. Up to six search attempts were made for each search; up to thirty of the first Google hits were reviewed to determine a match and content accuracy. The Google searches retrieved 702 (55%) of the URL citations with Humanities and Social Science accounting for the highest rates of accessibility. From the year 2000 citations, 47% (n=8) of the Humanities (n=17) and 67% (n=83) of the Social Science citations were located. In 2003, 60 (4%) of the 135 Humanities citations and 131 (43%) of the 202 Social Science citations were found. The searches produced 59 (45%) of 130 Humanities sources in 2006, and 274 (65%) of 424 Social Science sources. All of the Science search results yielded less than 50% for each of study years, with 12(38%) of 31 in 2000; 12 (27%) of 44 in 2003; and 63 (37%) of 172 in 2006 (see Table 4.18 and Figures 4.24 and 4.25).

Table 4.18 Number and Percentage of Accessible URL Citations via Google Searches by Academic Group across Study Years (2000, 2003, 2006)

Accessible URL Citations via Google by Academic Group across 2000, 2003, 2006		All URLs	Accessible	Percent
2000	Humanities	17	8	47%
	Science	31	12	38%
	Social Science	132	83	63%
	2000 Total	180	103	57%
2003	Humanities	135	60	44%
	Science	44	12	27%
	Social Science	202	131	65%
	2003 Total	381	203	53%
2006	Humanities	130	59	45%
	Science	172	63	37%
	Social Science	424	274	65%
	2006 Total	726	396	55%
	Total	1287	702	55%

Figure 4.24 Percentage of Accessible URL Citations via Google Searches by Study Years (2000, 2003, 2006) and Academic Group

118
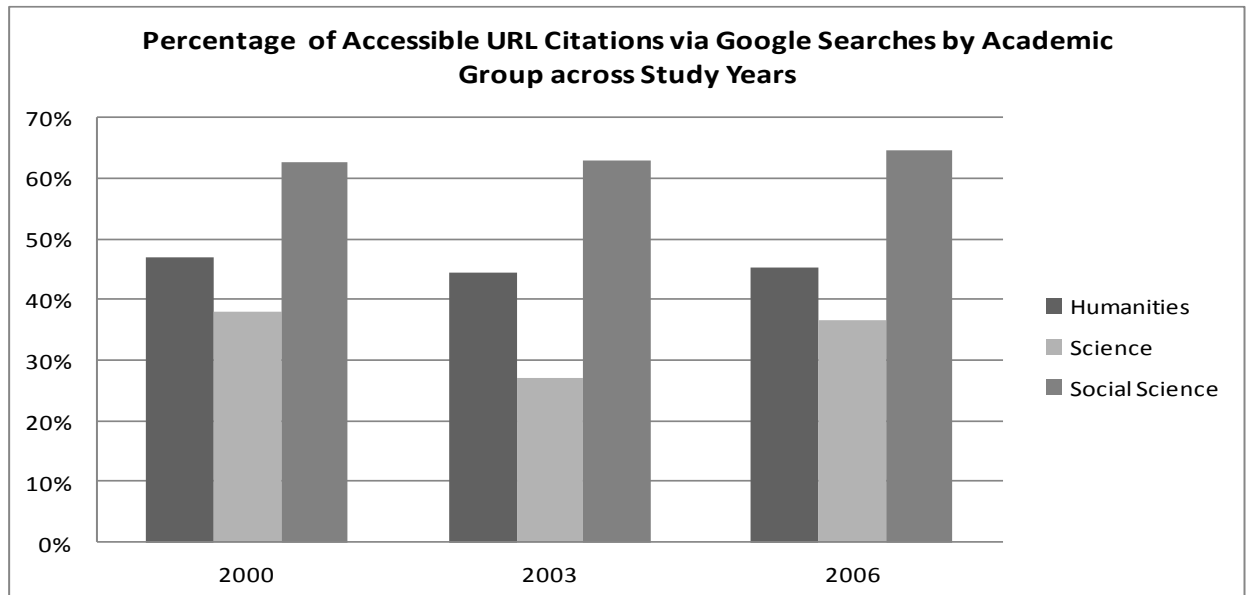
Figure 4.25 Percentage of Accessible URL Citations via Google Searches by Academic Group across Study Years (2000, 2003, 2006)
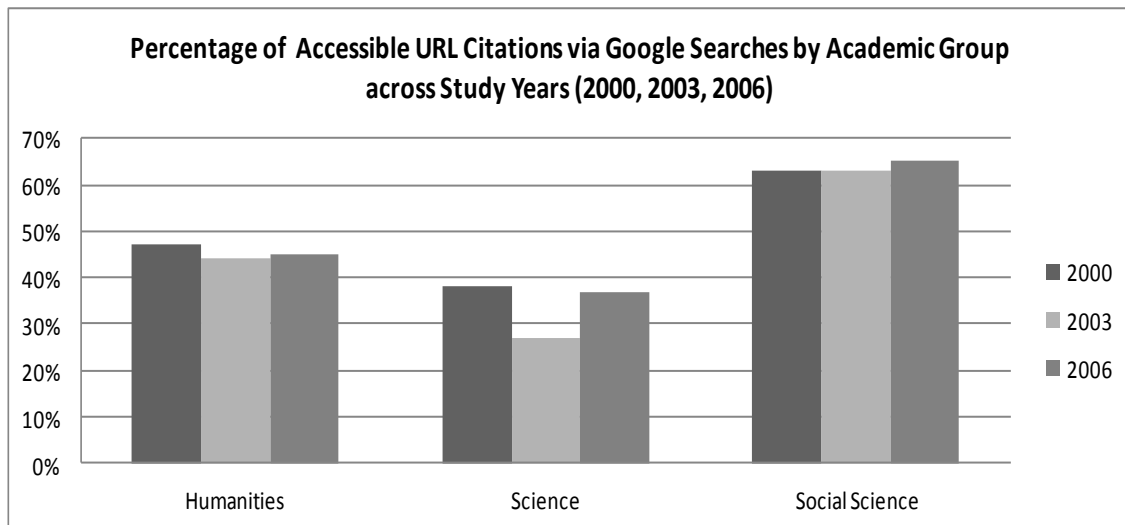
Error Messages for Google

Unlike the Web browser searches, which had distinct error messages for negative results, unyielding Google searches were mainly due to not being in the search engine's index, and thus did not have any distinguishing error notations. There were some unique characteristics and problems with the Google searches. One of the two major problems occurred because there was 'Not Enough Information' with some of the dissertation citations. Many of these citations contained only the URL and no other pertinent bibliographic information (e.g., author's name, website title, etc.) needed to retrieve the source through a search engine. This type of error occurred 144 (43%) times in the 333 errors. Seventy-nine Science URLs (56% of the 'Not Enough Information' errors) experienced this problem. The other most common reason for inaccessibility was due to the 'Only Cites' error. Since a search engine's primary function is to retrieve sources which

fulfill search string queries, there were hits in the Google results list which contained the search keywords. These were citations within bibliographies citing the same citation source as in the study. There were 187 (56%) 'Only Cites' comments, and occurred most frequently in Science and Social Science URL citations (see Table 19).

Table 4.19 Number and Percentage of Error Messages for Google Searches by Academic Group

	Error Messages for Google Searches by Academic Group							
	Humanities		Science		Social Science		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Not Enough Information	10	3%	79	24%	55	17%	144	43%
Only Cites (In bibliography)	36	11%	67	20%	83	25%	187	56%
Terminated	0	0%	1	0.3%	1	0.3%	2	0.6%
Total	46	14%	147	44%	139	42%	333	

As with the URL Web browser searches, the Google searches with atypical accessible results were documented. The majority of these were identified as 'Accessed via URL.' These were retrieved through searching for the URL in the search box since the URL was the only bibliographic element provided in the citation. The combination of 'Not Enough Information' results and the 'Accessed via URL' sources revealed 209(16%) URL citations that were poorly constructed citations. Another error in the accessible Google category had thirty-six (35%) URLs which retrieved main sites that contained the citation information (see Table 4.20).

Table 4.20 Number of Error Comments for Accessible Google Searches by Academic Group

Number of Error Comments for Accessible URLs via Google			
	Humanities	**Science**	**Social Science**
	Number	Number	Number
Accessed via URL	4	41	20
Main Website	12	3	21
New Server	1	67	1
Total	17	111	42

Since Google's main function is to access content, the results produced few hits with content which did not match the citation. Google's 55% rate of accessibility dropped only to 52% after verifying content accuracy. To determine which academic group performed best in Google a *t*-test was run. It identified that Social Science was significantly (*t*=2.6, *df*=5, p=.0004) better than the other two academic groups in retrieving URL citations using Google searches (see Table 4.21 and Figure 4.26).

Figure 4.26 Rate of Accessible URL Citations and Content Accuracy via Google Searches by Academic Group
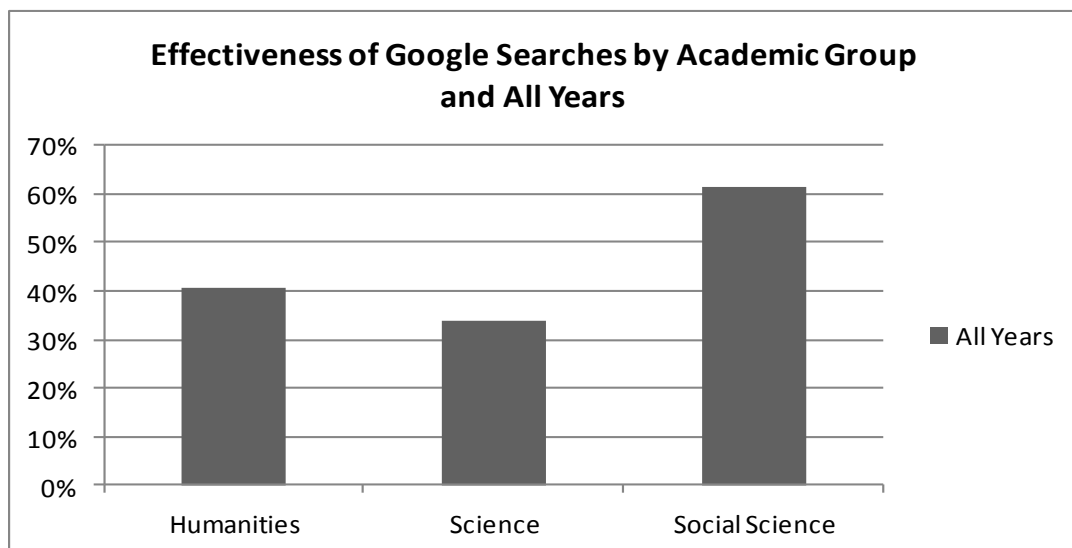
Table 4.21 Rate of Accessibility and Content Accuracy of URL Citations via Google by Academic Group across Study Years

Number and Percentage of Accessible URL Citations with Content Accuracy via Google				
		All URLs	Accessible	Percent
2000	Humanities	17	8	47%
	Science	31	12	38%
	Social Science	132	80	61%
	2000 Total	180	100	57%
2003	Humanities	135	55	41%
	Science	44	12	26%
	Social Science	202	124	61%
	2003 Total	381	191	50%
2006	Humanities	130	52	40%
	Science	172	60	35%
	Social Science	424	263	62%
	2006 Total	726	375	52%
	Total	1287	666	52%

Rate of Accessibility via Internet Archive

A possible solution for locating a URL source found to be inaccessible through a URL Web browser search or via a search engine attempt is through the Internet Archive. In previous studies testing URL accessibility (Casserly & Bird, 2003, 2008; Dimitrova, & Bugeja, 2007; Wu, 2009), the availability of the digital resources were also checked using the Internet Archive's Wayback Machine. The Internet Archive uses spiders to crawl the Internet, captures websites, and stores them. The Wayback Machine is the search utility for accessing the archived sources. Users enter a URL for access, if the URL is incorrect, the source cannot be retrieved. In this study all of the URLs from 2000, 2003, and 2006 were tested for accessibility rates.

122

The Internet Archive searches retrieved 600 (47%) of the URL citations

(n=1287). In 2000, five (29%) of the 17 Humanities, 16 (51%) of the 31 cited

Science URLs, and 57(43%) of the 122 Social Science URL citations were accessible.

In 2003, 35 (25%) of 135 Humanities citations, 30 (68%) of 44 Science citations,

and 87 (43%) of 202 Social Science citations were found. For 2006, the Wayback

Machine produced access to 55 (42%) of Humanities sources, 107 (62%) of the

Science sources, and 208 (49%) Social Science sources (see Table 4.22 and Figures

4.27 and 4.28)

Table 4.22 Number and Percentage of URL Citations Accessible via the Internet Archive by Academic Group across Study Years

Accessible URL Citations via the Internet Archive			
	All URLs	Accessible	Percent
2000 Humanities	17	5	29%
Science	31	16	51%
Social Science	132	57	43%
2000 Total	180	78	43%
2003 Humanities	135	35	25%
Science	44	30	68%
Social Science	202	87	43%
2003 Total	381	152	39%
2006 Humanities	130	55	42%
Science	172	107	62%
Social Science	424	208	49%
2006 Total	726	370	51%
Total	1287	600	47%

Figure 4.27 Percentage of Accessible URL Citations via the Internet Archive by Academic Group and Study Years (2000, 2003, 2006)

Percentage of URL Citations Accessible via the Internet Archive by Academic Group across Study Years

(bar chart: Humanities, Science, Social Science for years 2000, 2003, 2006)
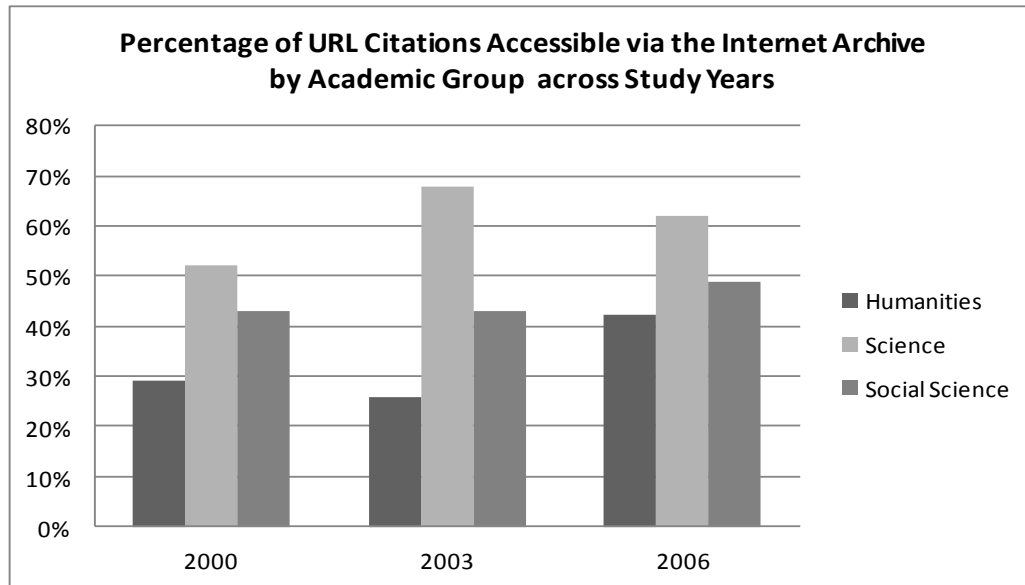
Figure 4.28 Percentage of URL Citations Accessible via the Internet Archive by Study Years (2000, 2003, 2006) and Academic Group

Percentage of URL Citations Accessible via the Internet Archive by Study Years and Academic Group

(bar chart: 2000, 2003, 2006 for Humanities, Science, Social Science)
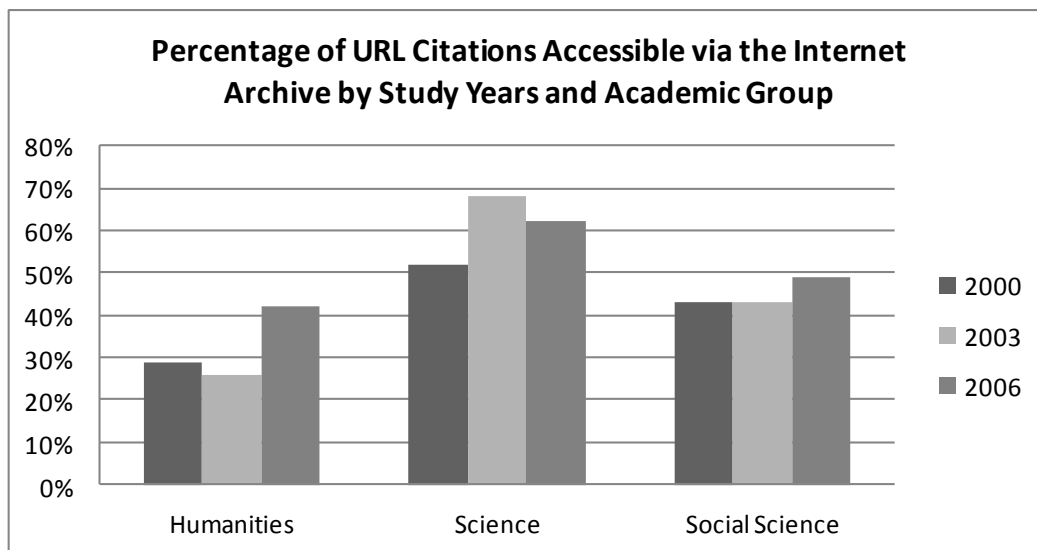
Error Messages for Internet Archive

Error messages from the Internet Archive searches were recorded for

sources that either retrieved an error page or had no results. There were 501 (39%)

124

URL citations (n=1287) which had no results. Another problem came with sources

that retrieved a 'No Retrieval Error' message, 10% (123) of the URL citations

received this error notice. The last error was a 'No Robots' post which indicated that

the cited URL Web page blocked the Internet Archive robots from capturing its

content. This is done through the Robots Exclusion Standard which is a "convention

to prevent cooperating web spiders and other web robots from accessing all or part

of a website. [I]nformation specifying the parts that should not be accessed is

specified in a file called robots.txt in the top-level directory of the website" (Robots

exclusion standard, p. 417). Sixty-three (5%) of the URL citations (n=687) retrieved

this error message (see Table 4.23).

Table 4.23 Number of Error Messages in the Internet Archive by Academic Group

Number of Errors for Internet Archive Searches by Academic Group				
	Humanities	**Science**	**Social Science**	**Total**
	Number	Number	Number	Number
No Resource	134	69	298	501
No Retrieval Error	19	22	82	123
No Robots	12	33	18	63
Total	165	124	398	687

Effectiveness of Internet Archive

The Internet Archive searches retrieved 187 (15%) sources (n=1287) with

content that did not match the cited URL. Taking this into account, the final count for

Internet Archive searches with effective, accessible URL citations with

corresponding content was 413 (32%) (see Table 4.24) .

Table 4.24 Rate of Accessibility and Content Accuracy of URL Citations via the Internet Archive by Academic Group across Study Years (2000, 2003, 2006)

Accessible URL Citations with Content Accuracy via the Internet Archive			
	All URLs	Accessible	Percent
2000 Humanities	17	5	29%
Science	31	16	52%
Social Science	132	41	31%
2000 Total	180	62	34%
2003 Humanities	135	27	20%
Science	44	26	59%
Social Science	202	67	33%
2003 Total	381	120	31%
2006 Humanities	130	30	23%
Science	172	48	28%
Social Science	424	153	36%
2006 Total	726	231	31%
Total	1287	413	32%

Rate of Accessibility via WebCite

In 2005, a similar Web archival tool, WebCite, was created by the International Internet Preservation Consortium. This resource allows users, including students, to archive freestanding resources. WebCite recommends that when citing Internet sources, authors should cite both the Web page URL and the WebCite ID (Ducut, Liu, & Fontelo, 2008; Eysenbach, 2008). None of the URL citations in the data contained WebCite identification.

The results of the WebCite searches determined that thirty-one (2%) sources were retrieved matching the URL citations. One of the issues was that nine (7%) of those sites were from the year 2000 (see Table 4.25 and Figure 4.8). As noted, the sources did match the cited work, though they would have been entered into the

archive much later than when the URL citations had been accessed and cited

(approximately in 2000). In addition to the 31 accessible sources, there were an

additional 26 (2%) URLs which retrieved a captured site that did not match the

content of the cited material (see Table 4.25 and Figure 4.29).

Table 4.25 Number and Percentage of Accessible URL Citations via WebCite by Academic Group across Study Years (2000, 2003, 2006)

Accessible URL Citations via WebCite			
	All URLs	Accessible	Percent
2000 Humanities	17	0	0%
Science	31	4	13%
Social Science	132	5	4%
2000 Total	180	9	5%
2003 Humanities	135	1	1%
Science	44	1	2%
Social Science	202	5	2%
2003 Total	381	7	2%
2006 Humanities	130	0	0%
Science	172	3	2%
Social Science	424	12	3%
2006 Total	726	15	2%
Total	1287	31	2%

Figure 4.29 Percentage of Accessible URL Citations via WebCite by Academic Group across Study Years (2000, 2003, 2006)

-

Percentage of URL Citations Accessible via the WebCite by Academic Group across Study Years
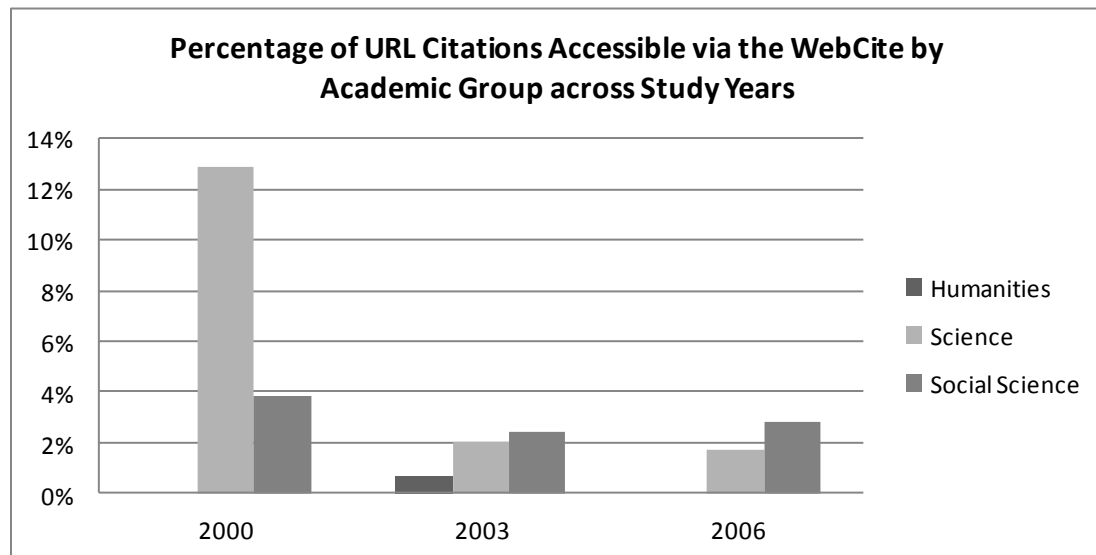
Effectiveness of the Web Access Tools by Academic Group

Several studies have indicated differences in the performance outcomes of the Google and the Internet Archive (Dimitrova, & Bugeja, 2007; Wu, 2009). In 2009, Wu found that when testing accessibility rates of cited URLs, Google had a better rate of retrieval. In another study, Dimitrova and Bugeja (2007) reported that the Internet Archive outperformed Google. The current study compared the retrieval modes (Web browser search, Google search and Internet Archive search) and determined that there are differences in accessibility rates by academic group. For Humanities, t-test results (t=1.94, df=6, p=.0011) found that Google had the highest rate of accessibility. In Science, a t-test indicated there were no significant differences (t=.69, df=6, p=.2686). Lastly, in Social Science Google performed better

128

(t=18.3, df=6, p=.0015) than the other search utilities (see Figures 4.30, 4.31, and 4.32).

Figure 4.30 Proportion of Accessible Humanities URL Citations with Accurate Content by Web Browser, Google, and Internet Archive Searches across Study Years (2000, 2003, 2006)
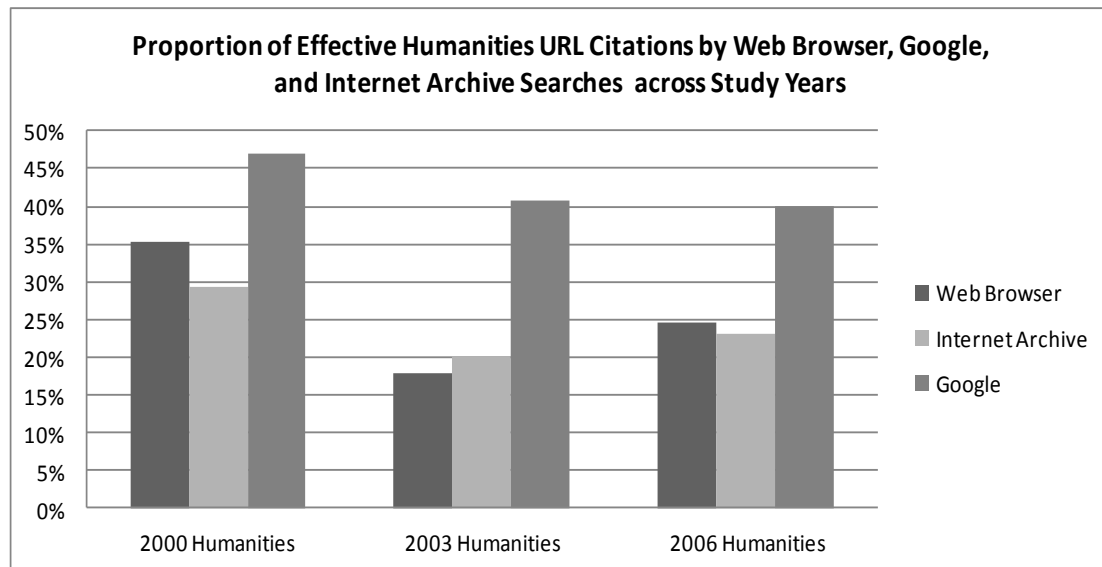
Figure 4.31 Proportion of Accessible Science URL Citations with Accurate Content by Web Browser, Google, and Internet Archive Searches across Study Years (2000, 2003, 2006)
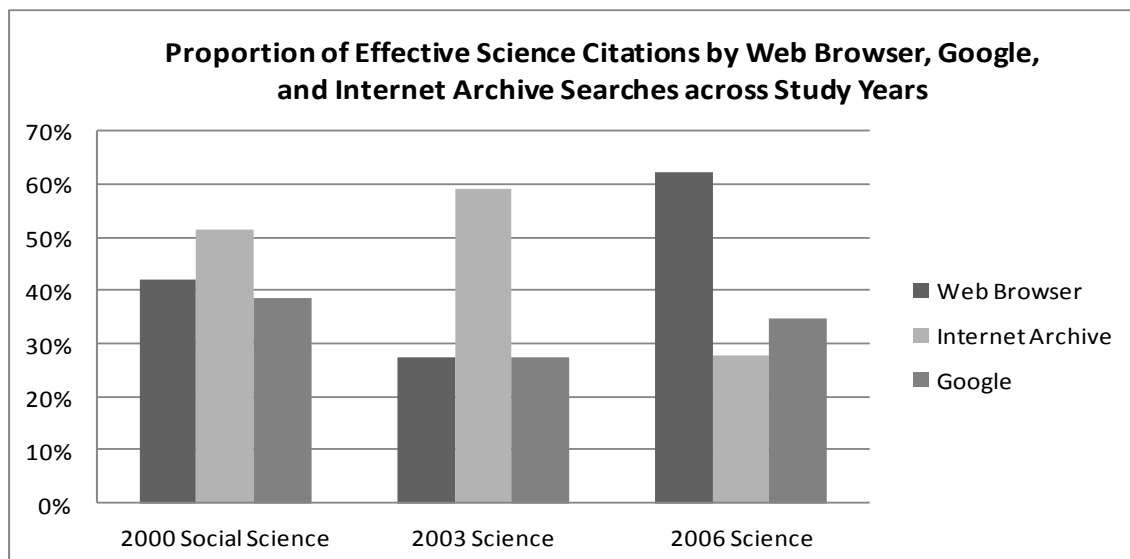
Figure 4.32 Proportion of Accessible Social Science URL Citations with Accurate Content by Web Browser, Google, and Internet Archive Searches across Study Years (2000, 2003, 2006)
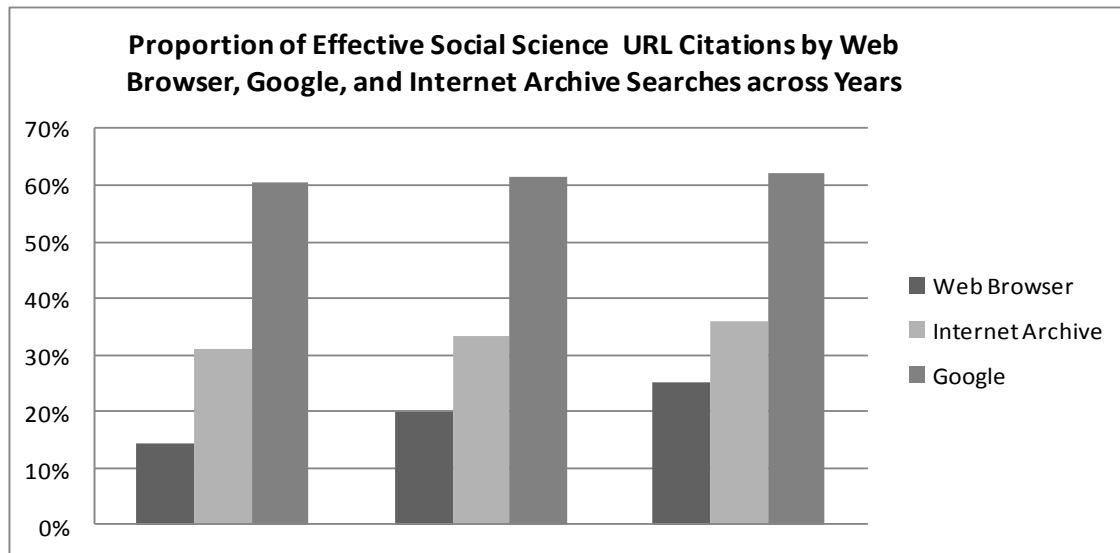
Summary of Chapter

This chapter reviewed the state of URL citations and their accessibility rates in University at Albany dissertations published over a twelve year period. This analysis focused on the quantitative aspect of the study. As follows is a brief summary of the findings:

- An analysis of the number and percentages of dissertations from 1996 through 2007 indicates that there is no significant difference (R^2=.957, F (3, 32) =235.39, p=.555) in the number of dissertation published across study years.

- A review of the references cited within the above dissertations showed a significant difference (R^2=.93, F (3, 32) =143.77, p=.0142) in the number of citations across years. However the slope was a slight 160.

130

- An investigation of the number and percentage of dissertations containing at one URL citation, shows that there is a significant difference (R^2=.949, F (5, 30) =112.1, p=.0004). In the Humanities there was no significant difference (p= .9805) across years. However, both Science and Social Science made significant (p=.0000 for both academic groups) contributions for the changes in URL citation use.

- After testing for accessibility the Web browsers searches retrieved 626 (49%) of the URL citation (n=1287) prior to accounting for errors in content. The balance of accessible URL citations after deducting for content was 359 (28%). The Google searches produced 702 (55%) of the original URL citations, with 666 (52%) remaining after adjusted for content loss. The Internet Archive accessed 600 (47%) URLs before content adjustment, 413 (32%) after. WebCite retrieved 31 (2%) of the study's 1287 URL citations.

- An examination of the three year URL citation subset shows that when conducting a Web browser search, Science performed significantly (t=2.11, df=8, p=.0361) better than the two other academic groups; in Google, Social Science did better (t=2.6, df=5, p=.0001) than Humanities and Science. The three academic groups perform the same in the Internet Archive.

- An assessment of the differences between access utilities (Web browser search, Google search and Internet Archive search) found that Google works best in Humanities (t=1.94, df=6, p=.0011) and in Social Science (t=18.3, df=6, p=.0015). However, in Science, a t-test indicated no significant differences (t=.69, df=6, p=.2686) in access utility performance.

Chapter 5 contains a review and analysis of the qualitative aspect of the research; focusing on the results of the doctoral department interviews. Those interviews provided a needed perspective from doctoral departments on scholarship, as well as the practices and policies regarding the dissertation process. Chapter 6 discusses the quantitative and quantitative findings, implications of these results, and recommendations for future research.

Chapter 5. Results

Part II – Doctoral - Granting Department Interviews

Overview

Analysis of the dissertation citation data as set forth in Chapter 4 provided valuable insights into the practices of doctoral students with respect to citation of URL references in their theses. This portion of the research explores how doctoral granting departments address the use of electronic resources in the dissertation process and the implications this has for scholarship. Interviews with doctoral department stakeholders provided an important perspective on departmental practices and policies in relation to these topics. The interviews were conducted with department chairs or their designated representatives for the purpose of discussing individual doctoral programs.

Guided by the study's research questions, an interview questionnaire consisting of ten questions and a follow-up request for interview recommendations was developed. Because this study involved human subjects, the research proposal was reviewed and approved by the IRB at the University's Office of Research Compliance on September 22, 2011 (see Appendix E for the IRB informed consent document). The interviews commenced in November 2011, and the last interview occurred on March 15, 2012.

The interview pool consisted of fifteen representatives, most of whom were chairs, from the five departments having the highest number of published dissertations in each of the three academic groups—Humanities, Science, and Social

Science. Initial contact with individual department chairs was established via email with the goal of scheduling an interview. Chairs who did not respond received a follow-up email and/or an in-person request encouraging their participation. In cases where an interview was denied, the next department in the respective academic group was contacted. Some department chairs expressed reluctance in participating in the study for various reasons. Several requests were denied based on the chairs' perceptions that their department's doctoral students were non-users of URL citations and the notion that an interview would thus yield little useful material. Other chairs noted that because their departments had no policy governing the use of electronic citations, the potential for discussion would be limited. In some cases, chairs opted to have a representative from their department participate in the interview. In the end, interviewees comprised eight department chairs and seven representatives recommended by department chairs. In all cases, the representatives were identified as individuals in the department who had been actively involved in the doctoral program.

The interviews were guided by the aforementioned questionnaire and a scripted introduction (see Appendix F). All interview participants were asked identical questions; clarification was provided, if requested. The interviews ranged from fifteen to thirty-five minutes in duration, and with the exception of one, all were conducted in the interviewees' offices. All interviews were audio recorded, transcribed, and coded by the researcher. Following each interview, the audio tape was transcribed, and open-ended questions were coded and reviewed to establish categories from the recording units (key terms from the transcript). The responses

to each of these questions were analyzed using Weber's (1990) standard content

analysis procedures. Major concepts were drawn from these terms and grouped

with similar words and phrases to establish categories of each question's theme.

Recording units were examined a second time to confirm that labeling of categories

was consistent. What follows is an example of the recording terms and category

selections for the first interview question, which addressed the respondents'

perspectives on scholarship (Interview Question #1):

Recording Terms	Category
It is taken for granted that [the dissertation is] original research	Original
[Students] will use standard scholarly materials	Literature
With 18th century documents there will be higher expectation	Expertise
The process [will be] experimentally-based	Methods
[U]se of data collection	Methods
[Students will] work to fill the gap	Contribution
Able to integrate findings to argue both sides of a problem	Synthesis

Scholarship

As underscored in Chapters One and Two, scholarship is a vital element in

the doctoral and academic processes. The question addressing the issue of

scholarship asked participants to focus on scholarship and its relevance to the

doctoral dissertation. The fifteen interview responses for this question contained a total of 182 recording units from which seven categories were generated. Many of these categories resembled the concepts highlighted in the doctoral and dissertation literature that are referenced in Chapter One. The final categories that emerged were contribution, expertise, language, literature, methods, original, and synthesis. These categories were defined as:

- Contribution: Significant addition to the body of knowledge and the academic community.
- Expertise: Demonstration of in-depth knowledge of a particular field of study.
- Language: Use of documents and verbal communication with a focus on language.
- Literature: Ability to read, critique, and understand the written research in a discipline, usually published in peer-reviewed articles, monographs, etc.
- Methods: Elements of research design and analysis.
- Original: Research that is a unique and represents a new idea and/or concept.
- Synthesis: Ability to process new knowledge and integrate it into a research or scholarly agenda.

An analysis of the 182 terms provided by doctoral department interviewees revealed that the top three most emphasized categories were Synthesis (23%),

Expertise (20%), and Methods (20%). These results reflect the departments' focus on the development of doctoral students as researchers. This priority was directly stated by one of the interviewees who noted that *"Each student is explicitly training to be a researcher, which is emphasized in the doctoral programs"* (Interviewee #9). The Literature category, although not in the top three categories, accounted for 35 (19%) of the 182 response terms. One interview participant, for example, observed *"[t]hat a student needs mastery of some part of the literature to demonstrate knowledge of the field. Since this field typically doesn't run experiments, it draws on tradition and the knowledge of the literature"* (Interviewee #6).

Interestingly, while many guides and manuals on dissertation writing frequently refer to the thesis as an original contribution, the category Original had the least amount of terms, accounting for only nine (5%) of 182 (see Table 5.1).

Table 5.1 – Frequency of Terms Used to Describe Scholarship

Frequency of Terms Used to Describe Scholarship				
	Number	Mean	Percent	Standard Deviation
Contribution	16	1.1	9%	1.2
Expertise	36	2.4	20%	2.4
Language	9	0.6	5%	1.6
Literature	35	2.3	19%	2.2
Methods	36	2.4	20%	2.2
Original	9	0.6	5%	0.5
Synthesis	41	2.7	22%	3.2
Total	182	12.1	100%	7.4

An analysis of the responses by academic group brought to light differences in frequency of the use of terms within the various groups. In Humanities, 85 (45% of total terms, n=182) terms were used to describe scholarship. The top three categories for this

academic group were Methods (26%), Expertise (21%), and Synthesis (20%) (see Table

5.2). Science, by contrast, yielded only 36 (20% of total terms, n=182) terms, which is

likely explained by the succinct responses provided by the participants representing this

academic group. Science's top categories were Expertise (25%) and Methods (19%),

followed by Synthesis and Contribution (each 17%) (see Table 5.3). Of the 61 (35% of

total terms, n=182) terms used in Social Science, the most frequent categories were

Synthesis (30%), Literature (26%), and Expertise (15%) (see Table 5.4).

Table 5.2 – Frequency of Terms Used to Describe Scholarship in Humanities

Frequency of Terms Used to Describe Scholarship in Humanities				
	Number	Mean	Percent	Standard Deviation
Contribution	3	0.6	4%	0.89
Expertise	18	3.6	21%	3.1
Language	6	1.2	7%	2.7
Literature	15	3	18%	2.3
Methods	23	4.6	26%	1.1
Original	3	0.6	4%	0.55
Synthesis	17	3.4	20%	4
Total	85	17	1	5.78

Table 5.3 – Frequency of Terms Used to Describe Scholarship in Science

Frequency of Terms Used to Describe Scholarship in Science				
	Number	Mean	Percent	Standard Deviation
Contribution	6	1.2	17%	1.1
Expertise	9	1.8	25%	1.6
Language	0	0	0%	0
Literature	4	0.8	11%	1.1
Methods	7	1.4	19%	1.1
Original	4	0.8	11%	0.45
Synthesis	6	1.6	17%	4.1
Total	36	7.2	1	5.4

Table 5.4 – Frequency of Terms Used to Describe Scholarship in Social Science

Frequency of Terms Used to Describe Scholarship in Social Science				
	Number	Mean	Percent	Standard Deviation
Contribution	7	1.4	11%	1.6
Expertise	9	1.8	15%	2
Language	3	0.6	5%	0.8
Literature	16	3.2	26%	2.1
Methods	6	1.2	10%	1.9
Original	2	0.4	3%	0.5
Synthesis	18	3.6	30%	3
Total	61	12.8	1	7.6

As noted in the literature review, link rot and citation erosion are not new issues and have been widely discussed in journals and other publications. To determine whether this conversation was occurring in the doctoral departments, a closed-ended question (Interview Question # 7) was posed asking whether the use of URL citations had come up in departmental meetings. All fifteen (100%) respondents answered that, to date, conversations on the persistence of URL citations had not occurred during department meetings. One participant volunteered that the topic of citation erosion would be on the next department meeting agenda.

Publication Experiences

Faculty members at the University at Albany are expected to engage in some type of formal publication, be it writing, editing, or mentoring writers. It is only reasonable to assume that the interviewees would have been engaged in some type of communication with editors or publishers. Some editors have expressed concerns

about link rot and the corrosive effects of URLs on publication citations (Denehy,

2005; Gould, 2011). Interview discussion of the interviewees' publication

experiences would therefore likely provide valuable insight into their connections

with scholarly publication venues and the ways in which these entities are

responding to URL citations. Three interview questions addressed this topic, two of

which were open-ended.

The initial question posed to interviewees with respect to their publication

experience addressed their use of electronic citations, specifically freestanding URLs

in publications (Interview Question #2). It also prompted the interviewees to

discuss their thoughts on the use of freestanding URLs for citations. While several

responses focused on changes in publication trends, one interviewee from Science

emphasized the persistence of traditional materials in the field:

"*For published work, the researcher cites articles in high quality journals, some*

second tier. Typically, [we] use published journals with good impact factors from Web

of Knowledge. I might cite a textbook, or a review article, or perhaps a government

document in those citations, but URLs, no" (Interviewee #12).

Another interviewee reinforced this perspective with the following statement:

> "*A vast majority have some tangible form, either hard copy or through a*
> *database. Hard copy still exists but I haven't been in stacks – ten years. I still*
> *get the hard copy. For citations, approximately 95% are refereed journal*
> *articles in the field. Although we've adapted to the electronic versions"*
> (Interviewee # 2).

Four respondents stated that they had used URL citations in formal publications. A few interviewees, however, had not, and their comments made it clear that they "*do not trust URLs*" because these sources access *"worthless information"* (Interviewees #8 and #4). The discussions from this question generated 145 recording units. Six categories emerged from these terms and were:

- Adapting technology: Ability to embrace traditional resources in a new medium (e.g., print journals to journals via online databases).

- Ephemeral Resources: Resources that are not permanent and may be unavailable after a certain time period.

- Publication Changes: Evolution of publishing formats (e.g., from print to online-only sources).

- Style standards: Use of style manuals for citation and paper formatting.

- Traditional Materials: Materials that have enjoyed a long standing in academia, such as books and journal articles.

- Web changes: Evolution of the Web.

Coding and labeling of terms (n=145) from this question's responses yielded the following top three categories: Traditional Materials with 32%, Adapting to Technology with 26%, and Publication Change with 17%. Style Standards and Web Changes were among the lowest categories in the replies, with ten (7%) and twelve (8%) terms, respectively (see Table 5.5).

Table 5.5 – Frequency of Terms Used to Describe Citation of Freestanding URL Sources

Frequency of Terms Used to Describe Practice of Citing Freestanding URL Sources				
	Number	Mean	Percent	Standard Deviation
Adapting Technology	38	2.5	26%	2.1
Ephemeral Resources	14	0.9	10%	1.3
Publication Changes	25	1.7	17%	2
Style Standards	12	0.8	8%	1.2
Traditional Materials	46	3.1	32%	3.3
Web Changes	10	0.7	7%	1.6
Total	145	9.7	1	7.3

A closer look at the distribution of terms by category for each academic group revealed that the top three categories in both Humanities and Social Science not only matched, but fell in the same order as the distribution for the analysis of all terms. In Science, although the top three categories were consistent with the other two academic groups, the Adapting Technology category had more terms, i.e., eleven (38%), than Traditional Materials, which had nine (31%) (see Tables 5.6, 5.7 and 5.8).

Table 5.6 – Frequency of Terms Used to Describe Citation of Freestanding URL Sources in Humanities

Frequency of Terms Used to Describe Practice of Citing URL Sources by Humanities				
	Number	Mean	Percent	Standard Deviation
Adapting Technology	17	3.4	27%	2.8
Ephemeral Resources	7	1.4	11%	1.3
Publication Changes	7	1.4	11%	1.7
Style Standards	5	1	8%	12
Traditional Materials	21	4.2	33%	2.6
Web Changes	6	12	10%	2.7
Total	63	12.6	1	4.8

Table 5.7 – Frequency of Terms Used to Describe Citation of Freestanding URL Sources in Science

Frequency of Terms Used to Describe Practice of Citing URL Sources by Science				
	Number	Mean	Percent	Standard Deviation
Adapting Technology	11	2.2	38%	1.6
Ephemeral Resources	1	0.2	3%	0.45
Publication Changes	5	1	17%	2.2
Style Standards	1	0.2	3%	0.45
Traditional Materials	9	1.8	31%	3
Web Changes	2	0.4	8%	0.89
Total	29	5.8	1	4.8

Table 5.8 – Frequency of Terms Used to Describe Citation of Freestanding URL Sources in Social Science

Frequency of Terms Used to Describe Practice of Citing URL Sources by Social Science				
	Number	Mean	Percent	Standard Deviation
Adapting Technology	10	2	19%	2
Ephemeral Resources	6	1.2	11%	1.8
Publication Changes	14	2.8	26%	2.2
Style Standards	5	1	9%	1.4
Traditional Materials	16	3.2	31%	5
Web Changes	2	0.4	4%	0.6
Total	53	10.6	1	10.5

Interviewees were asked about their publication experiences with respect to writing, editing, and working with formal publication venues, such as editorial boards and publishers. This question also queried how publication outlets had been responding to the use of URL citations (Interview Question #3). The 117 resulting recording units generated four categories: contributor, quality control, style standards, and technology changes.

These categories were defined as follows:

- Contributor: Interviewee actively contributes to publication venues, such as proceedings and referred articles.

- Quality Control: References to editorial boards, peer review and other mechanisms for maintaining high standards in publications.

- Style standards: Use of style manuals for citation and paper formatting.

- Technology Change: Technology adjustments in publication, such as open access and digital institutional repositories.

Of the 117 terms identified in connection with this question's responses, 52 (44%) focused on Quality Control and 30 (26%) fell within the Contributor category. Style Standards and Technology Changes accounted for less than 30% of the referenced terms (n=117) (see Table 5.9)

Table 5.9 – Frequency of Terms and Categories Used to Describe Publication Experiences

Frequency of Terms Used to Describe Publication Experiences				
	Number	Mean	Percent	Standard Deviation
Contributor	30	2	26%	2.5
Quality Control	52	3.5	44%	2.7
Style Standards	20	1.3	17%	1.6
Technology Changes	15	1	13%	1.5
Total	117	7.8	1	5.4

A review of the terms in these responses distributed over the three academic groups reveals no difference in terms of category distribution. Contributor and Quality Control had the highest number of terms for all three groups, while Style

Standards and Web Changes had the least (see Figure 5.10, Figure 5.11, and Figure

5.12).

Table 5.10 – Frequency of Terms and Categories Used to Describe Publication Experiences in Humanities

Frequency of Terms Used to Describe Publication Experiences in Humanities				
	Number	Mean	Percent	Standard Deviation
Contributor	6	1.2	26%	1.8
Quality Control	23	4.6	44%	3.2
Style Standards	11	2.2	17%	2.3
Technology Changes	7	1.4	13%	2
Total	47	9.4	1	7.3

Table 5.11 – Frequency of Terms and Categories Used to Describe Publication Experiences in Science

Frequency of Terms Used to Describe Publication Experiences in Science				
	Number	Mean	Percent	Standard Deviation
Contributor	9	1.8	25%	2.5
Quality Control	17	3.4	47%	2.8
Style Standards	5	1	14%	1.6
Technology Changes	5	1	14%	1.5
Total	36	2.8	1	5.4

Table 5.12 Frequency of Terms and Categories Used to Describe Publication Experiences in Social Science

Frequency of Terms Used to Describe Publication Experiences in Social Science				
	Number	Mean	Percent	Standard Deviation
Contributor	15	3	26%	3.7
Quality Control	12	2.4	44%	2.6
Style Standards	4	0.8	17%	1.1
Technology Changes	3	0.6	13%	0.9
Total	34	7.8	1	5.4

The closed-ended question addressing publication experiences (Interview

Question # 9) asked whether the interviewee had used Web archival tools, such as

145

the Internet Archive and WebCite. If the interviewee indicated having had

experience with a Web archival source, a follow-up question was asked to prompt

the respondent to provide the reason for using the source. Of the fifteen

participants, eleven (73%) had never used Web archival tools for retrieval or

submission of documents. The remaining four (one representing Social Science, one

Science, and two Humanities) stated they had used digital archives once or twice. All

four (27%) reported that while they had used the Web archival tools to locate lost

Web sources, they had not used them to submit Web citation resources for digital

storage purposes with the intent of accessing the sources in the future.

Use of URL Citations in Dissertations

The analysis of the URL citation dataset presented in Chapter 4 clearly

suggested an increase in the use of URL citations in dissertations over time. An

interview question addressing how doctoral students use URL citations in

dissertations was therefore posed to interview participants in order to gain an

understanding of the departments' view of the upward trend in digital citations and

the practice of URL citations (Interview Question #4). Many of the interviewees

responded by deferring to dissertation chairs and committee members as "the

advisors" who handle such technicalities. One Humanities interviewee commented

that the department "*leaves that to faculty advisors since there is enough variation in

the sources. Some use legal citations; others are more archival in nature. It's all left to

the committee chair*" (Interviewee #4). Another interviewee stated that "*this

discussion is between the student and their mentor*" (Interviewee #15). Although the

concept of deferring to dissertation committees or advisors on matters of citation practice yielded enough terms to generate a category, the number of recording units for this category was lower than that of other categories generated by the terms. From the 146 recording units, the six categories that emerged for this question were defined as follows:

- Citation Standards: Use of style manuals specifically for citation formatting.
- Defer to Committee: Notion that responsibility for bibliographies and the citation needs of doctoral students lies with the doctoral committee.
- No Concern: Absence of issues related to URL citations.
- Quality: Scholarship level of a dissertation.
- Source Types: Materials used to support an argument.
- Technology Changes: Technology adjustments in publications, such as open access and digital institutional repositories.

The three categories with the highest number of terms (n=146) were Citations Standards, (35%), Quality (21%), and Sources Types (17%). Deferring to the Committee and Technology Changes had the smallest percentage of term use, with 8% each. Sixteen (11%) terms from the participants expressed that No Concern was necessary with respect to the use of URL citation (see Table 5.13). Many of these comments came from interviewees who were unaware of their doctoral students' use of URL citations when writing dissertations.

Table 5.13 Frequency of Terms Used to Describe the Use of URL Citations within Departmental Dissertations

Frequency of Terms Used to Describe the Use of URL Citations within Departmental Dissertations				
	Number	Mean	Percent	Standard Deviation
Citation Standards	51	3.4	35%	2.5
Defer to Committee	11	0.7	8%	1.5
No Concern	16	1.1	11%	2.8
Quality	31	2.1	21%	5.6
Source Types	26	1.7	17%	2.6
Technology Changes	11	0.7	8%	1.7
Total	146	9.7	1	9.3

An analysis of the terms by academic group revealed that the Citation Standards category had the most terms for both Humanities (19 terms, 41%) and Science (18 terms, 52%). In Social Science, the top category was Quality, with 21 terms (32%). The category which reflected No Concern with URL citations in dissertations had zero (0%) terms in the Humanities, seven (21%) in Science, and nine (13%) in Social Science (see Table 5.14, Table 5.15, and Table 5.16).

Table 5.14 Frequency of Terms Used to Describe the Use of URL Citations within Departmental Dissertations in Humanities

Frequency of Terms for the Use of URL Citations within Departmental Dissertations in Humanities				
	Number	Mean	Percent	Standard Deviation
Citation Standards	19	3.8	41%	2.4
Defer to Committee	5	1	11%	1.7
No Concern	0	0	0%	0
Quality	10	2	21%	2.9
Source Types	11	2.2	23%	2.7
Technology Changes	2	0.4	4%	0.9
Total	47	9.4	1	5.9

Table 5.15 Frequency of Terms Used to Describe the Use of URL Citations within Departmental Dissertations in Science

Frequency of Terms Used for the Use of URL Citations within Departmental Dissertations in Science				
	Number	Mean	Percent	Standard Deviation
Citation Standards	18	0.6	52%	3.3
Defer to Committee	3	0.6	9%	1.5
No Concern	7	1.4	21%	1.5
Quality	0	0	0%	0
Source Types	3	0.6	9%	0.9
Technology Changes	3	0.6	9%	0.5
Total	34	6.8	1	2.6

Table 5.16 Frequency of Terms Used to Describe the Use of URL Citations within Departmental Dissertations in Social Science

Frequency of Terms for the Use of URL Citations within Departmental Dissertations in Social Science				
	Number	Mean	Percent	Standard Deviation
Citation Standards	14	2.8	22%	2.5
Defer to Committee	3	0.6	5%	1.5
No Concern	9	1.8	13%	2.8
Quality	21	4.2	32%	9.8
Source Types	12	2.4	19%	3.6
Technology Changes	6	1.2	9%	3
Total	65	13	1	15.23

Instruction and Advisement on Citation Awareness

An additional set of questions addressed departmental instruction and advisement for students learning how to correctly cite sources. Although many students enter graduate school having some familiarity with proper crediting of sources, the introduction of electronic media may necessitate a refresher in order that students properly reference new formats of the literature. Responses to the closed-ended question asking whether the department provided instruction to doctoral students in order to foster scholarly practices and citation persistence

(Interview Question #6) included eight (53%) in the affirmative, six (40%) in the negative, and one (6%) unsure response. In the case of the departments that acknowledged the provision of instruction, all interviewees indicated this instruction took the form of a core research course in the program. All respondents who answered "no" to this query stated that this material was not covered because it had not been an issue; their students did not cite URL sources.

A question regarding advisement strategies for limiting citation erosion (Interview Question #5) yielded eight (53%) "no" responses and seven (47%) "yes" responses. As in the case of the responses to the question about instruction, many respondents stated that the students in their departments were not using URL references. Most of the respondents who indicated that their departments offered advisement strategies acknowledged that mentoring was typically provided by the dissertation chair and/or committee members.

Policy

One of the closed-ended questions asked for information concerning departmental policies governing citations, both in traditional and electronic formats (Interview Question #8). The responses suggested the absence of any specific policies on this subject, with twelve (80%) of the interviewees confirming that their respective department had no policy in place. Once again, many of the interviewees deferred to the dissertation committees on this issue, commenting that "*citation is at the discretion of the chair/ committee*" (Interviewee #2) and *"under the direction of individual professors"* (Interviewee #14). Three (20%) participants acknowledged

that according to departmental policy, doctoral candidates are referred to the official style manual in their discipline.

Promotion and Tenure Policies

The final interview question addressed the issue of whether department had accepted electronic materials, such as e-journals and scholarly Web projects for tenure and promotion cases (Interview Question #10). With the conversion of academic publications to digital formats, there is growing acceptance of these types of resources as contributions to the dossier for the scholarship criteria of tenure and promotion processes (Andersen, 2004; Fountain, 2004; Kling & Callahan, 2003; Shatz, 2004; Vaughan & Shaw, 2003). Of the fifteen responses, fourteen (93%) indicated that their departments accepted digital sources, specifically e-journals, as appropriate for tenure and promotion submissions.

Several interviewees made additional comments, stressing the value of peer reviewed scholarship in the promotion and tenure process. One interviewee remarked that *"[I]f Web journals are accepted, the criteria must be that the publications are top-tier publications. If not, they will not help. A Web page is not a scholarly activity"* (Interviewee #8). Another participant noted that his department *"considers any journal publication, depending on how prestigious it is. Publications that are not peer reviewed are discouraged"* (Interviewee #11).

The foregoing interview responses provided important insight into how doctoral departments are handling the use of URL citations, as well as a glimpse at department perspectives on scholarship and publication. Chapter 6 offers a discussion of the findings

of testing URL dissertation citations and their accessibility rates using Web browser

searches, Google, and the Internet Archive. It also reviews the results of the departmental

interviews, particularly as they relate to the issue of accessibility of electronic resources.

Chapter 6. Discussion and Conclusions

Overview

The aim of this dissertation was to explore URL citation erosion in academic resources by means of an analysis of the similarities and differences between digital sources across time (dissertations published between 1996 and 2007) and discipline variables (Humanities, Science, and Social Science). To this end, a dataset consisting of University at Albany dissertation citations from 1996 to 2007 was created. Three years within the twelve year dataset (2000, 2003, and 2006) were selected to form a data subset from which URL citations were tested for accessibility and content accuracy using the Firefox Web browser (Could the URL still retrieve the cited document?), a Google search (Could the cited material be located through a search engine?), and two Web archival resources, the Internet Archive and WebCite (Was the citation source stored in a Web archive?).

A qualitative research component was incorporated to complement the quantitative methods outlined above. This portion of the research consisted of face-to-face interviews with members of doctoral programs at the University. The interviews yielded valuable information about departmental practice and policy governing the use of freestanding URL citations in dissertations and other academic publications.

Chapters 4 and 5 presented the results of this twofold analysis. This chapter discusses the outcomes of those findings, particularly as they relate to the accessibility rates of the URL resources and issues of content accuracy. It also offers

recommendations aimed at improving the citation practices of doctoral students and enhancing the preservation of digital resources cited in dissertations.

Part I – Citation Analysis

### Review of Dissertations and Citations

The number of dissertations published in the three academic groups, Humanities, Science, and Social Science remained stable across study years. An analysis of the number and percentage of dissertations published from 1996 through 2007 revealed no significant difference (R^2=.957, $F(3, 32)$ =235.39, p=.555) in output across years. Social Science, however, produced three to four times more dissertations than the other two academic groups (see Figure 6.1).

Figure 6.1 Number of Dissertations by Academic Group across Study Years
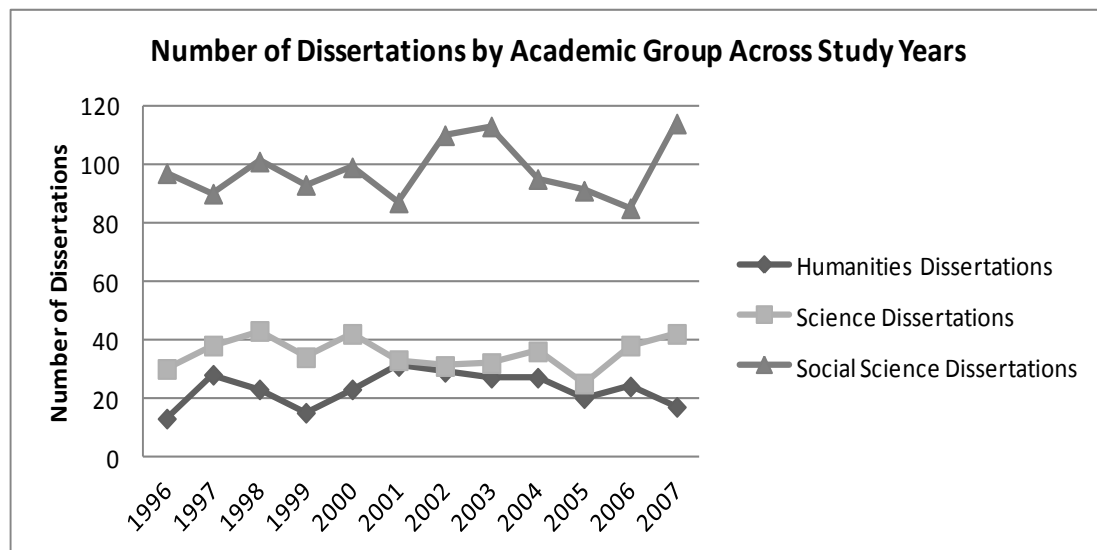
This disparity is accounted for by the fact that enrollment in Social Science doctoral programs (n=863) was three times as high as that of the other two academic groups (Humanities, n=287 and Science, n=316) (see Figure 6.2) for the period under examination. That the Social Science group sponsors twice as many doctoral departments (n=18) as Humanities (n=7) and Science (n=9) is an additional factor. Because the number and percentage of dissertations remained consistent over time, this provided a baseline for future comparison.

Figure 6.2 Number of Students Enrolled in Doctoral Programs (2011)
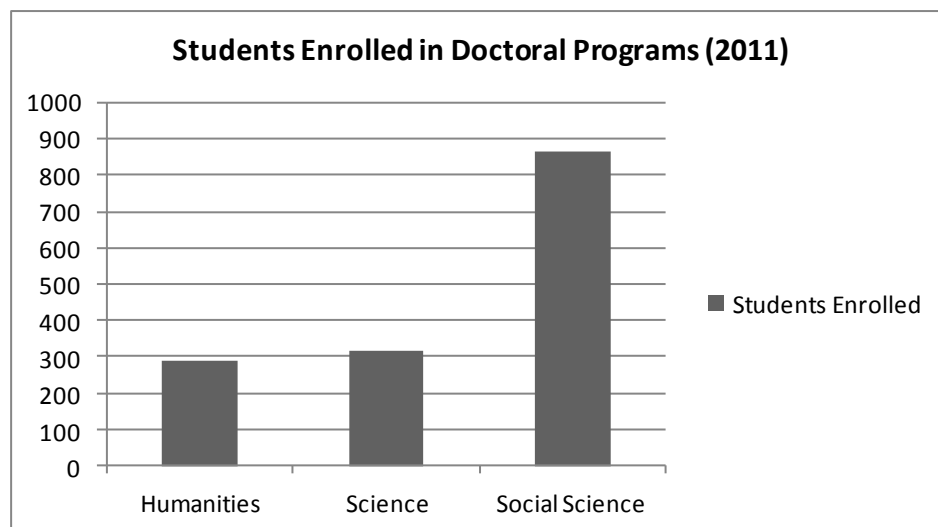
An analysis of the dissertation citations showed a significant disparity (R^2=.93, F (3, 32) =143.77, p=.0142) in the number of citations across years. The findings point to significant differences between the citations in Humanities (p=.0372), as compared with Science (p=.0000) and Social Science (p= .0000). Once again, this can be accounted for by the size of Social Science and its variation from the other two academic groups.

Since the study's focus was on URL citation patterns, it was important to investigate the number and percentage of dissertations containing at least one URL citation to determine trends in citation use. The results revealed a significant difference (R^2=.949, F (5, 30) =112.1, p=.0004) in the number of dissertations with at least one URL citation across study years. Two of the three academic groups, Science and Social Science (p=.0000 for both academic groups), contributed to the overall increase of URL citation use in dissertations. There was no significant difference in Humanities (p= .9805) over time. Figure 6.3 illustrates the trend for increased use of URL citations in dissertations across the study years. The slight drop in URL use to approximately 1% for both Science and Social Science in 2007 is worth noting. This drop is difficult to interpret and may in fact represent a random fluctuation. Clearly, there is need for future research which would include additional years of dissertation citation data in order to better understand long term trends in URL citation use.

Figure 6.3 Number of Dissertations with at Least One URL Citation by Academic Group and Year
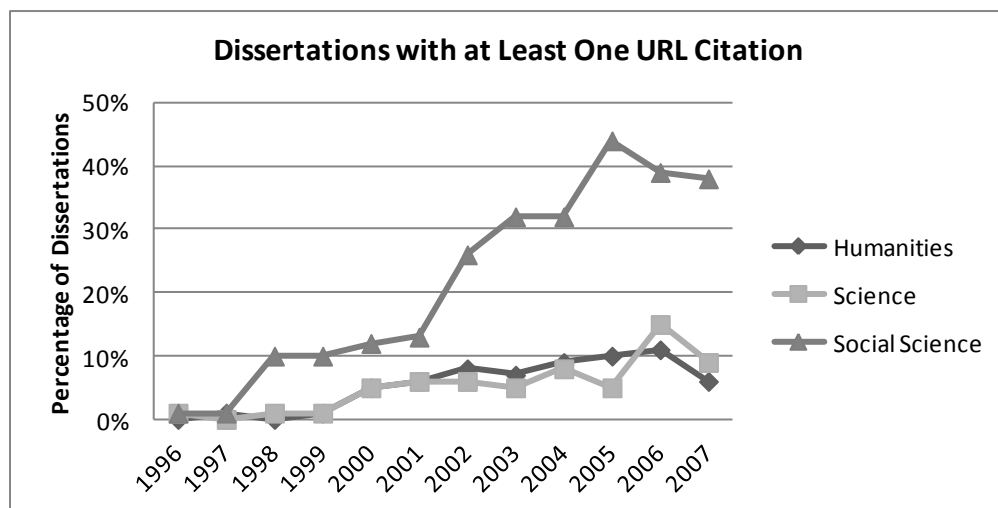
The analysis of dissertations with at least one URL citation demonstrated that this type of dissertation is increasing over time. This trend is consistent with the findings published in other studies investigating the use of URL citations in various types of academic publications, including journal articles (Denehy, 2005; Falagas, Karveli, & Tritsaroli, 2008) and open access scientific literature (Goh & Ng, 2007; Kurtz et al., 2005; Wouters & de Vries, 2004).

Review of Dissertation URL Citations

An analysis of the number of URL citations (n=4258) for the twelve study years indicated that the rate of URL citations in dissertations is increasing. On the other hand, URL sources accounted for less than 2% of the entire set of dissertation citations (n=263,727). While the results infer a small rate of use, this trend will change if the current increasing pattern of URL citation use continues. According to a review by academic group and year, the highest percentage of URL citation use for a single academic group was in Humanities in the year 2007, when the rate of URL citations was 6%. This finding is surprising given the scarcity in the Humanities literature of studies on this topic—a fact which points to the likelihood that few or no problems have surfaced in this regard.

None of the dissertation URL citations (n=4258) referenced Web archival tools such as the Internet Archive or WebCite. This suggests that no attempts were made to submit cited Web pages to a digital archive for preservation purposes or future access. Given that the Internet Archive was a viable archival tool throughout the study years and WebCite was accessible only during the study's last two years it

157

can be assumed that doctoral students were either unaware of these archival mechanisms or did not consider it important to use them. In any event, the potential benefits of storing cited URL materials were not realized. Since the majority (73%) of the department interviewees indicated they had never used either one of these tools, it is unlikely that the doctoral students would have received information about the resources within their departments. This gap opens the opportunity for future research which would include, first, interviews of doctoral students and past graduates in order to determine their experiences and knowledge of Web archival tools, and, second, an analysis of instructional materials for core doctoral research courses to determine the level of instruction provided on this topic.

Part II Accessibility Rates of URL Citations

Overview of URL Citation Subset

To test URL citations for rates of accessibility and content accuracy, a three year subset was created using the URL citations for dissertations from the years 2000, 2003, and 2006. This subset contained 1287 (30%) URL citations of the total dissertation freestanding URLs (n=4258). The distribution of URL citations for the academic groups was 22% for Humanities (n=282), 19% for Science (n=247), and 59% for Social Science (n=758). This distribution was consistent with the percentage of dissertations and citations from the full dataset.

An examination of frequency patterns was undertaken to determine the number of URL citations students were using in each of the doctoral dissertations.

According to the results of this analysis, the median number of cited URLs ranged between two and four across academic groups, indicating that more than 50% of the dissertations with URL citations each contained less than four URL citations. These figures may well explain the interviewees' responses highlighting the lack of URL citations in dissertations. The study's 1876 dissertations (1996-2007) contained a total of 263,727citations with a mean number of 140 citations in each dissertation. URL citations accounted for less than 3% of total citations in most dissertations, leaving room for the possibility that URL citations in dissertations were indeed overlooked in departments, as the interview responses suggested. In any event, more research of this type is warranted given that more than 60% of the dissertations currently use URL citations.

Rate of Accessibility via Web Browser Search

Each of the URLs in the digital citation subset was tested in a Firefox browser to determine rate of accessibility and check for content accuracy. These search attempts were conducted on three separate occasions at least one month apart to compensate for the possibility of down time for servers or other technical complications. The results showed that 626 (49%) of the URL citations (n=1287) were accessible through the browser tests. The distribution of the rates of accessible URL citations within the academic groups was 44% for Humanities, 64% for Science, and 45% for Social Science. These rates were lower than those found in other URL accessibility studies, which reported rates of access ranging between 55% and 95% (Casserly & Bird, 2003, 2008; Ducut, Liu, & Fontelo, 2008; Nagaraja,

Joseph, Polen, & Clauson, 2011; Riahinia, Zandian, & Azimi, 2011; Russell & Kane, 2008).

Because the focus of the aforementioned studies was on URL sources published in scholarly journals, the higher retrieval rates may be explained by an awareness on the part of editorial boards of the increased use of URL citations and their concern about the consequences of citation erosion, a motivating factor for stronger measures to avoid this problem (Denehy, 2005; Gould, 2011). In contrast to the high accessibility findings of these studies, a 2008 multi-method study of accessibility rates and content accuracy of footnotes in several history journals found a low retrieval rate for URL footnotes. Moreover, interview responses from members of the journals' review boards reflected a complacent attitude toward URL citation. Interviewees expressed that online sources were the exception rather than the rule (Bugeja, M., Dimitrova, D. V., & Hong, H. 2008).

Interestingly, the complacent responses noted in the aforementioned study paralleled a good number of the responses recorded in this study's interviews. In both cases, interview participants conveyed little concern about URL citation use, viewing its practice as minimal. In the current study, one interviewee noted that "*98% of citations in dissertations are traditional sources, peer-reviewed journal articles, books, and such. There are few URLs, maybe a government document. Websites are ephemeral*" (Interview #9). Another interviewee commented that, "*99.9% [of citations] are refereed journal articles*" (Interview #8). While the percentage of URL citations in the dissertations represented in the current study is still moderately low, the results nevertheless suggest a pattern of increased use.

Because of the perceived "*ephemeral*" nature of these source types, further research is necessary to measure and explore the accessibility and stability of cited digital resources.

The disparity between the retrieval results of this study and other studies may also be attributed to differences in methods. The unit of analysis for this study was restricted to freestanding URL citations, whereas other studies allowed for the inclusion of non-freestanding URL sources, such as subscription journal and newspaper articles, in the datasets (Nagaraja, Joseph, Polen, & Clauson, 2011; Russell, & Kane, 2008; Thorp, & Brown, 2007). Due to the formal distribution and archiving mechanisms of the latter sources (e.g., through a vendor database), these materials would likely maintain higher levels of persistence than freestanding URL sources.

Rate of Accessibility and Content Accuracy for Web Browser Searches

Since all of the links of this study were checked manually, the researcher had the opportunity to verify the accuracy of citation content (Did the content of the retrieved page match the referenced citation?). This was accomplished by comparing each retrieved Web document with its cited bibliographic information. Many retrieved URLs were redirected to other Web pages (e.g., a host site), frequently resulting in content that did not match the original citation elements. The manual approach to URL sources thus facilitated the discovery of inaccuracies. It is important to note here that the measurement of accessibility rates for URL sources is not always performed manually. A number of studies utilize methods incorporating the use of computer-generated checkers and other types of software

to test website accessibility (Goh & Ng, 2007; Strader & Hamill, 2007; Wren 2004, 2008). The drawback of using such an approach becomes evident when a URL source is redirected to a site other than the intended URL and returns a false positive. In such cases, when a site (although different than the cited URL) is accessed, no error message is sent back to the link checker or documented as an error. Website managers frequently use redirects on Web pages so that users will not retrieve error messages. After accounting for the 267 redirects with content inaccuracies, only 359 (28%) Web pages for the URL citations in the subset of the current study (n=1287) delivered the referenced sources.

It is vital that researchers who choose to study Web accessibility incorporate methods to ensure that URL sources are tested such a way as to verify both access and content. While it is hardly insignificant that a URL can retrieve a source, it is essential that the retrieved URL document contain the same information as the cited material. Since scholarship and citation rely heavily on knowledge gained from earlier works, it is imperative that users are able to access previously consulted resources.

The results for accessibility rates of URL sources in the current study differed among the three academic groups. To determine which academic group achieved the most effective results in the Web browser searches (after content accuracy was accounted for), a t-test was run; Science performed significantly better (p=.0361) than both of the other academic groups.

Rate of Accessibility and Content Accuracy via Google Searches

Another Web retrieval tool used in the current study was Google Search. This search engine allows the user to conduct keyword searches to retrieve desired Web documents. Keyword searches using the bibliographic information from the study's digital citations were conducted to verify the number of accessible sources via Google. The results produced 702 (55%) of the subset's URL citations (n=1287). Interestingly, Social Science performed significantly better (t=2.6, df=5, p=.0001) in Google than Humanities and Science. After adjusting for content loss, 666 (52%) of the Social Science URL citations remained accessible. It is important to note here that Google had a much lower rate of content loss (3%) than the other Web retrieval tools (i.e., Web browser search, Internet Archive). This can be explained by the fact that the Google testing incorporated additional citation elements (e.g., title of website, authors), as opposed to the singular website URL.

The Google searches became problematic when citations lacked sufficient bibliographic information. A number of the study's citations could not be accessed because bibliographic information was limited to a URL. These limited citations accounted for 11% of the subset's citations (n=1287). Science URL citations produced the highest percentage of 'Not Enough Information' errors, with 79 sources. As noted earlier, it is essential that source information is cited correctly in order that readers can access supporting documentation. If this material cannot be retrieved, the integrity of the work is at stake and its scholarship comes into question (Awrey, et al., 2011; Axel-Lute, 1982; Cronin, 1984; Kronick, 1985).

Most of the errors encountered in the URL citations could have been avoided with strict adherence to style guide requirements. Interview participants made continual references to the respective discipline's style manual as the source for doctoral students to consult when writing citations.

A review of the style guides from several disciplines confirmed that these manuals consistently instruct users to cite a number of bibliographic elements in citations to enable others to locate the cited work. In Science, the main style guide, published by The Council of Science Editors, states that:

> The basic rules for citing…do not differ markedly for homepages and other Internet items from what is required for print. There is still an author or organization with responsibility for the item, a title, a place of "publication," a publisher, a date of publication and the extent of the item (i.e., number of pages or the equivalent). Anyone preparing a citation to an Internet document should attempt to locate all of these elements. Simply adding a Uniform Resource Locator (URL) or other electronic address to a title is not sufficient. Internet sites disappear with great frequency, and users of a citation must be given some other identifying information if they are to locate it (Style Manual Committee, Council of Science Editors, 2006, p. 556).

Similarly, for public health and medical students, the style manual for the American Medical Association instructs users to cite the author(s), title of the Web source, URL, date of publication, updated date, and access date (Iverson, 2007). Several style guides are used in Social Science among them *The Chicago manual of style* (The University of Chicago Press, 2010), *The Bluebook: A uniform system of citation* (Harvard Law Review Association, 2010), and the *Publication manual of the American Psychological Association* (American Psychological Association, 2009). The most widely accepted style guide in Humanities is published by the Modern Language Association. While recommendations for citing URLs vary across

disciplines, each of the aforementioned manuals specifically instructs users to cite

the same elements (title, author(s), publication information) for online resources as

for other formats (e.g., print); each encourages the addition of as much electronic

information as possible in order that future readers will be able to retrieve the cited

document (American Psychological Association, 2010; Harvard Law Review

Association, 2010; Modern Language Association of America, 2009; The University

of Chicago Press, 2010).

Clearly, style guide regulations require the use of essential bibliographic

notations so that others can locate a cited source. Based on an analysis of the

department interviews, proper citation and the use of style guides were repeatedly

emphasized within individual departments. This is evidenced by the fact that the

interview responses generated Style Standards and Citation Standards as main

categories in three of four open-ended questions. Interview participants provided

further evidence that proper citation is a high priority among key members of the

doctoral granting departments. One interviewee stated that her discipline "*has a*

particular way to cite – different from other sciences – put authors, title, journal,

pages, just as the style manual notes" (Interview #1). Another commented that

"*UAlbany has no hard and fast style with dissertations. The department picks the style*

that fits best. Students must follow that style" (Interview #12).

The discovery of poor citing practices with respect to URL citations exposes a

disparity between the value departments place on scholarship and the seemingly

casual attitude of doctoral students toward proper citation and its implications for

scholarship. This disparity points to the need for improved instruction and advisement in the doctoral process.

This important issue can easily be addressed. Core courses in doctoral programs, for instance, might include the discussion of new citation formats (i.e., digital resources). It is also recommended that librarians, particularly bibliographers, work closely with doctoral programs to provide educational support. These individuals can augment the initial research/methods courses and enhance instruction by teaching proper citation formatting. Library instruction for doctoral students is an essential component of a rich academic experience. All students can benefit from some method of formal citation instruction, particularly in light of current developments affecting access to electronic resources (Herrington, 1998). Library instruction at the graduate level would provide students with much needed information for effective accessing of resources, and familiarity with proper citation styles may help increase the accessibility rate of dissertation URLs. Another component of this type of instruction might include the use of Web archival tools. This would likely prove valuable to students and faculty alike, as the interview responses indicated that Web archives are neither widely known nor used in doctoral programs.

Rate of Accessibility via the Internet Archive

While Google was able to retrieve more sources than the Web browser searches, the availability of other Web tools offers solutions for storage of digital information. Libraries, publishers, and archival institutions have traditionally

functioned to maintain and store print materials, such as books and journals. With the advent of digital objects in general, and freestanding Web pages in particular, preservation strategies have become more complex. Web archival tools in particular have introduced certain viable solutions. As previously noted, there are several Web archival sites that have captured and maintain Internet Web resources. In order to test the study's URL citations for access rates in Web archival resources, the study selected two specific tools, the Internet Archive and WebCite. The test of the Internet Archive showed that 600 (47%) of the subset's 1287 URL citations were available. The results were not without issues, however, in the case of some sources. Because the Archive did not have a matching link for date of access in the citation (see Appendix A for screen shot of the Internet Archive with links to captured dates), this meant that a date close to the cited access date was selected, which left room for error.

In addition, many of the links in the Archive produced results linking to pages that were no longer intact. One possible explanation for this may be that while the Archive's crawler was able at one point to capture a website's image, the owner of that site either subsequently requested that the Internet Archive disable the content because of copyright laws or placed a Robot Exclusion file into their Web coding. In the event of the latter, the Robot Exclusion file would have been detected when the Archive's crawler conducted a follow-up capture and the crawler would have left the source without capturing its content. This coding would also have prompted the crawler to disable any pages previously placed in the Archive (Panos, 2003).

Similar to the Web browser searches, many of the URL citations accessed in the Internet Archive retrieved documents that did not match the citation content. For the purposes of this study, these false positives were recorded as accessible and adjusted for content inconsistency. As stated above, 600 (47%) of the 1287 original subset URLs were accessible in the Internet Archive. After an adjustment of 187 URLs with inaccurate content, 413 (32%) citation URLs linked to an archived Web source containing the cited material.

Rate of Accessibility via WebCite

Since the Web retrieval tool, WebCite, was in existence only during the last two years of the study, it is not surprising that sources accessed using this tool were limited (Ducut, Liu, & Fontelo, 2008; Eysenbach, 2008). The searches conducted in WebCite retrieved 31 (2%) of the study's 1287 URL citations. A future study using the current subset together with the addition of several more years of doctoral dissertations may help to establish trends toward increased use of this archival utility.

Institutional Repositories

In 2010, the Internet Archive stopped authorizing permission allowing users to submit link requests. These requests had previously prompted the Archive's crawlers to capture a specific Web page for longevity purposes (see Appendix G for screen shot of message). WebCite, on the other hand, continues to be a viable Web archive which allows users to store cited Web materials. Academic institutions are

nevertheless well advised to take a proactive stance on the persistence of Web-based materials created or cited by their constituents, whether students, faculty, or researchers. Institutions of higher education need to take steps for the preservation of the digital content they generate. One avenue in this direction that academic institutions might pursue is the implementation of institutional repositories. In 2003, Clifford Lynch, the Coalition for Networked Information director, wrote that:

> A university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution. While operational responsibility for these services may reasonably be situated in different organizational units at different universities, an effective institutional repository of necessity represents a collaboration among librarians, information technologists, archives and records managers, faculty, and university administrators and policymakers (2003, p. 329).

The primary function of most institutional repositories is to serve as a repository of scholarship created by an institution's scholarly community (Giesecke, 2011; Lynch, 2003; Mercer, Koenig, McGeachin, & Tucker, 2011). Many repositories are now including additional resources, such as articles and white papers. Another potential function of an institutional repository might be to offer users the ability to capture their cited Web sources in order that these materials will enjoy longevity. If digital publication trends continue to grow to the point where the preferred medium of academic publications is electronic, then students should be encouraged to begin adding to institutional repositories at an early stage in their doctoral career. These submissions can include captured content from cited Web sources used in academic publications, including the dissertation.

Comparison of the Rates of Accessibility and Content Accuracy

The goal of this study was to determine relationships between rates of accessibility and content accuracy across academic group and years. The results varied among the academic groups. An analysis of the URL citations and the ability to access accurate content using the three Web access tools (Web browser search, Google, and the Internet Archive) showed that at least one of these tools successfully retrieved 42% of the Humanities URL citations (118/282), 61% of the Science URL citations (151/247), and 63% of the Social Science URL citations (479/758). More than 30% of Science and Social URL citations from the three study years could not be accessed. In Humanities the rate of loss was greater than 50%. Given that the dissertation demonstrates and communicates knowledge about a specialized topic and adds to the body of scholarly knowledge, its level of scholarship is somewhat dependent on the accessibility of works cited therein. Access to works cited must be simplified to eliminate the cumbersome strategy involving three different tools and multiple searches (the Google testing gave up to six) described above. Key members of doctoral programs need to be cognizant of access issues where URL citations are concerned and address the practical aspects of the situation that may impact the scholarship of doctoral dissertations and, quite possibly, future academic publications.

Departmental Interviews

"Web materials change, they are gone and subsequently c'est la vie" (Interview #7).

While this comment conveyed a nonchalant attitude about the use of URLs in dissertation citations, it echoed a sentiment shared by a number of interview

participants. These interviewees expressed little concern about the practice of URL

citation and pointed instead to overarching standards (i.e., style manual

publications) as the solution to all URL related issues.

This attitude runs contrary to the respect for scholarship communicated in

response to the interview question addressing scholarship in dissertations

[Interview Question #1]. Responses to this question generated several categories,

Synthesis among them. Synthesis is much in line with Boyer's (1990) scholarship of

integration. For Boyer, integration meant gaining new knowledge and synthesizing

it with other knowledge to produce new research and scholarship. This process

includes the use of other scholars' works to support ideas and theories. Pertinent

comments from the interviewees on synthesis highlighted that *"[s]tudents will gain*

mastery of the literature [in the field] and go beyond [by] balancing that knowledge

into their research" (Interview #6) and that "*expert students will be well versed and*

conversant with the literature as it relates to their topic. They will have developed the

skill to be parsimonious and weed out the extraneous. They will consume, synthesize,

and [then] articulate a problem to solve [in order to] fill gaps in the literature"

(Interview #2).

Because integration of the literature goes hand in hand with the citation

process, doctoral students need to be thoughtful when citing sources that facilitate

new ideas and research. Only in this way will others desiring to integrate that

scholarship be able to access support documents. Citation provides a fundamental

mechanism for the integration of these connections. If cited works cannot be

accessed citation erosion will slowly result in a corresponding erosion of scholarship.

One focus of the current study was to identify departmental policy governing the use of citation in dissertations. Quite possibly, departmental policies could help deter scholarly drift. An analysis of the departmental interviews made it clear that there were no policies in place regarding URL citation practices. Moreover, no urgent need for development or implementation of such policies was noted. Rather, responsibility for decisions regarding URL citation use was left "*at the discretion of the committee chair*" (Interview #6) or *"up to the project"* (Interview #11). For one interviewee, it was a simple matter of "a professor's choice" (Interview #7). Another interviewee commented that the department "*leaves[s] formal style of the footnotes and citation to the faculty advisor. She uses the style guide best for the project since there is enough variation in topics*" (Interview #14). Since the departments either had no policy or referred to the discipline's style guide as a policy proxy, it was impossible to identify any relationship between policy and the accessibility of URL citations.

Departmental policy changes regarding dissertation citation may not be the solution for citation erosion in dissertations. As one interviewee pointed out "*such policy would be like having policies for methods. Our faculty are experts and rigorous researchers. They have high standards for methods and citation*" (Interview #2).

Further Research

This study provides a baseline for the use of URL citations in dissertations at the University at Albany for the years 1996 through 2007. Further research

172

exploring subsequent years of dissertation citations (i.e., 2008, 2009, and 2010) will

provide valuable data to determine predictive measures for the rate of URL citation

use as well as URL citation accessibility. This will help establish whether or not the

trend of increased use of URL citations persists and make possible the identification

of accessibility and content access changes.

Additional analysis of the current dataset aimed at investigating the rate of

link rot in individual dissertations may identify disparities in the access rates of URL

citations by dissertation researchers rather than by aggregated groupings (i.e.,

Humanities, Science, and Social Science) as measured in this study. Given that one of

the findings of the current study was the difficulty of accessing citations due to

limited bibliographic information, further analysis might include a review and

scoring of citation construction to evaluate the use of proper citation style.

As previously suggested, a solution to poor citation practices and insight into

Web preservation may arise from instruction provided in the early stages of the

doctoral program. An assessment of such instruction alongside dissertation

citations could help determine whether the implementation of improved instruction

decreases citation erosion, thereby enhancing the scholarly output of the University.

Discussion Summary

The impetus for this dissertation was a growing concern over the

accessibility of scholarly material in digital formats. The use of URL citations in

academic resources has drawn much attention and generated considerable

discussion in the literature. Some articles have expressed apprehension regarding

the use of Web sources to support scholarship (Borgman, 2007; Eaton, 2005; Fialkoff, 2003; Waters, 2006; Weller, 2001) while others have eagerly embraced it as an exciting change (Campbell & Campbell, 1995; Herring, 2002; Ross, 1999; Shatz, 2004). The results of the current study highlight issues related to rates of accessibility for URL citations in dissertations. Depending on the Web access tool used (Web browser, Google, or the Internet Archive), the rate for accessing accurate URL dissertation citation content ranged from 28% in the Web browser searches to 52% in Google searches. Rates for successfully accessing a URL citation with accurate content from at least one of the three web tools ranged from 42% in Humanities to 61% in Science and 63% in Social Science. These rates of access include both link rot and citation erosion, both of which have strong implications for scholarly access, as they limit the ability to locate supporting documentation and the ability to replicate research.

It would be naïve to think that after transitioning scholarly materials from hard copy to electronic formats, we would revert back to print. Careful consideration must be given to the use of digital resources in order to sustain academic integrity and scholarship. Instruction and the use of web archival tools (e.g., an institutional repository) may reduce the risk of losing materials. Proactive measures to ensure the preservation of knowledge created through doctoral dissertations are vital. Further research is necessary to explore and implement strategies for preserving the scholarship of the institution. It is the role of the institution to support the intellectual output of its scholars in an ever-changing digital world.

Bibliography

Abbate, J. (1999). *Inventing the Internet.* Cambridge, MA: MIT Press.

Abilock, D. (2009). Guiding the gifted to honest work. *Knowledge Quest, 37*(3), 12-15.

Aldrich, K. (2008). Web cites: When courts cite to URLs: A study of Washington and
New York cases. *Legal Reference Services Quarterly, 27*(2/3), 203-240.

Alexander, G., Ghoneim, I., & Fadhli, A. (2004). [Letter]. 'Web reference' – will they
be erased with time? *BURNS, 31,* 251.

Alford, L. P. (Ed.). (1942). *Cost and production handbook.* New York : The Ronald
Press Co.

American Psychological Association. (2010). *Publication manual of the American
Psychological Association.* Washington, DC: Author.

Andersen, D. L. (2004). Introduction. In D.L. Andersen (Ed.), *Digital scholarship in
the tenure, promotion, and review process* (pp. 3-24). Armonk, NY: M.E. Sharpe.

Ansari, M. A., & Ahmad, S. M. (2007). Citation analysis of Ph.D. theses: A study of
doctoral theses submitted to Aligarh Muslim University during 1985-2005 in
statistics and operations research. *Singapore Journal of Library & Information
Management, 36,* 71-80.

Arms, W. (2001). Uniform resource names: Handles, PURLs, and digital object
identifiers. *Communications of the ACM, 44*(5), 68.

Aronsky, D., Madani, S., Carnevale, R.J., & Duda, S. (2007). The prevalence and
inaccessibility of Internet references in the biomedical literature at the time of
publication. *Journal of the American Medical Informatics Association, 14,* 232-
234.

Ashman, H. (2000). Electronic document addressing: Dealing with change. *ACM Computing Surveys*, *32*(3), 201-212.

Association of American Colleges. (1960). *A guide to graduate study: Programs leading to the Ph.D. degree.* Washington, DC: Author.

Austin, A. E., & McDaniels, M. (2006). Using doctoral education to prepare faculty to work within Boyer's four domains of scholarship. In J.M. Braxton (Series Ed.) *New directions for institutional research(129): Analyzing faculty work and rewards: Using Boyer's four domains of scholarship* (pp. 51-65). SanFrancisco: Jossey-Bass.

Awrey, J., Inaba, K., Barmparas, G., Recinos, G., Teixeira, P. R., Chan, L. S., & ... Demetriades, D. (2011). Reference accuracy in the general surgery literature. *World Journal of Surgery, 35*(3), 475-479.

Axel-Lute, P. (1982). Legal citation form: Theory and practice*. Law Library Journal, 75*, (1), 148-156.

Bar-Ilan, J. & Peritz, B.C. (2004). Evolution, continuity, and disappearance of documents on a specific topic on the Web: A longitudinal study of "informetrics." *Journal of the American Society for Information Science and Technology, 55*(11)980-990.

Bar-Ilan, J., & Peritz, B. C. (2009). A method for measuring the evolution of a topic on the Web: The case of "informetrics." *Journal of the American Society for Information Science & Technology, 60*(9), 1730-1740.

Badgett, R.G. (2006). [letter response] Sotos, J. G. Scholarship erosion. *Annals of Internal Medicine, 145*(1), 77.

Barger, C. M. (2002). On the Internet, nobody knows you're a judge: Appellate

    courts' use of Internet materials. *Journal of Appellate Practice and Process*, *4* (2)

    417-429.

Barnes, N. (2006). The use of U.S. government publications as bibliographic

    references in doctoral dissertations. *Journal of Academic Librarianship, 32*(5),

    503-511.

Barry, C. A. (1997). Information skills for an electronic world: Training doctoral

    research students. *Journal of Information Science, 23*(3), 225.

Baruth, B. (2000). Is your catalog big enough to handle the Web? *American Libraries,*

    31 (7), 56–59.

Beam, J., & Copeland, N. (2001). Electronic resources in union catalogs: URLs and

    accessibility issues. *Serials Review*, *27*(3/4), 33-47.

Beazley, M.B. (2006). *A practical guide to appellate advocacy*, (2nd ed.). New York:

    Aspen Publishers.

Beck, S., & Beck, R. (2006). Web citation: A proposal for standardized specification.

    Online, *30*(4), 31-34.

Benbow, S. M. P. (1998) File not found: The problems of changing URLs for the

    World Wide Web. *Internet Research: Electronic Networking Applications and*

    Policy 8, (3) 247-250.

Benjamin, A.C. (1960). The ethics of scholarship: A discussion of problems that arise

    in its application. *Journal of Higher Education, 31* (9), 471-480.

Benning, S.P., & Speer, S.C. (1993). Incorrect citations: A comparison of the library literature with medical literature. *Bulletin of the Medical Library Association, 81*(1), 56–58.

Berelson, B. (1960). *Graduate education in the United States.* New York, McGraw-Hill.

Bermejo, F. (2007). *The Internet audience: Constitution & measurement.*   New York: Peter Lang.

Berners-Lee, T. (2000). Universal resource identifiers in WWW: RFC 1630. In P. Loshin (comp.). *Big book of World Wide Web RFCs* (pp. RFC 1630. 1-28). San Diego: Morgan Kaufmann.

Berners-Lee, T., Masinter, L, & McCahill, M. (2000). Uniform resource locators: RFC 1738. In P. Loshin (comp.). *Big book of World Wide Web RFCs* (pp. RFC 1738. 1-25). San Diego: Morgan Kaufmann.

Birr, K. & Reda, V. (Ed.). (1994). *A tradition of excellence: The sesquicentennial history of the University at Albany, State University of New York, 1844 to 1994.* Virginia Beach, VA: Donning Co.

Block, M. (2003, July 15). Dealing with digital. *Library Journal, 128*(12), 40-43.

Bomba, C. B. (2000). *The Internet and higher education: Dissertations using Internet citations from 1989-1998 at Oklahoma State University* (Doctoral dissertation). Oklahoma State University, Stillwater, OK.

Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher, 34(*6), 3-15.

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet.*   Cambridge, MA: MIT Press

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual Review of Information Science and Technology*, (Vol. 36, 3-72). Medford, NJ: Information Today.

Boulding, K. E. (1983). The optimum utilization of knowledge: Some central concepts. In k.e. Boulding and L. Senesh (Eds.), *The optimum utilization of knowledge: Making knowledge serve human betterment* (pp. 1-19). Boulder, CO: Westview Press.

Boyce, B.R., & Banning, C.S. (1979). Data accuracy in citation studies. *RQ, 18*(4), 349–350.

Boydston, J., & Leysen, J. (2002). Internet resources cataloging in ARL Libraries: Staffing and access issues. *Serials Librarian*, *41*(3/4), 127-145.

Boyer, C.J. (1973). *The doctoral dissertation as an information source: A study of scientific information flow.* Metuchen, NJ: Scarecrow Press.

Boyer, E.L. (1990). *Scholarship reconsidered: Priorities of the professoriate.* Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Bracke, P. (2001). Access to remote electronic resources at the University of Arizona. *Science & Technology Libraries*, *20*(2-3), 5-14.

Bradford, J. T., Costello, B. & Lenholt, R. (2005). Reference service in the digital age: An analysis of sources used to answer reference questions. *The Journal of Academic Librarianship*, *31*(3), 263-72.

Brause, R. S. (2000). *Writing your doctoral dissertation: Invisible rules for success.* New York: Falmer Press.

Brewer, J. & Hunter, A. (1989). *Multimethod research: A synthesis of styles*. Newbury Park, CA: Sage Publications.

Brody, H. (1996). The Web maestro. *Technology Review, 99*(5), 32.

Bross, V. (2003). The PCC/CONSER PURL Project: Improving access to free resources. *Serials Librarian, 45*(1), 19-26.

Broughton, J. (2008). *Wikipedia: The missing manual.* Cambridge: O'Reilly.

Brown, A. (2006). *Archiving websites: A practical guide for information management professionals.* London: Facet Publishing.

Brown, C. (2004). Knowing where they're going: Statistics for online government document access through the OPAC. *Online Information Review, 28* (6), 396-409.

Brown, C. (2007). The role of Web-based information in the scholarly communication of chemists: Citation and content analyses of American Chemical Society journals. *Journal of the American Society for Information Science and Technology, 58*(13), 2055–2065.

Brown, C. C. (2011). Knowing where they went: Six years of online access statistics via the online catalog for federal government information. *College & Research Libraries*, *72*(1), 43-61.

Bryman, A. (2004). *Social research methods* (2nd ed.). New York: Oxford University Press.

Buchanan, A. L. & Hérubel, J.V.M. (1995). *The doctor of philosophy degree: A selective, annotated bibliography.* Westport, CT: Greenwood Press.

Buchanan, R. (2007). Science Citation Index expanded: The effect of journal editorial policies. *Journal of Academic Librarianship*, *33*(5), 532-539.

Budd, J. (2009). *Higher education's purpose: Intellectual and social progress.* Lanham, MD : University Press of America.

Bugeja, M., & Dimitrova, D.V. (2005a). Exploring the half-life of Internet footnotes. *Iowa Journal of Communication, 37* (1), 77-86.

Bugeja, M., & Dimitrova, D.V. (2005b). The half-life phenomenon. *Serials Librarian, 49* (3),115-123.

Bugeja, M., & Dimitrova, D. (2007). Reconsider the source: Implications of vanishing online citations in communication journals. *Journalism & Mass Communication Educator*, *62*(2), 212-218.

Bugeja, M., & Dimitrova, D. (2009). *Vanishing act: The erosion of online footnotes and implications for scholarship in the digital age.* Duluth, MN: Library Juice Press.

Bugeja, M., Dimitrova, D. V., & Hong, H. (2008). Online citations in history journals: Current practice and views from journal editors. *American Journalism, 25*(4), 83-100.

Burke, G., Germain, C.A & Van Ullen, M.K. (2003). URLs in the OPAC: Integrating or disintegrating research libraries' catalogs. *Journal of Academic Librarianship, 29* (5), 290-297.

Burke, K. (1969). *A rhetoric of motives.* Berkeley, CA: University of California Press.

Campbell, D. R. & Campbell, M. (1995). *The student's guide to doing research on the Internet.* Reading, MA: Addison-Wesley Pub. Co.

Candy, P. C. (2000). Knowledge navigators and lifelong learners: Producing graduates for the information society. *Higher Education Research & Development, 19*(3), 261-277.

Caplan, F. (1980). *The quality system: A sourcebook for managers and engineers.* Radnor, PA: Chilton Book Co.

Carlson, S. (2005, March 18). Scholars note decay of citations to online references. *The Chronicle of Higher Education*, A30.

Carmichael, O. C. (1955). *New York establishes a State University: A case study in the process of policy formation.* Nashville: Vanderbilt University Press.

Carnevale, R., & Aronsky, D. (2007, April). The life and death of URLs in five biomedical informatics journals. *International Journal of Medical Informatics, 76*(4), 269-273.

Casey, C. (1998). The cyberarchive: A look at the storage and preservation of Web sites. *College & Research Libraries, 59*(4), 304.

Casserly, M., & Bird, J.E. (2003). Web citation availability: Analysis and implications for scholarship. *College & Research Libraries, 64*(4), 300-17.

Casserly, M., & Bird, J.E. (2008). Web citation availability: A follow-up study. *Library Resources & Technical Services, 52*(1), 42-53.

CENDI Persistent Identification Task Group. (2004). Persistent identification: A key component of an e-government infrastructure. *New Review of Information Networking, 10* (1), 97-106.

Chandrakar, R. (2006). Digital object identifier system: An overview. *Electronic Library*, *24*(4), 445-452.

Chen, S. (2001). The paradox of digital preservation. *Computer, 34*(3), 24-28.

Cherniss, H. (1968). The history of ideas and ancient Greek philosophy. In *Studies in intellectual history* (pp.22-47). New York: Greenwood Press.

Cheung, J. (2001) Vanishing websites are the weakest link. *Nature, 414* (6859)*,* 15.

Childress, E., Jul, E., & Miller, E. (1998). Don't panic, it's a common disaster. *Journal of Internet Cataloging*, *1*(3), 5-8.

Clark, J. B., Leslie, W.B., & O'Brien, K. P. (Eds.). (2010). Introduction. *SUNY at sixty: The promise of the State University of New York*. Albany, NY: State University of New York Press.

 Clarke, M., & Oppenheim, C. (2006). Citation behaviour of information science students II: Postgraduate students. *Education for Information*, *24*(1), 1-30.

Cloonan, M., & Sanett, S. (2000). Comparing preservation strategies and practices for electronic records. *New Review Of Academic Librarianship*, *6,* 205-216.

Cloonan, M. V., & Sanett, S. (2005). The preservation of digital content. *portal: Libraries & The Academy, 5*(2), 213-237.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Conrad, F.G., & Schobe, M. F. (2008). New frontiers in standardized survey interviewing. In S.N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 173–188). New York: Guilford Press.

Controlling your Web site. (1998). *InfoWorld, 20* (30), 66-70.

Correia, A., & Teixeira, J. (2005). Reforming scholarly publishing and knowledge

    communication: From the advent of the scholarly journal to the challenges of

    open access. *Online Information Review, 29*(4), 349-364.

Coughlin, P.J. (1971, February 17). Senate awaits budget decision. *Albany Student*

    Press, p. 1.

Council of Graduate Schools. (1990). *The doctor of philosophy degree.* Washington,

    DC: Council of Graduate Schools.

Cox, J. (2008). Citation analysis of graduate dental theses references: Implications

    for collection development. *Collection Management, 33*(3), 219-234.

Coyle, K. (2006). Managing technology - identifiers: Unique, persistent, global.

    Journal of Academic Librarianship, *32*(4), 428-431.

Creedy, J. (2008). *Research without tears: From the first ideas to published output.*

    Northhampton, MA: Edward Elgar.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods*

    approaches. Los Angeles, CA: Sage Publications.

Crichlow, R., Davis, S., & Winbush, N. (2004). Accessibility and accuracy of Web page

    references in 5 major medical journals. *Journal of the American Medical*

    Association, 292 (22), 2723-2724.

Cronin, B. (1984). *The citation process: The role and significance of citations in*

    scientific communication. London: Taylor.

Cronin, B., Snyder, H. W., Rosenbaum, H., Martinson, A., & Callahan, E. (1998).

    Invoked on the Web. *Journal of the American Society for Information Science,*

    49(14), 1319-1328.

Davids, J. R., Weigl, D. M., Edmonds, J. P., & Blackhurst, D. W. (2010). Reference

    accuracy in peer-reviewed pediatric orthopaedic literature. *Journal Of Bone &*

    Joint Surgery, 92-A(5), 1155-1161.

Davis, G.B. & Parker, C.A. (1997). *Writing the doctoral dissertation: A systematic*

    approach (2nd ed.). Hauppauge, NY: Barron's.

Davis, P. (2003). Effect of the Web on undergraduate citation behavior: Guiding

    student scholarship in a networked age. *portal: Libraries & the Academy*, *3*(1),

    43-51.

Davis, P.M., & Cohen, S.A. (2001).The effect of the Web on undergraduate citation

    behavior 1996-1999. *Journal of the American Society for Information Science*

    and Technology, 52 (4), 309-14.

Dean, N. (2002). New world for WorldCat. *Information World Review*, (177), 12.

Deegan, M. & Taner, S. (2002). *Digital future; Strategies for the information age*. New

    York: Neal-Schuman Publishers.

Dellavalle, R. P., Hester, E. J., Heilig, L. F., Drake, A. L., Kuntzman, J. W., Graber, M., &

    Schilling, L. M. (2003) Going, going, gone: Lost Internet references. *Science*,

    302 (5646), 787-788.

Denehy, J. (2005). Web resources: The vanishing act. *Journal of School Nursing*,

    21(5), 253-255.

DiCarlo, J., Pastor, X., & Markovitz, B. (2000). The shadow uniform resource locator:

    Standardizing citations of electronically published materials. *Journal of the*

    American Medical Informatics Association, *7*(2), 149-151.

Dillon, M., & Jul, E. (1996). Cataloging Internet resources. The convergence of

    libraries and Internet resources. *Cataloging & Classification Quarterly*, *22*(3/4),

    197-238.

Dimitrova, D., & Bugeja, M. (2006). Consider the source: Predictors of online citation

    permanence in communication journals. *portal: Libraries & the Academy*, *6*(3),

    269-283.

Dimitrova, D., & Bugeja, M. (2007a). The half-life of Internet references cited in

    communication journals. *New Media & Society, 9* (5), 811-826.

Dimitrova, D. V., & Bugeja, M. (2007b). Raising the dead: Recovery of decayed online

    citations. *American Communication Journal*, *9*(2), 1-12.

Doty, P., & Bishop, A. P. (1994). The national information infrastructure and

    electronic publishing: A reflective essay. *Journal of the American Society for*

    Information Science, 45(10), 785-799.

Ducut, E., Liu, F., & Fontelo, P. (2008, June 11). An update on Uniform Resource

    Locator (URL) decay in MEDLINE abstracts and measures for its mitigation.

    BMC Medical Informatics and Decision Making, *8*, 23.

Duda, J.J. & Camp, R.J.(2008).Ecology in the information age: Patterns of use and

    attrition rates of internet-based citations in ESA journals, 1997-2005. *Frontiers*

    in Ecology and the Environment, 6 (3), 145-151.

Dunleavy, P. (2003). *Authoring a Ph.D.: How to plan, draft, write, and finish a doctoral*

    thesis or dissertation. New York : Palgrave Macmillan.

Eaton, J. (2005). Volatility and the long term archive…. *Managing Information,*

    12(10), 10.

Egghe, L. (2000). New informetric aspects of the Internet: Some reflections--many

problems. *Journal of Information Science*, *26*(5), 329-335.

Elphinstone, L. & Schweitzer, R. (1999). *How to get a research degree: A survival*

guide. Australia: Allen & Unwin Academic.

Emtage, A. (1994). The why and what of URLs and URNs. In E.F. Duranceau (Ed.),

Naming and describing networked electronic resources: the role of uniform

resource identifiers. *Serials Review*, *20*(4), 32-34.

Evans, M.P. & Furnell, S.M. (2001). The resource locator service: Fixing a flaw in the

Web. *Computer Networks, 37*, (3-4), 307-330.

Evans, M.P., Phippen, A.D., Mueller, G., Furnell, S.M., Sanders, P.W., & Reynolds, P. L.

(1999). Strategies for content migration on the World Wide Web. *Internet*

Research, 9 (1), 25–34.

Eysenbach, G. (2008). Preserving the scholarly record with WebCite

(www.webcitation.org): An archiving system for long-term digital

preservation of cited Webpages. In L. Chan & S. Mornatti (Eds.), *Open*

scholarship: Authority, community, and sustainability in the age of Web 2.0 -

Proceedings ELPUB 2008 Conference on Electronic Publishing, Toronto, Canada.

(pp. 378-389).

Eysenbach, G., Diepgen, T., & Coiera, E. (1998). Towards quality management of

medical information on the internet: Evaluation, labeling, and filtering of

information. *BMJ: British Medical Journal , 317*(7171), 1496-1500.

Falagas, M., Karveli, E., & Tritsaroli, V. (2008). The risk of using the Internet as

    reference resource: A comparative study. *International Journal of Medical*

    Informatics, *77*(4), 280-286.

Farrell, D., & Petersen, J. C. (2010). The growth of Internet research methods and the

    reluctant sociologist. *Sociological Inquiry, 80*(1), 114-125.

Feather, J. (2008). *The information society: A study of continuity and change* (5th ed.).

    London: Facet.

Featherstone, M. (2000). Archiving cultures. *British Journal of Sociology, 51*, 161–

    184.

Feezel, J., & Welch, S. A. (2000). What is new or different about the scholarship of

    teaching? *Journal of the Association for Communication Administration, 29*(3),

    250-256.

Feldman, S. (1997). `It was here a minute ago!': Archiving the Net. *Searcher*, *5*(9), 52-

    64.

Fialkoff, F. (2003). Too soon to ditch print. *Library Journal, 128* (9), 87.

Fiess, M. (2001, April 19). UAlbany establishes new school of Nanosciences and

    Materials. *University Update,* pp. 1, 7.

Florentino, L. (1979, Oct. 5). Chemistry Ph.D. Department awaiting reinstatement.

    Albany Student Press, p.1.

Fountain, K. C. (2004). To Web or not to Web? : The evaluation of World Wide Web

    publishing in the academy. In D.L. Andersen (Ed.), *Digital scholarship in the*

    tenure, promotion, and review process (pp. 44-60). Armonk, N.Y.: M.E. Sharpe.

French, W. M. & French, F.S. (1944). *College of the Empire State: A centennial history of the New York State College for Teachers at Albany.*

Fulda, J.S. (2000). The Internet as an engine of scholarship. *Computers and Society, 30* (1), 17-27.

Gale, R., & Golde, C. M. (2004). Doctoral education and the scholarship of teaching and learning. *Peer Review, 6*(3), 8-12.

Galloway, P. (2004) Preservation of digital objects. In B. Cronin (Series Ed.), *Annual review of information science and technology: Vol. 38* (pp. 549-590). Medford, NJ: Information Today, Inc.

Garcia, O., Nagarajan, S.V., & Croll, P. (2003). WLGauge — A Web link gauge to measure the quality of external WWW links. In C. Lovelle & J. Manuel (Eds.) *Web Engineering: International Conference, ICWE 2003 Oviedo, Spain, July 14–18.* Berlin: Springer.

Gardner, S. K. (2009). *The development of doctoral students: Phases of challenge and support.* In K. Ward & L.E. Wolf-Wendal (Series Eds.), *ASHE Higher Education Report* (Vol. 34 no. 6). San Francisco: Jossey-Bass.

Garfield, E., (1979). *Citation indexing: Its theory and application in science, technology, and humanities.* New York: John Wiley & Sons.

Garfield, E., Sher I. H. & Torpie R. J. (1964). *The use of citation data in writing the history of science.* Philadelphia: Institute for Scientific Information.

Garson, G. D. (2002). *Guide to writing empirical papers, theses, and dissertations*. New York: Marcel Dekker.

Gatten, R. (2010). A case study in reference list accuracy. *New Library World, 111*(1/2), 16-25.

Germain, C. A. (2000). URLs: Uniform resource locators or unreliable reliable resource locators? *College and Research Libraries, 61* (4), 359-365.

Giesecke, J. (2011). Institutional repositories: Keys to success. *Journal of Library Administration, 51*(5/6), 529-542.

Gilbert, G.N. (1977). Referencing as persuasion. *Social Studies of Science, 7*, (1), 113-22.

Glanville, L. (2010). Web archiving: Ethical and legal issues affecting programmes in Australia and the Netherlands. *Australian Library Journal, 59*(3), 128-134.

Glassick, C.E., Huber, M.T. & Maeroff, G.I. (1997). *Scholarship assessed: Evaluation of the professoriate* San Francisco. CA: Jossey-Bass.

Gluibizzi, A. (2002). Citation analysis in the decorative arts: A tool for shaping a collection development policy for serials. *Art Documentation: Bulletin of the Art Libraries Society of North America, 21*(1), 23-26.

Goh, D., & Ng, P. (2007). Link decay in leading information science journals. *Journal of the American Society for Information Science & Technology*, *58*(1), 15-24.

Gordon, M. (2007). What makes interdisciplinary research original? Integrative scholarship reconsidered. *Oxford Review of Education, 33*(2), 195-209.

Gorniak-Kocikowska, K. (2001). Revolution and the library. *Library Trends, 49* (3), 454-70.

Gould, T. P. (2011). Protocols and challenges to the creation of a cross-disciplinary journal. *Journal of Scholarly Publishing, 42*(2), 105-141.

Grafton, A. (1997). *The footnote: A curious history.* Cambridge, MA: Harvard
   University Press.

Grafton, A. & Jardine, L. (1986). *From humanism to the humanities: Education and
   the liberal arts in fifteenth- and sixteenth-century Europe.* Cambridge,
   MA: Harvard University Press.

Gresham, K. (1996). Preventing plagiarism of the Internet: Teaching library
   researchers how and why to cite electronic resources. *Colorado Libraries,
   22*(2), 48-50.

Griffin, F. (2003). 404 File Not Found: Citing unstable Web sources. *Business
   Communication Quarterly, 66*(2), 46-54.

Griffin, L. (2002). Webwatch. *Library Journal*, *127*(16), 32-34.

Guenther, R., & Myrick, L. (2006). Archiving Web sites for preservation and access:
   Mods, Mets and Minerva. *Journal of Archival Organization, 4*(1/2), 141-166.

Gumport, P.J. (2005). Graduate education and research: Interdependence and strain.
   In P.G. Altbach, R.O. Berdahl & P. J. Gumport (Eds.), *American higher education
   in the twenty-first century* (2nd ed*.,* pp. 396-426). Baltimore: Johns Hopkins
   University Press.

Gurm, B. (2009). Is all scholarship equally valued? Fusion of horizons on the
   definition and status of scholarship. *International Journal for the Scholarship of
   Teaching & Learning, 3*(2), 1-10.

Hakala, J. (2004). Archiving the Web: European experiences. *Program: Electronic
   Library & Information Systems, 38*(3), 176-183

Hall, D. (2007). *The academic community: A manual for change*. Columbus: Ohio State
University Press.

Harter, S., & Kim, H. (1996). Accessing electronic journals and other e-publications:
An empirical study. *College & Research Libraries*, *57*(5), 440-443.

Harvard Law Review Association. (2010). *The bluebook: A uniform system of citation*,
(19th ed.). Cambridge, MA: Author.

Hathaway, K. (2006). How the GPO got its groove back: Government Printing Office
and government information on the Internet. *Reference Librarian, 45*(93), 109-
128.

Hawking, D., Craswell, N., & Thistlewaite, P. (1999). Overview of the TREC-7 very
large collection track. In E. M. Voorhees, & D. K. Harman (Eds.), *Proceedings of
the Seventh Text Retrieval Conference* (pp. 91–104). Gaithersburg, MD: NIST
Special Publication 500-242.

Hawley, P. (2003). *Being bright is not enough: The unwritten rules of doctoral study*.
Springfield, IL: C.C. Thomas.

Heppner, P.P. & Heppner, M.J. (2003). *Writing and publishing your thesis,
dissertation, and research: A guide for students in the helping professions.*
Belmont, CA: Brooks Cole.

Herbst, J. (1965). *The German historical school in American scholarship: A study in the
transfer of culture.* Ithaca, NY: Cornell University Press.

Hernandez-Borges, A., Jiménez-Sosa, A., de Arcaya, M., Macías-Cervi, P., Gaspar-
Guardado, M., & Ruíz-Rabaza, A. (2005). Risk markers for disappearance of

pediatric Web resources. *Journal of the Medical Library Association*, *93*(3), 374-380.

Hernon, P. & Dugan, R.E. (2004), The US government and e-government: Two steps forward, one step backwards? In A. Woodsworth & W. D. Penniman (Eds.), *Advances in Librarianship* (vol. 28, 121-149). New York : Academic Press.

Herring, M.Y. (2007). *Fool's gold : Why the Internet is no substitute for a library.* Jefferson, N.C.: McFarland.

Herring, S. D. (2002). Use of electronic resource in scholarly electronic journals: A citation analysis. *College & Research Libraries*, *63*(4), 334-41.

Herrington, V. J. (1998). Way beyond BI: A look to the future. *Journal of Academic Librarianship, 24*(5), 381-386.

Hester, E., Heilig, L., Drake, A., Johnson, K., Vu, C., Schilling, L., ... Dellavalle, R.P. (2004). Internet citations in oncology journals disintegrating: A vanishing resource? *Journal of the National Cancer Institute, 96* (12), 969-971.

Hilse, H. & Kothe, J. (2006). *Implementing persistent identifiers: Overview of concepts, guidelines and recommendations.* London: Consortium of European Research Libraries.

Hofstadter, R. & Hardy, C.D. (1952). *The development and scope of higher education in the United States.* New York: Columbia University Press, for the Commission on Financing Higher Education.

Horowitz, I. L. (1991). *Communicating ideas: The politics of scholarly publishing*. New Brunswick, N.J.: Transaction Publishers.

Horrell, J. (2008). Converting and preserving the scholarly record: An overview. *Library Resources & Technical Services*, *52*(1), 27-32.

Hovde, K. (2007). You can't get there from here: Student citations in an ephemeral electronic environment. *College & Research Libraries*, *68*(4), 312-321.

Howell, B. (2006). Proving Web history: How to use the Internet Archive. *Journal of Internet Law*, *9*(8), 3-9.

Hswe, P., Kaczmarek, J., Houset, L., & Eke, J. (2009). The Web Archives Workbench (WAW) Tool Suite: Taking an archival approach to the preservation of Web content. *Library Trends, 57*(3), 442-460.

Hutchins, R. M. (1956). *Some observations on American education*. Cambridge: University Press.

Iverson, C. (2007). *AMA manual of style: A guide for authors and editors*. New York : Oxford University Press.

Jacobs, J. R., & Reich, V. (2010). Preservation for all: LOCKSS-USDOCS and our digital future. *Dttp: A Quarterly Journal of Government Information Practice & Perspective, 38*(3), 32-33.

Jacso, P. (2002). The digital object identifier. *Information Today, 19*(11), 30-31.

Jansen, H. (2006). Permanent access to electronic journals. *Information Services & Use*, *26*(2), 129-134.

Jatowt, A. A., & Tanaka, K. K. (2007). Towards mining past content of Web pages. *New Review of Hypermedia & Multimedia, 13*(1), 77-86.

Jencks, C. & Riesman, D. (1968). *The academic revolution*. Garden City, NY: Doubleday.

Johnson, K.R., Hester, E.J, Schilling, L.M., & Dellavalle, R.P. (2004, February 21). Addressing internet reference loss. *Lancet,* 363 (9409), 660-661.

Jones, H.M. (1979). The scholar and the world: Scholarship as ritual, contest, and civil independence. *Journal of Higher Education, 50* (4), 429-438.

Jonscher, C. (1999). *The evolution of wired life: From the alphabet to the soul-catcher chip- how information technologies change our world*. New York: John Wiley.

Jorgensen, C., & Jorgensen, P. (1991). Citations in hypermedia: Maintaining critical links. *College & Research Libraries, 52*(6), 528–536.

Jorgensen, J. (2006). The online government information movement: Retracing the route to DigiGov through the federal documents collection. *Reference Librarian*, *45*(94), 139-162.

Jul, E. (1997). Cataloging Internet resources: Survey and prospectus. *Bulletin of the American Society for Information Science & Technology, 24* (1), 6-9.

Jul, E. (1998). Cataloging Internet resources: An assessment and prospectus. *Serials Librarian, 34*(1/2), 91-104.

Juran, J. M. (1974). Quality policies and objectives. In J.M. Juran, F. M. Gryna, Jr., & R. S. Bingham (Eds.), *Quality control handbook,* (3rd ed., pp. 3-1- 3-12). New York, McGraw-Hill.

Kahle, B. (1997) Preserving the Internet. *Scientific American, 276* (3), 82-83.

Kendall, S. L., Bakke, C., & McDaniels, L. (2006). Tracking usage of e-government publications. *Serials Librarian*, *50*(3/4), 295-303.

Kennedy, S. (2004). DOI. *Computers in Libraries*, *24*(2), 19.

Kiernan, V. (2004, March 19). Company to track citations of online scholarship. *Chronicle of Higher Education*, *50*(28), A31.

King, D., Boyce, P., Montgomery, C., & Tenopir, C. (2003). Library economic metrics: Examples of the comparison of electronic and print journal collections and collection services. *Library Trends*, *51*(3), 376-400.

Kirchhoff, A. J. (2008). Digital preservation: Challenges and implementation. *Learned Publishing, 21*(4), 285-294.

Kling, R & Callahan, E. (2003). Electronic journals, the Internet, and scholarly communication. In B. Cronin (Ed.), *Annual review of information science and technology,* (Vol. 37, pp. 127-177). Medford, NJ: Information Today, Inc.

Kling, R. & Spector, L. (2004). Rewards for scholarly communication. In D.L. Andersen (Ed.), *Digital scholarship in the tenure, promotion, and review process* (pp. 78-103). Armonk, NY: M.E. Sharpe.

Knutson, A. N. (2009). Proceed with caution: How digital archives have been left in the dark. *Berkeley Technology Law Journal, 24*(1), 437-473.

Kochen, M. (1987, March). How well do we acknowledge intellectual debts? *Journal of Documentation, 43,* 54-64.

Koefod, P. E. (1964). *The writing requirements for graduate degrees.* Englewood Cliffs, NJ: Prentice-Hall.

Koehler, W. (1999). An analysis of Web page and Web site constancy and permanence. *Journal of the American Society for Information Science, 50* (2), 162-180.

Koehler, W. (2002). Web page change and persistence - a four-year longitudinal study. *Journal of the American Society for Information Science and Technology, 53* (2), 162-171.

Król, D., & Litwin, K. (2010). Quality analysis of domain-dependent Internet archives. *Cybernetics & Systems, 41*(1), 75-85.

Kronick, D.A. (1985). *The literature of the life sciences*. Philadelphia, PA: ISI Press.

Kronman, A. T. (2007). *Education's end: Why our colleges and universities have given up on the meaning of life.* New Haven: Yale University Press.

Kucan, L. (2011). Approximating the practice of writing the dissertation literature review. *Literacy Research & Instruction, 50*(3), 229-240.

Kunze, J. (2004). ARK. *Computers in Libraries*, *24*(2), 18.

Kurtz, M., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. (2005). The effect of use and access on citations. *Information Processing & Management*, *41*(6), 1395-1402.

Kuruppu, P., & Moore, D. (2008). Information use by PhD students in agriculture and biology: A dissertation citation analysis. *portal: Libraries and the Academy*, *8*(4), 387-405.

Kushkowski, J. (2005). Web citation by graduate students: A comparison of print and electronic theses. *portal: Libraries & the Academy*, *5*(2), 259-276.

Landesman, B. (1997). Keeping the Jell-O ® nailed to the wall: Maintaining and managing the virtual collection. *Serials Librarian, 30* (3/4), 137-47.

Langston, M., & Tyler, J. (2004). Linking to journal articles in an online teaching

    environment: The persistent link, DOI, and OpenURL. *Internet and Higher*

    Education, 7(1), 51-58

Lawrence, S., & Giles, C.L. (1999). Accessibility of information on the Web. *Nature,*

    400, 107–109.

Lawrence, S. S., Coetzee, F. F., Glover, E. E., Flake, G. G., Pennock, D. D., Krovetz, B. B.,

    & ...Giles, L. (2000). Persistence of information on the Web: Analyzing citations

    contained in research articles. In A. Agah, J. Callan, & E. Rundensteiner (Eds.),

    Proceedings of Ninth International Conference on Information and Knowledge

    Management (CIKM), 2000 (pp. 235-42). McLean, VA: ACM.

Lawrence, S. S., Pennock, D. M., Flake, G. W., Krovetz, R. R., Coetzee, F. M., Glover, E.

    E., & ... Giles, C. L. (2001). Persistence of Web references in scientific research.

    IEEE Computer, *34*(2), 26-31.

Lehnert, W. G. & Kopec, R. L. (2008). *Web 101* (3rd ed.). New York : Pearson

    Education, Inc.

Leuf, B. (2006). *The semantic Web: Crafting infrastructure for agency.* West Sussex,

    England: John Wiley & Sons, Ltd.

Lewandowski, D. (2008). The retrieval effectiveness of web search engines:

    considering results descriptions. *Journal of Documentation, 64*(6), 915-937.

Lewis, D. (2007). A strategy for academic libraries in the first quarter of the 21st

    Century. *College & Research Libraries*, *68*(5), 418-434.

Liu, Z. (2008). *Paper to digital: Documents in the information age*. Westport, CT:

    Libraries Unlimited.

Locke, L. F. & Spirduso, W.W. (1976). *Proposals that work: A guide for planning research.* New York: Teachers College Press, Teachers College, Columbia University.

Lovitts, B. E. (2005). How to grade a dissertation. *Academe, 91*(6), 18-23.

Lovitts, B.E. (2007). *Making the implicit explicit: Creating performance expectations for the dissertation.* Sterling, VA: Stylus Publishing.

Lovitts, B.E., & Wert, E.L. (2008a). *Developing quality dissertations in the humanities: A graduate student's guide to achieving excellence.* Sterling, VA: Stylus Publishing.

Lovitts, B.E., & Wert, E.L. (2008b). *Developing quality dissertations in the social sciences: A graduate student's guide to achieving excellence.* Sterling, VA: Stylus Publishing.

Lyman, P. (2002). Archiving the World Wide Web. In *Building a national strategy for digital preservation: Issues in digital media archiving* (pp.38-51)*. Washington, D.C.: Council on Library and Information Resources and Library of Congress.

Lynch, C. (1998). Identifiers and their role in networked information applications. *Feliciter, 44*(2), 31-35.

Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *portal: Libraries & the Academy, 3*(2), 327-336.

Lynch, P.J. & Horton, S. (2009). *Web style guide: Basic design principles for creating Web sites* (3rd ed.) New Haven, CT: Yale University Press.

Lyons, S. (2005). Persistent identification of electronic documents and the future of footnotes. *Law Library Journal, 97*(4), 681-694.

McCain, K. W. (2000). Sharing digitized research-related information on the World Wide Web. *Journal of the American Society for Information Science, 51*(14), 1321-1327.

McClure, M. (2006). Archive-It 2. *EContent*, *29*(8), 14-15.

McCown, F., & Nelson, M. (2009). Lazy preservation and Warrick. *Communications of the ACM, 52*(11), 141-145.

McCracken, G. D. (1988). *The long interview*. Newbury Park, CA: Sage Publications.

McMillan, S.J., & White, C. (2001). Survivors: A qualitative study of Web sites that endure. In C.R Taylor (Ed.), *The Proceedings of the 2001 Conference of The American Academy of Advertising.* American Academy of Advertising (pp. 101-107). Villanova, PA: Villanova University.

Machrone, B. (1999). Yes, you have link rot, and your customers hate it. *PC Week*, *16*(7), 81.

Madsen, D. (1983). *Successful dissertations and theses: A guide to graduate student research from proposal to completion* (2nd ed.). San Francisco, CA: Jossey-Bass.

Maniatis, P., Roussopoulos, M., Giuli, T., Rosenthal, D., & Baker, M. (2005). The LOCKSS peer-to-peer digital preservation system. *ACM Transactions on Computer Systems*, *23*(1), 2-50.

Marill, J.L. (1999). A survey of standards for identifying serial items on the Internet. *Acquisitions Librarian 21* (21): 83-91.

Markwell, J., & Brooks, D. (2002a). Broken links: The ephemeral nature of educational WWW hyperlinks. *Journal of Science Education and Technology, 11* (2), 105-108.

Markwell, J., & Brooks, D. (2002b). Online. *Chronicle of Higher Education, 48* (34), A31.

Markwell, J., & Brooks, D. (2003). 'Link rot' limits the usefulness of Web-based educational materials in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education, 31* (1), 69-72.

Markwell, J., & Brooks, D. (2008). Evaluating Web-based information: Access and accuracy. *Journal of Chemical Education, 85*(3), 458-459.

Martin, K., & Eubank, K. (2007). The North Carolina State Government Website Archives: A case study of an American government Web archiving project. *New Review of Hypermedia & Multimedia*, *13*(1), 7-26.

Martin, S.K. (1996). Organizing collections within the Internet: A vision for access. *Journal of Academic Librarianship 22,* (4), 291-92.

Masanès, J. (2005). Web Archiving methods and approaches: A Comparative Study. *Library Trends*, *54*(1), 72-90.

Matusov, E., & Hampel, R. (2008). Two approaches to tenure and promotion criteria. *Academe, 94*(1), 37-39.

Matthews, V. M. & Owings, M.A. (1999). URL instability and content changes in science journal web sites. *Serials Review, 25* (3), 25-34.

Mauch, J. & Park, N. (2003). *Guide to the successful thesis and dissertation: A handbook for students and faculty.* New York : M. Dekker, Inc.

Mendenhall, W. & Sinich, T. (2003). *A second course in statistics: Regression analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Mercer, H., Koenig, J., McGeachin, R. B., & Tucker, S. L. (2011). Structure, features, and faculty content in ARL member repositories. *Journal of Academic Librarianship, 37*(4), 333-342.

Meszaros, R.L. (2002). The Internet, scholarly communication, and collaborative research. In C.F. Thomas (Ed.), *Libraries, the Internet, and scholarship: Tools and trends converging* (pp. 31-44). New York: Marcel Dekker.

Milstead, J., & Feldman, S. (1999). Metadata projects and standards. *Online*, *23*(1), 32–40.

Modern Language Association of America. (2009). *MLA handbook for writers of research papers.* New York : Author.

Mogge, D. (1999). Seven years of tracking electronic publishing: *The ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists*. *Library Hi Tech, 17* (1), 17–25.

Moghaddam, G. (2008). Preserving scientific electronic journals: a study of archiving initiatives. *Electronic Library*, *26*(1), 83-96.

Molholm, K. N. (2006). Standards and interoperability. *Information Services & Use, 26*(1), 29-37.

Morrisey, L. (2002). Bibliometric and bibliographic analysis in an era of electronic scholarly communication. *Science & Technology Libraries, 22*(3/4), 149-160.

Nagaraja, A., Joseph, S. A., Polen, H. H., & Clauson, K. A. (2011). Persistence and attrition of uniform resource locators (URLs) in an open access medical journal. *Program: Electronic Library & Information Systems, 45*(1), 98-106.

National Research Council (U.S.) International Standards, Conformity Assessment, and U.S. Trade Policy Project Committee. (1995). Standards*, conformity assessment, and trade: Into the 21st century*. Washington, DC: National Academy Press.

Neacsu, E. D. (2002). Legal scholarship and digital publishing: Has anything changed in the way we do legal research? *Legal Reference Services Quarterly*, *21* (2-3), 105-122.

Nelson, M., McCown, F., Smith, J., & Klein, M. (2007). Using the web infrastructure to preserve Web pages. *International Journal on Digital Libraries, 6*(4), 327-349.

Nettles, M.T. & Millett, C.M. (2006). *Three magic letters: Getting to Ph.D.* Baltimore: Johns Hopkins University Press.

Neuendorf, K.A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

Nicholls, G. (2005). *The challenge to scholarship: Rethinking learning, teaching and research*. New York: Routledge.

Notess, G. (2003). Unlocking URLs: Extensions, shortening options, and other oddities. *Online*, *27*(3), 54-56.

Ó Dochartaigh, N. (2007). *Internet research skills: How to do your literature search and find research information online*. Thousand Oaks, CA: Sage Publications.

O'Hara, M., Carter, C., Dewis, P., Kay, J & Wainwright, J. (2011). *Successful dissertations : The complete guide for education, childhood and early childhood studies students.* New York: Continuum

O'Leary, M. (2000). OCLC opens Web catalog to all comers. *Information Today*, *17*(2), 47.

O'Mahony, D.P. (1998). Here today, gone tomorrow. In F.C. Lynden and E.A.

Chapman (Eds.), *Advances in Librarianship* (Vol. 22, 107-22). New York:

Academic Press.

Oder, N. (2000, October). Cataloging the Net: Two years later. *Library Journal, 125*

(16), 50-51.

Oder, N., Hinman, H. & Leita, C. (1998, October). Cataloging the Net: Can we do it?

Library Journal, 123 (16), 47-51.

Oermann, M., Nordstrom, C., Ineson, V., & Wilmes, N. (2008). Web citations in the

nursing literature: How accurate are they? *Journal of Professional Nursing,

24*(6), 347-351.

Okerson, A. (2000, Spring). Are we there yet? Online e-resources ten years after.

Library Trends, *48*(4), 671-693.

Olfson, E., & Laurence, J. (2005). Accessibility and longevity of Internet citations in a

clinical AIDS journal. *AIDS Patient Care & STDs*, *19*(1), 5-8.

Online searching with Internet. (1992). *Library Journal, 117*(21), 102.

Oren, G., & Watson, M. (2009). Accuracy of references in the ophthalmic literature.

Journal of the Medical Library Association, 97(2), 142-145.

Pancheshnikov, Y. (2007). A comparison of literature citations in faculty

publications and student theses as indicators of collection use and a

background for collection management at a university library. *Journal of

Academic Librarianship, 33*(6), 674-683.

Panos, P. (2003). The Internet Archive: An end to the digital dark age. *Journal of

Social Work Education, 39*(2), 343-347.

Partridge, G. E. (1912). *Genetic philosophy of education; An epitome of the published educational writings of President G. Stanley*. New York: Sturgis & Walton Company.

Paskin, N. (1999). The digital object identifier system: Digital technology meets content management. *Interlending and Document Suppl,y 27* (1), 13–16.

Paulsen, M., & Feldman, K. (1995). Toward a reconceptualization of scholarship: A human action system with functional imperatives. *Journal of Higher Education, 66* (6), 615-40.

Pelikan, J. (1983). *Scholarship and its survival: Questions on the idea of graduate education.* Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Phelps, T.A. & Wilensky, R.(2004). Robust hyperlinks: Cheap, everywhere, now. In P. King & E.V. Munson (Eds.), *Digital documents: Systems and principles. 8th International Conference on Digital Documents and Electronic Publishing, DDEP 2000 , 5th International Workshop on the Principles of Digital Document Processing, PODDP 2000, Munich, Germany, September 2000, revised papers* (pp. 28-43). New York: Springer.

Phillips, M. (2005). What should we preserve? The question for heritage libraries in a digital world. *Library Trends*, *54*(1), 57-71.

Plano Clark, V.L., Creswell, J.W., O'Neil Green, D., & Shope, R.J. (2008). Mixing quantitative and qualitative approaches: An introduction to emergent mixed methods research. In S.N. Hesse-Biber & P. Leavy (Eds.), *Handbook of emergent methods* (pp. 363-387). New York: Guilford Press.

Pool, R. (2011). Store for safekeeping. *Research Information*, (56), 12-13.

Pope, N.R. (1992). Accuracy of references in ten library science journals. *RQ, 32*(2), 240–243.

Potts, K. (2007). *Web design and marketing solutions for business websites.* Berkeley, CA : Friends of Ed.

Preserving Bush Websites. (2008, October). *American Libraries, 39* (9), 42.

Press, L. (1994). Commercialization of the Internet. *Communications of the ACM, 37*(11), 17-21.

Quint, B. (2008). 'Just Do It': The Net newbie way. *Information Today*, *25*(9), 7-8.

Radia, P., & Stapleton, P. (2009). Unconventional sources as a new convention: The shifting paradigm of undergraduate writing. *Internet and Higher Education, 12*(3-4), 156-164.

Reich, V. (2008). CLOCKSS--It takes a community. *Serials Librarian*, *54*(1/2), 135-139.

Rhodes, S. (2010). Breaking down link rot: The Chesapeake Project Legal Information Archive's examination of URL stability. *Law Library Journal, 102*(4), 581-597.

Riahinia, N., Zandian, F., & Azimi, A. (2011). Web citation persistence over time: A retrospective study. *Electronic Library, 29*(5), 609-620.

Rice, R.E. (2005). "Scholarship reconsidered": History and context. In K. O'Meara and R.E. Rice (Eds.), *Faculty priorities reconsidered: Rewarding multiple forms of scholarship* (pp. 17-31). San Francisco, CA: Jossey-Bass.

Riley, M., Wood, R.C., Clark, M.A., Wilkie, E., & Szivas, E. (2000). *Researching and writing dissertations in business and management.* London: Thomson Learning.

Robbins, S., McCain, C., & Scrivener, L. (2006). The changing format of reference collections: Are research libraries favoring electronic access over print? *Acquisitions Librarian*, *18*(35/36), 75-95.

Roberts, J.M. (1993). *History of the world.* New York: Oxford University Press.

Robots exclusion standard. (2007). In J. Dong (Ed.), *Network dictionary* (p. 417). Saratoga, CA: Javvin Technologies, Inc.

Rosenzweig, R. (2003). Scarcity or abundance? Preserving the past in a digital era. *American Historical Review, 108*(3), 735-762.

Ross, S. (1999). Preservation and networking in aid of research. In T. Coppock (Ed.), *Information technology and scholarship: Applications in the humanities and social* (pp.311-328). Oxford: Oxford University Press.

Rossi, P. H. (1966). Researchers, scholars and policy makers: The politics of large scale research. In R. S. Morison (Ed.), *The contemporary university: U.S.A.* (pp. 110-129). Boston: Houghton Mifflin.

Rudestam, K. E. & Newton, R.R. (2007). *Surviving your dissertation: A comprehensive guide to content and process.* Los Angeles, CA: SAGE Publications.

Rudolph, F. (1962). *The American college and university: A history.* New York: Knopf.

Rudolph J. & Brackstone, D. (1990, April 11). Too many scholars ignore the basic rules of documentation. *Chronicle of Higher Education, 36,* A56.

Rumsey, M. (2002). Runaway train: Problems of permanence, accessibility, and stability in the use of Web sources in law review citations. *Law Library Journal, 94* (1), 27-39.

Russell, E., & Kane, J. (2008). The missing link: Assessing the reliability of Internet citations in history journals. *Technology & Culture, 49*(2), 420-429.

Russell, K. (2002). Libraries and digital preservation: Who is providing electronic access for tomorrow? In C.F. Thomas (Ed.), *Libraries, the Internet, and scholarship: Tools and trends converging* (pp. 1-30). New York: Marcel Dekker.

Ryan, T., Field, R., & Olfman, L. (2003). The evolution of US state government home pages from 1997 to 2002. *International Journal of Human-Computer Studies*, *59*(4), 403-430.

Sandys, E. (1906). *A history of classical scholarship: From the sixth century B.C. to the end of the Middle Ages* (2nd ed.). Cambridge: Cambridge University Press.

Scharf, D. (2002). The DOI is coming. *Information Outlook*, *6* (9), 6 –10.

Schieber, P. (2009). Chronology: Noteworthy achievements of the Cooperative 1967-2008. *Journal of Library Administration, 49*(7), 763-775.

Schilling, L., Wren, J., & Dellavalle, R. (2004). Bioinformatics leads charge by publishing more Internet addresses in abstracts than any other journal. *Bioinformatics*, *20*(17), 2903-2903.

Schmidt, K., Shelburne, W., & Vess, D. (2008). Approaches to selection, access, and collection development in the Web world. *Library Resources & Technical Services*, *52*(3), 184-191.

Schmiede, R. (2009). Upgrading academic scholarship: Challenges and chances of the digital age. *Library Hi Tech, 27*(4), 624-633.

Schneider, K. (2007). Lots of librarians keep stuff safe. *Library Journal*, *132*(13), 30-31.

Scholarship. (1989). In J.A. Simpson & E.S.C. Weiner (Eds.), *The Oxford English dictionary* (2nd ed., Vol. 14, p. 630). New York: Clarendon Press.

Schroeder, K. (2005). Persistent Identification for the Permanent Referencing of Digital Resources– The Activities of the EPICUR Project: Enhanced Uniform Resource Name (URN) Management at Die Deutsche Bibliothek. *Serials Librarian, 49* (3), 75-87.

Seadle, M. (2006). A social model for archiving digital serials: LOCKSS. *Serials Review*, *32*(2), 73-77.

Select Committee on Academic Programs and Priorities, State University of New York at Albany. (1975). *Select Committee on Academic Programs and Priorities: The general report.* Albany, NY: Author.

Select Committee on Academic Programs and Priorities, State University of New York at Albany. (1990). *Select Committee on Academic Programs and Priorities: Assessing a decade 1980-1990: A self-study report for the Middle States Association of Colleges and Schools.* Albany, NY: Author.

Sellitto, C. (2004). A study of missing Web-cites in scholarly articles: Towards an evaluation framework. *Journal of Information Science, 30*(6), 484-495.

Sellitto, C. (2005).The impact of impermanent Web-located citations: A study of 123

scholarly conference publications. *Journal of the American Society for*

Information Science and Technology, 56 (7), 695-703.

Shafer, K. E., Weibel, S.L. & Jul, E. (200l). The PURL Project. *Journal of Library*

Administration, 34(1/2), 123-125.

Shannon, C. E. & Weaver, W. (1948). *The mathematical theory of communication*.

Chicago, IL: University of Illinois Press.

Shatz, D. (2004). *Peer review: A critical inquiry.* Lanham, MD: Rowman & Littlefield.

Shuy, R. W. (2004). In-person versus telephoning interviewing. In J. A. Holstein & J.F.

Gubruim (Eds.), *Inside interviewing: New lenses, new concerns* (pp. 175- 193).

Thousand Oaks, CA: Sage Publications.

Singer, C. (2003). The transition of U.S. government publications from paper to the

Internet: A chronology. *Internet Reference Services Quarterly*, *8*(3), 29-45.

Small, H.G. (1978). Cited documents as concept symbols. *Social Studies of Science, 8*

(3), 327-40.

Smith, A. (2003). *New-model scholarship: How will it survive?* Washington DC:

Council on Library and Information Resources.

Smith, A. (2007). Valuing preservation. *Library Trends*, *56*(1), 4-25.

Smith, J. (1995, October 13). Citation takes fast track to a dead end. *The Times Higher*

Educational Supplement, 119, 6.

Smith, W. (2005). Still lost in cyberspace? Preservation challenges of Australian

Internet resources. *Australian Library Journal*, *54*(3), 274-287.

Snyder, P.J., & Peterson, A. (2002). The referencing of Internet Web sites in medical and scientific publications. *Brain and Cognition, 50*(2), 335-337.

Sollins, K.R. (1994). The hard problems are not all solved. In E.F. Duranceau (Ed.), Naming and describing networked electronic resources: The role of uniform resource identifiers. *Serials Review*, *20*(4), 34-37.

Sotos, J.G. (2006). [Letter]. Scholarship erosion. *Annals of Internal Medicine, 145*(1), 77.

Spinellis, D. (2003). The decay and failures of Web references. *Communications of the ACM, 46* (1), 71-77.

State University of New York. (1961). *State University of New York: The master plan, revised 1960*. Albany, NY: Author.

State University of New York at Albany. (1963). *Graduate School of Public Affairs: Preliminary announcement1962-63.* Albany, NY: Author.

State University of New York at Albany. (1963). *State University of New York at Albany general bulletin 1963-64: One hundred nineteenth year.* Albany, NY: Author.

State University of New York at Albany. (1966). *State University of New York at Albany graduate bulletin 1966-67.* Albany, NY: Author.

State University of New York at Albany. (1967). *State University of New York at Albany graduate bulletin 1967-68.* Albany, NY: Author.

State University of New York at Albany. (1968). *State University of New York at Albany graduate bulletin 1968-69.* Albany, NY: Author.

State University of New York at Albany. (1969). *State University of New York at Albany graduate bulletin 1969-70.* Albany, NY: Author.

State University of New York at Albany. (1970). *SUNYA graduate bulletin 1970-71.* Albany, NY: Author.

State University of New York at Albany. (1971a). *Self-study report in preparation for the accrediting visitation by Middle States Association of Colleges and Secondary Schools.* Albany, NY: Author.

State University of New York at Albany. (1971b). *SUNYA graduate bulletin 1971-72.* Albany, NY: Author.

State University of New York at Albany. (1990). *SUNYA graduate bulletin 1990-91.* Albany, NY: Author.

Steneck, N. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science & Engineering Ethics, 12*(1), 53-74.

Stimson, D. (1968). The History of Ideas Club. In Johns Hopkins History of Ideas Club (Ed.), *Studies in intellectual history* (pp. 174-96)*.* New York: Greenwood Press.

Storr, R. (1953). *The beginnings of graduate education in America.* Chicago: University of Chicago Press.

Strader, C., & Hamill, F. (2007). Rotten but not forgotten: Weeding and maintenance of URLs for electronic resources in the Ohio State University online catalog. *Serials Librarian*, *53*(1/2), 163-177.

Style Manual Committee, Council of Science Editors. (2006). *Scientific style and format: The CSE manual for authors, editors, and publishers.* Reston, VA: Council of Science Editors and Rockefeller University Press.

Sweetland, J.H. (1989). Errors in bibliographic citations: A continuing problem. *Library Quarterly, 59*(4), 291–304.

Swetnam, D. (2004). *Writing your dissertation: How to plan, prepare and present successful work* (3rd ed.). Oxford: How to Books.

Sykes, W. & Collins, M. (1988). Effects of mode of interview: Experiments in the U.K. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 302-320). New York: Wiley.

Szydlowski, N. (2010). Archiving the Web: It's going to have to be a group effort. *Serials Librarian, 59*(1), 35-39.

Tenopir, C., Baker, G., Robinson, W., & Grogg, J. (2006). Renovating this old house. *Library Journal, 131*(9), 32-36.

Thelin, J. R. (2004). *A history of American higher education.* Baltimore: Johns Hopkins University Press.

Thelwall, M. (2000). Commercial Web sites: Lost in cyberspace? *Internet Research, 10* (2), 150–159.

Thelwall, M., & Vaughan, L. (2004). A fair history of the Web? Examining country balance in the Internet Archive. *Library & Information Science Research*, *26*(2), 162-176.

Thomas, J. (2004). Web Site demise and graduate research: Persistence of Web pages cited in social work theses. *Behavioral & Social Sciences Librarian, 22*(2), 67-78.

Thompson, C. (2003). Information illiterate or lazy: How college students use the Web for research." *portal: Libraries and the Academy, 3* (2), 259-268.

Thorp, A., & Brown, L. (2007). Accessibility of Internet references in *Annals of Emergency Medicine*: Is it time to require archiving? *Annals of Emergency Medicine*, *50*(2), 188-192.e33.

Thwing, C. F. (1928). *The American and the German university: One hundred years of history*. New York: The Macmillan Company.

Tomaiuolo, N. (2006). DOIs, URLs, LoCKSS, and missing links. *Searcher, 14*(7), 18-23.

Tonta, Y. (2005). Internet and electronic information management. *Information Services & Use*, *25*(1), 3-12.

Trotter, A. (2002, December 4). Too often, educators' online links lead to nowhere. *Education Week, 22*(14), 1.

Trumbo, C., & Trumbo, J. (2002). Newspapers' use of URLs in daily content has grown steadily since mid-1990s. *Newspaper Research Journal, 23*(1), 78-82.

Tyler, D.C. & McNeil, B. (2003). Librarians and link rot: A comparative analysis with some methodological considerations. *portal: Libraries and the Academy, 3* (4), 615-632.

United States. Government Printing Office: Office of Electronic Information Dissemination Services. (2001). *GPO Access training manual*. Washington, DC: Government Printing Office.

University at Albany. (2004). *University at Albany President's report 2002-3003*. Albany: Author.

The University of Chicago Press. (2010). *The Chicago manual of style* (16th ed.). Chicago: Author.

Vallmitjana, N., & Sabaté, L. (2008). Citation analysis of Ph.D. dissertation references as a tool for collection management in an academic chemistry library. *College & Research Libraries, 69*(1), 72-81.

Van Leunen, M. (1992). *A Handbook for scholars* (rev. ed.). New York: Oxford University Press.

Van de Ven, A. H. (2007). *Engaged scholarship: A guide for organizational and social research.* New York: Oxford University Press.

Vaughan, L., & Shaw, D. (2003). Bibliographic and Web citations: What is the difference? *Journal of the American Society for Information Science & Technology*, *54*(14), 1313-1322.

Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science & Technology, 56*(10), 1075-1087.

Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology, 54*(1), 29–38.

Veronin, M. (2003). 'URL 404 File Not Found': Dealing with the transient nature of the Web. *Journal of Audiovisual Media in Medicine*, *26*(4), 153-155.

Veronin, M. (2005). One step forward, two steps back: The World Wide Web as an information resource. *Pharmacy Education*, *5*(3/4), 209-214.

Vogt, W.P. (1984). *The State University of New York at Albany, 1844-1984: A short history.* Albany, NY: State University of New York at Albany.

von Bubnoff, A. (2005). Science in the Web age: The real death of print. *Nature*, *438*(7068), 550-552.

Wagner, C., Gebremichael, M., Taylor, M., & Soltys, M. (2009). Disappearing act: Decay of uniform resource locators in health care management journals. *Journal of the Medical Library Association, 97*(2), 122-130.

Walker, G. (2008). Doctoral education in the United States of America. *Higher Education in Europe, 33*(1), 35-43.

Waters, D. (2006). Managing digital assets in higher education: An overview of strategic issues. *ARL: A Bimonthly Report on Research Library Issues & Actions*, (244), 1-10.

Waugh, C., & Ruppel, M. (2004). Citation analysis of dissertation, thesis, and research paper references in workforce education and development. *Journal of Academic Librarianship, 30*(4), 276-284.

Weber, R. P. (1990). *Basic content analysis* (2nd ed.). Newbury Park, CA: Sage Publications.

Weiss, T. R. (2001). Lost in cyberspace. *Computerworld, 35*(47), 10.

Weller, A.C. (2001). *Editorial peer review: Its strengths and weaknesses*. Medford, NJ: Information Today.

West, J. (2007, April 16). Saving digital history. *Library Journal*, *132*, 2-6.

What's up with the Wayback Machine? (2010). *Library Journal, 135*(9), 18.

Whisner, M. (2005). Cool Web sites. *Law Library Journal, 97*(4), 721-728.

Whittaker, J. (2004). *The cyberspace handbook.* New York: Routledge.

216

Wilkerson, W. (2006). The emergence of Internet citations in U.S. Supreme Court opinions. *The Justice System Journal, 27* (3), 323-339.

Woodward, J. (1996). Cataloging and classifying information resources on the Internet. In M.E. Williams (Ed.) *Annual Review of Information Science and Technology* (Vol. 31, pp.189-220). Medford, NJ: Information Today.

Woody, C. (1948). Language requirements for the doctor's degree. *Journal of Higher Education, 19*(2), 75-107.

Wouters, P., & de Vries, R. (2004). Formally citing the Web. *Journal of the American Society for Information Science & Technology*, *55*(14), 1250-1260.

Wren, J. D. (2004). 404 Not Found: The stability and persistence of URLs published in Medline. *Bioinformatics, 20*(5), 668–672.

Wren, J. D. (2008).URL decay in Medline--a 4-year follow-up study. *Bioinformatics, 24* (11), 1381-1385.

Wren, J., Johnson, K., Crockett, D., Heilig, L., Schilling, L., & Dellavalle, R. (2006). Uniform resource locator decay in dermatology journals: Author attitudes and preservation practices. *Archives of Dermatology*, *142*(9), 1147-1152.

Wu, Z. (2009). An empirical study of the accessibility of web references in two Chinese academic journals. *Scientometrics*, *78*(3), 481-503.

Wyles, D. (2004). Citation errors in two journals of psychiatry: A retrospective analysis. *Behavioral & Social Sciences Librarian, 22*(2), 27-53.

Zhang, Y. (2001). Scholarly use of Internet-based electronic resources. *Journal of the American Society for Information Science and Technology, 52*(8), 628–654.

Appendix A – Screen shot of Internet Archive

Appendix B – Screen shot of WebCite

Appendix C - Interrater Reliability Instructions

1. Using Access bring up the IRRtool.
2. From the Options next to the Security Warning Select enable.
3. Bring up one citation at a time. I recommend that you bring up three windows:
   a. Browser – to search URL
   b. Internet Archive (IA) – archive. org enter URL in Wayback Machine window
   c. Google Search – google.com
4. For each citation URL:
   a. Browser search – enter the URL into the location window
      i. If it retrieves an error message or blank screen select No in the Find IT? Drop down. Note the error in the Not Found drop down.
      ii. If it retrieves a source note Yes in the Find IT? Drop down. Make any notes in the Found Comment drop down (e.g., it is not the cited work)
   b. Internet Archive (IA) – archive. org enter URL in Wayback Machine window
      i. Check the dates closest to the cited work (up to 3)
      ii. If it retrieves an error message or blank screen, select No (with possible reason) in the IA? Drop down
      iii. If it retrieves a source note Yes in the IA? Drop down
   c. Google Search – google.com
      i. Conduct a Google search using the citation information
      ii. Look through the first 3 pages of screen shots for the cited source.
      iii. You may try up to 6 searches.
      iv. Note in the On Google? Drop down whether you were able to locate the source and comments. (NEI means Not Enough Information)
      v. If you were able to find the source put your Google Search in the Last Google window
   d. Add any additional comments

INTERVIEW QUESTIONNAIRE
UNIVERSITY AT ALBANY
DOCTORAL GRANTING DEPARTMENTS

Interview introduction: "Thank you so much taking the time to speak with me today. I would like to begin our conversation by asking you some questions about your department's doctoral practices with a focus on scholarship and citation. Your name and any information that might identify you as an individual will not be used in any publications. I would like your permission to tape this interview. As I mentioned when I set up this interview via email, I am doing research for my dissertation in Informatics. My research focuses on citation patterns in doctoral dissertations written at the University at Albany, with an interest in digital resources and rate of access. The quantitative component of my research includes an analysis of dissertation bibliographies. The interviews with members of Ph.D. granting departments will provide insight into the practices and policies at the University."

Permission to quote anonymously from the transcripts to support points made in the dissertations _____

"Do you have any questions before we start?"

Interview Checklist :

___Above blanks filled in   ___Recorder on   ___Tape labeled

Maintaining Persistent Scholarship

_____1. I'd like you to tell me your perspective on scholarship, in particular scholarship and doctoral dissertations.

_____2. When you write articles, books, other materials, do you use URL citations? What are your thoughts on the use of electronic citations, mainly free Web sites and pages?

_____3. Can you provide some details of your engagement with formal publication, such as writing and editing, and how those venues are responding to URL citation?

_____4. Since this is a doctoral granting department, dissertations are regularly published. Can you describe the citation practices within departmental dissertations and the use of URL citations within theses?

_____5. Do you and other department members use any methods in advising students to limit citation erosion?

_____6. Does the department provide any instruction to doctoral students to foster scholarly practices and citation persistence?

_____7. In departmental meetings, has the issue of citing URL sources within academic works, specifically dissertations, been discussed?

_____8. Does your department have any policies regarding citations within dissertations? (if available, request a copy)
          a. If there are policies:
          _____Age
          _____Usefulness
          _____Changes
          _____URL limits
          _____Archival requirements
          b. If there are no policies:
          _____a. Past policy
          _____b. Reasons for no policy
          _____c. Future policy

_____9. Have you used Web archival utilities, such as the Internet Archive's Wayback Machine or WebCite, to archive cited Web resources?

_____10. Does your department currently accept electronic publications, such as online-only journal articles, Web projects or sites, for the tenure and promotion evaluation process?

_____11. Can you recommend any one or two people in your department who have expressed interest and/or concern with the use of electronic citations in dissertations

Appendix E IRB Informed Consent

<u>Informed Consent</u>

Thank you so much for agreeing to participate in the study**, *Maintaining persistent scholarship: Case of University at Albany dissertations.* ** This research study looks at the use of freestanding URL citations within University at Albany dissertations published between 1996 and 2007. These will be tested for accessibility. In addition, the primary investigator will conduct interviews with key members of Ph.D.-granting departments to identify policies and practices relating to citation in dissertations. More specifically, the interview questions will focus on your usage of URL resources in publication; the practice of electronic resource citations within departmental dissertations; and any policies the department has developed regarding the use of Web-based resources.

We do not anticipate any risk in your participation other than you may become uncomfortable answering some of the questions. This interview will be taped on an audio-recorder. Please sign below if you are willing to have this interview recorded (specify audio or video). You may still participate in this study if you are not willing to have the interview recorded**.**

Your participation in this project is voluntary. Even after you agree to participate in the research or sign the informed consent document, you may decide to leave the study at any time without penalty or loss of benefits to which you may otherwise have been entitled. I will retain and analyze the information you have provided up until the point you have left the study unless you request that your data be excluded from any analysis and/or destroyed. You may choose not to answer any questions and may refuse to complete any portions of the research you do not wish to for any reason. Pertinent quotes from the interviews will be anonymous.

This project has been approved by the University at Albany Institutional Review Board. Approval of this project only signifies that the procedures adequately protect the rights and welfare of the participants. All information obtained in this

study is strictly confidential unless disclosure is required by law. In addition, the Institutional Review Board and University officials responsible for monitoring this study may inspect these records.

Although you may not receive direct benefit from your participation, others may ultimately benefit from the knowledge obtained from this research. We believe the results of this study will provide important information regarding citation practices in dissertations and will lead to best practice and a stronger scholarship.

Copy of Consent: One copy of this document will be kept together with the research records of this study. Also, you will be given a copy to keep.

I have read, or been informed of, the information about this study. I hereby consent to participate in the study.

Consent to participate interview_____

Constant to anonymous quote_____

Consent to have this interview taped via audio_____

Primary Investigator :   **Carol Anne Germain**
                                     Department of Informatics, Student
                                     518-442-3590
                                     cg219@albany.edu

Advisor :   **Deborah Lines Andersen**
                                     Department of Informatics
                                     518-442-5122
                                     dla@albany.edu

If you have any questions concerning your rights as a research participant that have not been answered by the investigator or if you wish to report any concerns about the study, you may contact the University at Albany's Office of Regulatory Research Compliance at 518.442-9050 or orrc@uamail.albany.edu."

Appendix F Doctoral Granting Departments

Department	University at Albany Doctoral Programs		
	Faculty (FTEs)	Doctoral enrolled	Year*
Anthropology	19.5	68	1971
Atmospheric and Environmental Sciences	15.3	39	1967
Biology	33.5	39	1965
Biomedical Sciences	10.5	48	1985
Business Organizational Studies (suspended)	46.2	6	1985
Chemistry	22.9	44	1965
Computer Science	14.4	28	1982
Communications	16	14	2009
Criminal Justice	19.5	47	1965
Economics	23.3	52	1969
Educational Administration and Policy Studies	13	90	1960
Educational and Counseling Psychology (ECPY)	37.2	118	1962
Division - Curriculum and Instruction	with ECPY	with ECPY	1963
Division - Education Psychology and Methodology	with ECPY	with ECPY	1964
Division - School Psychology	with ECPY	with ECPY	1985
Educational Theory & Practice	15.7	82	1994
English	53	64	1963
Environmental Health & Toxicology	10.5	21	1985
Epidemiology & Biostatistics	13.9	50	1988
History	27.1	58	1963
Humanistic Studies (suspended)	0	0	1984
Informatics	7	33	1990
Languages, Literatures and Cultures	27.1	35	1968
Latin American, Caribbean, and US Latino Studies	8.3	20	1995
Mathematics	31.3	39	1968
Nanoscale Science and Engineering			2001
Philosophy	13.3	42	1970
Physics	17.8	36	1965
Political Science	22.1	51	1966
Psychology	32.4	89	1963
Public Administration and Policy	23	48	1962
Reading	8.4	30	1963
Social Welfare	28.7	67	1984
Sociology	36.4	108	1970
Total	647.3	1466	