

Nicole M. Downing. UNC Law Library's Redaction of its Digitized Collection of North Carolina Supreme Court Briefs: A Case Study. A Master's Paper for the M.S. in L.S degree. April, 2014. 58 pages. Advisor: Anne Klinefelter.

This study evaluates the digital redaction process as undertaken by the University of North Carolina Kathrine R. Everett Law Library as part of digitizing their collection of North Carolina Supreme Court briefs. New privacy concerns are raised by digitizing court documents and making them available online. Libraries have an interest in digitizing their print collections of court documents for public access on the Internet, but have received no clear guidance on how to proceed in the face of legal concerns. The purpose of this research is to inform libraries of the legal, ethical, and practical situation surrounding redaction of digitized court documents faced by the UNC Law Library under North Carolina law and to provide libraries with an example of redaction procedures. This research was accomplished through a case study of the UNC Law Library and composed of interviews with the law librarians and observation of the redaction process.

Headings:

Digitization

Right of privacy

Public records

Law libraries

UNC LAW LIBRARY'S REDACTION OF ITS DIGITIZED COLLECTION OF
NORTH CAROLINA SUPREME COURT BRIEFS: A CASE STUDY

by
Nicole M. Downing

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

April 2014

Approved by

Anne Klinefelter

Table of Contents

Introduction.....	2
Background Statement.....	2
Purpose.....	5
Research Questions.....	6
Literature Review.....	6
Methodology.....	13
Research Design.....	13
Data Collection Methods.....	14
Data Analysis.....	16
Expected Benefits and Summary.....	17
Results and Discussion.....	18
Legal Considerations.....	20
Constitutional Law.....	20
Tort Law.....	22
Federal Law.....	23
State Law.....	24
Ethical Considerations.....	28
Information Protected by Laws.....	29
Censorship.....	31
Knowledge and Predictability.....	32
Practical Considerations.....	33
Viewpoints against Redaction.....	35
The Final Decision.....	35
The Redaction Process.....	37
BitCurator.....	37
Adobe Searches.....	39
Reflecting on Adobe Searches.....	43
Future Steps.....	46
Conclusion.....	48

Introduction

Background Statement

Over the last decade, documents of all kinds have gone through the digitization process and been made available online. Digitization is the process of changing or converting data or images to digital form. Businesses of all types have even been overhauling their operations so that many, if not all, of their documents exist only in digital form. Benefits of digitization are numerous, but there is one large concern with the process of digitization. When forms are digitized and made available online, privacy issues are of increased importance due to wider discoverability and ease of access. Documents often contain detailed, sensitive information about a person and the information may be the type one does not want made available to the general public. This is of specific concern when the documents are the type required by law to be made available to the public. Each state has its own law that sets out what documents are considered public records. Generally, public records are documents that are not considered to be confidential. This means that anyone may request access to these documents.

An example of public records, and the focus of this research, are the court documents included in a case docket, specifically case briefs. A case docket is a summary of all of the proceedings of a particular case and includes briefs as well as various writs, memorandums, and documents submitted as evidence. Dockets often contain a great deal of highly sensitive information, including social security numbers, driver's license

numbers, and home addresses. As dockets are public records, any citizen may view them, including the sensitive information they may potentially contain. Traditionally, there was a great deal more control over this process. Court documents were in paper form and people had to request individual documents from a clerk of court or locate the document in a library. In either case, these steps of having to go to a physical location and go through a process to access court documents resulted in practical obscurity. While the documents were publicly available, the steps required to access them made it so only the most determined people would do so.

However, there is now an interest in digitizing court documents to make them available online and allow a greater number of people access to the documents. Many courts have transitioned to e-filing, whereby court documents skip the paper step and are only available in digital form. This is another reason more documents are being digitized, as it is part of the courts' transition to e-filing. All of these digital documents create a problem, because of the high level of sensitive information they include without practical obscurity standing as a barrier to help protect this information.

This problem was first addressed when Congress passed the E-Government Act of 2002 (E-Government Act of 2002). This act required that the federal courts maintain a web site that allows the public access to certain court documents. This website is now known as PACER, which stands for Public Access to Court Electronic Records. The act also provided some guidance on the privacy issue this entailed, first by acknowledging that there were privacy concerns with this action and second by requiring the courts to establish rules to protect privacy for electronic court records. As a result, Federal Rule of Civil Procedure 5.2 and Federal Rule of Criminal Procedure 49.1 were created. These

rules set out what information must be redacted from e-filed and paper-filed documents: social security numbers and tax payer identification numbers, birth dates, name of individuals known to be minors, financial account numbers, and home addresses of individuals. These rules also required that the filer of court documents redact the documents before they were filed.

As e-filing has grown in state courts as well, individual states have been passing laws to address the issue for state courts. As of today, the majority of states have guidelines for e-filing, most of which include some type of redaction model (Hulse, 2009). There are three main types of redaction model: (1) filers should redact specified sensitive information prior to filing, (2) use of a confidential cover sheet which includes all of the confidential information that is redacted from the file, or (3) manual redaction of a file by the clerk after a request for the file is made (Hulse, 2009).

All of this guidance on redacting digital court documents applies to the courts and clerks of courts. However, many law libraries across the country are in possession of these same court documents that include sensitive information. The UNC Law Library is in possession of the North Carolina Supreme Court briefs, which no publisher has digitized and made available through a subscription service or otherwise. The trend in libraries has been to make as many library collections as possible available in digital form. It is not surprising then that libraries may wish to digitize their court documents and make them available for patrons on the Internet, especially in the face of concerns over practical obscurity blocking access to collections. But if these documents include the same sensitive information that many courts are redacting in the digital format as required by e-filing guidelines, how should libraries respond to digitizing court

documents? How does that process fit in with the law, if the relevant laws do not specifically address libraries? Are libraries also required to redact court documents before making them available on the Internet?

Currently, these printed court documents are available in un-redacted form in libraries and courthouses. However, libraries have traditionally concerned themselves with issues of reader privacy. Where does this form of content privacy fit in with traditional library notions of privacy? It may follow that if the courts consider this information too sensitive to be made publicly available, so should libraries. Private and public libraries may wonder if they should redact regardless of how the law applies to them due to these privacy concerns. Whether or not it is determined that redaction by libraries is mandated by law, many libraries interested in digitizing their court documents may be interested in redacting them as well. Some step by step guides on how to generally redact PDF's are available from state court websites, but there has been no formal guidance on how this procedure would work for libraries taking on a mass digitization project.

This research will study the undertaking by the University of North Carolina Kathrine R. Everett Law Library to digitize its collection of North Carolina Supreme Court briefs. It will explain the legal landscape the Law Library encountered in North Carolina, discuss why the library has decided to redact these briefs as part of the digitization process, and then evaluate the redaction process itself.

Purpose

The purpose of this research is to inform libraries of the legal, ethical, and practical considerations that may be involved in redacting court documents by looking at

the specific situation faced by the UNC Law Library and to show libraries an example of how procedures designed to accomplish this task may look. The findings related to the technological process and procedures for redacting court documents can inform all libraries in possession of court documents as well as librarians who may be facing the digitization of other forms of sensitive information. The findings as to the UNC Law Library's decision to go through the redaction process will be specific to a public academic law library in North Carolina, but their considerations may prove of interest to other public or private libraries making the decision. Currently, many libraries are unaware of their legal requirements on this issue or their options under the law. This study will explain the law surrounding this issue in North Carolina and discuss many of the privacy concerns raised by digitization. Even in the presence of a clear policy on redaction, undertaking a digitization and redaction project is a large and complicated process. Without guidance, libraries may err on the side of not digitizing court documents. Lack of knowledge about how to approach this project should not be a reason libraries hold back from digitization, when digitization serves the libraries' goals of broad access to information.

Research Questions

1. What considerations went into the UNC Law Library's decision to redact its digitized North Carolina Supreme Court Briefs collection?
2. Why did the UNC Law Library ultimately decide to redact?
3. What procedure was used to redact the digitized collection?

Literature Review

There has not yet been a great deal of scholarly literature studying the issues involved in digitizing a library's collections of court documents. However, there has been research conducted on the problems created by a library's physical collection of court documents. Also heavily addressed in legal literature is the complicated and problematic movement by state courts to make the switch to digital court records. Most relevant is the literature that addresses the reasons to digitize collections of court documents and the privacy issues raised by the digitization of these documents. These privacy issues were a crucial part of the UNC Law Library's decision of whether or not to redact their digitized collection of court documents.

In 1993, Leary wrote one of the earliest articles addressing the problem of a hard copy collection of court briefs by discussing the problem of "disappearing briefs". Briefs are the written arguments submitted by counsel to the court in a case. They provide the foundation for the opinion rendered by the court. Court briefs aren't technically published, but instead are produced in very small quantities of fewer than fifty copies. They are then deposited with selected libraries and courthouses. The briefs are of great importance during the appeal process which they address, but as time passes they become of less immediate importance. Leary writes that their value then resides in their historical precedence, citing authority, legal arguments for lawyers, and their research value for legal and nonlegal scholars. The article points out that there has been dissent within the library community as to the importance of holding onto court briefs. After studying the problem of federal court briefs that have been lost over the years and compiling all of the federal court brief holdings in libraries across the nation, Leary reaches the conclusion that briefs should all be converted to microform. This conversion idea was never adopted

by libraries on a wide scale, but the thought process also supports a conversion to digital form. Digitization allows for the preservation of and easy access to the material without requiring the space that hard copy holdings require, just as microform does.

The value of court briefs was addressed in Leary's article as it relates to the substantive legal history contained in the briefs. There is another reason to preserve and make widely available court briefs and that reason is government transparency. Studies measuring government transparency have been conducted using access to public records to determine the level of transparency. Armstrong's 2010 research consisted of a case study of Florida's government websites. He analyzed what public records were available on Florida's court websites, concluding that with this information public access advocates can turn their attention toward getting online access to those public records which are lacking. Court briefs were not among the records available on the studied Florida websites, even though they are public records. Ganapati and Reddick's 2012 study takes a broader look at open e-government initiatives by surveying state Chief Information Officers to determine the level of transparency in each state. This study found that there wasn't a great deal of consistency for transparency from state to state, although most Chief Information Officers believed their state government was working on becoming more transparent. These two studies show a concern for government transparency through online access to public records and they show that there is still a long way to go in obtaining complete government transparency and access. Libraries digitizing their collections of court briefs may be a step forward in attaining this goal.

In 2010, Byrne addressed the dichotomy created by the need to make public records available to the public online and the privacy concerns this raises for the

government. The study first examines two state public record laws, including access provisions. Then findings are presented from a case study of local government websites from those two states. Government documents available freely online were studied to determine what sensitive information they contained. The study found there was little consistency in the redaction of sensitive information, but there was not a lot of high-level sensitive information present in the documents. However, the study also concluded that the type of records studied weren't the type to have high-level sensitive information to begin with. Court briefs contain a great deal of highly sensitive information, raising the privacy concerns and the need for strict redaction guidelines.

These studies show that the value of digitizing court briefs extends beyond value to the legal community to value to the public. The solution to the problem of hard copy briefs and a demand for transparency has begun to be addressed through digitization of court records. As holders of court briefs, libraries are potentially in the position to take on this project. However, with digitization comes a host of privacy and implementation problems. This has briefly been addressed specifically in relation to putting appellate briefs on the web. Whiteman's 2005 study again begins by emphasizing the value of court briefs to lawyers and scholars. He addresses the increased value in court briefs if they were transitioned to the web and addresses the point that it is no longer "if" court documents should be digitized but "how". While he mentions that librarians are a part of this issue, his focus is on courts transitioning to an e-filing system for court records. He studied and compiled all the various court rules for e-filing and what websites currently offer online access to court briefs. He also studied concerns with this implementation, which come in the form of preservation, privacy, and copyright.

While Whiteman briefly addresses digitization concerns specifically in relation to putting court briefs on the Internet, the majority of the information on these problems has been studied in relation to digitization of court records through state government transitions to e-filing. Numerous studies have been conducted on the privacy issues raised by the various state court methods for an e-filing system that uses all digital documents. While this research is not directed at libraries, these studies inform libraries, because e-filing poses the same privacy issues as digitizing hard copies.

In 2004, Blankley studied the recent movement of courts putting their public records on the Internet in searchable databases. She explains the concept of “practical obscurity” that previously protected sensitive material contained in court documents. While court documents have always been legally available to the public, privacy was not a great concern, because so many roadblocks existed to accessing the print material in hard copy format, creating limited exposure to the content. There also wasn’t the danger of easily aggregating information when it is only contained in hard copy format. In light of her identified privacy danger in digitizing and putting court documents on the web, Blankley studied three court system policies: Ohio, Florida, and Maryland. Her study concludes that legislatures should enact statutes that prohibit the online publication of personally identifying information and that the burden should be on lawyers to remove sensitive information in preparation for court documents to be published online. In the library environment, the parallel burden would be on librarians to redact the personally identifying information in the documents before putting them on the Internet.

Jones’s 2013 assessment of the Florida court system studies the ongoing development of court rules and policies for public access to electronic court records. It

reviews common approaches for providing and limiting access and determines that these approaches do not adequately address privacy concerns. Jones conducted a case study of the Florida court's approach to electronic records, an approach which seeks to minimize the inclusion of personal information in court documents. The article rules out redaction as an option for courts, but only because of the logistical, cost, and feasibility problems of having state courts responsible for redaction. As a solution to privacy concerns, it is considered a good option. As will be investigated in this study, libraries may be better able to overcome the many logistical problems with redaction through factors such as flexible funding and flexible timelines.

In 2012, Nissenbaum and her fellow authors studied the transition from a paper-based system of processing and keeping court records to a digital records system for court records. The article includes an empirical study of how the digitization of court records and placing them on the Internet affects the flow of personal information contained in the records. The study searched court records using two systems that provided online access to court records: PACER and Google Scholar. The study also conducted searches at two physical courthouses: the Superior Court County Clerk's office in Trenton, New Jersey, and the Superior Court of New Jersey, Hudson Vicinage. The findings support the conclusion that costs of retrieving various types of personal information differ significantly between the online and physical access systems. Cost differences include getting to the location of the information system, the query interface and indexing mechanism, linking multiple information sources, access restrictions, format of records and human factors. One information source did not overcome the other, although the article concluded that practical obscurity prevailed at the courthouse. The comparison of

the two systems informs on privacy and access concerns for making the decision to transition to digital records.

Sudbeck's 2006 article presents an extensive study of state court electronic access policies in an effort to recommend a proposal for South Dakota's courts. The study focuses on case law governing public access and privacy interests involving court records. A survey of South Dakota clerks of courts, its citizens, and the attorneys practicing law in South Dakota was used to collect information on the current access to court services such as public record access and preferred methods of accessing these services in the future. A survey was also conducted of the members of the Conference of State Court Administrators to collect data regarding their states' electronic access policies and the processes used to develop those policies. Additional research included an in-depth study of the Minnesota, Florida, Maryland, and New York electronic access policies and processes. Interviews were conducted with people in those states who were part of the development of those access policies, which is similar to the interviews which will be conducted for the proposed research. Given the research and analysis conducted by Sudbeck, she concluded that the best way to respond to the issue of balancing judicial accountability with public trust and confidence regarding electronic access to court records is to: limit information in the case record, vary levels of access for different users, limit internet access to court-generated documents, provide access to case records at public terminals in the clerks' offices, and provide electronic storage of case records.

Many of these studies have supported the conclusion that digitization is a realistic and ideal option, and although there are privacy concerns, it is a transition that is occurring in spite of them. The conversation has turned from whether to digitize to how

to digitize. The next step in the conversation is redaction. As has been discussed, court policies on redaction have already been undertaken. However, now libraries may want to digitize their collections, but they have to consider all of the privacy, ethical, and practical concerns that have been raised by the courts. What is the law governing libraries' interested in redacting court documents? What are their specific concerns? Are there guidelines for how a library should proceed with the large-scale redaction of digitized court documents? This study addresses that void in the literature by discussing the specific landscape faced by the UNC Law Library's effort to determine whether they could redact their digitized North Carolina Supreme Court briefs and how they are attempting to accomplish that redaction.

Methodology

Research Design

This study will use qualitative research to accomplish a case study. Qualitative research usually involves a researcher collecting data in the field (Creswell, 2009). A case study is a type of qualitative study "in which the researcher explores in depth a program, event, activity process or one or more individuals". A case study typically focuses on a single concept, makes interpretations of data, and creates an agenda for change. It can employ a variety of data collection procedures, such as open-ended questions or evaluation of text data.

This case study will focus on the redaction process undertaken by the University of North Carolina Kathrine R. Everett Law Library as they digitize their collection of North Carolina Supreme Court briefs. Qualitative research on this particular process will allow the researcher to present a full description of how and why this process occurred.

While describing the reasoning and process uncovered in the research, discussion and interpretation of the results will expand upon the research to form a complete picture of the situation.

The Kathrine R. Everett Law Library is a public, academic law library located at the UNC School of Law in Chapel Hill, North Carolina. The library houses a collection of over 300,000 volumes of law and law-related materials as well as access to a number of legal subscription databases. The library supports the UNC Law School community and the larger University of North Carolina-Chapel Hill community. The law school has about 750 enrolled students every year and over 70 faculty members. As a public library, it is also open to lawyers and other members of the public.

The setting for the case study is the University of North Carolina Kathrine R. Everett Law Library because this is the library that is undertaking the project of digitizing and redacting its collection of North Carolina Supreme Court briefs. The Law Library has a staff that includes 16 librarians, divided into departments including administration, public services and collection services. Of those working at the library, five librarians from administration and collection services were involved in the decision to digitize and redact the collection of North Carolina Supreme Court briefs and/or were involved in the process created following the decision. Those librarians were Anne Klinefelter, Director of the Law Library, Sara Sampson, Deputy Director of the Library, Steven Melamut, Information Technology Services Librarian, Julie Kimbrough, Assistant Director for Collection Services, and Donna Nixon, Electronic Resources Librarian. Anne Gilliland, UNC Scholarly Communications Officer, was also consulted on the decision.

Data Collection Methods

The data collection methods for this case study were comprised of two parts: observation and interviews. The events observed were the actions involved in the physical process of redacting briefs. The process itself was observed to create a detailed guide on every step involved. The actors interviewed included the UNC law librarians and other UNC employees who either took part in the decision to redact the digitized collection of briefs and/or helped to design the digitization and redaction process and procedures.

Observation was chosen as the method for this study because it allowed the researcher to have first-hand experience with the process, leading to a more detailed report. The researcher can record information as it occurs and note any unusual or important aspects during the observation. Observation allows the researcher to act as both participant and observer, which was the situation in this case study as the researcher is one of the people involved in carrying out the redaction process for the Law Library. However, there are also some limitations to participant observation. The researcher may observe private information that cannot be reported and the researcher may seem intrusive due to involvement in many aspects of the studied process.

Observation protocol was maintained by recording information as it is being observed. Descriptive notes were kept detailing the physical setting, particular events, and participants. Reflective notes were kept detailing the researcher's thoughts, feelings, problems, impressions, and possible biases.

While observation was necessary to create a detailed description of the redaction process, it did not allow for any insight into the reasoning behind the Law Library's decision to take on the redaction process. Interview was chosen as a method for this study

because it allowed the researcher to gain insight into the reasoning for decisions made related to the studied project. As discussed in the background, redaction is an unsettled area of law and as will be shown through a detailed description of the redaction process, digital redaction of a whole collection of court documents is a very large undertaking. The interviews helped inform others on why they may also choose to undertake a similar project. Interviews with those involved in the fulfillment of the project gave a fuller picture of how the day-to-day operations of the project ran.

The interview included four questions. The first question asked the interviewee to explain their role in the project, the next two asked the interviewee for their understanding of the reasoning behind the project, and a final follow-up question was asked to determine if they had any final comments. The interview questions for the Processing Assistant were modified slightly to take into consideration the fact that he was only involved in digitization and not redaction. The interviews were all digitally recorded with the permission of the interviewee. Additionally, hand-written notes were taken during the interview. Interviews took between 10 and 30 minutes each.

Interview subjects included those who participated in the digitization and redaction decision and procedures for the Law Library: Anne Klinefelter, Director of the Law Library, Steven Melamut, Information Technology Services Librarian, Julie Kimbrough, Assistant Director for Collection Services, Donna Nixon, Electronic Resources Librarian, Jesse Griffin, Processing Assistant, and Anne Gilliland, UNC Scholarly Communications Officer.

Data Analysis

Analysis of the observation data involved an ongoing process of reflection. At the end of the research, the results of the observed redaction procedures were compiled into charts reflecting the redaction that occurred. Analysis of the interview data took place once all of the interviews had been conducted. All responses to the questions were compared to create a complete picture of the institution's reasoning.

Expected Benefits and Summary

Libraries of all types throughout the country are in possession of hard copies of court documents and may be interested in digitizing these documents. Digital redaction is an issue the library would face in this situation and it is an issue without clear guidance from the law. Digital redaction is also a complicated undertaking in the technical sense with limited guidance on how the actual process would function in a library. This study will explain the specific legal, ethical, and practical issues faced by the UNC Law Library so that other libraries will have an idea of what sort of issues they may face. A library will also be able to read through the reasoning of a law library in coming to their decision to redact, obtained through interviews with the librarians in charge of that decision at the Law Library. This study will explain how the digital redaction process can be undertaken by a library with an instructional guide for the process, produced through observation of the Kathrine R. Everett Law Library's experience digitally redacting its collection of North Carolina Supreme Court briefs. With this study, libraries can be more informed on digital redaction for court documents and able to make an informed decision on what is best for their own library if the issue of digital redaction arises. In the case that a library chooses to redact their collection of court documents, this study can provide some

guidance on how to begin undertaking that process. It also stands as an example of a unique project being undertaken by the Law Library.

Results and Discussion

In 2012, the UNC Law Library received \$85,000 in one-time funding from the University of North Carolina to begin the project to digitize its print collection of North Carolina Supreme Court briefs. The Law Library wanted to digitize this collection to make these documents widely and easily available to the public, based on historical requests for access, especially from law firms in the state of North Carolina. The Law Library decided to digitize just the court briefs and not the records. First, this cut down on the amount of digitization required. Second, the briefs hold information more valuable to lawyers, clerks, judges and the public. Finally, it was predicted that there would be fewer practical issues with digitizing the briefs, which are more standardized, both in content and format. The print collection of briefs consists of selected documents ranging from 1873 through 2000.¹ After 2000, the North Carolina courts switched to e-filing and print copies were no longer acquired by the Law Library. This meant the Law Library was taking on the large-scale digitization process of converting approximately 100 years of printed briefs into digital files.

As the Law Library does not have the equipment or the staff to digitize these documents, they contracted with legal publisher Hein to conduct the digitization, as Hein had experience in digitizing library records and offered a range of services related to digitization. The Law Library staff prepares one reporter case volume at a time and sends the original copy to Hein for digitization, then Hein returns the original copy to the Law

¹ The Law Library is not in possession of every brief during this time period. The Law Library does not know why they are not in possession of all case files.

Library along with the digital copy. The digital copies are transferred to the Law Library's servers and the original copies are returned to the print collection. Under the terms of service the Law Library has with Hein, Hein does not keep a digital copy of the briefs. Hein takes approximately six to eight weeks to digitize each batch of print materials sent to them. As of March 2014, digitization had been occurring for approximately two years and approximately 15 years of briefs had been digitized.

Digitization projects by libraries are generally not unique. What makes this digitization project unique is that it is a mass-digitization project of public records containing sensitive information conducted by a public library. These factors combined to create a discussion of whether redaction of sensitive information was legally required or ethically required by the library before these digitized documents could be made publically available on the Internet. The Law Library concluded that they were going to take on the project of redacting sensitive information from the newly digitized collection. This decision involved consideration of the law, ethics, and practical matters. Additionally, there are arguments by certain groups against the redaction of records. The Law Library weighed all of these factors in their decision. Once the decision was made, more discussion surrounded how the redaction process would be conducted until procedures were ultimately developed.

The results of the research conducted on this case are divided into six sections. Section One discusses the law potentially implicated by the redaction, including constitutional, tort, federal, and state law. Section Two discusses the ethical concerns raised by the redaction in the face of a legal landscape that doesn't require redaction. Section Three addresses the practical issues faced by the library in taking on the redaction

process. Section Four considers the concerns raised by groups traditionally against redaction and the Law Library's response to those concerns. After a summary explanation of the Law Library's decision to redact, Section Five discusses the planning stage for redaction, while Section Six explains the actual procedure implemented for redaction at the Law Library.

Legal Considerations

The legal issues faced by the Law Library when they were considering redaction can be divided into four areas of privacy law: constitutional law, tort law, federal statutory law, and state statutory law. In analyzing each potentially implicated area of law, the Law Library had to consider whether the law applied to the library as a public institution, whether the law applied to court records, and whether the law applied to the action of digitizing court records. The issue was whether the particular law or area of law either required the Law Library to redact the documents or required the Law Library not to redact the documents. Part of the analysis involved determining how the laws interact with each other. There was initial concern that the Law Library could be in an impossible situation: that there may be law preventing them from putting the digitized content on the Internet without redacting and that there may be law preventing them from putting the digitized content on the Internet with redactions. The Law Library ultimately concluded that the various areas of law that would either force them to redact or prevent them from redacting did not apply to them under the circumstances. They believe they can legally, in good faith, redact the digitized briefs and eventually post them on the Internet.

Constitutional Law. The area of privacy law has constitutional roots. A general right to privacy has been read into the constitution through the First Amendment (U. S.

Constitution) and the Fourth Amendment (U. S. Constitution). Case law has also supported a right to privacy in various aspects of people's lives throughout the years.² A more narrow constitutional right to privacy is thought to be at issue in the particular instance of publishing these digitized briefs on the Internet and that is a constitutional right to informational privacy. Alan Westin, a respected privacy scholar, explained a right to informational privacy in the following way: "Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967). This right is then potentially implicated in the present case as the briefs contain a great deal of information about the lives of individuals. The right of informational privacy advocates for the right of those people to determine the appropriate dissemination of that information to others. In the case of these briefs, the people discussed within should be able to decide how information about themselves may be made available to the public, whether in print or digital format. They should also be able to decide how much information is made available to public, or in other words, whether they want certain information redacted before it is made available.

The courts have addressed this right to informational privacy. One of the major cases on the topic is *Whalen v. Roe* (1977), in which a New York state statute was challenged for allowing the state to collect and store the name and address of any person receiving a certain drug prescription. The statute was challenged on the grounds that it violated the constitutional right to informational privacy. While the court did not strike

² For additional United Supreme Court cases that have addressed a constitutional right of privacy, see *Griswold v. Connecticut*, 381 U.S. 479 (1965), *Roe v. Wade*, 410 U.S. 113 (1973), *Nixon v. Administrator of General Services*, 433 U.S. 425 (1977), and *Bowers v. Hardwick*, 478 U.S. 186 (1986).

down the statute, they did include a discussion in their opinion on privacy rights. The Court's analysis included recognition of "at least two different kinds of [privacy] interests. One is the individual interest in avoiding disclosure of personal matters, and another is the interest in independence in making certain kinds of important decisions" (Whalen v. Roe, 1977). The Court is here confirming that informational privacy can be a legitimate constitutional interest, even if they did not feel it had been violated in the particular circumstances of the case.

While the Law Library does not at all discount a person's constitutional right of informational privacy,³ they consider it a weak claim to be made against their project. Informational privacy would seem to come in more as argument against digitization, than redaction. However, in either case, the information is already available to public in the form of a printed brief. There is an argument to be made that the "how" is changed when going from a print to digital copy, but the information in that brief is not changing when it is digitized. That same brief is currently available to the public.

Tort Law. Tort law, which is a civil area of law where a wronged party can make a legal claim based on harm or injury, also includes areas of privacy law in the form of four privacy torts. The four privacy torts are public disclosure of private fact, false light, intrusion upon seclusion, and misappropriation. Under the public disclosure of private fact tort, a plaintiff must establish that private facts a reasonable person would find offensive were disclosed to the public. A plaintiff can make a claim under false light when they accuse someone of publicly spreading false information about the plaintiff. An intrusion upon seclusion claim may arise when someone intentionally intrudes on the

³ Instead, it is an interest in privacy that put the Law Library in the situation addressed in this case study.

private affairs of another person, causing mental pain or suffering. Misappropriation occurs when someone uses the name and likeness of another person for commercial purposes without their consent.

The exact elements of each tort claim vary by state. The Law Library was particularly concerned with the privacy tort law in North Carolina. In North Carolina, the state does not recognize a tort claim for public disclosure of private fact or false light (Daye & Morris, 2012). That leaves intrusion upon seclusion and misappropriation as the two potential privacy tort claims, both of which the Law Library concluded were weak claims under the circumstances of the digitization project. For example, as far as the intrusion upon seclusion tort is concerned, the Law Library is not intentionally trying to intrude on anyone's privacy. In the case of misappropriation, there aren't any commercial purposes to making the cases available on the Internet, as they will be freely available to the public. Additionally, as was argued in the case of the right of informational privacy, all of this information is currently publicly available in print format.

Federal Law. The Law Library had to consider whether any federal law governed redaction of the digitized collection. While there are federal laws that consider privacy, none of them governed this process. For example, HIPPA is a federal law concerned with health insurance and national health care standards (The Health Insurance Portability and Accountability Act of 1996). Title II of HIPPA governs standards of security and privacy for health data. However, these standards apply to specific covered entities and business associates of covered entities, of which the Law Library is not.⁴ FERPA is another federal

⁴ Covered entities are defined in the HIPAA rules as (1) health plans, (2) health care clearinghouses, and (3) health care providers who electronically transmit any health information in connection with transactions for which HHS has adopted standards. A business associate is a

law concerned with privacy (The Family Educational Rights and Privacy Act of 1974). This law governs privacy rights relating to children's educational records. In this case, the briefs are not "education records" of the type FERPA is meant to protect.⁵ While the Law Library took a range of other federal privacy statutes into consideration, none of these were found to apply to the situation they were encountering.

State Law. The laws that were most concerning for the redaction process and seemed to have earned the closest attention by the Law Library were some potentially applicable North Carolina General Statutes. One of the most basic concerns faced by the Law Library was whether they were legally unable to redact the documents because they are public records under North Carolina law. Under North Carolina public record law, a public record includes

all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, artifacts, or other documentary material, regardless of physical form or characteristics, made or received pursuant to law or ordinance in connection with the transaction of public business by any agency of North Carolina government or its subdivisions. (N. C. Gen. Stat. § 132-1(a))

North Carolina Supreme Court briefs are documents made pursuant to law in connection with the North Carolina courts, an agency of North Carolina government. The briefs are also arguably records of the UNC Law Library, as they are held by the library "pursuant to law or ordinance in connection with the transaction or public business". As these documents are public records, they are "the property of the people. Therefore, it is the policy of this State that the people may obtain copies of their public records and public information free or at minimal cost unless otherwise specifically provided by law" (N. C.

person or entity conducting certain functions on behalf of a covered entity (The Health Insurance Portability and Accountability Act of 1996).

⁵ The full definition of what constitutes an "educational record" under FERPA is available at 20 U.S.C. § 1232g(4)(B).

Gen. Stat. § 132-1(b)). In other words, as these briefs are public records, they belong to the people and should be openly available to the people. The issue then faced by the Law Library was whether they could redact these documents before making them publicly available to the people. Does the redaction interfere with the right of the people to obtain copies of their public records? Does it interfere with the theory of government transparency behind such a law?⁶

The Law Library ultimately determined that the redaction of the digitized copies does not violate public record law. First and foremost, the un-redacted copies are still available to the public in the courts, which have a more central obligation to provide access to the public. Second, the Law Library continues to provide access to its collection of un-redacted print copies, at least while shelf space remains available for it. Finally, the procedures for courts e-filing in the North Carolina court system calls for redaction of court documents.⁷ This means that the courts are already currently redacting public records in the case of sensitive information.

The other side of the issue required researching whether the Law Library was legally required to redact the court documents before making them available on the

⁶ Also to be considered are provisions in public record law against destroying public records (N.C. Gen. Stat. § 132-3). While it may seem obvious that redaction is not the same thing as destruction, this type of provision shows an interest by the government in protecting government records from harm.

⁷ These rules can be found in the Supplemental Rules of Practice and Procedure for the North Carolina eFiling Pilot Project. Rule 6.3 on Private Information states: “Except where otherwise expressly required by law, filers must comply with G.S. 132-1.10(d) to exclude or partially describe sensitive, personal or identifying information such as any social security, employer taxpayer identification, drivers license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code or passwords from documents filed with the court. In addition, minors may be identified by initials, and, unless otherwise required by law, social security numbers may be identified by the last four numbers. It is the sole responsibility of the filer to omit or redact non-public and unneeded sensitive information within a document. The clerk of superior court will not review any document to determine whether it includes personal information.” These rules are available at <https://www.efiling.nccourts.org/manual/fiCourtRules.htm>.

Internet. This primarily involved studying two North Carolina laws concerned with the privacy of North Carolinian's personal information: the social security number provision in the public record law and the Identity Theft Protection Act. Both of these laws are concerned with protecting people from the financial harm that could result from the release of sensitive personal information.

North Carolina's public record law includes a section on "social security numbers and other personal identifying information", in which the state recognizes the harm that comes with the release of a person's social security number. The law states "when State and local government agencies possess social security numbers or other personal identifying information, the governments should minimize the instances this information is disseminated either internally within the government or externally with the general public" (N. C. Gen. Stat. §132-1.10(a)(3)). It further sets out special rules on how the government may collect and use social security numbers, specifically stating that "no agency of the State or its political subdivisions...shall do any of the following...(5) Intentionally communicate or otherwise make available to the general public a person's social security number or other identifying information" (N. C. Gen. Stat. §132-1.10(b)(5)).⁸ The Law Library, as a public library, is an agency of the State and would seem to fall under the provisions of this statute. In which case, making the digitized collections available on the Internet with social security numbers un-redacted would be

⁸ Identifying information under this section includes: "(1) Social security or employer taxpayer identification numbers. (2) Drivers license, State identification card, or passport numbers. (3) Checking account numbers. (4) Savings account numbers. (5) Credit card numbers. (6) Debit card numbers. (7) Personal Identification (PIN) Code... (8) Electronic identification numbers, electronic mail names or addresses, Internet account numbers, or Internet identification names. (9) Digital Signatures. (10) Any other numbers or information that can be used to access a person's financial resources. (11) Biometric data. (12) Fingerprints. (13) Passwords. (14) Parent's legal surname prior to marriage" (N. C. Gen. Stat. 14-113.20(b)).

against this law as “intentionally communicating” the information. However, the statute includes the following exception. The provision does not apply “to any document filed in the official records of the court” (N. C. Gen. Stat. §132-1.10(c)(7)). All of the briefs were filed with the court, so they fall under the exception to the rule. Under this section, the Law Library does not appear to be legally required to redact the documents. Also of interest in this law is subsection (f1), which states:

Without a request made pursuant to subsection (f) of this section, a register of deeds or clerk of court may remove from an image or copy of an official record placed on a register of deeds' or clerk of court's Internet Web site available to the general public, or placed on an Internet Web site available to the general public used by a register of deeds or clerk of court to display public records, a person's social security or drivers license number contained in that official record. Registers of deeds and clerks of court may apply optical character recognition technology or other reasonably available technology to official records placed on Internet Web sites available to the general public in order to, in good faith, identify and redact social security and drivers license numbers. (N. C. Gen. Stat. 132-1.10(f1))

This section gives the Court permission to redact the public record before making it publicly available on the Internet. This offers support for the Law Library not being liable under public records law for redacting sensitive information from the briefs before making them publicly available on the Internet.

North Carolina statutes also include an Identity Theft Protection Act, which contains a provision protecting the public from security breaches. Under this law,

[a]ny business that maintains or possesses records or data containing personal information of residents of North Carolina that the business does not own or license, or any business that conducts business in North Carolina that maintains or possesses records or data containing personal information that the business does not own or license shall notify the owner or licensee of the information of any security breach immediately following discovery of the breach, consistent with the legitimate needs of law enforcement as provided in subsection (c) of this section. (Identity Theft Protection Act)

There was concern that the Law Library, as a possessor of records containing the personal information of North Carolina residents, would be liable for a breach of data security if they were to post the records in the form of the briefs online. Would the Law Library than be required to track down and notify every person in the briefs of the “breach”? However, it was concluded that once again the law does not apply to the Law Library, as the Act further states that “[b]usiness shall not include any government or governmental subdivision or agency” (Identity Theft Protection Act). As a government agency, the Law Library does not fall under the definition of a “business” governed by the Act. The Act also states that

[i]t shall be a violation of this section for any person to knowingly broadcast or publish to the public on radio, television, cable television, in a writing of any kind, or on the Internet, the personal information of another with actual knowledge that the person whose personal information is disclosed has previously objected to any such disclosure. (N.C. Gen. Stat. § 75-66(a))

Under this provision, there is also an issue with the phrase “actual knowledge” and what sort of knowledge it covers. The Law Library has no actual knowledge that any person has previously objected to the dissemination of these briefs online. Once again, the Law Library is not legally required to redact the digitized collection before putting it on the Internet.

Ethical Considerations

The Law Library took constitutional, tort, federal statutory, and state statutory privacy law into account when making their redaction decision. While they did encounter laws that required analysis to determine whether they would govern their redaction decision, they ultimately decided that there wasn’t law directly applicable to their unique position. Consideration of all of these laws contributed to the library concluding that they

were not legally prevented from redacting and that they were not legally required to redact. This meant the redaction decision became a value judgment. The Law Library had to ask itself why it would want to redact outside of legal requirements. Commercial publishers that have made briefs submitted to the courts available in databases have not added redactions, and they have done this without liability. By deciding to redact, the Law Library is stepping outside of the norm and their reasons for doing so can largely be found in these ethical considerations.

Information Protected by Laws. While the laws that were discussed in the previous section of the paper do not require redaction by the Law Library, they do indicate types of potentially private information that the government has made some effort to protect. Personally identifying information, children, and rape victims are all statutorily protected under different circumstances of state and federal law. The personally identifying information is protected to prevent a financial harm to the people attached to the information, whereas the children and rape victims are protected to prevent a dignity harm.

The North Carolina Identity Theft Act protects a whole range of what it terms “identifying information”, such as social security numbers, driver’s license numbers, and credit card numbers.⁹ The provision in the public records law on social security numbers protects the same “identifying information”. This shows that both the government and businesses are tasked with protecting the privacy of this information. It would seem that the state of North Carolina has placed value on keeping this information private, even if a public library does not have a specific statutory requirement to protect it in the form of court records the way other government agencies and businesses have to protect the

⁹ For a full list of “identifying information” under the act, see Footnote 8.

information under other circumstances. This same law can also show what sort of personal information has been given less consideration of being the type of information that should be kept confidential. The Identity Theft Act, in defining “personal information”, also defines what it does not consider “personal information”: “Personal information does not include publicly available directories containing information an individual has voluntarily consented to have publicly disseminated or listed, including name, address, and telephone numbers...” Names, addresses, and telephone numbers do not rank as high a priority as social security numbers.

While names are generally not the type of information that is considered highly personal, there are certain circumstances where a name is considered particularly sensitive. Children are a group that society has determined generally require a greater deal of sensitivity and protection. A statutory example of this is the federal Child Online Privacy Protection Act, which governs the online collection of personal information about children under the age of 13 years and protects the confidentiality of the information gathered. This act demonstrates that at a federal level, the government is concerned with the privacy of children. Arguments have also been made for protecting children’s names in press coverage when they have been victimized, as it has been concluded that publication of the children’s names can have negative mental health ramifications for the child (Jones, Finkelhor & Beckwith, 2010). A court brief about a case dealing with a child victim often includes both the name of the victim and a great deal of detail about the crime the child was involved in. If the child’s name is protected from the press coverage for ethical reasons and concerns, it would seem the same may translate to the online publication of a brief by a law library.

Rape victims are another group that has gained varying levels of statutory protection of their privacy, in part through the restriction of the publication of their names by the media. Laws that contain such provisions as well as other rules regarding protection for rape victims are termed rape shield laws, and they exist at both the federal and state level.¹⁰ However, many of the provisions restricting what the media can print have been challenged and ultimately been struck down by the courts as unconstitutional.¹¹ As a result, it is often considered more of a professional ethic maintained by journalists not to print the name of rape victims, just as they often don't print the names of child victims.¹² As children and rape victims are the named subjects of some of the North Carolina Supreme Court briefs, the Law Library had to weigh the privacy ethics of redacting those names in the briefs with the value of having a completely whole record available on the Internet.

Censorship. Another ethical issue that the Law Library had to consider was whether redaction sounds too much like censorship. The American Library Association (ALA) defines censorship as “the suppression of ideas and information that certain persons—individuals, groups or government officials—find objectionable or dangerous”. Is redacting sensitive personal information a suppression of the type to be labeled censorship? Librarians have their own ethical code, which places a high value on fighting censorship. The ALA Code of Ethics states, “We uphold the principles of intellectual

¹⁰ The federal Rape Shield Law is contained in the Violence Against Women Act of 1994. North Carolina's Rape Shield Law can be found in N. C. Gen. Stat. § 8C-412.

¹¹ Examples of cases on this issue include *Florida Star v. B.J. F.*, 491 U.S. 524 (1989) and *Cox Broadcasting Corporation v. Cohn*, 420 U.S. 469 (1975).

¹² For example, the Society of Professional Journalists includes the following statement in their Code of Ethics: “Journalists should: Show compassion for those who may be affected adversely by news coverage. Use special sensitivity when dealing with children and inexperienced sources or subjects...Be cautious about identifying juvenile suspects or victims of sex crimes”.

freedom and resist all efforts to censor library resources.” The Law Library had to determine whether the redaction of these court briefs was censorship or an act of content privacy.

Also part of the Code of Ethics is a provision protecting privacy: “We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted.” This provision specifically deals with protecting patron privacy, but it illustrates that privacy and confidentiality are strong values in the profession, just like fighting censorship.

Knowledge and Predictability. Another ethical issue that had to be considered was the knowledge level of the people with information contained in the digitized briefs. All of the briefs being digitized are from the year 2000 and earlier. Could people before 2000 have been able to predict what the availability of records would look like in 2014 and beyond? Could they have foreseen that their court records would be widely available in a forum like the Internet? Or were they counting on the practical obscurity of court records to maintain a level of privacy about the proceedings. Could they have anticipated that data elements such as social security numbers would have such a high risk for financial harm such as identity theft? The Law Library believes that these people couldn't have foreseen this use of the briefs or the resulting financial or dignity risks, and instead they would have expected the barriers to access to remain throughout the years.

Considering this issue in determining whether to redact is seen reflected in the law, even if that law doesn't apply to the Law Library. The public record provision on social security numbers stresses that no agency shall “[f]ail, when collecting a social security number from an individual, to provide, at the time of or prior to the actual

collection of the social security number by that agency, that individual, upon request, with a statement of the purpose or purposes for which the social security number is being collected and used” (N. C. Gen. Stat. § 132-1.10(b)(3)). Additionally, no agency shall “[u]se the social security number for any purpose other than the purpose stated” (N. C. Gen. Stat. § 132-1.10(b)(4)). This reflects an interest by the courts that people be aware of how and for what purpose their personally identifying information is being used. The North Carolina Identity Theft Protection Act has a similar concern seen reflected in the provision dealing with actual knowledge discussed earlier:

It shall be a violation of this section for any person to knowingly broadcast or publish to the public on radio, television, cable television, in a writing of any kind, or on the Internet, the personal information of another with actual knowledge that the person whose personal information is disclosed has previously objected to any such disclosure. (N.C. Gen. Stat. § 75-66(a))

The government wants people not only to know how their information is being used, but also to confirm that there is actual knowledge of the use. Knowledge of the exact use of their information is something the people featured in the briefs before 2000 do not possess, even if the government has made an exception for court documents from the provisions on this issue.

Practical Considerations

The complicated legal and ethical issues aside, the Law Library had to look at the practicalities of taking on a redaction project of this magnitude. With over 100 volumes of briefs, even if there was a minimum amount of information to be redacted, it was still going to be a large-scale project. Time seemed to be the factor that was most on the mind of the Law Library. No matter what the procedure ended up being if they decided to redact, it was going to be a time-consuming and slow process. It also had the potential to

be an expensive process. While the Law Library received money for the initial digitization, the redaction was going to require some sort of additional staff commitment to complete.

The Law Library also had to consider the fact that they would be forging new territory with this project. There were no standard procedures to guide their redaction efforts. Planning would be necessary to create a feasible process. Even with a great deal of planning and research into redaction strategies, the Law Library predicted that it would still be an imperfect process. The process had the potential to need modifications along the way and still it could have errors. Additionally, one of the imperfections could be the private information that slips through the cracks of the redaction process. Once the Law Library was looking at the briefs with redaction in mind, the library staff saw many ways private information could be in the documents and even remain in documents after the redaction efforts.

Another factor to consider is that all of these records are currently freely available in the building. This has been a counter argument to many of the arguments raised against redacting. Yes, these documents are being redacted, but the originals are still available in the building for anyone who would like to see them. However, the fact that the originals are in the building can also be a practical argument against redacting. Why would the library redact all of these documents in the face of concerns over time, money, and imperfections of process, when the documents are all available in the library?

This is not to say that the Law Library didn't have practical considerations in their favor. The biggest seemed to be that there was no definite timeline for this project. This project didn't have a date by which time the briefs had to be made available on the

Internet, so there was no deadline for when all the redaction had to be completed. The Law Library could set its own markers for progress. Additionally, by taking on this project, the Law Library paved new territory for other libraries interested in taking on similar projects.

Viewpoints against Redaction

Journalists and archivists have both expressed concerns about similar redaction efforts. Journalists are very enthusiastic about moving as much information as possible to the Internet, as it strengthens and helps their ability to fact check and report. Their particular concern is with redacting identifiers, such as birth dates (Raggi, 2010). Without birthdates in court records, journalists are concerned about confirming the identity of the people they are reporting on in the briefs. The counter argument here seems to be that if a journalist is unable to figure out from the Internet document if they have identified the correct person, they can still track down the un-redacted hard copy of the brief from a court or library.

Archivists are interested in maintaining an accurate and complete record. They are concerned that redaction is a change to the record that could interfere with its historical value. It has been suggested that as a society we have no idea what could be of value in the future, so we may be redacting information that will be valuable in the future. The redaction could interfere with the ability to analyze the data in the briefs in the future for a public good or it could prevent advocacy for groups whose story is hidden. In the face of all these arguments against redaction, it can again be repeated that these documents are still available to the public un-redacted in print form.

The Final Decision

In considering the legal, ethical, and practical concerns raised by redaction, the Law Library believed they were faced with three options for moving forward: (1) not digitize, (2) redact, or (3) delay and continue to seek more clarity on choices related to redaction. In spite of the fact that there were no legal requirements for the Law Library to redact the sensitive information in the digitized briefs and that there would be some practical hurdles to redaction, the Law Library made a value judgment. Due to the ethical implications raised by the law and the overriding librarianship interest in privacy, the Law Library came to the conclusion that they should be redacting the briefs. Just because the law hadn't caught up to the situation the Law Library was facing didn't mean they shouldn't redact.

The decision to redact was just the beginning of a whole new round of discussion on what and how to redact. The law that had been so thoroughly studied as part of the legal consideration provided an ethical guide on what sort of information should be redacted from these documents. Based on society's concern for protecting children's privacy and rape shield laws, the Law Library decided to redact the names of children and rape victims from all of the briefs. Based on a government concern for protecting against financial harm, the Law Library decided to redact birth dates and social security numbers from all of the briefs. Credit card numbers were another area of concern, but they have not yet been added to the identified list of information to be redacted, largely because of the practical challenges in locating occasions of use in the briefs. Names, addresses and telephone numbers were also determined to not require redaction at the present time.

As part of the decision on what to redact, the Law Library had to consider when to redact. These briefs go back over 100 years, so the Law Library had to decide whether it wanted to redact that far back. In concluding that it would redact all of the digitized briefs, the Law Library felt that it came down to preventing a dignity harm. The harm from a brief from 1930 would not be financial, but people connected to that brief may still be alive and affected by it. Once all of these preliminary decisions were made, the Law Library was able to move onto the actual process of redacting the briefs.

The Redaction Process

After researching redaction options and strategies used on a smaller scale, the Law Library created a three-prong approach to identifying information in need of redaction: (1) use the BitCurator tool, (2) conduct adobe searches, and (3) conduct sample readings by librarians. Once sensitive information was identified, Adobe Acrobat Pro XI software was used to redact the PDF files of the digitized briefs. This section will explain the Law Library's experience with the BitCurator tool for identifying sensitive information, then explain the Adobe search strategy that was predominantly used to identify sensitive information. It will also explain the mechanics of redaction, including steps and procedures taken throughout the Adobe redaction process.

BitCurator. BitCurator is a suite of digital forensics and data analysis tools designed to help improve digital preservation efforts. It was developed in collaborative efforts by the School of Information and Library Science at the University of North Carolina at Chapel Hill and Maryland Institute for Technology in the Humanities at the University of Maryland. BitCurator can perform a number of functions including creating forensic disk images, analyzing files and file systems, extracting file system metadata,

and identifying and removing duplicate files. Its proposed function that was to be used by the Law Library was its ability to identify and redact sensitive information in digital files.

The Law Library developed a procedure for using BitCurator to quickly and easily identify sensitive information in the form of dates of birth and social security numbers, which would then be manually redacted. Faculty and graduate students at the School of Information and Library Science at the University of North Carolina at Chapel Hill provided the Law Library with the BitCurator program and support for its use. The program was run on all of the briefs that had been digitized at that time.

The BitCurator program generated several text files containing the results of its search for sensitive information. The files identified the sensitive information, what document it was in, and the text surrounding the sensitive information, as well as how many instances of the identified sensitive information were in the document. These files were used to create a spreadsheet listing the BitCurator identified items and whether the items were redacted. If a manual observation of the identified information proved to be something requiring redaction, the Adobe redaction procedures were followed to redact the information.

The BitCurator program only identified eight instances of sensitive information in three PDF documents among the hundreds of PDF documents searched. All of the identified sensitive information was the type the Law Library was interested in redacting. However, scans of the documents in the folders revealed additional sensitive information in the form of dates of birth and social security numbers that the BitCurator tool did not identify as the Law Library had hoped.

Adobe Searches. The next step in the redaction procedure was the ongoing process of conducting Adobe searches. Using Adobe Acrobat Pro XI¹³, the Law Library has developed a process for searching all the digitized briefs to identify sensitive information that needs to be redacted. The digitized briefs are organized on the Law Library's server in folders of between 400 and 1,400 briefs. Each folder includes the batch of digitized files that Hein sent to the Law Library. The batches are organized by the date the Law Library received the files from Hein. The files are in reverse chronological order as that is how the digitization has been occurring, starting with the most recent briefs and working back in time. The Adobe searches were conducted on one folder of digitized briefs at a time, starting with the first folder received from Hein.

Step One: Conduct Search

The first step in the redaction process was to select a folder of briefs and conduct a search on that folder. The Law Library has created three searches to run on each folder.¹⁴ All of the steps listed here were completed for one folder at a time, before moving on to the next the folder. To run a search in Adobe Acrobat Pro XI, the following steps were taken:

1. Open Adobe Acrobat Pro XI.

¹³ Adobe Acrobat Pro XI was selected as the redaction tool for its search features, saving options, and redaction function. Adobe Acrobat Reader can search the documents, but it cannot save the results of a search and it does not have the redaction function. Reader may allow you to black out text, but Pro's specific redaction tools are required to ensure the security of the redacted document.

¹⁴ The Law Library decided to run three different searches to identify sensitive information by type of information. The first search attempts to identify children and sexual assault victims with the terms "girl", "boy", "child", "children", "minor", "molest", "rape", "youth", and "student". The second search attempts to identify more possible children with the terms "daughter", "son", "father", "mother", and "born". The third search attempts to identify driver's license numbers and social security numbers with the terms "license", "DOB", "birth", "SS#", "social", and "security".

2. Choose Edit → Advanced Search.
3. Choose Show More Options at the bottom of the search screen that opens up.
4. Choose Match Any of the Words under “Return results containing:”¹⁵
5. Check the box next to Stemming.¹⁶
6. Choose a folder to search in the “Look in:” tab, then “Browse for Location”.
Select the folder.
7. Enter search terms.
8. Run search.

Step Two: Save Search

Once the search had completed running,¹⁷ the search was saved. An Adobe XI advanced search feature allows the user to save the batch or results of a search as a CSV file in Microsoft Excel. To save the search results, the “Save” option in the bottom corner of the search box was clicked. Then the prompted steps were followed to save the excel file and maintain the saved file as a record of the results of the folder search.

Step Three: Open Search Results

The completed Adobe XI search results consisted of a list of all the files in the folder with terms matching the search string. Clicking on a file expands to show all of the instances of the search terms located. Clicking on a specific instance of a located search

¹⁵ The third search for driver’s license numbers and social security numbers uses a specially designed “Boolean Query” instead, with the following search string: license OR DOB OR birth OR SS# OR social AND security. This was necessary to search for the two-word term “social security”.

¹⁶ Stemming ensures that search terms with variations of the words, such as “girls” or “molested” are also identified by Adobe XI.

¹⁷ The length of time it takes the search to run depends on the amount of files in the folder and the amount of results being located. Searches took on average between three and five minutes to complete.

term opens up the file directly to the instance selected. Starting at the beginning of the list of results, each file was individually opened in Adobe XI, one at a time.

Step Four: Redaction Determination

When one file was opened, the file was browsed to determine whether there was sensitive information of the type previously determined to require redaction. There were two ways the browsing generally occurred. The first was to go through the document reading pieces of the brief around the highlighted terms. Sometimes it was immediately obvious that the brief features a minor or rape victim. If nothing was immediately apparent, it was necessary to skim the first few pages of the brief to determine if the case was not about a child or sexual assault victim. Information located on dates of birth and social security numbers was generally more apparent and located near the highlighted search terms. This part of the process relied on the judgment of the searcher in ultimately making a decision about whether the document needed to be redacted. If the document did have sensitive information needing redaction, the next action taken was Step Five. If the document did not have sensitive information needing redaction, the next action taken was Step Six.

Step Five: Redact

Adobe XI's redaction tool is used for the actual redaction process. Before redacting a brief, the Law Library duplicated the file. This allowed the library to maintain an original, un-redacted digital copy of the brief. The duplicated file was saved with “_Redacted” added to the file name. The following steps were taken to redact the duplicated file:

1. Choose Tools → Protection.

2. Click Mark for Redaction.
3. Go through the PDF and highlight the text to be redacted.¹⁸
4. Click Apply Redactions.¹⁹
5. Click Sanitize Document → OK.²⁰
6. Type a name for the file and click Save.²¹

Step Six: Record

The final step was to record the action taken in relation to the search results. To record this information, a separate excel document was created. Each brief in the results list is recorded in one row of the excel document, along with whether the brief was redacted and what information was redacted from the brief.²² Steps three through six were repeated for every file in the results list of a folder search.

Step Seven: Confirmation

Once every file in a results list has been checked for sensitive information, redacted if necessary, and recorded in the Excel document, a confirmation step concluded the completed search. This involved comparing the number of files in the excel document with the number of files in the folder search results to make sure that this number was the same and ensure that no files were skipped. Additionally, a scan of the list of file titles in

¹⁸ After highlighting information to be redacted, a red box is placed around the information. By hovering over the red box, the user can see what it will look like blacked out to ensure all of the information is covered.

¹⁹ The user must make sure all of the sensitive information is highlighted before applying redactions, as this cannot be undone. However, additional text can be redacted after applying redactions. Sometimes after applying redaction, Adobe XI will ask if you want to remove any hidden information. You can remove hidden information at that step, or do it with the next step of sanitizing the document.

²⁰ This step removes any hidden information and metadata from the file to ensure security.

²¹ This prompt is automatic after sanitizing the document. You can save the document under its previous name and replace the original file with the sanitizing file.

²² Columns in the UNC Law Library excel spreadsheet were: file name, case title, page numbers, search terms located in brief, redacted (Y/N), and notes on what type of information was redacted.

the Excel document was conducted to ensure that files weren't listed twice. When this final step was complete, the process returned to step one to conduct a new search.

Reflecting on Adobe Searches. The redaction procedure had been completed for two folders over the course of observation.²³ Problems with the procedures have already been identified and adjustments were made throughout the process. For example, the original search for children and sexual assault victims did not include the word “children”. As stemming was checked, it was assumed that “children” would be in the search results due to the root term “child”, which was being searched. However, even though the program picked up terms like “licensing” for “license” and “molested” for “molest”, it did not pick up “children” for “child”. This prompted the Law Library to add this term to its search string for future searches.²⁴ Another search that was adjusted was the one aimed at identifying driver's license numbers, which originally only searched for the word “license”. It was expanded to include date of birth and social security number searches, since they were not all successfully identified by the BitCurator tool. Trial searches were run until a search string was developed that seemed to pick up the sensitive information the Law Library was looking to redact.

Another problem with the search procedure was one the Law Library doesn't have much control over. Adobe XI is not always as successful at searching pages that have a landscape layout instead of the traditional portrait layout. Some of the pages also have

²³ Two folders with over 1,000 documents combined were searched. It took approximately 6 months to complete the redaction procedures on the two folders, with one person working anywhere between 3 and 10 hours per week on redaction. The results of the redaction procedure on these two folders were analyzed and those results are described in the tables in Appendices B and C.

²⁴ A search for the word “children” was conducted on the folder to confirm that no additional files were returned than those returned for the previously run searches. There were no additional files. The word was still added to the list of search terms to aid in identifying sensitive information when looking at a brief.

print that is too small, smudged, or unclear for Adobe XI to recognize the words. Both the landscape pages and the unrecognizable words could include sensitive information. To help address this, a scan of documents that contained highlighted search terms can help identify pages with either of those characteristics. However, they can be hard to identify from a quick look. Since such small amount of pages meet this criteria, it was determined that this is not something that a great deal of labor should be devoted to at this time.

The issue of false positives was a concern noted while designing the searches, but it became even more apparent once searches had been run. Every search term was not expected to result in a file that needed to be redacted. However, some search terms ended up revealing even more false positives than originally considered. For example, “license” returns results any time a brief mentions that an attorney for the brief is licensed. The second search that is run looking for children includes terms like daughter, son, mother, and father. All of these terms can and do result in cases about children, but they are also accurate descriptors about people involved in a case that are all adults. The issue was that the majority of files returned with sensitive information about children had been identified by the previous search. This has led to discussion of eliminating these words from the current searches and creating only two searches to be run on each folder. This change would drastically cut down on the time devoted to each folder and cut down on another issue, which is documents that appear in multiple searches. Every time a file appears in the second search for children, a library staff member would have be

determined whether it was already redacted in the previous search and that the previous search found all the sensitive information.²⁵

The most time consuming part of the process, outside of the actual redaction, is recording the page numbers of each page instance where a search term appears in a brief in the excel document. When a brief is 300 pages and uses the word “child” on 200 of the pages, it requires a lot of time and attention to note each page number. This also seems like a lot of time devoted to a potentially unnecessary record, as the previously saved file of search results includes a list of every term and the page number. This has prompted discussion of either eliminating this column from the excel spreadsheet where the files are being recorded or using the saved file of search results to note which files are being redacted.

It should also be addressed that although searches were designed to identify particular types of sensitive data, if sensitive data of another type was observed, that data was redacted. For example, if a birth date was found in a file returned for the key word “rape”, that birth date was redacted at that time. If the name of a child was located in a file returned for the key word “born”, that child’s name was redacted from the document at that time. On a related note, if a file was redacted in an earlier search, but was again returned in a later search and additional sensitive information was located, the file was further redacted. For example, if the name of a rape victim was redacted from a file and the file was also a result of the third search for a social security number, the redacted file was opened and the social security number was additionally redacted. These

²⁵ The proposal is that moving forward each folder would be searched twice, instead of three times. The first search would include the terms “girl”, “boy”, “child”, “children”, “minor”, “molest”, “rape”, “youth” and “student”. The second search would include the terms “license”, “DOB”, “birth”, “born”, “SS#”, “social”, and “security”.

opportunities for additional redaction were available because library staff members were actively engaged in the redaction step.

Overall, while a redaction procedure was created, there were still revisions to the procedure made once enacted and more revisions were predicted for the future. As more redaction takes place, more issues were discovered. However, as more redaction takes place, the procedure was corrected and improved.

Future Steps. The final step in the redaction process will be to have librarians review briefs to potentially identify private information that was not redacted. This stage of the redaction process has not yet been reached. The Law Library has another future step in mind to safeguard against some of the imperfections of the redaction process. Once the Law Library has reached a point that they begin to place the digitized briefs on the Internet, they plan on including a message to the public. This message will invite people to let the library know if they find information the Law Library missed or information they want taken down.

At the very end of the research process, a new program was identified by the Law Library with the potential to be incredibly valuable to the redaction process. Identity Finder is a program designed to manage sensitive information and ultimately prevent data breaches by locating the personal information on a computer so that action can be taken to secure that information. Identity Finder has been marketed to prevent identity theft and data breach to various industries, including healthcare, banking, and higher education. The University of North Carolina at Chapel Hill recently acquired the program as part of an effort to locate social security numbers on campus computers. However, an

investigation into the product has revealed that it may be a valuable tool in locating sensitive private information in the digitized briefs.

Identity Finder allows the user to search specific file folders on a computer, such as a folder with the digitized briefs. Identity Finder can search many different file formats, including Adobe PDFs. The user can search for sensitive information in the form of social security numbers, credit card numbers, passwords, bank accounts, driver's licenses, dates of birth, and passport numbers all in one search. Rather than using key words to locate these numbers, it uses the patterns of these numbers to locate them within the documents. Identity Finder also has a separate keyword search function. When a search is conducted, the results of the search are displayed in on the screen in three panels: (1) a list of files in a Search Results Panel, (2) a Preview Pane giving a look at the sensitive information located, and (3) a Properties Pane giving details on the file. You can open an Adobe file directly from the search results, however you still have to go in and manually redact the sensitive information. Just as in Adobe, you have the option to save all of the search results in a CSV file.

Test searches run with the program have had very promising results, as evidenced by the amount of sensitive information they located in the digitized briefs. Since Identity Finder doesn't rely on key words surrounding sensitive numbers to locate those numbers, it seems like it could have greater success in that area, as the words surrounding a private number can vary a great deal. It also saves a great deal of time in relation to the many false positives that are found when searches are run for these numbers using key words. It seems that it could prove very similar in functionality to Adobe when it comes to the search for children and rape victims.

Conclusion

The UNC Law Library took on a large and ambitious project with their decision first to digitize their collection of North Carolina Supreme Court briefs and then to redact the digitized collection. The decision to redact required multiple considerations and balancing concern for privacy. While the Law Library concluded that they could post the newly digitized collection of briefs un-redacted without legal liability, ethical considerations inspired by trends in the law had the Law Library making a value judgment to redact. The Law Library had to consider the fact that maybe the law had not caught up with the project they were undertaking. While the Law Library reached a decision on redaction and enacted procedures to complete the redaction, there are still many future considerations for this project. The Law Library already had to think to the future and what sort of position they would be in if privacy law advances to a point where all court records are required to be redacted to be available online. Redacting now seems like the better course of action. But what happens when these documents are all available on the Internet and someone comes to the library seeking out a print brief with ill will stemming from their identification of redacted material that they believe has some value to them? Is there some way to stop that sort of behavior? Another future consideration may be the creation of standard guidelines for how to complete redaction. This would require a great deal of experience redacting and thought about what the best practices may be. Until that point in time, what the Law Library has currently accomplished paves the way to open up discussions about the potential future for libraries and the digitization of court documents.

Bibliography

Armstrong, C. L. (2011). Providing a clearer view: An examination of transparency on local government websites. *Government Information Quarterly*, 28(1), 11-16.

Retrieved from

<http://www.sciencedirect.com/science/article/pii/S0740624X10001048>.

Blankley, K. M. (2004). Are public records too public? Why personally identifying information should be removed from both online and print versions of court documents. *Ohio State Law Journal*, 65(2), 413-450. Retrieved from

<http://lawschool.westlaw.com/shared/westlawRedirect.aspx?task=find&cite=65+Ohio+St.+L.J.+413>.

Byrne, D. S. (2010). Access to online local government public records: The privacy paradox. *Legal Reference Services*, 29(1), 1-21. Retrieved from

<http://search.proquest.com/docview/742889554?accountid=14244>.

Conley, A., Datta, A., Nissenbaum, H., and Divya, S. (2012). Sustaining privacy and open justice in the transition to online court records: a multidisciplinary inquiry.

Maryland Law Review, 71(3), 772-847. Retrieved from

<http://lawschool.westlaw.com/shared/westlawRedirect.aspx?task=find&cite=71+Md.+L.+Rev.+772>.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.

- Daye, C. E. and Morris, M. W. (2012). *North Carolina Law of Torts*. New Providence, N.J. : LexisNexis.
- Family Education Rights and Privacy Act, 20 U.S.C. § 1232g (1974).
- Ganapati, S. and Reddick, C. G. (2012). Open e-government in U.S. state governments: Survey evidence from Chief Information Officers. *Government Information Quarterly*, 29(2), 115-122. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0740624X12000056>.
- Health Information Privacy Protection Act, Pub. L. 104-190 (1996).
- Hulse, R. (2009). E-filing and privacy. *Criminal Justice*, 24, 14-18. Retrieved from <http://lawschool.westlaw.com/shared/westlawRedirect.aspx?task=find&cite=24-SUM+Crim.+Just.+14>.
- Jones, D. R. (2013). Protecting the treasure: An assessment of state court rules and policies for access to online civil court records. *Drake Law Review*, 61(2), 375-422. Retrieved from <http://lawschool.westlaw.com/shared/westlawRedirect.aspx?task=find&cite=61+Drake+L.+Rev.+375>.
- Jones, J. M., Finkelhor, D., & Beckwith, J. (2010). Protecting victims' identities in press coverage of child victimization. *Journalism*, 11(3), 347-367.
- Leary, M.A. (1993). The case of the disappearing briefs: A study in preservation strategy. *Law Library Journal*, 85, 357-384. Retrieved from <http://heinonline.org/HOL/Page?handle=hein.journals/llj85&collection=journals&id=367>.
- Raggi, R., Reidenberg, J., Hedges, R., Winn, P., Dalglish, L., Morris, C. G., & Marcus,

- M. (2010). Panel One: General Discussion on Privacy and Public Access to Court Files. *Fordham Law Review*, 79, 1-25. Retrieved from <http://lawschool.westlaw.com/shared/westlawRedirect.aspx?task=find&cite=79+Fordham+Law+Review+1>.
- Sudbeck, L. E. (2006). Placing court records online: balancing judicial accountability with public trust and confidence: an analysis of state court electronic access policies and a proposal for South Dakota court records. *South Dakota Law Review*, 51(1), 81-121. Retrieved from <http://heinonline.org/HOL/Page?handle=hein.journals/jusj27&collection=journals&id=294>.
- Westin, A. (1967). *Privacy and Freedom*. New York: Atheneum.
- Whalen v. Roe, 429 U.S. 589 (1977).
- Whiteman, M. (2005). Appellate court briefs on the web: Electronic dynamos or legal quagmire. *Law Library Journal*, 97(3), 467-491. Retrieved from https://auth.lib.unc.edu/ezproxy_auth.php?url=http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=17888145&site=ehost-live&scope=site.

Appendix A
Interview Questions

Anne Klinefelter, Director of the UNC Law Library, Steven Melamut, Information Technology Services Librarian at UNC Law Library, Julie Kimbrough, Assistant Director for Collection Services at UNC Law Library, and Donna Nixon, Electronic Resources Librarian at UNC Law Library, Anne Gilliland, Scholarly Communications Officer at UNC

Question 1. Can you please discuss your role in the UNC Law Library's project to digitize its collection of North Carolina Supreme Court briefs? Were you part of the decision making process and what are your current responsibilities in relation to the project?

Question 2. What is your understanding of why the UNC Law Library decided to redact its collection of North Carolina Supreme Court briefs as part of the digitization process?

Question 3. Do you know of any objections raised to redaction? If so, why were those objections overruled?

Jesse Griffin, Processing Assistant at UNC Law Library

Question 1. Can you please discuss your role in the UNC Law Library's project to digitize its collection of North Carolina Supreme Court briefs? What are your current responsibilities in relation to the project?

Question 2. Can you discuss/explain the current digitization procedure?

Question 3. What do you see as the goal of the project to digitize the UNC Law Library's collection of North Carolina Supreme Court briefs?

Appendix B
Results of the 2012-12 Folder²⁶ Redaction²⁷

Search One

Search Terms	girl boy child minor molest rape youth student
Files in Folder	457
Files in Result List	128
Files Redacted	
Included Child's Name	26
Included Sexual Assault Victim	6
Date of Birth	0
Social Security Number	0
Driver's License Number	0
Total	34

Search Two

Search Terms	daughter son father mother born
Files in Folder	457
Files in Result List	175
Files Redacted	
Included Child's Name	2
Included Sexual Assault Victim	0
Date of Birth	4
Social Security Number	0
Driver's License Number	0
Total	6

²⁶ This folder included the North Carolina Supreme Court case briefs from Volume 196 (1928) and Volume 352 (2000). This was the first folder and the only folder to have two such different volume years. After this point, all folders are in reverse chronological order, working backward from the year 2000. The title of the folder denotes that the digitized copies of this batch of briefs were received from Hein in December 2012.

²⁷ The total number of files redacted is not a sum of the individual types of types of information that needed to be redacted. For example, some files had both rape victims and children or both a social security number and birth date.

Search Three

Search Terms	license ²⁸
Files in Folder	457
Files in Result List	30
Files Redacted	
Included Child's Name	0
Included Sexual Assault Victim	0
Date of Birth	1
Social Security Number	1
Driver's License Number	0
Total	1

²⁸ As discussed in the paper, the original search was just for driver's license number, before search modification. After the search was modified for the second folder, quick searches were run on this folder to ensure no social security numbers or birth dates were missed, using the updated search string. No additional sensitive information was located. However, these additional searches are not part of the original recorded results.

Appendix C
Results²⁹ of the 2013-01 Folder³⁰ Redaction³¹

Search One

Search Terms	girl boy child children minor molest rape youth student
Files in Folder	668
Files in Result List	349
Files Redacted	
Included Child's Name	89
Included Sexual Assault Victim	37
Date of Birth	26
Social Security Number	0
Driver's License Number	1
Total	111

Search Two

Search Terms	daughter son father mother born
Files in Folder	668
Files in Result List	276
Files Redacted	
Included Child's Name	1
Included Sexual Assault Victim	0
Date of Birth	9
Social Security Number	0
Driver's License Number	0
Total	9

²⁹ The results of the searches run on the 2013-01 folder are very different than those run on the 2012-12 folder. One of the reasons for the differences is likely the dates of briefs redacted. The cases from the early 1900s seemed to have possessed less sensitive information than more recent cases. Quick searches of other digitized folders support the supposition that the results of this search are much more typical of what the redaction numbers will look like for other folders.

³⁰ This folder included the North Carolina Supreme Court case briefs from Volume 350 (1999-2000) and Volume 351 (1998-1999). The title of the folder denotes that the digitized copies of this batch of briefs were received from Hein in January 2013.

³¹ The total number of files redacted is not a sum of the individual types of types of information that needed to be redacted. For example, some files had both rape victims and children or both a social security number and birth date.

Search Three

Search Terms	license DOB birth ss# social security
Files in Folder	668
Files in Result List	304
Files Redacted	
Included Child's Name	0
Included Sexual Assault Victim	0
Date of Birth	20
Social Security Number	4
Driver's License Number	0
Total	21