The Dissertation Committee for Craig Erben Blaha Certifies that this is the approved version of the following dissertation:

FACEBOOK FOREVER: PRIVACY, PRESERVATION AND SOCIAL NETWORKING RECORDS

Committee:

Philip Doty, Supervisor

William Aspray

Patricia Galloway

David Spence

Yan Zhang

FACEBOOK FOREVER: PRIVACY, PRESERVATION AND SOCIAL NETWORKING RECORDS

by

Craig Erben Blaha, B.A.; M.A.Ed.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2013

Dedication

To Isabelle, Sophia, and Cyndie. Without your love, support, and fantastic advice, I would still be working on “the next draft”.

Acknowledgements

I would like to acknowledge the help and support of my committee: Philip Doty, Pat Galloway, Yan Zhang, Bill Aspray, and David Spence. Without your help, support, guidance, and patience this would not have been possible. I would especially like to thank Philip Doty for his herculean efforts in reading and commenting on each and every draft, generously continuing to look forward and believe the next draft would be better, and pushing me to make sure it was.

I would like to thank the Institute of Museum and Library Services for both the personal financial support and the foresight to invest in archives and preservation research. The work I have seen from my IMLS colleagues is impressive and I am proud to be an IMLS Fellow.

I would also like to thank my friends for their kind words, support, and encouragement. Especially the Blackburns for the good bottle of scotch, which will be gone by the time you read this.

Most of all I would like to thank my family. My daughters Isabelle and Sophia, for putting up with me when I lacked sleep and patience, had my head stuck in a book or fingers on the keyboard, and for offering advice that every writer should follow, “don’t write too much, and don’t use the same words over and over again.” My wife Cyndie for her love, understanding, patience, support, and for occasionally asking the tough question.

Facebook Forever: Privacy, Preservation and Social Networking Records

Craig Erben Blaha, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Philip Doty

For the first time in history one billion subscribers are creating records using a single software platform: Facebook. Subscribers create historically significant Facebook records every day, yet there is no concerted effort to preserve these records. Archivists do not agree on whether or how these records will continue to exist, nor do they agree on the best way to preserve these records. At the same time, privacy advocates are concerned that social networking records will continue to exist "forever" and therefore have serious privacy implications.

In this study I examine the seemingly opposing viewpoints of privacy scholars and archivists. I find that privacy scholars are concerned that the lack of subscriber control over social networking records threatens privacy over time. Archivists address this lack of control through the concepts of donor agreements and the trusted digital repository, but the application of these concepts to the long-term preservation of Facebook records depends on who will preserve these records.

I explore four different ways Facebook records may be preserved. I examine whether the U.S. federal government can and should play a role in encouraging Facebook to preserve records. I find that the U.S. federal government is unlikely to take action.

I take a first step in empirically examining the likelihood that individual Facebook subscribers will preserve their own records using both an online survey (n = 144) and focus group to ask Facebook subscribers what they expect to happen to their Facebook records. I find that Facebook subscribers do not trust Facebook, do not think about preservation when they use Facebook, and do not expect their Facebook records to exist forever.

This research makes four contributions to existing literature: a discussion of the value of social networking records and whether they should be preserved, a close examination of the differing opinions of archivists and privacy scholars about these records, a discussion of the role public policy might play in the preservation of Facebook records and privacy in the United States, and an empirical exploration of the attitudes and behaviors of a small group of Facebook subscribers related to preservation and privacy.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: Introduction

For the first time in history a single corporation, Facebook, has over 1 billion subscribers (Facebook a, n.d.) using a single information and communication technology (ICT) to correspond with one another. The volume of records being created is staggering, and the various social, political, economic, historical, and privacy implications of the use of this service are receiving considerable attention from both scholars and the popular press. Archivists fear that these digital social networking records will not continue to exist if archival practice does not change. Privacy scholars fear these records will continue to exist forever, threatening the privacy of the records creators. In this study I discuss the historical and research value of Facebook records, explore the seemingly opposing viewpoints of privacy scholars and archivists, and discuss who might preserve valuable Facebook records.

Records created using Facebook may be of historical, cultural, personal, and research value. Facebook records may be valuable to society as cultural heritage artifacts, to researchers as a large sociological dataset, to historians as first person accounts of historical events, and to individuals for personal memory and identity. I discuss in detail whether Facebook records are valuable and should be preserved.

Privacy scholars fear that the persistence of social networking records is a threat to privacy. They use the terms "preservation" and "archive" to refer to the minimum time period after which a record can be destroyed. The ability to delete a record is important to the protection of privacy, and this ability depends on the record creator's control of access to the social networking record. That control, and the privacy that depends on it, is threatened by the creation of multiple copies. Subscribers make copies of Facebook

records for their own personal use, and Facebook makes copies to provide the service, create customer profiles, and for advertising purposes. Each of these copies represents a risk that the record will be accessed without the knowledge or permission of the record creator.

The longer a record persists, according to privacy scholars, the greater the likelihood that some other individual or group has made a copy of that record. The record creator has no way of knowing whether a copy of that record has been made either by the service provider, another service subscriber, or a third party entity such as the Internet Archive. Over time it is increasingly difficult for an individual subscriber to have any confidence that she has control over who has access to the records she has created using Facebook and nearly impossible to be sure she can remove all copies of a record from public view in order to protect her privacy.

In addition, Facebook has been known to make changes to their privacy policy and the technological affordances of the software. These changes are designed to encourage the subscriber to share more information. Facebook's fiscal health and shareholder satisfaction depend on advertising revenue, and the more information individual subscribers share, the more valuable advertising space on Facebook becomes.

Archivists have a different set of concerns related to the preservation of Facebook records. Archivists worry that the use of proprietary formats, the rapid pace of hardware and software versioning, the rapidly changing business model and potentially short life-span of technology companies, and the lack of awareness and concern on the part of subscribers all threaten the persistence of valuable social networking records.

Archivists have a long history of protecting the privacy of records donors while preserving records and allowing for access to those records. A new set of standards has

emerged that allows a digital repository to prove that it is "trustworthy," that it follows prescribed guidelines to ensure that records will persist, and that the policies, procedures, and organization of the institution providing the preservation of records will protect the record and the privacy of the record donor. While this type of organization does not address all of the concerns of privacy scholars, it can serve as an example of what a trusted information communication technology provider might look like.

Given the opposing viewpoints of archivists and privacy scholars, it is unclear who might best protect privacy while preserving Facebook records. I outline four general approaches to preserving Facebook records and protecting privacy: Facebook subscribers can preserve their own records, Facebook can act as a trusted digital repository, Facebook can partner with a trusted third party, and a trusted third party can act independently to download and preserve Facebook records. I examine each of these possible solutions using Lessig's (2006) four regulatory constraints: computer code, law, norms, and the market. I also discuss some of the problems with each approach.

Facebook is the service provider and the repository for all records created using the Facebook software. If we are to preserve one billion Facebook subscribers' records, changes in public policy may be necessary to ensure that Facebook takes appropriate steps to preserve these records, or to allow subscribers or third parties to preserve these records. I examine the role public policy might play in encouraging Facebook to act as a trusted digital repository, a platform that will provide for the persistence of and access to these records. I use the market imperfection theory of government regulation to discuss whether government regulation is warranted in the context of social networking records. If the relationship between Facebook subscribers and Facebook indicates the existence of

a natural monopoly, a public good, an externality, asymmetry of information, or a moral hazard, then the federal government may have cause to regulate.

It is important to note that this study was conducted in 2011 – 2013. The policies, technological affordances, and interface of Facebook have not only changed over the eight-year existence of Facebook, but over the two years this study has been conducted. Specific references to the interface, policies, and affordances throughout this study are either accompanied by dated screenshots or by citations that include the date. I have also included, where appropriate, screenshots of previous Facebook interfaces and a timeline of major changes to Facebook over time in Appendix A.

Cause to regulate does not mean that elected or appointed government officials will necessarily have the will to impose regulation. I use Kingdon's (2003) garbage can model of public policy to understand whether and how the federal government might be encouraged to act to preserve social networking records and subscriber privacy. I examine existing and proposed laws and regulations that might indicate a willingness on the part of government to regulate interpersonal communication in the context of social networking.

If Facebook cannot be encouraged to act as a trusted digital repository, the preservation of Facebook records will most likely be the responsibility of each of the one billion Facebook subscribers. Subscribers' expectations are important for us to consider when discussing the preservation of Facebook records. If, as Cox asserts (2008), individual records creators are the foundation of future archives, we need to begin to understand how and whether these individual records creators consider preservation in the context of Facebook records. On the one hand, if subscribers expect these records to persist, and service providers do not ensure this persistence, records of potential historical

value could be lost. On the other hand, if Facebook subscribers do not expect these records to persist, we will need to determine whether subscribers are taking action to preserve their own records and whether subscriber-led preservation efforts would enable the dual goals of preservation: persistence and access.

Examining records creators' expectations is also an important step in understanding how best to train and educate "citizen archivists" for the task ahead. My research empirically addresses the preservation expectations of Facebook subscribers by asking a small group (n = 144) of School of Information students at a major research university about their expectations related to the preservation of Facebook records. The individuals participating in the online survey and focus group are different from the average Facebook subscriber in at least one way; they are part of a community that is concerned with, and at least to some extent educated about, archives and preservation. I expect these participants to be more aware of the privacy and preservation issues raised by Facebook records than Facebook subscribers that are not affiliated with an academic program that includes a preservation concentration.

My analysis of survey and focus group responses addresses two main questions:

1. What are Facebook subscribers' preservation expectations? In other words, do Facebook subscribers expect the records they create using Facebook to persist over long periods of time? Further, do they expect to be able to control and access these records over that period of time?
2. To what extent are subscribers' preservation expectations associated with disclosure behavior? Are subscribers more or less likely to disclose personal information if they do not expect records to persist over time?

To my knowledge, no empirical studies have examined records creators' preservation expectations. Facebook has existed for only nine years, and the use of social networking services to create the large volume of records that we are seeing today is a new phenomenon. The short existence of Facebook is one reason little research has been done to address the question of whether social networking records will be preserved and what records creators expect to happen to their records. Some privacy scholars have asserted there should be a right to delete, but there is a lack of empirical analysis of how long users expect records to exist.

In the next section I discuss my use of the term "record" and the value of Facebook records. A close examination of the balance between privacy and preservation of Facebook records needs to begin with a clear understanding of whether the records created using Facebook are valuable and worth preserving. If these records are not valuable, then our path to protecting privacy is far less complex. Before turning to the discussion of the value of Facebook records, I clarify what I mean when I use the term "record" in the context of Facebook.

WHAT IS A FACEBOOK "RECORD"?

Facebook is complex software that creates a variety of record types as subscribers and third-party systems interact with it. Facebook lists four types of information they receive about subscribers (Facebook, n.d. b):

1. Registration information
2. Information that subscribers choose to share
3. Information that others choose to share about subscribers
4. Other information.

The "Other information" category includes records that security researcher Bruce Schneier calls "behavioral data." Schneier describes behavioral data as "data the site collects about your habits by recording what you do and who you do it with" (Schneier, 2009).

For the purpose of this study I am interested in those records that most closely resemble paper-based correspondence, e.g. letters that are either hand or typewritten or sent via telegraph. I consider correspondence records in Facebook to include the records a subscriber knowingly creates to communicate with other subscribers: status updates, photos, videos, likes, and comments. Facebook refers to these records as "information subscribers choose to share." In the next section I will discuss how Facebook records differ from paper-based records. Making the connection between paper-based records and Facebook records will help to create context and to tie the current study to the extensive body of research on paper-based records.

HOW FACEBOOK RECORDS DIFFER FROM PAPER-BASED RECORDS

Facebook records differ from paper-based correspondence in four ways that are relevant to the preservation of Facebook records. These four differences are the scale of Facebook, the relationship of Facebook subscribers to each other, the relationship of Facebook subscribers to the records they create, and the clarity of expectations in traditional archives. I have already briefly discussed how for the first time in history, one billion subscribers are using the same information and communication technology platform to communicate with one another. In the next chapter I will discuss this scale of Facebook in more detail and how the volume of records being created makes these

records valuable to researchers. I discuss the other three reasons Facebook records are different from paper-based records in turn below.

Facebook changes the relationship between individuals corresponding with one another. Before the widespread adoption of electronic information communication technologies, interpersonal correspondence relied on the post office, telegraph, or telecommunications companies to carry messages between parties. The post office and telecommunications companies did not necessarily store the record in their own facilities; the original record remained the property of the recipient. In the case of the telegraph, a copy of the original or the original record itself may have been retained by the telegraph company. Importantly, the technology behind the delivery of a message did not require the messaging service provider to store the original record on property owned by the service provider, as Facebook does when it stores messages on a Facebook server. This concept of control is a prominent theme in the focus group discussion and critical to understanding the threat to both preservation and privacy that is posed by Facebook.

The ownership of the original records in correspondence changed with the creation of Usenet bulletin boards in the late 1970's (Abbate, p. 202). Usenet bulletin boards can be considered a precursor to social networking services. Messages posted by Usenet subscribers were both created and stored on a server that was usually not owned by either the message creator or the subscriber viewing the message. Many current iterations of email operate this way by default; Gmail, Yahoo, Apple's Icloud Mail, and Hotmail are all examples.

The default protocol for most of these services is the Internet Message Access Protocol (IMAP), which leaves a copy of the message on the central server and either displays that message through a browser client or downloads a copy to the desktop client.

While most of these services can be configured to use the Post Office Protocol (POP), which allows subscribers to download messages from the central mail server to the subscriber's local hard drive and then delete the messages on the central mail server, we know the default settings of software are powerful manipulators of human behavior (Kesan & Shah, 2006, p. 598). When a Gmail subscriber creates an email message, for example, the message is created and stored on Google's server. If the recipient is also a Gmail subscriber, when the recipient logs into their Gmail account they view the message hosted on Google's server through their browser, the record is not downloaded to the subscriber's desktop. Gmail, Yahoo, and Hotmail each depend on advertising revenue to offset the cost of providing the service and encouraging subscribers to leave email messages on the central servers is an important component of the business model. These email providers scan the content of the email messages that are received by the subscriber and use this content to decide which ads to show the subscriber. The more email a subscriber retains, the more information the service provider has available to them to use to decide which advertisements that particular subscriber is likely to click on.

The service provider retains a copy of the message created by the subscriber, which is the equivalent of the original record created when an individual writes a handwritten letter that will be sent through the post office. The recipient and the creator have access to or complete control over only a copy of the original record. Control over the original record has shifted from the two parties engaged in correspondence to the third party offering the communication service, in this case, Facebook. By using Facebook, we entrust to the company the original record of all correspondence conducted using the service.

In addition to this change in relationship between Facebook subscribers is the change in the relationship between the Facebook subscriber and the record she has created. As described above, the record creator and the recipient of the record have access only to a digital copy presented to them by Facebook. Correspondence using Facebook is different from paper-based correspondence where the original record is kept by the recipient, and the creator and the recipient control that original copy in turn. If the recipient destroys the original, both parties can be certain that it has been destroyed. When using Facebook, a record creator does not have control over the original record; Facebook maintains the record on Facebook's servers.

The trusted relationship is no longer between two individuals known to each other with a "trusted" delivery service, such as the post office, in the middle. The relationship has expanded to include Facebook (and numerous intermediaries, such as Internet Service Providers), that will retain a copy of the record, allowing other parties access to those records depending on the privacy settings and options the record creator has set at that time. This last point should be emphasized, Facebook has changed privacy policies and the technical ability individuals have to control access to their records enough times to be audited by the Federal Communications Commission (FCC) for privacy violations for the next twenty years (Reuters, 2011).

In addition to the three reasons Facebook records differ from paper-based records already mentioned (the scale of Facebook, the change in relationship between two (or more) parties involved in correspondence, and the change in relationship between the record creator and the record she has created), the archives profession has historically included a well-defined process for managing the expectations of records donors, a process that will necessarily need to be revised given the three changes mentioned above.

The deed of gift is a legal document meant to clarify the relationship between a donor of records and an archives and to set preservation expectations. “The relationship between you, as a donor, and a repository must be based on a common understanding of your wishes and the ability of the repository to carry out its mission and responsibilities” (Weideman, n.d.).

Norman Mailer sold his records to the Harry Ransom Center at the University of Texas at Austin for $2.5 million two years before his death in 2007 (Vertuno, 2008). While unusual in the dollar value placed on the records, the process followed for the donations of these records is time honored. The donor and the organization agree to a donation, after which both parties sign a deed of gift that allows for the exchange of property rights and defines access limits. A representative of the organization then processes, organizes, and catalogs the materials.

The act of donating one’s correspondence to Facebook is not accompanied by the same clarity of agreement. Because Facebook does not consider itself an archives, and its audience is current Facebook subscribers, there is no recognition of the long-term nature of the relationship established when the subscriber posts her first comment or status update.

The Facebook equivalent of the deed of gift is the terms of use, which includes Facebook’s privacy policy. Facebook has made considerable changes to their privacy policy over time. Facebook instituted a practice in early 2012 such that, if 7,000 people complain about an impending policy change, Facebook will consider a vote on that change. That policy was revised in late 2012 to eliminate subscribers’ opportunity to vote on policy changes.

These policies make it clear that Facebook is not maintaining an agreement between an individual record donor and that donor's desire to grant access to records in a particular way over the lifetime of those records, as is the case with a deed of gift.

The volume of interpersonal correspondence conducted via Facebook, the changed relationship among the corresponding parties using Facebook, the changed relationship between the correspondence creator and the record she has created, and the lack of clearly defined preservation expectations between Facebook and its subscribers are all new challenges to digital preservation of interpersonal correspondence. Since these challenges are relatively new, little empirical research exists about how archivists plan to preserve important cultural heritage and research records created using Facebook. This study takes a first step in addressing this research shortfall. In the next section I will discuss three different reasons Facebook records may be considered valuable.

THE VALUE OF FACEBOOK RECORDS

To examine the conflicting opinions of privacy experts and archivists and discuss who might preserve Facebook records, I will first need to discuss whether these records are worth preserving. In this section I will examine a variety of different ways Facebook records may be considered worth preserving, a process called appraisal that determines whether a record or group of records has "enduring value" (Pearce-Moses, n.d.). This section is intended as an overview of potential sources of value and a description of the preservation issues these value sources raise in the context of Facebook. Rather than attempting to designate specific Facebook records or groups of records as significant, I offer a general discussion of whether these records may be worth preserving.

The process of determining the value of a record or group of records has a long tradition in archives. I explore three different possible sources of value: cultural heritage value, personal value, and the volume of Facebook records for research value. These three areas are not mutually exclusive; valuable cultural heritage records may also be of personal value, for example. By distinguishing among the three types of value I am better able to discuss who might preserve records of each type.

In many ways, Facebook is simply a new way for individuals to correspond with each other. Asking whether Facebook records should be preserved is akin to asking whether postal mail or telegrams should be preserved, a question that has received considerable attention in the past. Digital communications such as email and blogs have received similar attention to paper-based communications from archivists, and I will build on these previous discussions when I discuss the cultural heritage and personal value of Facebook records. The third source of value, the research value of the volume of Facebook records, includes a discussion that is relatively new to both archives and research, the discussion of "big data" and the use of large data sets to conduct humanities research.

Records of Cultural Heritage Value

Archivists are charged with preserving and making available original sources that collectively serve as part of our cultural heritage. The definition of original source has included everything from written records such as correspondence and manuscripts, to objects such as photographs and other artifacts. In this section I offer a brief history of the approaches professional archivists in the United States have taken to determine whether records are worth preserving. This brief overview will help us to determine whether

Facebook subscribers produce records archivists would find of value to the historical record, and to determine whether the use of Facebook is comparable to other, long-standing records creation practices.

Cultural or national heritage is difficult to define precisely because it connects personal memory to group history. Lowenthal notes that reasons for admiring the past are "vague or perfunctory" (p. 36) and "[e]qually ineffable is the concept of national heritage, normally evoked with sub-lyrical vagueness" (p. 36). The United States Cultural Heritage Center considers a range of projects important to cultural heritage from the restoration of historic buildings to documentation of craft techniques to improved storage conditions for archives (Bureau of Educational and Cultural Affairs, n.d.).

Smith defines cultural heritage as content that performs the functions of "recording experience, shaping perceptions of our world, adding or subtracting meaning, providing pleasure or inflicting pain" (2007, p. 14). Lloyd discusses three reasons why items are selected as valuable cultural heritage artifacts: they are unique, they represent intellectual endeavor, and they record an activity at a point in time (2007, p. 54). Lloyd points out that the designation of an object or record as representing cultural heritage is an exercise of power; an act of defining history, knowledge, and culture that privileges dominant groups.

As broad as the definition of cultural heritage might be, the importance of cultural heritage is even broader, but perhaps better explored. Lowenthal uses George Perkins Marsh's book *Man and Nature* (1864) as one example of how cultural heritage serves scientific research. Lowenthal calls *Man and Nature* "the fountainhead of conservation consciousness" and focuses on Marsh's emphasis on the preservation of "the artifacts of everyday life rather than the great monuments of antiquity" (xvii). These artifacts of

everyday life are represented in part by the records that 1 billion Facebook subscribers from all over the world are creating on a daily basis.

When discussing the value of preservation, Smith says we need to look beyond the "aesthetic, documentary, evidential, forensic, and economic or market values of content" (2007, p. 10) and instead focus on the role this content plays in the formation of meaning and the value of that content in creating greater social cohesion (p. 14). Cultural heritage records, according to Smith, are important because direct access to reliable and authentic cultural heritage records protects liberty and the ethical and moral values of a democratic society, and because "information is a constitutive force in society, all aspects of its integrity, completeness, authenticity, and accessibility are profoundly important" (p. 18).

Archivists have traditionally looked for cultural heritage value in the records of organizations and historically significant individuals. There exists a rich tradition of selection and appraisal methods by which archivists determine whether a set of records is worth including in an archival collection and preservation efforts.

O'Toole and Cox (2006) point to two traditions in early American archives: the public archives tradition (p. 53) and the historical manuscript tradition (p. 55). In the public archives tradition, public institutions held records maintained by government officials. These records included land records, birth and death records, and journals of legislative proceedings. The public archives tradition, as the name implies, made the records held by the archives institution publicly available. Access to the records facilitated transactions and increased transparency by making business transaction records available to the general public (p. 54).

In the historical manuscript tradition, individuals or local collecting institutions such as historical societies collected manuscripts, documents, and other objects and artifacts. These individuals and institutions focused primarily on the records of historically significant individuals and collected and maintained these records in order to support and promote historical research and the preservation of cultural heritage. Therefore, these collecting institutions had at least two important aspects to their mission: preserving and providing access to the records in the collection.

From the late 1800's to the early 1930's, organizations in the historical manuscripts tradition were taking two different approaches to preservation. They were collecting everything they could find of historical value and either storing and cataloging the records, or editing and publishing the text of these records. Both approaches were seen by the collecting institutions as ways to increase interest in history and to support historical research.

The Society of American Archivists' (SAA) contemporary definition of an archives reflects both the historical manuscript tradition and the public archives tradition because the definition includes both personal and organizational records:

> Materials created or received by a person, family, or organization, public or private, in the conduct of their affairs and preserved because of the enduring value contained in the information they contain or as evidence of the functions and responsibilities of their creator, especially those materials maintained using the principles of provenance, original order, and collective control; permanent records. (Pearce-Moses, n.d.)

In addition to describing a broad range of records creators, the SAA definition of an archives allows for a wide variety of reasons to permanently retain records.

We can see from this definition that not all Facebook records would necessarily be considered part of an archives, only those records that have enduring value. The term

"enduring value" is vague and disputed within the field of archives. A brief overview of selection and appraisal, which follows, will not only help us clarify the term "enduring value," but will also add to our understanding of how records of "enduring value" can support the claims of privacy scholars, archivists, or both in the discussion of preserving Facebook records and the concomitant threats to privacy.

Both the public records tradition and the historical manuscript tradition emphasize the dual role of the archive, preserving records and facilitating access to those records over time. Both traditions also assume value has to be determined based on the activities, historical significance, or the role of the record creator or organization. While these considerations do apply to Facebook records, Facebook records are also different from paper-based records because of the volume of the records being created and the tools and technology that are available to analyze these records. I will discuss this difference in more detail later in the chapter, but it is important to note that the historical foundation of appraisal and selection still applies to Facebook, along with new considerations required by the volume and digital nature of Facebook records.

Within these two traditions of early American archives we find distinct approaches to developing an archive. In the public archives tradition, the archivist is described as actively appraising which records or groups of records are considered valuable, but does not actively seek records to include in the archive. In the historical manuscript tradition, the archivist is described as passively accepting any and all records related to an individual or historic event, but actively seeking sources for these records. These descriptions are generalizations rather than accurate descriptions of these two traditions, but the differentiation between active and passive appraisal and active and

passive collection will be important as we discuss who should preserve Facebook records.

Appraisal Theory

The process of determining which records or groups of records represent valuable cultural heritage objects has a long and interesting history. A brief overview of this history will help us understand how Facebook records might be evaluated and whether these records should be considered valuable.

One of the earliest English-speaking theories of appraisal comes from Hilary Jenkinson. According to Jenkinson, appraisal and selection of records were activities performed only by the creator of those records during the course of regular business. "[B]ut for an Administrative body to destroy what it no longer needs is a matter entirely within its competence" (1922, p. 149). Variation from these principles imperils the "unquestioned impartiality" (p. 147) of the records.

I interpret Jenkinson's theory of appraisal of records to afford complete control over the records to the record creator, and defer to the appraisal and selection choices of the records creator. The level of control and deference advocated by Jenkinson allows the records creator great influence over the appraisal and selection process. Lyle describes this approach to appraisal as "passive appraisal" (2004, p. 2), emphasizing the reliance of the archivist on the record creator to determine which records are worth keeping.

This influence allows the organization, or in the case of social networking, the individual, to determine which records should be destroyed prior to preservation, allowing for greater privacy protection. The ability to destroy records, as I explain later in the discussion of the "Right to Delete," allows the record creator to limit the risk that

their privacy will be violated over time through their ability to make that record unavailable to anyone. When using social networking, the records creator does not presently control the destruction of records. When the subscriber deletes a record, access to that record by the subscriber may be removed, but it is unlikely that the record is removed from the company's servers, and there is no way for the service subscriber to know whether other service subscribers or third parties have made a copy of that record.

Theodore Schellenberg is a key figure in American archival theory of particular interest to this study. Schellenberg posited the ideas of primary and secondary value (1956). The term "primary value" indicates the value to the creator of the use of the records in the conduct of business. Secondary value refers to the use of the records for research and historical purposes after their primary purpose has been completed. This dichotomous description of records value helps us understand how social networking service providers might conceive of preservation. A record is worth retaining if the record creator or the service provider is still using it. There is no business reason for social networking service providers to assess secondary value in the sense that Schellenberg used the term.

Facebook records complicate Schellenberg's concept of primary and secondary value in that records created using Facebook are used by the record creator to communicate with "friends" and by the service provider to increase advertising revenue. Primary value for the record creator is found in the interaction and communication with "friends" in their network. Depending on the type of information shared, this value may last a few minutes, when a subscriber shares her location information in order to facilitate a face-to-face meeting, for example, or indefinitely, as when a subscriber shares the photograph of her child's first birthday or wedding day.

Facebook has a different way of assessing primary value that includes both short- and long-term value simultaneously. As the service subscriber shares her location information with a friend, there is immediate value to the service provider in allowing for location-based advertising. If the subscriber is trying to meet someone for a cup of coffee, an advertisement for a coffee shop in the general vicinity might persuade the subscriber to propose that coffee shop as a meeting location. In this scenario, the subscriber, service provider, and coffee shop have collaborated to provide a convenient service to the subscriber. The coffee shop pays the service for the advertising opportunity; the subscriber trades personal information for access to the service, and the immediate value of the social networking record is realized.

The long-term value of records shared by service subscribers is in the service provider's ability to develop a profile of the subscriber. If Facebook knows a particular subscriber sets up meetings at ten o'clock on Sixth and Brazos Streets every morning, Facebook has an opportunity to target advertising to that individual. Of potentially more value to Facebook is the ability to aggregate these profiles and analyze consumer behavior trends over time. The ability to predict what advertisements a subscriber will react to will increase the value of Facebook advertising to advertising companies. For example, if Facebook maintains data on subscribers who post a birth announcement and the effect of that birth announcement on the advertising response of their "friends," Facebook can sell highly targeted advertising space to companies that sell baby products.

As we see from this discussion of Schellenberg's bifurcation of records value into primary and secondary value, neither the service provider nor the subscriber is focused on the secondary value of the records.

Lyle describes Schellenberg's approach as "active appraisal" (2004, p. 2), emphasizing the role of the archivist in determining what records should be kept. The characterization of Jenkinson as advocating for "passive" appraisal and Schellenberg advocating for "active" appraisal may be overstated by Lyle, but I think these two concepts are useful in understanding who might preserve valuable Facebook records. The Library of Congress partnership with Twitter is an example of passive appraisal, and the idea of "active" appraisal is one that can be applied to collecting archives that choose to preserve Facebook records.

Both Jenkinson and Schellenberg were operating in the context of government archives, but both of their ideas can apply to Facebook and other social networking records. The development of information communication technologies, from the typewriter and affordable copying technology to the widespread adoption of the Internet, have led archivists to question this focus on organizational records. In the 1970's, archivists and historians began to question the role of the archivist in supporting existing power structures through the appraisal and selection of cultural heritage records. The challenge to archivists was to clearly articulate their appraisal process to improve transparency and limit the chance of unwittingly marginalizing minority and oppressed groups. The resulting appraisal theories emphasize the value of records generated by individuals and marginalized, oppressed, or radical groups as contributors to cultural heritage, all of whom can be represented in Facebook records.

In a speech to the Society of American Archivists in Washington D.C. in 1970, historian Howard Zinn questioned the neutrality of the archival profession (Quinn, 1977, p. 11), and challenged archivists to focus more on the records created by ordinary people, calling this collection a necessary step in creating a real democracy (Zinn, 2001, p. 173).

Zinn asserted "those with the most power and wealth in society will dominate the field of knowledge, so that it serves their interests" (p. 167). According to Zinn, the daily routine of archivists, the unquestioning selection, appraisal, and cataloging of government, business, and military records, supported the wealth and power structure that was being questioned through protests, civil rights legislation, and political reform.

Gerald Ham in his own speech to the Society of American Archivists, "The Archival Edge" (1975), points to Zinn and other critics and asks why archivists are failing to provide "an informed selection of information that will provide the future with a representative record of human experience in our time" (p. 5). Ham singles out Cornell historian and archivist Gould Colman's accusation that archivists' poor selection practices were focused on unrepresentative aspects of contemporary culture, and these practices were distorting the historical record.

Ham describes a number of challenges to the archival profession that he says require archivists to change from passive recipients and custodians of records to active collectors of records that represent the "human experience of our time" (p. 5). One of the challenges Ham articulated was a structural change in society that reflects greater participation of "non-elite" (p. 8) population groups. Non-governmental organizations that influence government decisions, such as protest groups, were creating more records, but those records were not finding their way into archives. Ham also discussed the increased volume of records created by various photocopying and print reproduction technologies (p. 9) and the decreased quality of records. Ham pointed out that the telephone is more likely to be used to communicate important or sensitive information than any form of writing. Ham cites technology and instant archives, or records that are destroyed soon after they are created, as the last challenges that will require archivists to

change from a passive collector of records to an active gatherer of records. Ham warned that archivists must actively seek out and collect material such as magnetic tape, photographs, and database documentation. Ham's call for archivists to change from "passive" to "active" collectors mirrors Lyle's description of appraisal as either a "passive" or "active" process.

Each of the challenges articulated by Ham in 1975 still exists. Loosely related groups such as Occupy Wall Street or Anonymous influence government decisions, the volume of both digital and paper records created by organizations and individuals continues to increase, limited storage communication tools such as the telephone, text message, and encrypted communication are used to communicate sensitive or important information, and "instant archives" and technology have become almost synonymous with one another as new information and communication technology services are created at a rapid pace.

One example of the creation of individuals and loosely related groups influencing government using Facebook is the participation of Google executive Wael Ghonim in the Egyptian revolution of 2011. Ghonim created a Facebook account to document the brutality of Egyptian police (CBS News, 2011). While multiple people managed the account, Ghonim was arrested for using Facebook to encourage the protests that eventually overthrew the Egyptian government. If the revolution had failed, the account may have been deleted or the important historic records changed or removed from that account, eliminating a primary source of information for an important historical event.

Ham said archivists needed to "change old habits and attitudes" (p. 12) and recommended a collaborative national collection system (p. 12) that would offer

guidelines and concepts that would help archivists determine what to collect and how better to manage that collection (p. 12).

Ham built on these ideas of cooperation and active collection when he called the era leading up to the "information revolution" the "custodial era" (1981, p. 207). He describes that time as a time of passive engagement by the archivist with the documentary record. With society creating fewer records, the archivist, according to Ham, was able to focus on passively collecting and maintaining records that were made available to the archives by organizations and individuals. Ham says this low level of document creation allowed archivists to become overly focused on maintaining existing collections (p. 207).

Ham refers to the early 1980's as the post-custodial era, and says, "chemistry and electronics have forever altered the archivist's placid world" (p. 208). The use of videotape, photographs, film, and electronic databases and other electronic information storage devices had increased the amount of data we create, use and store, creating new problems and opportunities for the archivist. The volume of data and the accompanying problems and opportunities to which Ham referred in the 1980's has increased dramatically since it has been coupled with the prevalence of consumer computing systems, broad access to the Internet, and a billion Facebook subscribers.

Ham outlines a five-point agenda that archivists need to address in order to meet the challenges posed by chemistry and electronics:

1. We must develop coherent and comprehensive acquisition programs at all levels, national, regional, and local.
2. We must utilize the benefits of modern technology to provide easy and centralized access to increasingly complex and decentralized holdings.
3. We must deal with the impact of modern technology on the creation of information, and devise programs for its selective preservation and use.

4. We must participate in resolving the conflict between the freedom of information and the right to privacy as they affect the quality and content of the archival record, and access to that record.
5. We must make better use of the limited (and, I might add, diminishing) resources available for archival activity nationwide. (p. 211)

The overarching call to action is important. Ham was advocating inter-institutional cooperation such that established archives help less-established archives improve their holdings, records management processes, and acquisition strategies. He was calling on archivists to work together to preserve the historical record writ large, not just the records relevant to their collecting institutions. By helping each other with planning, research, staffing, and management, and by sharing information about the holdings of each institution, archivists, Ham believed, would work together to preserve important records without regard to the collecting institution responsible for those records. "Only archivists and their profession can determine whether the post-custodial era will be one of archival abdication or of planned response and integration" (p. 216).

It is also important to point out that Ham was emphasizing active collection and active appraisal. The second point in his agenda mentions the selective preservation and use of information, emphasizing his belief that the role of the archivist would continue to be one of determining the value of records and deciding which records should be preserved. Some Facebook records would clearly match Ham's determination of value.

Samuels builds on the idea of post-custodialism in "Who Controls the Past" (1986) in which she develops the idea of a documentation strategy. According to Samuels, a documentation strategy is "a plan formulated to assure the documentation of an ongoing issue, activity, or geographic area" (p. 115). Samuels asserts that "individuals and organizations do not exist separately" and that "records mirror the society that creates them" (p. 111).

Building on Ham's concept of the post-custodial era, Samuels highlights the idea that documents pertaining to any particular topic are created, preserved, and maintained by multiple institutions and individuals. She separates the documentation of a topic from the archival institution collecting that documentation, saying "each collection becomes a part of a larger collection" (p. 124), that these records will exist in multiple locations in multiple forms, and that it is the archivists job to link these records together by first understanding the topic to be documented and the documentation that should exist to provide future generations with an adequate record of the topic (p. 122). Archivists then set about creating finding aids that link existing records together, rather than collecting those records together in one place.

Documentation strategy would clearly include Facebook records as a possible source of value. People increasingly use Facebook to communicate with each other about a particular issue, activity, or geographic area, to plan and organize events, and to document their own reactions.

Over twenty years after Samuels, in a review of attempts to implement documentation strategy, Malkmus found that organizations that participated in documentation strategy experiments lacked adequate intellectual control of their collections and "could not or would not amend their collecting policies" (2008, p. 388). In addition, Malkmus found general agreement that the initial analysis of what should be collected in creating a documentation strategy was expensive to conduct (p. 409).

Despite budget and organizational challenges, documentation strategy was effective when seen as "more than an exercise in collection analysis, it has enormous potential to bring excitement, energy, and expertise from the community being documented to the challenge of collecting a representative record" (2008, p. 409).

Documentation strategy was also effective when the topic of the strategy was coincident with the mission of the organization managing the strategy.

By analyzing existing attempts at documentation strategy, Malkmus helps us understand that asking an existing archive to participate in a resource-intensive project that is outside of that archive's mission is not likely to succeed. Establishing an organization that will work with the community that is being documented, and that enjoys reliable funding, will work not only to identify and preserve historically significant records, but will help to promote understanding of the challenges of digital preservation and the importance of the historical record.

Cook offers a different approach to appraisal in a post-custodial era. Instead of focusing on a particular topic and the records that a panel of experts has agreed would adequately document that topic, Cook posits a theory of records appraisal called macro-appraisal that focuses less on the individual record and instead on the "context of their creation and contemporary use" (2005, p. 102). Macro-appraisal is a "meta" theory that has developed over time at the National Archives of Canada and offers a framework for managers, future generations, and others to evaluate the appraisal practices of archivists (p. 159).

Cook requires the archivist to consider both the position of the record creator within the organization and the anticipated use of the records by researchers in the appraisal of a record. Cook describes macro-appraisal as an appraisal theory that combines an analysis of the structural-functional roles of the organization and the culture of the organization in order to determine which records to retain:

> [M]acroappraisal assesses the societal value of both the functional-structural context and work-place culture in which the records are created and used by their

> creator(s), and the interrelationship of citizens, groups, organizations – ‘‘the public’’ – with that functional-structural context. (p. 101)

In his description of macro-appraisal, Cook focuses on the tension between the individual and government, and human agency and the structure of organizations, and suggests that these “hot spots” are the best places to look for evidence of the functioning of government.

Some of these “hot spots” will clearly be found on Facebook. Many governmental organizations including the U.S. Supreme Court, the U.S. Senate and House of Representatives, The White House, elected officials such as President Barack Obama, former Presidents George W. Bush and Bill Clinton, Congressmen Lamar Smith and Eric Cantor, Senator Harry Reid, and organizations and committees such as the U.S. House and Senate Judiciary Committees, and the FBI are represented on Facebook. This very brief list describes the range of individuals and organizations that maintain an active Facebook presence and interact with the general public, creating records that may be considered valuable by archivists.

Zinn asserted archivists’ claim to neutrality was a fraud favoring the collection of records created by the wealthy and the powerful, and Colman claimed that archivists were collecting and preserving records that did not represent society, skewing our documentary heritage in favor of the wealthy and powerful. Ham responded in part by declaring a new era, which he referred to as post-custodialism. In this era, archivists had a responsibility not only to their own collections and the mission of their collecting institutions, but to records that represented our documentary heritage, regardless of where those records may exist.

Ham called for a change in the archives profession that would de-couple the preservation of records from the institution that maintained these records, and for more

active participation of archivists in the creation of our documentary heritage. Samuels asserted that archivists should identify the records that would adequately document a particular topic for future generations, regardless of who creates those records or where they may exist. Cook emphasized the interaction of government agencies, organizations, and individual citizens to identify a context in which valuable records are likely to be created. Each of these approaches separates the creation, collection, and preservation of valuable records from the organizations (such as governments and archives) that are charged with the collection and preservation of the records. At the same time, each of these approaches to appraisal came to emphasize the role of the individual record creator, whether as part of an organization or as an individual citizen, in ways that had not been recognized in the past.

Based on this brief discussion of appraisal theory, we see that Facebook is used to create individual records of cultural heritage value and historical significance every day. Some are the records of, or correspondence with, important individuals, groups, and organizations. Other records are first-person accounts of historical events. In the next sections I will discuss the other two possible sources of Facebook records: why Facebook records may be worth preserving because they represent records of personal value and why the volume of Facebook records enhances their research value.

Records of Personal Value

Despite post-custodial appraisal theory, including documentation strategy and macro-appraisal, dissatisfaction exists with the recognition of the importance of personal records within the archival community. Hyry and Onuf describe personal papers as a range of items including video and sound recordings that are "created for personal

reasons, be they communication, artistic endeavor, or other activities not necessarily linked to the production of commodities and services" (1997, p. 38). Hyry and Onuf assert these records are some of the most valuable records in archives (p. 38). Thomas concurs, emphasizing that personal records are an important part of our cultural memory (2007, p. 1). Cox underscores the importance of personal records in *Personal Archives and a New Archival Calling,* "the evidence once looked for in libraries, archives, and museums might be more readily found on various individual and organizational Web sites" (2008, p. 245). At the same time, Cox emphasizes the lack of focus by archivists on Web-based personal archives, "archivists and records managers need to figure out just what their responsibility ought to be with regard to the World Wide Web and the range of people making use of it" (p. 246).

O'Sullivan points to the "recognition on the part of the archival community of the importance of these materials" (2005, p. 137), but clarifies that the appraisal of personal papers has been neglected as an area of study. Hobbs concurs, "recent debates concerning the acquisition and appraisal of records have centred on administrative or government records models" (2001, p.126). Hobbs calls for an appraisal theory and methodology that are specific to personal archives. Pollard also emphasizes the dearth of appraisal theory that relates to personal papers, "the professional literature addressing the appraisal of personal papers is both scant in quantity and lacking in specific guidelines addressing key theoretical questions" (2001, p. 140).

Beagrie (2005) specifically connects archives and the collections of individual records creators when he refers to the use of artifacts and the urge to "express individuality and creativity" as the "foundation and lifeblood of most museum, library, and archive collections" (p. 2). Beagrie discusses the difficulty of preserving born-digital

materials using Internet-based services and the ease of creating large collections due to cheap storage and easy-to-use software. Personal digital collections, according to Beagrie, are "likely to be as significant for future users of historic collections as their paper equivalents are today, providing it survives for future access" (p. 8).

Scholars Richard Cox and David Lowenthal both help us understand the value of Facebook records by emphasizing the connection between personal archives and history, and the transient nature of digital records. Cox specifically addresses the assumption of permanence when he says "we used to believe that the documents we created would outlast us, but now, when we use technologies such as the Web, we have to contemplate much more carefully just what might lie in store for us, and our personal and family archives" (2008, xvi).

Lowenthal further clarifies the connection between memory and history, "The death of each individual totally extinguishes countless memories, whereas history (at least in print) is potentially immortal" (xxii). Lowenthal explores the benefits of these written records and their contribution to cultural heritage by discussing six different categories: familiarity and recognition, reaffirmation and validation, individual and group identity, guidance, enrichment, and escape. (1985, p. 38) Each of these categories underscores the value of Facebook records through the contribution of these records to documenting our times.

Facebook records could help future generations make sense of their present by reminding them how we communicate with each other, whether through the pictures, milestones, comments, or status updates we share on Facebook. Preserved Facebook records can also validate attitudes and actions in the future by allowing for the continuation of traditions or by reminding people of traditions that once existed.

The sense of identity of future generations can also be supported by Facebook records. The ability to look back and see how people connected with one another, to understand familial connections and friendships helps to validate ones sense of identity and "gives existence meaning, purpose, and value" (p. 41).

Facebook records can also offer guidance and enrichment, sharing mistakes, successes, and improving future generations' appreciation of the world around them by connecting them with the emotions and history of the past, lengthening and deepening their experience of the present.

Lastly, Facebook records can allow future generations the opportunity to imagine a different present. Access to the pictures, videos, and social interactions of the past can allow individuals the freedom to imagine changes to present circumstances.

Individual Facebook subscribers determine value as they work with their personal digital archive; creating, capturing, organizing, keeping and destroying records over time (McKemish, 1996, p. 28) based on personal criteria the individual record creator establishes.

Some of these records are examples of subscribers' communicating with each other in intensely personal ways, whether the correspondence is related to general or mental health, family matters, or sexual or romantic activity. Anecdotal evidence and a survey from a UK divorce Web site shows that Facebook has been cited as a reason in 30% of UK divorces in 2011 (Divorce Online, 2011). Clearly Facebook subscribers are using the service to change their relationship status and find new partners.

Last year in the United States, Facebook and the National Suicide Prevention Lifeline launched a partnership that will allow subscribers who see a comment or suspect

a friend is depressed to contact a suicide counselor directly or to have a suicide counselor contact the friend (Facebook, 2011).

At this moment a group of Facebook friends could be disrupting a potential suicide attempt, planning a revolution, or becoming complicit in a sexual affair. All of these types of communication have the potential to create what privacy scholar Daniel Solove refers to as "privacy problems," a concept I will discuss in detail in the next chapter. The sensitive nature of these records, the first-person documentation of historical events, and the variety of individual voices and recorded emotions make Facebook records valuable personal records as well as valuable additions to our cultural heritage and worth preserving.

The Research Value of the Volume of Facebook Records

In addition to the cultural heritage and personal reasons listed above, Facebook records are worth preserving to support a range of different research activities. Some of these activities have already been discussed because the research methods involved have existed before Facebook and other social networking services. Other research activities are consistent with previous research conducted in a variety of academic disciplines, social networking in these cases expands existing research opportunities. Murthy summarized the role of social media in ethnographic research as including gatekeepers with chains of friends who are potential research respondents, as collections of data from a wide range of social groups, offering researchers the ability to observe "invisibly," and the structure of relationships represented by social media software are worthy of research (2008, p. 845). In this section I will focus on the new research challenges and

opportunities created by the volume of digital records that are being created by Facebook subscribers.

An increasingly high volume of correspondence is being conducted using social networking services, a volume of correspondence that is unique in the history of communications technology. Young people from ages 18 – 29 are using social networking sites at rates of 86% (Brenner, 2012), and the use of social networking among adults aged 50 and over has doubled from 22% in 2009 (Madden, 2010) to 50% in 2012 (Brenner, 2012). Among those who use social networking services, 92% use Facebook (Brenner, 2012). Facebook is the most popular social networking service in the United States and quickly becoming more popular in other countries at the time of this writing (Alexa, n.d.). Facebook has over one billion subscribers worldwide (Facebook, n.d. a) and approximately 156 million subscribers in the U.S., with over half of these subscribers logging in each day. Subscribers have a mean number of 130 friends, spend an average of 15 hours and 33 minutes on Facebook per month, and create "90 pieces of content each month" (Facebook, n.d. c).

When a billion people are using the same information communication technology to correspond with one another, a large, consistent data set is being created. A single company is determining the programming languages, documentation, server types, and other infrastructure used to support the software, the types of data and metadata that are being collected, and how these data are being stored. This consistency is beneficial to researchers by making it less expensive to search across a broad set of data maintained by Facebook than it would be to combine multiple different data sources to achieve the same effect. Consistency in this context can also create problems for researchers in that the infrastructure designed to handle large volumes of subscribers are often designed to

accomplish very specific tasks, limiting the flexibility a researcher may need to investigate new connections that were not part of Facebook's corporate priorities.

Researchers and archivists are just beginning to understand the value in such a large dataset. The data analysis tools that researchers have used in the past are inadequate, and the volume of these data challenges the assumptions and contributions of traditional academic disciplines. Until researchers and academics understand the value of these data to their disciplines, archivists will have a difficult time appraising the value of the entire corpus of Facebook and other social networking records.

Private corporations and government agencies have been developing tools that allow them to analyze these large volumes of social networking data, sometimes referred to as "big data." Facebook's primary source of revenue, as I discussed earlier, is selling advertising space based on subscribers' behavior. Facebook collects and analyzes subscribers' data and sells advertising space based on the demographic and behavioral characteristics of subscribers.

Raytheon has developed software that can analyze large social networking datasets for surveillance purposes (Gallagher, 2013). The software is named Rapid Information Overlay Technology, or RIOT, and is part of a new category of software called extreme-scale analytics. The software tracks subscribers' actions across social networks, including Facebook, Twitter, and Foursquare, and allows analysts to predict future behavior. RIOT is a tool that is geared toward law-enforcement surveillance and the prevention of terrorist attacks, but the development of this type of military technology has implications for the future possibilities of digital humanities research such as analyzing status updates to understand reactions to historical events, to better understand citizen access to government services, or to determine the most effective approach to

educate young people about protecting their privacy and preserving their records while using social networking services.

"Big data" has been used for a variety of different "real-time" and predictive analysis. Algorithmic criminology is one field that is developing computer algorithms to predict the likelihood that an individual will commit a crime in the future. Pennsylvania recently passed a statute that allows for the development of a tool to provide judges with "quantitative forecasts of risk" to use in sentencing (Berk & Bleich, 2013, p. 1). Google's flu trends application uses search terms submitted by Google users to determine where a flu outbreak may be taking place, sometimes a week to ten days faster than the Center for Disease Control can reach the same conclusion (Helft, 2008). One problem with these tools and the corporate analysis of "big data" is that context can be ignored, and in the rush to find a fast answer to a simple question, corporate researchers can end up misunderstanding the results. In the case of Google flu trends, the estimate of actual flu infections can be off by 200% because the algorithm did not take into account people searching for flu symptoms because they are worried about the flu, not because they actually have the flu (Bilton, 2013). Corporations and government agencies are still working on improving the analysis tools, but the potential value of the volume of records generated by social media to predict crime, terrorism, health risks, pregnancy, and insurance risks is being recognized.

Some progress is being made in determining the value of these records for research. The recent partnership between the Library of Congress and Twitter offers an example of the value and challenges of a large data set of social networking records for research. Twitter has contracted with a third party, the Library of Congress, in the interest of preserving records. Twitter is a micro-blogging service that allows subscribers to send

140-character status updates called “tweets.” Twitter has agreed to donate all public tweets, which amount to over 50 million per day, to the Library of Congress for preservation (Library of Congress, 2010). The Library of Congress’ collection of Tweets “are likely to be of considerable value to future historians. They contain more observations, recorded at the same times by more people, than ever preserved in any medium before” (Stross, 2010). Costello and Priem claim there have already been some minor successes, such as predicting movie revenue based on Tweets (2011, p. 8).

Despite these minor successes, it is clear that the research tools necessary to analyze large datasets are still being developed. As the Library of Congress’ Erin Allen points out, “Even the private sector has not yet implemented cost-effective commercial solutions because of the complexity and resource requirements of such a task” (Allen, 2013).

Tools to analyze large data sets for humanities research are still in their infancy. A team of researchers at the Texas Advanced Computing Center recently developed a set of visualization tools that allow researchers to work with very large data sets. This new set of tools is specifically geared toward the humanities researcher because existing tools “can be complicated and ill-suited to humanities research” (TACC, n.d.). This toolset resulted in the team receiving the “Best Digital Humanities Visualization” award for 2012 from the European Association for Digital Humanities, underscoring the influence of large datasets on ongoing changes in humanities research, and the need for new tools to support this research.

Not only are these data sets challenging the technical abilities of researchers, they are forcing researchers to ask tough questions about the value of their academic discipline and how these data may be changing the role of research. Savage and Burrows asserted

that the generation of social media data by subscribers and service providers, and the use of these data by service providers and advertisers, calls into question the role of sociologists as researchers (2007, p. 886). Savage and Burrows called for sociologists to forego the use of sampling techniques and instead study whole populations since those data are now routinely collected. The authors emphasize, "in the current situation, where data on whole populations are routinely gathered as a by-product of institutional transactions, the sample survey seems a very poor instrument" (2007, p. 891).

At a 2008 Sociology journal board meeting, board members were surprised by the number of downloads and citations of the Savage and Burrows article and asked for a review of the article and subsequent responses. McKie and Ryan conducted the review and asserted that social media offer methodological opportunities and challenges for sociology (2012, p. 1). Social media are generating new kinds of transactional data that were not available before, creating new social categories and constructs, enacting populations and producing subjects (p. 2). The increased use of social media poses a number of challenges to sociological research including the range and availability of data, the pace of change in social media and the comparatively lengthy funding process for research projects, and the increased number of non-sociologists who are conducting sociological research (p. 2).

The Library of Congress offers some insight into the types of questions that will be asked of the Twitter corpus, "Twitter will enable future researchers access to a fuller picture of today's cultural norms, dialogue, trends and events to inform scholarship, the legislative process, new works of authorship, education and other purposes" (Allen, 2013). To gain access to the Twitter archive, a researcher needs to submit a research

proposal to the Library of Congress, a step that helps protect the privacy of the record creators by limiting access to the records.

Library of Congress has received research inquiries related to predicting stock market trends, tracking flu outbreaks, analyzing elected officials' communications related to public policies, tracking vaccination rates, and analyzing access to the court system. The Library of Congress highlights two questions that researchers would not be able to address without the Twitter archive:

1. A master's student is interested in understanding the role of citizens in disruptive events. The student is focusing on real-time micro-blogging of terrorist attacks. The questions focus on the timeliness and accuracy of tweets during specified events.
2. A post-doctoral researcher is looking at the language used to spread information about charities' activities and solicitations via social media during and immediately following natural disasters. The questions focus on audience targets and effectiveness (Allen, 2013).

Scientists are just now creating the tools necessary to analyze large data sets, and the creation and analysis of these data by social networking and advertising firms is causing theorists in a variety of academic disciplines to ask difficult questions about the role and value of academic research. If Facebook were to stop operations today, the data this corporation has collected over the last eight years alone would represent one of the largest data sets available to researchers from a wide variety of fields. The number of Facebook subscribers and the volume of records created using Facebook make it a unique communication technology that warrants research attention.

While these questions continue to be discussed, the corpus of records from a growing list of major social networking services such as Facebook, LinkedIn, and MySpace are at risk of being lost. While it may not seem like a company with one billion subscribers will go out of business any time soon, it is possible that Facebook subscribers could be convinced to migrate from Facebook to some other service, which is what happened to MySpace when Facebook was introduced. MySpace managed to survive this massive migration in part because Rupert Murdoch's News Corporation purchased it just before the launch of Facebook but its value dramatically dropped: News Corporation purchased MySpace for $580 million in 2005, and sold it for $35 million in 2011 (Rushe, 2011).

Social networking records face other threats as well. In chapter five I discuss how the U.S. Department of Justice shut down the file sharing service MegaUpload, immediately disabling access to all of the company's servers and freezing the company's assets. As a result, the server hosting company threatened to turn off the servers, essentially destroying the data. While MegaUpload is not considered a social networking company, similar copyright concerns and risks exist for companies like Facebook that allow subscribers to upload documents, images, and videos.

Another, more likely risk of loss comes from more subtle threats. Social networking services face fierce competition from large, existing companies to smaller start-ups. This competition requires social networking services to adjust existing service offerings and to bring new services to market very rapidly. These software changes are not developed with archives and preservation in mind, the primary focus is rapidly adding new services or changing existing services. As existing services are retired or changed, the records created using the "old" version may not be accessible. As MySpace changes

from a social networking company for teenagers to a music file sharing service to a social networking company for thirty year-olds, records created with each version of the software are threatened by the transition.

As researchers work to gain a better understanding of whether and how the volume of records created by social networking services will affect the future of academic disciplines, archivists should work to preserve these records while they are still available.

Facebook records may be valuable as cultural heritage objects, personal records, and the volume of Facebook records may be valuable for research. I offered a brief overview of appraisal theory that included a discussion of Jenkinson and the idea of passive appraisal, which is similar to the approach the Library of Congress has taken with Twitter. I discussed Schellenberg and the ideas of primary and secondary value, and how the idea of the secondary value of Facebook records may not be receiving attention. I also discussed a variety of different methods archivists have used to appraise records in the past that may be applicable to Facebook records, as well as the idea of post-custodialism, which will allow us to consider methods of preserving Facebook records without having to have all of those records located in a single repository.

The archival community has called for a closer examination of the value of personal digital records, and, while research related to blogs and Web sites has been published, very little research exists regarding the value of social networking records.

With one billion subscribers creating records using a single software platform, Facebook and other social networking tools have raised new technical and research questions for scholars in many different disciplines. We do not know how long Facebook will continue to exist as a company, or how long current Facebook records will continue

to exist as a corpus of records without intervention. Since researchers and software developers are creating new tools and imagining new questions, the value of the corpus of Facebook records is still being determined, but these records should be protected while the discussion of value takes place.

This chapter offers one of the four major contributions of this research; a discussion of the value of Facebook records and why they might be worth preserving. Privacy scholars claim that the persistence of social networking records represents a threat to privacy. This chapter serves to complicate this claim by exploring how the destruction of these records may threaten valuable cultural heritage objects, personal memory, and potentially valuable data set. Archivists have a long history of protecting privacy while preserving important records, but even the most privacy protective preservation program would not directly address the threat to privacy represented by the loss of access control over social networking records. As soon as a record is created using Facebook, that record may be copied, published, or preserved by anyone who has access to that record without the record creator's knowledge. The concerns of privacy scholars and the protection of privacy by archivists will be explored in more detail in Chapter 3.

In the next chapter I will discuss relevant social networking and privacy research. I will discuss the types of information subscribers share and the various reasons subscribers use social networking services. I will explore a variety of different ways to define privacy, and discuss three different definitions in more detail: defining privacy based on the privacy problems that one faces, defining privacy as a way to control access to records, and a brief discussion of the right to delete as a privacy protecting mechanism. I will also discuss subscribers' perception of risk and the different ways subscribers protect their privacy when using social networking services.

Chapter 2: Social Networking and Privacy

In the last chapter I discussed how I will use the term "record" throughout this study and how Facebook records differ from paper-based records. I then established that Facebook records are worth preserving for three reasons: Facebook records represent cultural heritage objects, Facebook records have personal value to individuals, and the volume of records created using Facebook is valuable for research. In this chapter I discuss the relevant social networking research that will help us understand why people use social networking and how social networking subscribers understand privacy and risk. I review ways of understanding privacy, and discuss in more detail two approaches to privacy that have been closely related to social networking and digital communication technology. This chapter serves to ground my study in previous research and to more clearly define the context of social networking and privacy for the next chapter where I will discuss the differing opinions of archivists and privacy experts on the preservation of Facebook records and the concomitant threat to privacy.

SOCIAL NETWORKING

There exists a substantial body of academic research related to social networking as an information and communication technology. In this section I have chosen to discuss those studies that help us understand social networking as a tool for correspondence, what some might call the next iteration of postal mail or the telegraph. While there is little published research specifically examining social networking and digital preservation, these studies help us connect the existing social networking literature with existing archival theory.

Pew Senior Research Analyst Amanda Lenhart describes social networks as "web spaces where individuals can post information about themselves, usually by creating a profile or website, and where they can connect with others in the same network" (2006, p. 2). A number of researchers have found that Facebook subscribers tend to use Facebook to communicate with people with whom they already have an established relationship (Gross & Acquisti, 2005; Lamp, Ellison, & Steinfield, 2007; Subrahmanyam & Greenfield, 2008). This finding is important because it ties the use of Facebook for interpersonal communication to older forms of correspondence such as paper-based, postal correspondence. These older forms of correspondence have a rich history of study by archives and privacy researchers. Scholars have also explored the use of Facebook correspondence for identity creation and exploration (Subrahmanyam & Greenfield, 2008), activities that underscore the importance of Facebook records as both personal memory and cultural heritage objects.

Lampe, Ellison & Steinfield of Michigan State University conducted a three-year study of a random sample of students between 2006 and 2008, with 288, 468, and 419 respondents each year respectively and follow up interviews with a subset of respondents each year. Respondents used Facebook at a relatively constant rate over the three-year period, and they used Facebook for three main reasons: to keep in touch with people they knew previously, to learn more about someone they met outside of Facebook, or to learn more about other students in their classes. Respondents' expectations of who would view their profiles remained relatively constant as well; with most respondents expecting someone they had met face-to-face outside of their use of Facebook as most likely to view their profiles.

This study by Lampe et al. helps us understand why people use Facebook and why they share different types of information. Specifically, the respondents in this study were using Facebook in order to communicate with people they had met face to face. Facebook acted as a common communication platform for existing social groups.

Subrahmanyam and Greenfield relate social networking records directly to cultural heritage when they equate adolescent exploration online to identity construction. Subrahmanyam and Greenfield build on psychologist John Hill's framework for early adolescence, which includes four developmental tasks (identity, autonomy, intimacy, and sexuality) and the variables and factors that influence these tasks (2008, p. 124). The researchers assert that social media link the physical and virtual worlds psychologically and act as a playground for development. "Thus understanding how online communication affects adolescents' relationships requires us to examine how technology shapes two important tasks of adolescence—establishing interpersonal connections and constructing identity" (p. 124). Subrahmaynam and Greenfield find that most adolescents use social media to communicate with individuals they are familiar with in real space.

Seounmi Youn (2005) conducted an online survey of 326 high school students looking at whether respondents' perception of risk affected their disclosure behavior. Youn used Rogers' (1975, 1983) protection motivation theory as the theoretical foundation for her study. Rogers' protection motivation theory asserts that individuals are motivated to protect themselves from risk or harm if they perceive that risk or harm is likely to happen to them, likely to be severe, the recommended protection is effective, and the individual believes in her ability to execute the recommended response to the risk or harm. Youn determined that the respondents in her sample were less willing to provide information to Web sites that the respondents perceived as a risk to privacy, and at the

same time more willing to provide information if they perceived a benefit to disclosing that information. "Subsequently, as teenagers were less likely to give out their information, they tended to engage in several risk-reducing strategies such as falsifying information, providing incomplete information, or going to alternative Web sites that do not ask for personal information" (Youn, 2009, p. 86) (see Rogers, 1983, pp. 167–172).

Facebook subscribers are exposed to two different and conflicting risks: a risk that their privacy will be violated and a risk that their records will not continue to exist for long periods of time. Two criteria from Rogers' protection motivation theory are directly relevant to individuals acting to preserve their Facebook records – the perceived likelihood that the threat will occur to them and the perceived severity of the threat. Since a clear recommendation of how a social networking subscriber should preserve her records does not exist, the last two criteria, the effectiveness of the solution and an individual's assessment of her ability to implement that solution, are not applicable.

The perception of risk and the severity of risk are important in discussing individual subscribers' expectations and whether public policy should play a role in the preservation of social networking records. If social networking subscribers do not perceive the risk to the cultural record posed by the potentially short life span of corporations such as Facebook, the proprietary format of digital social networking records, or the complexity of long-term preservation of digital records, the need for public policy to protect these records is clearer. A change in public policy may also be required if an asymmetry of information exists where Facebook subscribers are unaware of how Facebook manages their records or how the long-term management of records might create privacy problems for them in the future.

The body of research related to social networking offers a number of useful concepts upon which this study builds. Previous research finds that social networking subscribers use these services to maintain existing relationships and that they choose how they interact with these services based on their perceptions of risk. Younger subscribers also use these services to explore and develop their identities. These findings allow us to see more clearly that social networking records are not an entirely new type of record; they are similar to the correspondence records that archives have been managing for centuries, albeit with some new challenges.

PRIVACY

In this next section I explore a variety of ways to understand privacy and highlight the definitions and concepts of privacy that are most relevant to social networking and digital preservation. While there certainly is no consensus about what privacy is, a brief discussion of different ways of understanding privacy and how some of these ways of understanding have evolved will add context to the discussion of the preservation of digital records and the implications of those records for privacy. In addition, this section offers background information on privacy to archivists in order to facilitate conversations between archivists and privacy scholars. By sharing the terminology and concepts used by privacy scholars with the archival community, I hope to improve the quality of discourse between the two groups.

From social norms to statutory and case law to privacy frameworks, we see that our understanding of privacy changes over time, and often changes in response to changes in technology or the use of technology in society. I discuss three different concepts in my review of the privacy literature: the definition of privacy, the protection

of privacy, and threats to privacy. These three concepts are certainly interrelated; Daniel Solove uses threats to privacy, or to use his terminology, "privacy problems" (2008, p.9) to define privacy. The definition of privacy and threats to privacy are one and the same, according to Solove. The interrelation of the definition of privacy and the protection of privacy is true to different extents with each of the authors I discuss, but the separation of the definition of privacy from threats to and protections of privacy will be useful when I discuss possible approaches to both preserving Facebook records and protecting the privacy of Facebook subscribers.

After this brief overview, I discuss three privacy concepts useful to the analysis of preservation and privacy: the privacy framework described by Daniel Solove (2008), the framework first described by Moor (1997) and further developed by Tavani (2007), and the concept of the right to delete. Moor and Tavani will help us separate the definition of privacy from the protection of privacy, allowing the discussion of what privacy is to be separated from and how it should be protected, a separation that allows privacy protecting tools and policies to be responsive to changing definitions of privacy. Solove helps us by discussing what privacy should be protected from, and the right to delete is a concept that has not historically been an integral part of the privacy literature, but is increasingly recognized as a tool in protecting privacy.

The Social Construction of Privacy

There are several ways to understand privacy, such as social norms, statutory law, case law, and "privacy frameworks," but what we regard as privacy is socially constructed. Different cultures understand privacy differently, an understanding that changes over time and that is not necessarily internally consistent nor generally accepted

within a culture at any particular time. The concept of privacy is also constantly negotiated among individuals and among groups, and the results of these negotiations are situation dependent. For example, standing close to a stranger on the subway with your hand touching hers as you hold the sway bar may be acceptable during rush hour, but, at 11:00 AM when the subway is nearly empty, this same physical proximity would most likely be regarded by others as a privacy intrusion.

I discuss social norms, statutory law, and case law related to privacy in more detail in the following section. Social norms can represent the values of a culture and shape statutory and case law. Statutory and case law that protects privacy can also be said to define privacy. These laws circumscribe a range of activities or states that are normatively or descriptively considered "private." I will refer to multifaceted discussions of privacy by legal, privacy, or information scholars as privacy frameworks.

Social Norms

Social norms can be said to represent the values and beliefs of a culture and to vary over time. Changes in social norms can indicate or precipitate changes in the concept of privacy. New statutory and case law may shape and influence social norms, just as changing social norms may inspire legislators to craft new statutes and judges to interpret existing laws differently, especially in the common law tradition.

Literature, popular culture, and the news media can have a strong influence on establishing norms. Our shared experience with popular culture often influences the range of topics we are comfortable discussing in public. Academic research can also influence social norms and the definition of privacy, particularly if that research finds its way into

popular culture. The first Kinsey report, *Sexual Behavior in the Human Male* in 1949 (Pomeroy, Martin, & Kinsey, 1949), is one such example.

Events that garner national attention, such as terrorist attacks, also can set or change social norms. The attacks of September 11, 2001, had a profound effect on the general perception of safety and security among United States citizens. There were a number of anecdotal reports of changes in behavior that could be indicators of changes in social norms: a shortage of American flags, increased sales of gas masks, increased sales of duct tape, and an increase in the creation of neighborhood watch programs.

Less sudden, but equally profound changes in social norms are currently taking place according to Palfrey and Gasser in *Born Digital* (2008). Palfrey and Gasser talk about the effects that the Internet, social networking, and pervasive access to digital technology such as cell phones and digital video cameras are having on the current generation of young people, whom they refer to as Digital Natives. "Digital Natives' ideas about privacy, for instance, are different from those of their parents and grandparents" (p. 7). It may be that these new ideas about privacy become the social norms of the near future, or it may be that these young people have new ideas about technology and how social networking can be used to explore the same developmental milestones that have made past generations of parents shake their heads and say "kids these days."

Statutory Law

In addition to, and often in conjunction with, social norms, statutory law helps us to understand privacy, and this section discusses a small number of such statutes in the United States. The statutes I have chosen to discuss are examples of the range of privacy

statutes in the U.S. Some of these statutes are reactions to specific events like the disclosure of video records or terrorist attacks, others are omnibus bills that attempt to address a range of privacy and communication challenges. The ideas of privacy and defending privacy from threats have deep roots in human history. Societies have used law to protect what we call privacy since biblical times, and businesses, governments, and individuals have violated these laws for equally as long. Carpenter and Meriweather explored the roots of privacy in their 2000 Supreme Court brief in *Kyllo v. United States*. Looking as far back as the Code of Hammurabi the authors found reference to special protections of the home. Article 21 of the Code of Hammurabi states "If any one break a hole into a house (break in to steal), he shall be put to death before that hole and be buried" (King, 2004, p. 5). The protection of the home as private makes up one part of our contemporary legal protections of privacy. I offer a brief overview of the range of contemporary privacy protection statutes below. This list is far from exhaustive, but serves to represent the variety of different ways the United States has chosen to legally protect privacy.

The Family Educational Rights and Privacy Act (FERPA) of 1965 (20 U.S.C. § 1232, 1965) protects the privacy of students' educational records, giving students increased control over the use and release of their educational records and personal information.

The Freedom of Information Act (FOIA) of 1966 (5 U.S.C. § 552, 1966) created a mechanism for citizens, citizen groups, and others to request access to government records, thereby increasing government transparency. Included in the FOIA are two exemptions that are meant to protect the privacy of subjects identified in the records. Government records are not required to be released under the FOIA if those records were

collected for law enforcement purposes or as part of a personnel or medical file. The Omnibus Crime Control and Safe Streets Act of 1968 (42 U.S.C. § 3789, 1968) protected privacy by requiring federal agents to acquire a warrant prior to installing electronic surveillance devices and created other regulations governing how electronic surveillance could be conducted.

In 1978 both the Foreign Intelligence Surveillance Act (FISA) (50 USC § 36) and the Right to Financial Privacy (12 U.S.C. § 35) were passed by Congress. FISA created guidelines for surveillance of electronic communications and established the FISA court. The initiation of surveillance of any suspected intelligence agents working within the United States for a foreign power required prior approval of the FISA court in the form of a warrant. The Right to Financial Privacy Act sets the requirements for the federal government to access financial records of individuals. The Right to Financial Privacy established that financial records were the property of the bank, not the subject of the records. The Right to Financial Privacy was preceded in 1976 by *United States v. Miller*, where the Supreme Court held that individuals did not have a reasonable expectation of privacy in regard to their financial records because they have already disclosed these records to a third party (Solove, 2011, p. 104).

The Electronic Communication Privacy Act (ECPA, 18 USC § 2510) was established in 1986 and provides "strong protections of privacy" (Solove b, p. 73). ECPA distinguishes between protecting electronic communications in transit and stored communications, and provides greater privacy protection for communications in transit. Law enforcement agents are required to submit proof that surveillance will reveal evidence of a crime, will have minimal effect on the privacy of subjects that are not under investigation, and that other means of surveillance are not available to law enforcement.

In addition, the subject under investigation will eventually be notified that she was the subject of surveillance.

The 1988 Video Privacy Protection Act (VPPA, 18 USC § 2710) was signed by President Ronald Reagan after Supreme Court nominee Robert Bork had his video rental history published by a local Washington, D.C., newspaper (EPIC, n.d.). The VPPA protects personal video rental history from being shared without the renters consent, requires law enforcement to obtain a search warrant or court order to obtain these records, but allows for third party advertising agencies to obtain genre information for marketing purposes. The law allows for civil remedies if there is a violation, and requires evidence obtained in violation of the law to be excluded from any court proceedings.

The Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism or USA PATRIOT Act (PL 107-56) is another statutory law that helps us understand privacy. The USA PATRIOT Act is a broad law that changes a variety of existing legislation and allowed for increased authority of law enforcement, especially in relation to surveillance. Law enforcement agents have been granted the authority to include "any tangible thing" they may encounter while executing a search warrant, even if the search warrant did not include that particular thing. The USA PATRIOT Act also included a gag order rule that required secrecy by limiting the information parties served by a warrant could reveal. The USA PATRIOT Act also allows for the sharing of information across organizational boundaries (Relyea, 2002), which increases the information an agency may have access to.

The list of examples above, while only indicative and far from exhaustive, offers a brief overview of different statutes that can help to understand privacy.

Case Law

Case law is yet another way to understand privacy. Case law codifies the societal interpretation of privacy in the context of the court case that is being decided. While this interpretation can change over time, the decisions made by courts help privacy scholars and society understand the boundaries of privacy. I include both constitutional law and privacy torts in my review of case law, and I focus specifically on a few examples that demonstrate the protection of privacy related to communication and access to information.

American Constitutional law is a contract made between the federal government and the citizens of the United States, defining the relationship between the federal government and the citizens. The Supreme Court has established Constitutional privacy law through a long series of often-conflicting decisions.

The U.S. mail offers an early example of communication technology leading to a reassessment of the legal protection of privacy. The first sealed envelopes became widely available in 1845, increasing the public expectation of postal privacy (Pope, 1997). In 1878, the Supreme Court tested this privacy expectation in *Ex Parte Jackson* (96 U.S. 727). The court ruled that sealed letters, packages, and envelopes are protected by the Fourth Amendment from search, whereas open format mail such as newspapers, magazines, and pamphlets that can be read without intrusion into a sealed envelope are not protected.

One of the first major Supreme Court cases directly related to privacy was *Olmstead v. United States* in 1928 (277 U.S. 438). *Olmstead* focused on whether a wiretap without court oversight was a violation of the defendant's Fourth and Fifth Amendment rights. Justice Taft wrote the opinion of the court stating that the method of

acquisition of evidence did not have an effect on whether that evidence can be used at trial. The court upheld the conviction; evidence obtained through wiretapping without judicial oversight was admissible.

The common approach to evidence gathering that law enforcement had taken to enforce prohibition laws was affirmed by the Supreme Court; wiretapping was not an illegal search and seizure since the wires that carried the conversation were outside the home or office. The Fourth and Fifth Amendments would cover only the private physical space; the conversation was not protected. Justice Brandeis wrote a dissenting opinion in this case that would later be used in *Katz v. United States* (389 U.S. 347, 1967) to overturn this precedent.

The Supreme Court decision in *Katz* changed the legal definition of privacy from a protection of physical space as defined in *Olmstead* to a protection afforded the individual regardless of location. This protection is limited and determined by whether the individual has made an effort to protect the privacy of the conversation, by stepping into a phone booth, for example, and whether he enjoys an expectation of privacy that society would find reasonable.

Clearly this standard is problematic, relying on the court system to determine what "society" feels is a reasonable expectation of privacy invariably privileges some groups over others, and as Facebook subscribers become more aware of the threats to privacy they encounter by using Facebook and other social networking tools, their expectations of privacy diminish, which would consequently reduce the protection of privacy.

Griswold v Connecticut in 1965 tested a Connecticut law that prohibited married couples from the use of contraception and others from counseling married couples about

the use of contraception. Justice Douglas wrote the opinion for the court in which he indicated, "The Connecticut statute forbidding use of contraceptives violates the right of marital privacy which is within the penumbra of specific guarantees of the Bill of Rights" (381 U.S. 479). This was one of the first Supreme Court decisions that enumerated a right to privacy in the constitution, even though the right to privacy was not specifically articulated in the Constitution. *Eisenstadt v. Baird* in 1972 (405 U.S. 438) extended the Griswold decision to unmarried couples, and in 1973 *Roe v. Wade* (410 U.S. 113) extended privacy to a woman's right to choose an abortion during the early part of her pregnancy.

These brief examples of Supreme Court cases offer a broad overview of the ways case law can help us understand privacy. These examples represent only a limited number of U.S. Supreme Court cases. There are far more Supreme Court cases and a broad variety of court cases that have been decided before reaching the Supreme Court and many others that have been decided at various levels of the different state court systems. As we see from these limited examples of case law, the judicial branch interprets statutory law, creating an interpretation of privacy. The meaning of a statute and how that law should be applied changes over time, often influenced by social norms and case law. These three elements, social norms, statutory law and case law, have been combined by privacy scholars in different ways to explain privacy in what I refer to as privacy frameworks.

Privacy Frameworks

A number of scholars have developed explanations of privacy that include many of the components that we have already discussed. These privacy frameworks are

valuable to the current study because they provide archivists some examples of how privacy scholars combine these components, allowing for a more robust and complex discussion of privacy than any interpretation of privacy discussed so far. The work of the scholars I have chosen to discuss represents a variety of different explanations of privacy that have been accepted at different times through history. This list is by no means a comprehensive one, but rather illustrative of the kinds of arguments that privacy scholars erecting frameworks typically make.

One of the most well-known privacy frameworks is discussed in an article written by Samuel Warren and Louis Brandeis for the Harvard Law Review in 1890. This article laid out what was popularly referred to as the "right to be let alone" (Warren & Brandeis, 1890) as a definition of privacy. Warren and Brandeis explain that the common law protection of privacy is the inevitable result of the progress of civilization, beginning with a time when "the 'right to life' served only to protect the subject from battery in its various forms; liberty meant freedom from actual restraint; and the right to property secured to the individual his lands and his cattle" (p. 1).

To define the limits of privacy and the remedies for invasion of privacy, Warren and Brandeis refer to existing "legal analogies already developed in the law of slander and libel, and in the law of literary and artistic property" (p.10). Combining and reinterpreting existing legal precedent to address a new privacy problem, such as the snap camera and the increase in "gossip" publications is the same process used to create other "privacy frameworks" listed below. This process is not the inevitable result of progress, as Warren and Brandeis claim, but a temporary legal stance that is negotiated by powerful actors such as attorneys, judges, and elected officials, and to some extent reflects the social norms of U.S. culture at a particular point in time.

In William Prosser's discussion of privacy in the *California Law Review* (1960), Prosser reviews the court cases and articles in legal journals that had followed the 1890 Warren and Brandeis article. Prosser finds four distinct torts have emerged to protect privacy: intrusion upon solitude, public disclosure of embarrassing private facts, publicity that places the plaintiff in false light, and appropriation of likeness (p. 389).

Prosser discusses the evolution of these four torts through case law and concludes that in the years since the Warren and Brandeis article was published, the process of developing privacy "has gone on without any plan, without much realization of what is happening or its significance, and without any consideration of its dangers" (p. 423). Prosser asserts that, until the decades prior to the publication of his article, the law has been concerned with establishing whether there is a right to privacy, but ignored the question of how to reasonably protect this right.

We are left to wonder what Prosser would say about the protection of privacy in the use of social networking when he says "the question may well be raised whether there are not some limits, and whether, for example, a lady who insists upon sun-bathing in the nude in her own back yard should really have a cause of action for her humiliation when the neighbors examine her with appreciation and binoculars" (p. 422). Of course, Prince William and Kate Middleton might disagree.

In his forward to Alan Westin's *Privacy and Freedom* (1967), Oscar Ruebhausen refers to the rapid pace of technological change and the strain this pace places on the institutions of society necessitating a thorough examination of existing definitions of privacy and the threats to privacy imposed by new technologies. "It is no surprise that our social and political institutions are now sorely pressed to find the flexibility to utilize new technology effectively while, at the same time, preventing its abuse" (viii). Westin

defined privacy using four different elements: intimacy, anonymity, solitude, and reserve, but it is the process of recognizing a change in the use of technology by society that is important to my research. The adoption of Facebook as a communication tool by hundreds of millions of people around the world is a change of similar magnitude that necessitates the re-examination of privacy, and in the case of Facebook, our approach to the preservation of cultural heritage artifacts.

There were a number of other frameworks of privacy developed in the 1990's. In 1992, Ken Gormley described privacy in the *Wisconsin Law Review* (1992) and emphasized that definitions of privacy are tied to historical events, "the key to understanding legal privacy as it has developed over 100 years of American life, it will be argued, is to understand that its meaning is heavily driven by the events of history" (p.1340). Gormley's description of privacy combined existing legal terms (tort privacy, Fourth Amendment privacy, First Amendment privacy, fundamental-decision privacy, and state constitutional privacy) into a multi-faceted definition of privacy.

In 1997 and 1998 we see two important attempts to define privacy that are in part a response to the growing adoption of email, online databases and e-commerce. Judith DeCew published *In Pursuit of Privacy: Law, Ethics and the Rise of Technology* (1997) where she asserts three categories of privacy that combine to create an overall description of privacy: informational privacy, accessibility privacy, and expressive privacy. In the *Stanford Law Review*, Jerry Kang published "Information Privacy in Cyberspace Transactions" (1998) where he discussed three overlapping concepts that he used to explain privacy: physical space, choice, and flow of information.

James Moor (1997) and Herman Tavani (2007) describe privacy using the concepts of Restricted Access and Limited Control (RALC). The RALC framework

distinguishes the articulation of the concept of privacy from the justification and management of privacy. Moor and Tavani define privacy as an affirmative state; one has privacy in a particular situation if one is protected from "intrusion, interference and information access by others" (Moor, 1997, p. 30).

The RALC framework separates the description of privacy from the protection of privacy. This separation is useful in that it also supports the separation of concern for privacy from protection of privacy, which allows privacy scholars and archivists to discuss mechanisms that allow for the ongoing control of records without having to agree on a unified description of privacy that will endure the decades and centuries archivists would like these records to persist. Explanations of privacy have changed and will most likely continue to change over time. Creating a technical or procedural mechanism that affords records creators ongoing control over access to their records is one way to address the concerns of both privacy scholars and archivists. We have seen this separation in a number of the studies I have discussed above.

The RALC framework allows for the articulation of whether privacy was lost, invaded, or reduced, depending on a number of different contextual elements. In addition, the use of the term "situation" in the definition of privacy recognizes the social construction of privacy, allowing definitions to be context dependent.

The RALC privacy framework will be useful in analyzing whether the preservation of social networking records implicates privacy to the extent that the framework allows for the separation of the definition, justification, and protection of privacy. This theoretical separation is consistent with the social networking research we have already discussed, and allows me to compare the competing claims made by archivists and privacy scholars with the expectations of social networking subscribers.

Solove's Taxonomy of Privacy

In *Understanding Privacy* (2008), Daniel Solove proposes an inductive approach to describing privacy. Rather than discuss privacy deductively by starting with a broad concept and determining whether examples of specific acts are privacy violations according to that concept, Solove takes what he refers to as a "bottom up" (p. 9) approach. He uses the concept of "family resemblance" (p. 42), taken from Ludwig Wittgenstein, to characterize his description of privacy.

This inductive method allows for a pluralistic description of privacy. Each example of a privacy violation does not need to be included in a monolithic, unifying, top-down conception of privacy. Rather, the range of activities that Solove identifies as constituting privacy all have, according to Solove, something in common that allows them to be collectively labeled "privacy."

Solove refers to pragmatists John Dewey and William James when he discusses what should be considered privacy. These "classical pragmatists" (p. 46), as he refers to them, focus on specific situations and embrace pluralistic, context-specific concepts. As Misak emphasizes, "we are always immersed in a context of inquiry, where the decision to be made is a decision about what to believe from here, not what to believe were we able to start from scratch" (Misak, 2007, p. 3). Solove uses pragmatism when he decides to focus first on privacy problems rather than a totalizing definition of privacy. By focusing on problems rather than a top-down definition, context becomes important to describing privacy. An action taken in one context may be an egregious violation of privacy, whereas in another context the action may not be a privacy violation at all.

According to Solove, a privacy problem exists when particular "activities" (p. 9) are disrupted. Solove's taxonomy facilitates the discussion of the role of context in

privacy problems, which will be particularly useful as we examine whether privacy is implicated by digital preservation of social networking records.

Solove proposes four "groups of activities" (p.10) that create privacy problems: information collection, information processing, information dissemination, and invasion. Solove uses the term "activity" to refer to both the action that violates privacy, such as surveillance, and the act that is infringed upon or inhibited, such as personal correspondence. "Each group encompasses a variety of activities that can create privacy problems." (p. 10):

1. Information Collection
    a. Surveillance
    b. Interrogation
2. Information Processing
    a. Aggregation
    b. Identification
    c. Insecurity
    d. Secondary use
    e. Exclusion
3. Information Dissemination
    a. Breach of confidentiality
    b. Disclosure
    c. Exposure
    d. Increased accessibility
    e. Blackmail
    f. Appropriation
    g. Distortion
4. Invasion
    a. Intrusion
    b. Decisional interference.

Solove's approach to privacy is useful because it first reminds us that privacy is multi-faceted; there are many different activities that can be considered "private." In addition,

the problemitization of privacy allows us to look for specific issues and to clearly define policy goals by addressing specific problems that are "privacy related."

I use Solove's description of privacy to help explain the concerns of privacy scholars related to the persistence of Facebook records in the following chapters. Solove use of privacy problems to describe privacy allows us to explore specifically how privacy might be threatened by the loss of access control over Facebook records.

The Right to Delete

The concept of "the right to delete" is related to privacy in the context of social networking and, therefore, important to my study in that it is useful in analyzing the implications of digital preservation of social networking records for privacy. Within the print-based correspondence methods we have discussed earlier, there is an assumption that some records may be deleted or destroyed by the records creators. This act can be intentionally thwarted, such as when an email management system creates a record of emails that have been deleted, but in social networking there are multiple layers of service providers with copies of the records that can be deleted.

Therefore, once the subscriber deletes a record he has access to on the service, copies of the records may still exist in multiple places. In addition, access to the record may be denied to the service subscriber or the public, creating the illusion that the record has been deleted, but the record may still exist on backup tapes and other hardware, including the service subscriber's hard drive, and can create a persistent threat to privacy. Ohm discusses the right to demand deletion (Ohm, 2005; see also Palfrey, 2008, p. 291), a fourth amendment protection based on the property right to destroy. This right to

destroy cannot be exercised if a copy of the original record is stored somewhere out of control of the original creator.

The longer these records exist, the greater the cumulative threat to privacy. Each additional day of storage could expose those data to an additional request or data-mining attempt, such as the chance that the service provider will share records with law enforcement agents or other third parties.

A subscriber's ability to delete a Facebook record, and any copies of that record, is a powerful privacy protecting tool. The idea that a subscriber should be able to maintain this level of access control over her records illustrates the threat to privacy inherent in the use of social networking tools. I explore the right to delete as it relates to privacy scholars' concern with the persistence of digital records in more detail in the next chapter.

I have briefly discussed existing privacy research as background to the current study, and as a resource for archivists to better discuss privacy scholars concerns over the persistence of Facebook records. In the next section I offer a review of research that looks at both social networking and privacy.

SOCIAL NETWORKING AND PRIVACY RESEARCH

Substantial research has been conducted into how Facebook subscribers decide to disclose information (Christofides, Muise, & Desmarais, 2009; Buchanan, Paine, Joinson, & Ulf-Deitrich, 2007) and when and how they try to protect their privacy (Buchanan, Paine, Joinson, & Ulf-Deitrich, 2007; Gross & Acquisti, 2005; Youn, 2009), and the risks facing youth and the policy implications of protecting privacy when using social networking services (Palfrey, 2008).

Gross and Acquisti (2005) were among the first researchers to publish a study of social networking. The researchers, affiliated with Carnegie Mellon University, used this affiliation to access and download 4,540 Facebook profiles of Carnegie Mellon students. At the time, Facebook was available to students at only a few universities and not to the general population. Gross and Acqusiti recorded the type of information the students had shared on their profiles. The researchers then asked students from other universities to log in to Facebook and review the sample CMU profiles to determine whether their sample had changed the default Facebook privacy settings. The researchers conducted a third privacy review to determine whether CMU subscribers in their sample had restricted access to other CMU users. Illustration 1 below is a screenshot of a Facebook profile from that time period. Appendix A includes screenshots of profiles from 2005 – 2011.

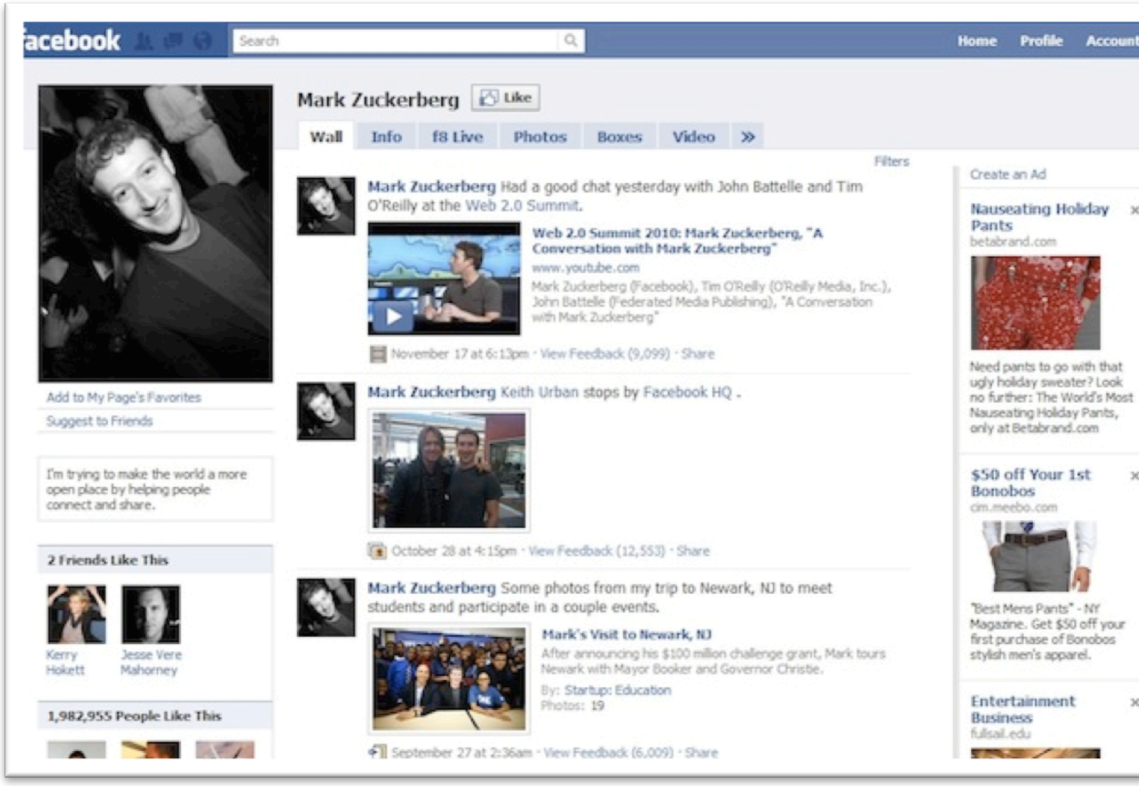
Illustration 1: *Facebook Profile Page from 2005 (Buck, 2011)*

Gross and Acquisti found that the majority of individual users reveal personal information including a profile photo, their birthdate, current residence, current relationship status, and political views. The researchers created an inventory of personal information disclosed by CMU students. I used this inventory to develop the disclosure questions for the current study. Gross and Acquisti found that 1.2% (p. 7) of users changed their profile settings to hide information from students affiliated with other institutions. According to their review of CMU profile settings, the researchers also found that only three CMU students, or 0.06%, (p. 7) had changed their profile settings to

restrict the information other CMU users could find. Most users accept the default privacy settings, and most make substantial amounts of personal information available.

Christofides, Muise, & Desmarais, in their 2009 study of Facebook use, separated information disclosure from information control. Using a sample of 343 undergraduate students, the researchers conducted an online survey that included a personality test. The researchers found that, while participants said they valued privacy, information disclosure was not significantly correlated with information control. Participants who revealed more information about themselves did not use privacy settings to assert more control over the information they shared. Moreover, the researchers found that information disclosure was correlated with the need for popularity. Information control, on the other hand, was associated with levels of self-esteem and negatively associated with levels of trust in Facebook. The researchers conclude, "disclosure and control on Facebook are not as closely related as expected but rather are different processes that are affected by different aspects of personality" (Christofides et al., 2009, p. 341). Illustration 2 below is a screenshot of a Facebook profile from 2009.

Illustration 2: *Facebook Profile Page from 2009 (Buck, 2011)*



My study expands on the Christofides et al. study by examining another possible reason Facebook subscribers choose to disclose information. I add preservation expectations to the range of possible explanations for the need to share information with others using social networking, and the study reported here examines whether preservation expectations are associated with information disclosure behavior. Christofides and colleagues separation of disclosure behavior from information control is invaluable in understanding the results of my data analysis. I will discuss this point in more detail in chapter six, but I find that Facebook subscribers understand the concepts of persistence of records and control of records as two different phenomena.

As part of a multi-state working group called the Internet Safety Technical Task Force, legal and privacy scholar John Palfrey worked with fifty state Attorneys General to better understand how to mitigate risk in adolescents' use of social networking. Relevant to our purposes, the task force studied the problems young people face online and linked privacy, social networking technology, and policy as a means to address these problems:

> Technology can play a helpful role, but there is no one technological solution or specific combination of technological solutions to the problem of online safety for minors. Instead, a combination of technologies, in concert with parental oversight, education, social services, law enforcement, and sound policies by social network sites and service providers may assist in addressing specific problems that minors face online (Palfrey, 2008, p. 6).

Buchannan, Paine, Joinson, and Reips (2007) created three different scales related to privacy. The first scale measured what they referred to as "Privacy Concern," the second measured general privacy behavior, and the third measured the privacy protections subscribers put in place when using the Internet, which the researchers referred to as "Technical Protection." The researchers found that respondents who expressed privacy concerns were more likely to take general steps to protect privacy, but not more likely to engage in technical protection:

> This is logical: We would only expect a positive correlation with Technical Protection for those individuals who had the relevant technical awareness and skills—and in fact, they might do these things as a matter of course, irrespective of their privacy concerns (p. 163).

This study is relevant to our purposes in that Buchannan et al. demonstrate that subscribers are likely to take steps to protect their privacy if they are both concerned about privacy and have the necessary technical knowledge to take steps to protect their privacy. Buchannan finds privacy concern and general protection are related, leading him

to conclude respondents take the steps they are capable of taking to protect their privacy. In the case of my study, Facebook subscribers are capable of deciding whether to share information using Facebook. Conversely, if subscribers do not understand how long Facebook records may continue to exist, their efforts to protect their privacy may not adequately address the privacy risks that are present when records exist for long periods of time. A lack of understanding of preservation can hamper subscribers' ability to protect privacy, just as limited technical knowledge hampers their ability to take technical protection steps.

Christofides et al. (2009) found that not only are young people concerned about privacy, but the activities of disclosure and control are correlated to the need for popularity and levels of self-esteem, respectively. This finding is contrary to the common refrain in the news media asserting that there is a generational difference in the definition and concern about privacy. Hoofnagle et al. claim "large percentages of young adults (those 18-24 years) are in harmony with older Americans regarding concerns about online privacy, norms, and policy suggestions" (2010, p. 10). In addition, Hoofnagle et al. find that young adults are more likely to believe there are laws in place to protect their privacy both online and off (p. 20), and assert that this combination of concern and ignorance indicate a need for improved privacy protection and education rather than a changing definition of privacy.

Boyd and Marwick agree, "participation in such networked publics does not imply that today's teens have rejected privacy as a value" (2011, p. 1). Livingstone points out that the social process of self-actualization that is common for teenagers of any age is now mediated, at least in part, by the technological and privacy affordances of social networking sites (2008, p. 407). It may be that the limited definition of "friends" and the

manner in which privacy protection options are offered are influencing this process of self-actualization and the way teenagers decide to explore identity and risk taking.

Hopefully this brief overview offers a nuanced discussion of social networking and privacy that will allow both privacy scholars and archivists to engage in a discussion of the persistence of Facebook records that addresses both the value of these records and the privacy threats inherent in the records persistence. I have discussed different ways to define privacy, focusing in particular on Solove's discussion of privacy problems, Moore and Tavani's concept of restricted access and limited control, and the right to delete. Each of these privacy concepts will be included in later chapters as I discuss the preservation of Facebook records and the resulting privacy implications. As we can see from this brief overview, previous research separates disclosure behavior from privacy concern and the perception of risk. This separation is useful to my study in understanding how disclosure behavior and preservation expectations might relate to each other. We also see that subscribers' ability to understand and use social networking software not only affects their ability to protect their privacy, but also may have implications for the social interactions this software enables.

The overview of disclosure behavior demonstrates that Facebook subscribers mostly use Facebook for interpersonal correspondence, a form of record that has a long history of preservation in the archives profession. In addition, we see that social networking service subscribers change their disclosure behavior in association with recognized risk. What has not yet been studied is whether changes in disclosure behavior are associated with Facebook subscribers' preservation expectations, a question I will discuss in detail in chapter six.

Chapter 3: Archivists and Privacy Scholars

In the last chapter I reviewed existing research into why people use social networking and how social networking subscribers perceive and understand privacy and risk. I then discussed a variety of ways to describe privacy, emphasizing Moor and Tavani's idea of restricted access and limited control theory, the right to delete, and Solove's approach of describing privacy based on the privacy problems we encounter. I find that privacy is multi-faceted and changes over time, but that we can separate the description of privacy from the protection of privacy, an idea that helps researchers to understand that privacy protection needs to be responsive to changes in the ways privacy is described. Lastly, I discussed some of the research related to social networking and privacy. I find that subscribers to social networking services often use those services to communicate with people they have met face to face. I also find that subscribers to social networking services may be concerned about privacy without acting to protect privacy, often because they lack the technical knowledge to take appropriate action.

In order to better understand how the preservation of Facebook records might implicate privacy, I next review some of the archival and preservation research, theories, practices, and vocabulary that are relevant to my study. I discuss why some archivists believe it is extremely difficult to ensure the persistence of digital records over time, and why privacy scholars fear digital records implicate privacy because they will continue to exist "forever." I first discuss the different vocabulary used by archivists and privacy scholars to discuss preservation. I then explore the different assumptions and perceived challenges of privacy scholars and archivists related to the preservation of Facebook records and the protection of privacy.

DIFFERENT VOCABULARY

Archivists use the term "preservation" to refer to specific theories, technologies, and practices that ensure valuable records will continue to exist over long periods of time and at the same time will be accessible to others. Moore describes preservation as an environment that "manages communication from the past while communicating with the future" (2008, p. 63). The dual goals of preservation, persistence and access, work in tandem. It is not enough to simply ensure the records continue to exist if the record creator, researchers, or future generations cannot access those records. The records cannot be used for research, personal, or cultural heritage purposes if they are inaccessible.

Privacy scholars use the term "data retention" (Ackerman et al., 1999, p. 5; Calabrese, 2009; Palfrey, 2008, pp. 271, 6, 9) to refer to the persistence of digital records over time, borrowing the term from the records management field which focuses on the active use of organizational records as well as the destruction of most records. The particular period of time to which these scholars refer varies depending on legal requirements, the perception of legal requirements, or the needs of the corporation.

Privacy scholars have experiences with and approaches to the persistence of digital records different from those of archivists. Privacy scholars are often more interested in data retention than preservation, where an attitude of the "shorter the better" helps to ensure privacy. The timelines for retention of data are often determined by federal or state legal requirements, standards established by professional organizations or accrediting agencies, contractual obligations with other companies or service providers, and internal business purposes including auditing and tax record compliance. Both preservation and data retention call for the persistence of records through the useful life

of the record, but the meaning of "useful life" varies depending on whether secondary value is considered. For the purposes of this study, I use the term "preservation" to refer to the persistence of records, and "data retention" to refer to the legal obligation to maintain records for an established minimum period of time to fulfill law enforcement or accounting requirements.

When archivists talk about "forever," they recognize the complexity of digital preservation in the context of the historical record, which is counted in centuries rather than decades. Archivists are concerned that digital records created by social networking sites will not be available as part of the historical record due to the fast pace of hardware and software versioning, the unpredictable and often short lifespan of technology companies, the use of proprietary software formats, and lack of awareness and concern about the importance of preservation on the part of records creators. As more people make social networking of all types part of their daily communication routines, and as more businesses and government agencies use social networking tools to communicate with customers, employees, and business partners, the likelihood that historically significant records are being created increases.

> We're heading toward a world where an extensive trail of information fragments about us will be forever preserved on the Internet, displayed instantly in a Google search. We will be forced to live with a detailed record beginning with childhood that will stay with us for life wherever we go, searchable and accessible from anywhere in the world (Solove, 2008, p. 17).

Like archivists, privacy scholars are concerned with the preservation of digital records. Unlike archivists, privacy scholars such as Daniel Solove are concerned that these records will persist "forever," regardless of the intentions of the record creator. The use of the term "forever" by privacy scholars as a description of a privacy problem refers to a lack of access control over the lifespan of the record. As I discussed in the previous

chapter, the right to delete is really an assertion of control over the persistence of a record. Privacy scholars often use the term "forever" if a record exists longer than a subscriber would like, or if the record garners unwanted public attention after the subscriber would prefer it had been deleted. Solove refers to this privacy problem as exposure.

This discussion of the different vocabulary used by archivists and privacy scholars underscores the fact that archivists assume preservation will be conducted by a preservation expert in the context of an established archives. Privacy scholars assume preservation, as they use the term, will happen as the result of the actions of individual Facebook subscribers in the context of Facebook. These two different assumptions lead privacy scholars and archivists to different conclusions about the privacy implications of preserving Facebook records.

DIFFERENT ASSUMPTIONS AND CHALLENGES

In addition to different vocabularies related to the preservation of records, archivists and privacy scholars also have different perspectives on the challenges and implications of preserving social networking records.

When archivists think about preservation and social media, they understand preservation in the context of archives and the long history of protecting primary, authentic, reliable records, protecting the privacy of records donors and subjects, and meeting the challenges that digital media present to archival traditions and professional practice. Archivists assume that a competent institution well versed in protecting privacy and enforcing access requirements will preserve Facebook records.

Privacy scholars assume nefarious individuals, oppressive governments, or corporations motivated by profit will preserve these records. Privacy scholars understand preservation as a result of loss of access control over records created using a service that is provided by a private corporation and supported by advertising revenue.

The assumptions of both privacy scholars and archivists can be correct, but, given the current lack of a concerted effort by archivists to preserve Facebook records, only privacy scholars have accurately described the current state of affairs concerning access control of Facebook records over time.

Despite these seemingly opposing assumptions, individuals, corporations, and archives can preserve Facebook records, and individual privacy would benefit greatly from a trustworthy preservation organization that is interested in preserving authentic and reliable social networking records. In the next section I will more closely examine the motives and technical challenges described by archivists and privacy scholars based on the different assumptions and challenges they face.

Archivists Assumptions and Challenges

When archivists discuss digital preservation, they do so aware of the history, traditions, and professional practice of the archival profession. As I discussed earlier, one of the goals of archivists is to ensure the persistence of, and access to, valuable cultural heritage and research records for future generations. I discussed the different ways Facebook records are considered valuable, for cultural heritage and personal reasons, and the value of the corpus of Facebook records for research. I also discussed the deed of gift, the archival equivalent of the Facebook terms of service agreement, designed to allow the record donor and the archives to negotiate access and ownership rights. In addition to

these concerns, archivists also have a set of assumptions about social networking records that help to explain why privacy scholars and archivists seem to have such divergent opinions about the risk and likelihood of preserving social networking records.

In addition to the challenges I have already discussed, archivists are concerned that hardware and software versions can change, social networking companies can have limited life-spans, and digital records can be created using proprietary software. Each of these concerns contributes to the overall technical complexity of preserving digital records.

Archivists are concerned about the reputation of the archives they manage. In order for patrons to trust the authenticity and reliability of the records, and for donors to continue to donate valuable records, archivists have developed policies that allow for controlled access, consistent and reliable metadata, and planned persistence. These policies are consistent over time and focused on protecting records, the privacy of record donors, and the reputation of the archives. The ability of an archives to acquire valuable records often depends on donations from individuals or organizations. Donors of both records and financial gifts select an archives based on the reputation of that archive. The financial stability and the value of an archive's records often depends on the reputation of the institution, a reputation built on the responsible management of records and following agreements made with records donors.

There are a number of different reasons for archivists to be concerned about the preservation of social networking records. Born-digital records are created using software programs such as Microsoft Word, Eudora, or Twitter. These software programs are iteratively changed; each new version or subversion changes the code of the previous version. Over time, the current version and previous versions often become incompatible,

making it difficult to open files that were created in the earlier versions of the software, or at the very least changing the appearance and experience for the viewer of that record. Hardware and operating systems face challenges similar to software; companies regularly release new hardware and operating systems that obviate previous versions of the hardware or software.

Additionally, social networking companies have limited lifespans. Some companies may be created solely to be acquired by larger companies once these smaller companies have gained an adequate market share. Users are fickle, migrating to the latest hot site. MySpace is one example, where in 2008 the number of unique visitors per month was 75.9 million, and in 2011 that number fell to 34.8 million (Gillette, 2011). Many smaller startup companies may not survive very long, and larger established companies may not last for the many decades that archivists expect historically significant records to be preserved.

Social networking companies often create proprietary formats for the records their software creates in an attempt to secure a competitive advantage over similar companies. The technical specifications of the formats are often closely guarded secrets and protected by the Digital Millennium Copyright Act, increasing the complexity of preservation in a number of ways. When a company goes out of business, or the software version of the record is four or five generations behind the current or last version, gaining access to an older document becomes an exercise in digital archaeology. It is an exercise that includes searching for and assembling retired hardware that is still functioning; determining, acquiring, and launching the operating system and software version that was used when the document was created; and hoping that these pieces hold together until one has had a chance to make the best possible accessible copy.

The increase in the use of software of all kinds to create born-digital records with very little public awareness or concern for the long-term existence of these records worries some archivists and historians that parts of our historical record may be lost if conscious effort is not made to preserve historically valuable records and the software code used to create these records.

The twin goals of ensuring the long-term existence and continued access to digital records have led to a discussion in the archival community about how best to avoid what some scholars previously referred to as the "digital dark ages" (Kuny, 1997, p. 1). More recently, O'Sullivan compares digital and non-digital records, concluding:

> Paramount to these differences is the appreciably shorter life span electronic records have without some form of human intervention due to the gradual obsolescence of hardware and software environments. Their survival, however, involves a great deal more than the mere capture, migration, and storage of information (2005, p. 54).

This description of the complexity of the preservation of digital records stands in stark contrast to the assumption by privacy scholars that these records are certain to last forever.

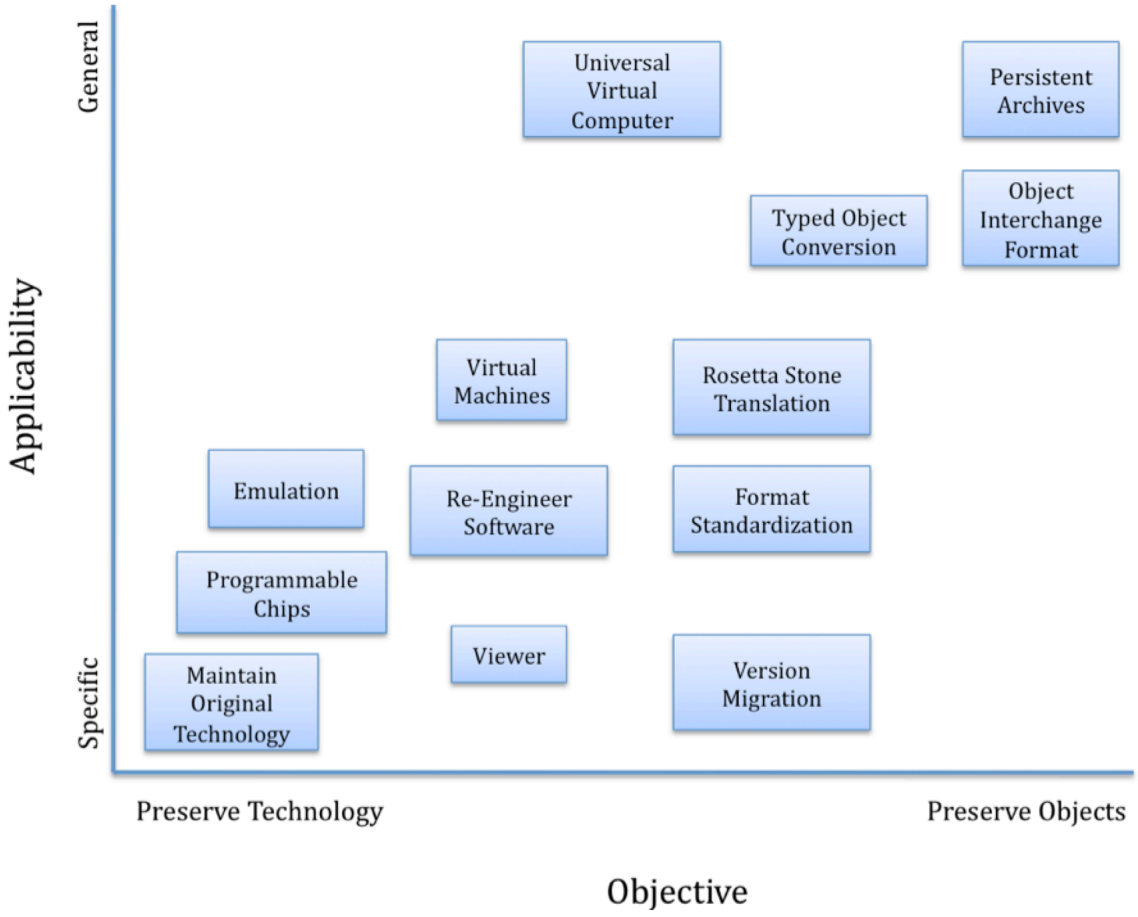
Thibodeau (2002) underscores the complexity of retaining, managing, and preserving digital records indefinitely when he summarizes the then - current range of methods available for digital preservation of government records. The details of the methods Thibodeau describes and discussed here are not important in their specificity, but I use them to illustrate the complexity of the preservation of digital records in a known, controlled environment. The preservation of personal records adds to this complexity because individuals managing their own records may not follow consistent records management processes and will most likely use a wide variety of software to

accomplish their tasks, without considering the consistency, interoperability, and ability to preserve the format of the records they create.

Thibodeau places the methods he has selected on the "two-way grid" (p. 15) in Figure 1 below. The x-axis describes the objective of the preservation method, whether the focus of the preservation effort is to ensure the method used to access the information is preserved, which Thibodeau refers to as preserving the "technology", or the bits that represent the information are preserved, which Thibodeau refers to as the data "object". For example, to preserve a Microsoft Word record an archivist may choose to preserve a computer running Windows '95 and Office '95, allowing access to the Word document over time. Alternatively, the archivist may choose to preserve only the bits that represent the Word document, using a number of different strategies to ensure that data object will be accessible over time.

The y-axis describes the applicability of the preservation method, or whether the method is appropriate to preserve a single, specific, technology or a wider variety of technologies or technological environments. From the example above, preserving the computer running Windows '95 is a specific solution, a computer running virtualization software that is able to imitate Windows '95, Windows '98, and Windows 2000 would be considered generally applicable to preservation.

Figure 1: *An Example of the Technical Complexity of Digital Preservation (Thibodeau, 2002, p. 15)*



I use this chart simply to underscore the fact that there exist a wide variety of approaches to digital preservation, the complexity of digital preservation, and the lack of consensus among archivists as to the most effective long-term approach to digital preservation.

The assumption that these records will be available "forever" is clearly not supported by current digital preservation research. While we may be able to preserve a particular bit stream for long periods of time, our ability to render that bitstream over time is a much more complex task, especially given the prevalence of proprietary data formats. Indeed, the complexity and range of possibilities for preservation indicates the

effort and expertise required if an organization were to decide intentionally to preserve digital records.

Privacy Scholars Assumptions and Challenges

I have discussed the assumptions and concerns archivists have about ensuring valuable digital records will be available for future generations including the rapidly changing and incompatible hardware and software versions, the potentially limited lifespan of social networking companies, and the use of proprietary formats. Privacy scholars have a different set of assumptions and concerns including the potential for social networking records to exist "forever," threatening the privacy of social networking subscribers. The concerns and assumptions of privacy scholars are important for us to discuss in detail in order to determine whether valuable social networking records can be preserved while protecting the privacy of individual subscribers.

Facebook uses the interaction of subscribers to gather a portfolio of personal information about each subscriber in order to increase the value of advertising space on Facebook. Advertisers pay for an opportunity to have their ad displayed to specific groups of Facebook subscribers, groups the advertiser believes are most likely to become customers. The more Facebook knows about each subscriber, the more advertisers are willing to pay for ad space.

Facebook has consistently looked for ways to encourage people to share more information about themselves, "Facebook's mission is to make the world more open and connected" (Facebook, n.d., a). One example of a change Facebook made to increase sharing was the introduction of the news feed "feature" in 2006. Prior to 2006, in order to view the activity of one of your friends on Facebook you had to visit that person's

Facebook wall. A subscriber would see only what others had written on their wall, not an accounting of each action their friends had taken. In 2006 Facebook turned on the news feed feature for every subscriber; no one had the ability to opt out. The news feed kept a constant record of each action one of your friends had taken, making subscribers suddenly aware of all of the activity within their Facebook network of friends.

Boyd uses the term "exposure" to describe this change in a way that is very similar to Solove's description of exposure that I discussed in chapter 2. Boyd uses the metaphor of trying to communicate with someone in an extremely loud room (2008, p. 14). While the music is playing, you may need to scream in order for someone standing next to you to hear what you are saying. You may choose to have a personal conversation and share sensitive information in this context, knowing only those people who are physically close to you will be able to hear what you are saying. If the room suddenly goes silent, your conversation is audible to far more people than you had expected, creating a privacy problem.

Creating this problem was, and still is Facebook's goal. If handled in a way that does not offend the subscribers enough to encourage them to leave, a decrease in privacy protection and an increase in sharing can increase participation. This increased participation adds to the data Facebook is collecting about subscribers and increases the value of Facebook advertising space.

Privacy scholars are concerned that the creation and use of digital records requires the creation and distribution of multiple copies to a multitude of different individuals and groups, all of whom have different motivations for, and claims to, the records created by the service subscriber. Privacy scholars are especially concerned that subscribers are

unaware of the fact that multiple copies of their records exist, and unaware of the privacy implications of the persistence of these records.

Subscribers to social networking services often believe that they are creating records in a "private" space (Palfrey, 2008, p. 284). This assumption that subscribers to these services are sharing information in a "private" online space that they have control over can lead subscribers to post intimate or embarrassing details about their lives, feelings, and activities.

There are two factors that contribute to this assumption of privacy. The first is that social networking services make an effort to create a customizable interface that service subscribers feel they control, creating an illusion of privacy. The second is most likely a lack of understanding of the infrastructure and technical mechanisms required to post information to these services. Most of the technical workings of the Internet, such as encapsulation, routing, storage, backup storage, and access rights, are not familiar to the users of social networking. The digital copies that are made of these records at each of these stages, and the potential to create digital copies as the information the service subscriber has submitted travels from her laptop to be displayed on her Facebook wall, are unfamiliar to most users. In addition, these records are used to develop profiles of the social networking subscribers in order to increase the value of advertising space. If subscribers were more aware of the number of service providers and copies of their records that were created in order to post and share the record, subscribers might be less likely to assume their records and correspondence are private.

In addition to the copies that are made for technical, business, and procedural reasons, other subscribers with access to the messages posted by the service subscriber can copy and paste those messages, distributing them to a wider audience than the

original creator intended. The original copy of the message a service subscriber posts can also be made available when the social networking service changes security or privacy settings. We have seen security and privacy setting changes expose previously "private" records when Facebook changed the default privacy settings, making most of the information in a user's profile public by default (Bilton, 2010). Once this information becomes available to a broader audience, it can be copied and shared rapidly among millions of users, with no chance that the original subscriber can regain control of that information.

Privacy scholars are also concerned that privacy protections, whether through laws, corporate privacy rules and procedures, or social norms, have not kept pace with technological change, undermining privacy rights and leaving service subscribers vulnerable to privacy violations.

From this brief overview of privacy scholars' concerns with preservation, we see that control of social networking records and trust that social networking service providers will act in accordance with the desires of social networking subscribers are important privacy considerations when subscribers use social networking services. I discussed these two components as part of Moore (1997) and Tavani's (2007) restricted access and limited control theory of privacy, and we will see these concerns again as we discuss the results of the online survey and focus group discussion.

Ackerman, Cranor, & Reagle's study of trust in e-commerce sites (1999) helps to summarize the problem privacy scholars have with the preservation of Facebook records. Ackerman et al. found that an important factor in determining whether a respondent would use an e-commerce site was whether the site has published a data retention policy. However, respondents rated the existence of a data retention policy as less important than

other privacy considerations, including whether the site disclosed why personal data were being collected or whether the service would share the collected data with a third party (p.5). Ackerman et al. concluded from the comments of their respondents that this relative lack of concern was "due to a distrust that companies will actually remove people from their databases and a belief that it will be impossible to remove information from all the databases it may have propagated to" (p. 5).

Privacy scholars perceive a privacy threat in the preservation of Facebook records in large part because Facebook is not trustworthy. Facebook's business model and mission encourages the leaders of the organization to make increased profit due to increased sharing, and the software and policy changes reflect this priority. As we have seen so far, archives can act as a model of what an information and communication technology service provider that respects privacy and preserves records might look like. The concept of the trusted digital repository represents that model.

On the one hand, archivists would like to intentionally preserve social networking records, ensuring the persistence of the volume of records created and allowing each record creator to have limited control over access to the records she creates. On the other hand, privacy scholars are concerned that the process of creating and sharing social networking records threatens privacy because of the potential for any individual record to persist outside of the control of the record creator. From the reliance on a private company that depends on advertising revenue to survive, to the multiple copies created to store and share records, and the ability for any subscriber to create additional copies of a record, the untrustworthy process of creating social networking records threatens both the persistence of the corpus of social networking records and the record creator's ability to control access to any individual record. The one billion subscribers to Facebook are

already accepting the risk to privacy inherent in using a service whose business model requires increased sharing and decreased privacy protection. Archivists can act to keep valuable records created using Facebook from being lost, and can do so in a way that limits any further risk to subscribers' privacy.

I discuss four approaches to preserving Facebook records in more detail in the next chapter. So far I have discussed the value of social networking records and why they should be preserved, and the conflicting opinions between privacy scholars and archivists. Privacy scholars are concerned that social networking services are designed to encourage the subscriber to share in order to increase advertising profit, that the use of social networking creates multiple copies of the record, and allows for other subscribers and third parties to duplicate the record, resulting in a loss of control over the record by the subscriber and increasing the threat to privacy. Archivists are concerned that the technical complexity, proprietary formats, hardware and software versioning, short lifespan of social networking companies, and lack of subscriber understanding and concern about preservation all threaten valuable cultural heritage records.

Of the four general approaches I will discuss in the next chapter, the only one that will address the concerns of both privacy scholars and archivists is if Facebook were to act as a trusted digital repository. Each of the other solutions would still allow Facebook to prioritize advertising profit over the persistence of and access to the records according to the subscribers' wishes. In each of the other three scenarios, copies of the record made by other subscribers or third parties can be an indefinite-term threat to the privacy of the record creator. This does not mean that the other three solutions are not valuable; important cultural heritage artifacts will be protected by the other strategies, and in most cases, subscribers will become more aware of the cultural heritage value of their record

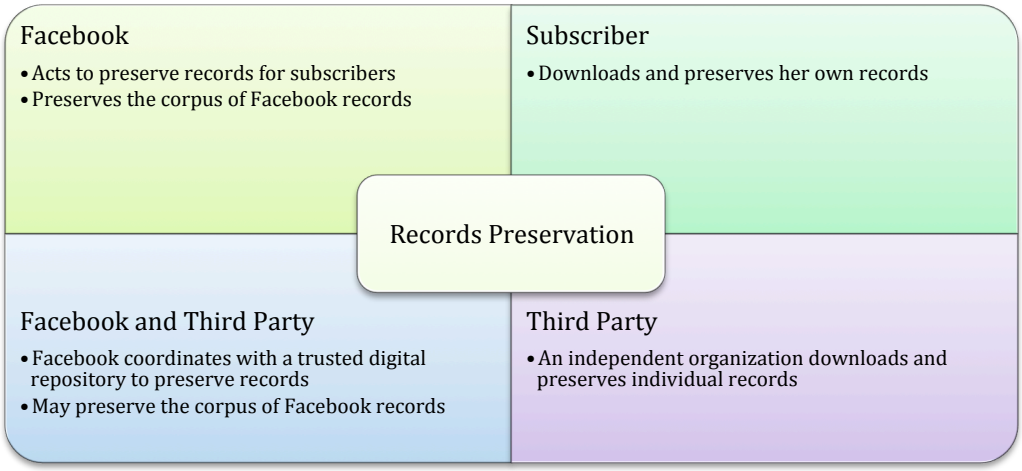
creating activities. A preservation solution offered outside of Facebook that prioritizes the record creators' needs while ensuring the persistence of and access to the records can also act as a counterpoint to Facebook's policies and procedures, empowering subscribers by reminding them that records can be managed in a different way.

Chapter 4: Facebook Preservation Options

In this chapter I discuss four types of solutions that could be used to preserve Facebook records and to protect subscribers' privacy. Not all of the solutions I discuss require changes to federal policy, but I discuss the role policy might play in each solution.

Subscribers' preservation expectations implicate the protection of the cultural record. If Facebook does not preserve records indefinitely into the future and make those records available to future generations, the preservation of the records is wholly dependent on the individual record creator. However, the record creator is not the only actor who can preserve records. Facebook could take steps to preserve records either by contracting with a third party or implementing a "Facebook Archives" that includes a preservation responsibility. A fourth option for preserving Facebook records is for an independent entity to download Facebook records for preservation without a formal agreement with Facebook. These four approaches can occur simultaneously and they can be combined. Illustration 3 summarizes these four options.

Illustration 3: *Four Approaches to Preserving Facebook Records*



I previously discussed the four general ways of addressing the problem of preserving social networking records: each individual subscriber can download and preserve her own records, the social networking service provider can preserve records created using that service, the social networking service provider can team up with a third party and donate records to that third party, or a third party can "scrape" records from the social networking service provider. There is some overlap with each of these potential solutions; an individual can download her own social networking records and donate them to a third party, for example, or a social networking service can collaborate with a trusted digital repository while a different third party is downloading and preserving records.

Of these four solutions, the corpus of digital records created using Facebook can be preserved only if Facebook chooses to act either on its own or in cooperation with a digital repository. The other solutions can preserve only a portion of the records created using Facebook, limiting the research value described in the discussion about the volume of Faceboook records.

In this chapter I will discuss each of the four general solutions using the four regulatory constraints on subscriber behavior articulated by Lessig: architecture, law, social norms, and market (2006, p. 234). Lessig explains that "architecture" can be understood in the context of cyberspace as computer code, and code acts as a constraint that "regulates some behavior by making other behavior possible or impossible. The code embeds certain values or makes certain values impossible" (p. 125). Code writers determine the technological affordances of Facebook, and these affordances not only regulate behavior, but the affordances are deliberately selected and subject to change.

I have discussed law as a regulatory constraint in the context of the preservation of Facebook records throughout this study. In this chapter I will discuss how each of the four solutions would benefit from changes in law or policy in order to achieve the goal of preserving Facebook records and protecting privacy.

Social norms also regulate behavior, according to Lessig. We have discussed social norms in the context of privacy, but in this chapter I discuss education and awareness efforts as a way of encouraging Facebook subscriber's to act in ways that will help preserve Facebook records while protecting privacy.

Lessig uses the term "market" to refer to pricing and availability of a service (p. 124) and to the government's ability to encourage or constrain the behavior of organizations and individuals through tax incentives and subsidies (p. 116). In this chapter I discuss the use of tax incentives and other market influences that might help to encourage the preservation of Facebook records and the protection of subscribers' privacy.

Lessig asserts that a complete view of regulation must include a discussion of all four of these constraints (p. 123). While the list of four constraints Lessig offers may not

be comprehensive, it is a very useful lens through which to analyze potential solutions to preserving Facebook records and protecting subscribers' privacy.

My aim is to create a high-level overview of possible ways to preserve Facebook records and point out some of the problems with each of those solutions in order to provide a starting point for discussions about both preserving social networking records and privacy, using Facebook as the example.

SUBSCRIBER PRESERVES RECORDS

Facebook currently allows subscribers to download their own information from Facebook using three different options: "downloaded info," "expanded archive," and "activity log" (Facebook, n.d. d). The "downloaded info" option includes timeline information that I have already described as the focus of this dissertation, the "expanded archive" option includes things such as logins and cookies, and the "activity log" includes "a history of all your activity from posts you've commented on or liked, to apps you've used, to anything you've searched for" (Facebook, n.d. d). I have included Facebook's detailed description of the data included in each download type in Appendix B. These download options do not allow a subscriber to download information other people have posted, whether that information is about the subscriber or in response to a subscriber's post.

The code for this solution is already in place; Facebook has created a way for individuals to download their profile information. The amount and type of information is incomplete since subscribers can download only the information they have posted, and the information is downloaded in html format, which is not ideal for long-term preservation or for creating algorithms that would be able to create links between

downloaded profile information, essentially recreating a Facebook friend network for future research and cultural heritage purposes. Preservation experts are already dealing with large numbers of html files, so these problems are not new, and solutions applied to other html files can be applied to these files as well.

The law constraint discussed by Lessig applies to Facebook policies related to the format and technological affordances offered by Facebook. Over the course of the last two years, Facebook has added the categories of "expanded archive" and "activity log" to the download option. These additional sources of information are a welcome expansion, but just as Facebook has the ability to expand the download affordances, these affordances can be altered or removed in a way that would decrease the compatibility of Facebook downloads over time. If future archives are to rely on individual records as Cox asserts, success in creating a collection will depend in part on the consistency of the format of these records over time. By committing to a consistent, backward-compatible, open format for the download, Facebook could take an important step in facilitating the preservation of Facebook records for generations to come.

Education programs could encourage individual Facebook subscribers to download and preserve their Facebook records, influencing social norms. Facebook advertisements could encourage subscribers to download their records and take steps to preserve these records. Online courses could be offered that help individuals understand the importance of their records, the technical steps they should take to preserve their records and protect their privacy at the same time. These programs could be offered by Facebook, but would more likely be offered by the company that provides the trusted digital repository (TDR) storage service, or by not for profit organizations such as the

Library of Congress that have preservation of cultural heritage as an institutional responsibility.

One way for the market to encourage individual subscribers to download and preserve their own Facebook records is to subsidize personal digital storage space, making it cost effective for individuals to upload records to a space that acts as a trusted digital repository. Storage space providers could be offered a tax incentive to provide storage that meets the TDR specifications at a discounted rate to individual subscribers.

The benefit to this approach is individual subscribers will retain control over the records they have downloaded, making it possible for the subscriber to decide what she feels comfortable sharing with others over time. The ability to download one's own information from a service such as Facebook also makes it possible to create a backup of that information in case the service goes out of business or decides not to maintain older files.

Individuals could download and preserve their own records. Cox emphasizes the importance of personal archives both to the individuals who create the records and their families, and to society at large, "personal and family archives are a critical aspect of our society and culture" (2008, p. 290). Cox sees the shift to digital documentation, especially the World Wide Web, blogs, and email, as an opportunity for archivists to expand their field by training and educating records creators. The volume and transience of records created by individuals requires this expansion if valuable records are to be preserved.

Cox emphasizes the radical change wrought by Web technologies such as Web pages, blogs, and social networking:

> [A]rchivists need to be mindful that the nature of personal and family archives is transforming in ways that require some fundamental re-imagining of archival practice. Archivists need to expand archival documentation efforts, like the old documentation strategy model, to include individuals and families as key players. (2008, p. 309)

Cox underscores the role individual record creators, using Web technology such as blogs, will play in the future of archives. He calls on archivists to embrace the change in their profession that is mandated by the increase in the creation of documents using World Wide Web technologies, saying "Archivists need to help individuals maintain personal and family archives, only collecting those of special or extraordinary significance when they are endangered" (viii) and simply "we need to build more citizen archivists" (284).

Relying on individuals to download their own records causes problems for the preservation of Facebook records. When records are removed from the context of Facebook, links to other records that may exist about the same topic or event are destroyed. The inclusion of the activity log and expanded archive options means the download option offered by Facebook includes more than just the individual messages posted by subscribers, but it does not allow for a particular subscriber to have a record of the reactions to her messages. These reactive records allow for a complete view of a correspondence, and allow for archivists and future researchers to better understand the context of any individual post.

Allowing for individual downloads does not create a situation where a collection of records exists in one place, or links between existing records can be created. As I discussed earlier, documentation strategy does not require all of the records that exist on a topic or event to exist in one place, as long as links among those records can be established and maintained. This idea of a distributed collection is an important part of the post-custodial era according to Ham (1981). If each individual is downloading her

own comments, photos, videos, and correspondence and keeping these on her own hard drive, the links between her records and the records of other members of her network will not exist and will be difficult to re-establish.

Another possible solution to preserving Facebook records similar to the individual download is a third party application that allows subscribers to download their profile information and activity log to a third party server. This approach would allow for multiple profiles of multiple individuals to exist in one place, increasing the likelihood that links among multiple individuals or groups who discuss a particular topic or event could be established.

Relying on individuals to download their Facebook records to a third-party service for preservation could provide privacy protection by allowing the subscriber to determine the access rights to the records they are uploading to the archive service. A subscriber could decide to make their records available in 50 years, for example, or keep some records private and make other records available to the public.

One drawback to relying on subscribers to download their records to a third-party for preservation is the fact that Facebook has one billion subscribers. Encouraging enough of these subscribers to allow for a relatively complete picture of any event or conversation would be difficult. The third-party download approach would allow archivists to apply documentation strategy to these records. A researcher or archives could identify individuals or Facebook groups that are important to a particular event or topic and encourage these individuals to upload their Facebook archives to the service.

FACEBOOK PRESERVES RECORDS

A second potential solution to the preservation of Facebook records and the protection of privacy is for Facebook to act as the trusted digital repository (TDR) for Facebook records. All records created using the Facebook platform would be stored and preserved by Facebook or a subsidiary of Facebook, charged with the long-term preservation of, and provision of access to, these records.

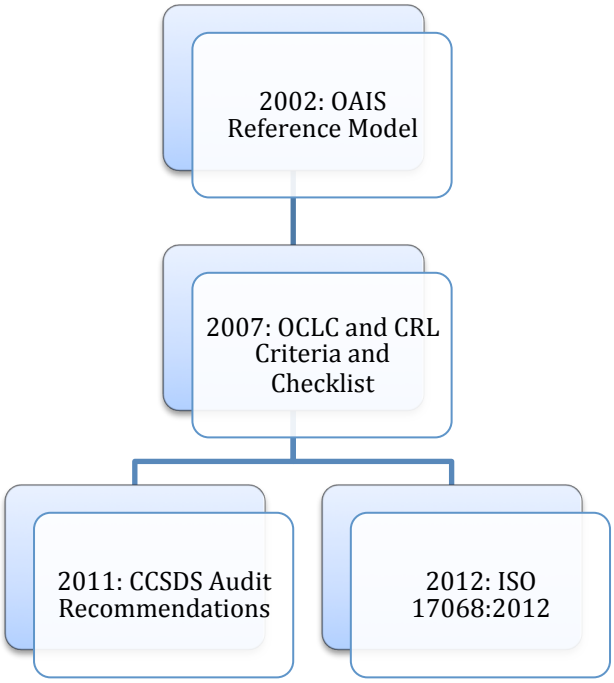
A trusted digital repository is responsible for ensuring the authenticity, availability, and integrity of records over time in accordance with the agreement made by the archives with the records creator. With digital social networking records, the subscriber has little control over the current and future privacy, availability, and preservation policies of the service provider, and should not expect to be aware of how these records might be used or shared by other subscribers once the records are made public. To act as a preservation authority, Facebook would need to follow established best practices for a trusted digital repository.

A trusted digital repository is defined, at minimum, as an organization with "a mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future" (Research Libraries Group, 2002). The National Archives and Records Administration (NARA) in collaboration with the Research Libraries Group (RLG) established a set of criteria based on the Open Archival Information Systems (OAIS) Reference Model, to certify organizations capable of providing long-term, trusted storage and access to digital collections. These certification criteria are called "Trustworthy Repositories Audit & Certification: Criteria and Checklist" and are organized into the broad categories of organizational infrastructure,

digital object management, and technologies, technical infrastructure, and security (OCLC and CRL, 2007).

The audit and certification recommendations from NARA and RLG were then used as a basis of the audit recommendations of the Consultative Committee for Space Data Systems (Consultative Committee for Space Data Systems, 2011) and the recent ISO/TR standard 17068:2012 "Information and Documentation – Trusted Third Party Repository for Digital Records" (International Standards Organization, 2012). Illustration 4 summarizes the milestones in developing the TDR ISO standards.

Illustration 4: *An Overview of the Development of TDR ISO Standard*



As we can see from this brief overview of trusted digital repository certification, for over a decade international groups of archives professionals have been formally developing criteria to determine when an organization should be trusted with the long-term preservation of digital records. The concept of the trusted digital repository is critical in resolving the apparent conflict between privacy experts and archivists. Illustration 5 demonstrates the different perceptions of privacy experts and archivists about the preservation of social networking records. Privacy experts believe the fewer records we preserve, the more we protect privacy. Some archivists believe we need to preserve more records to protect cultural heritage, and the views of privacy experts and archivists seem to be in opposition.

Illustration 5: *Collecting Records; Threatening Privacy or Cultural Heritage?*

The concept of the trusted digital repository (TDR) is one way to resolve this apparent conflict. I discussed Moor's (1997) definition of privacy as Restricted Access and Limited Control (RALC), and Tavani's (2007) revision of this definition to apply to Internet transactions. This definition of privacy is particularly helpful in resolving the claims of privacy experts and archivists because privacy is protected, according to Moor and Tavani, by allowing the subject of records to have a limited ability to restrict access to those records. A TDR would allow exactly this, a preservation agreement between the creator of the Facebook records and the repository that would allow the creator limited control over the records she creates using the Facebook platform. The record creator would be able to trust the social networking platform, in this case Facebook, because an independent third party verifies the policies and procedures of that platform.

As I discussed earlier, collecting archives have historically used a combination of the deed of gift agreement and their public reputation to reassure donors, protect privacy, and protect records. The difference between a collecting archives and Facebook, in the context of preserving records, is Facebook does not have the preservation of records and the protection of donor privacy as their primary business goals. Collecting archives would quickly go out of business if they were to change their privacy policies and expose records that had been donated with the expectation of privacy over a set period of time. Facebook has routinely made changes that undermine subscribers' trust, yet Facebook has continued to grow at an astounding rate.

Facebook has demonstrated a distinct lack of consistency in policy, especially related to privacy. The Electronic Frontier Foundation's Kurt Opsahl offers evidence of this inconsistency in his report of six revisions to Facebook's privacy policy between the years of 2005 and 2010 (2010). A commitment to the preservation of records according to established standards for trusted digital repositories would improve the transparency of Facebook's records management processes, and increase the level of trust subscribers could reasonably have in Facebook. Since Facebook has not elected to either contract with a third party or provide for the preservation of Facebook records using their own resources, however, the preservation of important Facebook records currently depends on the efforts of individual subscribers. Table 1 below summarizes the discussion above comparing the preservation environment created by Facebook as it currently exists and the environment offered by the idea of a trusted digital repository.

Table 1: *Facebook Compared to Trusted Digital Repository*

	Facebook	Trusted Digital Repository
Motive	Profit	Preservation of cultural heritage
Policies	Change over time to increase sharing	Remain consistent to protect existing agreements and subscriber trust
Persistence	Depends on the life-span of Facebook and the utility of the record to support advertising revenue	Depends on the funding and relationship to Facebook
Privacy	Depends on access policies and control of copies	Depends on deed of gift type agreement

In chapter five I will discuss the general records schedule established by the National Archives and Records Administration, and the Federal Records Act that requires all federal agencies to comply with the General Records Schedule (GRS) and to set and seek approval of their own records schedules for records descriptions that are not found

in the GRS in detail. Given that correspondence is one type of record covered by the GRS, NARA could require Facebook to agree to a set of preservation practices that would at least allow for the transfer of records to the NARA. This change would address the law constraint mentioned by Lessig, and would allow for these records to be preserved by a trustworthy repository.

In support of compliance with the GRS, NARA could create a list of certified trusted digital repositories that can be used by government employees. This list of trusted digital repositories could include Facebook if they agree to conform to the ISO certification procedures for trusted digital repositories. This step would improve the preservation practices of Facebook by requiring Facebook to review their own records management procedures and submit to a policy and procedure review by a third party.

NARA has already taken a similar step by establishing a list of social networking services that have modified their terms of service to eliminate requirements that are problematic for government agencies. NARA was able to negotiate with Facebook to change some of Facebook's standard terms of service provisions to make those terms acceptable to federal agencies, allowing agencies to sign the terms of service agreement and create Facebook pages.

NARA could add the trusted digital repository requirements to the list of requirements for a company to be pre-approved for use by the federal government. The inclusion of Facebook on this list is a market incentive for Facebook to seek the TDR designation. A trusted digital repository is responsible for ensuring the authenticity, availability, and integrity of the records over time in accordance with the agreement made with the records creator. With digital social networking records, the subscriber has little control over the current and future privacy, availability, and preservation policies of the

service provider, and should not expect to be aware of how these records might be used or shared by other subscribers once the records are made public. To act as a preservation authority, Facebook would need to follow established best practices for a trusted digital repository.

If Facebook agrees to apply for trusted digital repository certification, the process of certification may have a "trickle down" effect, improving the preservation practices of Facebook and creating an opportunity for all Facebook subscribers, not just employees of the federal government, to create and retain records according to preservation best practices. Part of this "trickle down" effect comes from the code or architecture required either to transfer Facebook records to a separate, Facebook-provided TDR, or to treat the Facebook platform as a TDR.

The Federal Trade Commission (FTC) could propose regulation that would require social networking companies to preserve records. The FTC could build on the current pursuit of privacy regulations related to Facebook (FTC, 2011a) and Google, (FTC, 2011b) by emphasizing the importance of preservation in the protection of customers' records.

The benefit to the FTC's requiring social networking companies to preserve customer's records is that Google and Facebook are already responsive to FTC regulation of privacy, so the FTC may be able to encourage preservation without additional Congressional authorization. The involvement of a Federal organization such as the FTC in encouraging long-term preservation of social networking records would raise privacy concerns and create mistrust between Facebook and subscribers, but it could have the effect of requiring a broad range of U.S. companies to preserve social networking records. The other drawback to the FTC's requiring social networking companies to

preserve customers' records is that it would increase the barrier to entry in creating a social networking company, decreasing the likelihood that a giant social networking company such as Facebook would face competition. Given the lack of preservation legislation in place or on the policy agenda as discussed in the last chapter, this solution is unlikely.

Despite the fact that Facebook is a U.S.-based company and U.S. law and regulations have influence over Facebook's behavior, over 80% of Facebook's one billion subscribers are from outside of the U.S. (Facebook, n.d. d) Because Facebook's subscribers are a global audience, international organizations such as the International Telecommunications Union (ITU), United Nations Educational, Scientific, and Cultural Organization (UNESCO), or the European Commission for Communications Networks, Content and Technology may be able to persuade Facebook to act as a TDR.

ITU is a United Nations agency that counts 193 countries and 700 private sector entities as members (ITU, n.d.). ITU sets standards, sponsors partnerships between public and private entities, and works to encourage:

> [E]ffective regulatory strategies and policies and an understanding of future trends: technical, social and economic. ITU brings partners together to discuss these issues, share insights and best practice, and lay the groundwork for long-term industry growth (ITU, n.d. a).

ITU is not in a position to mandate changes to social networking records preservation, but it is an organization that might facilitate a conversation about the preservation of historically significant records that are being created by a global community of subscribers.

UNESCO is an organization established by 20 member countries in 1946 as a response to World War II. (UNESCO, n.d.) UNESCO was formed to solidify the

"intellectual and moral solidarity of mankind" in an attempt to prevent future global conflict through communication, education, and cultural understanding. UNESCO has an established "Preservation of Documentary Heritage" program whose mission is to preserve a global documentary heritage that:

> [R]eflects the diversity of languages, peoples and cultures. It is the mirror of the world and its memory. But this memory is fragile. Every day, irreplaceable parts of this memory disappear forever (UNESCO, n.d. a).

The preserving documentary heritage program of UNESCO is an international program that has recognized the value of records and is working to preserve digital records by establishing the UNESCO charter on the preservation of digital heritage. This charter offers advice about the preservation of digital records including deciding what to keep, how to protect data and considerations of intellectual property rights (UNESCO, 2006). These guidelines are established to help individual archives both work with records creators and manage their own archival programs, but the international effort represented by these guidelines indicates that UNESCO is an organization that could influence Facebook and other social networking services to follow a set of preservation standards. Interestingly, UNESCO guidelines do not specifically cover protecting records creators' privacy.

The European Commission for Communications Networks, Content and Technology manages the "Digital Agenda" of the EU, a comprehensive plan with a mission to:

> [H]arness information & communications technologies in order to create jobs and generate economic growth; to provide better goods and services for all; and to build on the greater empowerment which digital technologies can bring in order to create a better world, now and for future generations (DG Connect, 2012).

DG Connect, as the EU Commission calls itself, includes projects that aim to improve privacy in the context of the Internet and to preserve digital records. While the preservation focus of this committee seems to be on digitizing paper records and protecting digital libraries, it is possible that the European Digital Agenda could include preservation of Facebook records. Facebook has been receptive to European influence in the past, such as changing privacy practices to satisfy Irish regulators (Halpin, 2012). While Facebook's change in privacy practices was in response to European privacy law, the European Digital Agenda could be a first step in creating similar requirements for preservation.

If the FTC, UNESCO, ITU, or European Commission manages to encourage Facebook to act as a trusted digital repository, social norms would have to be established that would encourage the long-term management of access to the records created using Facebook. As we learn from the focus group responses which I will discuss in detail in Chapter 6, Facebook subscribers do not use Facebook with preservation in mind. If these records persist over long periods of time, subscribers will need to manage access to records both during their lifetimes and after they die. With over one billion subscribers, a large number of Facebook subscribers pass away each day, 8,000 per day, by some estimates (Koetsier, 2012). This type of control over access to the records over the long-term will be similar to the deed of gift discussed in chapter three.

Federal regulation of social networking services for the preservation of interpersonal communication allows for self-regulation and certification given existing standards, practices, and infrastructure. This solution also preserves important governmental records created by government agencies using Facebook, without requiring each government agency to download and preserve their own records. The certification of

Facebook as a trusted digital repository would also address the lack of "trust" identified by the focus group described in chapter 6. As the ISO Certification points out, "Communicating audit results to the public—transparency—will engender more trust, and additional objective audits, potentially leading towards certification, will promote further trust in the repository and the system that supports it" (Consultative Committee for Space Data Systems, 2011, p. 2-1). Part of building this trustworthiness is consistency in providing access to the "designated community." This consistency, coupled with public documentation of the way Facebook manages records, documentation that would be verified by a third party audit, would improve the trustworthiness of Facebook.

Privacy would still be an important concern related to the long-term retention of Facebook records. But if Facebook subscribers were able to trust that Facebook was managing their records in a manner consistent with subscribers' use of privacy controls in the software, the protection of privacy would be a decision left in the hands of the subscribers, not the service provider. By including a third-party audit and a public set of processes to verify that Facebook is taking steps to act as a trusted digital repository, the information asymmetry expressed by the focus group respondents would be decreased. There would exist a reporting mechanism, ISO certification, that would allow subscribers to verify that the actions they have asked Facebook to take to protect their records were in fact being taken. This change in control over the records would reduce the amount of confusion and mistrust subscribers feel about Facebook, allowing subscribers to make more informed choices about what service they trust with their sensitive records and important memories.

FACEBOOK DONATES RECORDS TO A THIRD PARTY

Facebook could also partner with a third party to preserve Facebook records and protect privacy. The transfer of Facebook records could occur automatically after a pre-determined period of time, all records could be transferred, and the privacy settings in place prior to the transfer could be respected. Privacy could also be protected in this scenario if the collecting institution could provide Facebook subscribers a way to change privacy settings once the files were transferred, although ongoing access and account management would greatly increase the cost of providing preservation services.

Similar to the Twitter partnership discussed earlier, the Library of Congress (LOC) is one example of an organization that Facebook can partner with to preserve Facebook records. Facebook partnering with LOC is a more likely solution than NARA working with Facebook to change Facebook policy, given NARA's role as a government records archive and the LOC's role as a collecting archives. Facebook partnering with a third party as a preservation solution would have defined transfer times, similar to LOC's approach with Twitter. After the Library of Congress has had a message, post, video, or comment for six months, it could make that message available to researchers who request access for particular research projects. This six-month delay is the same time period that is currently in place for Twitter records.

This arrangement would help protect the privacy of individuals by respecting the privacy settings of Facebook's records creators and limiting the availability of these records. The records would be available only to researchers who apply for access through the Library of Congress. The Library of Congress has over 167 terabytes of information collected from the Web (Library of Congress, 2010), a collection that underscores their ability and commitment to preserving digital records and protecting privacy. If Facebook

did partner with the LOC, downloaded records that might be of historic value would be protected in the event that Facebook goes out of business or decides not to retain records older than a particular date.

The code used in the partnership between LOC and Twitter could be used as an example to allow the transfer between Facebook and the LOC. I am suggesting that Facebook should transfer all records, not just publicly available records, because Facebook has far more complex privacy rules than Twitter. A Twitter subscriber can decide to make her Tweets either "public" or "protected." A protected tweet is shared only with Twitter followers the subscriber has approved to view her tweets. Facebook posts are typically shared with all of one's "friends," which for some subscribers could number in the thousands. Facebook posts could also be "public," making them available to anyone with Internet access, including search engines, or shared with a range of smaller collections of Facebook subscribers from ad hoc groups to existing, predefined groups and "networks." Using a dichotomous setting of "public" or "private" to determine which records should be transferred would necessarily ignore the complexity and relationships demonstrated when Facebook subscribers use the software to communicate.

Restricting access to the Facebook archive and requiring anyone who does have access to respect the privacy of Facebook subscribers and publish only aggregate data is probably the most effective privacy protection strategy currently available. The LOC could use architecture to protect privacy by replacing the Facebook profile name with a unique identifier as the Facebook profiles are transferred. This additional step would be inadequate as a privacy protection strategy because profile pictures, address, and age information alone would be enough to identify most Facebook subscribers.

Legal and policy changes that might encourage a partnership between Facebook and a third party TDR are few and unlikely. One legal change that might encourage this type of partnership is for the Federal Communications Commission to regulate any communication service provider with more than one hundred million U.S. subscribers. While this number is arbitrary, it seems sufficiently large to satisfy the FCC's mission to "promote competition and reduce regulation in order to secure lower prices and higher quality services for American telecommunications consumers and encourage the rapid deployment of new telecommunications technologies" (47 U.S.C. §151).

While this new set of regulations would not require a service to be regulated as a monopoly, large information and communication service providers such as Facebook and Twitter would be subject to a specific set of rules that require them to establish and follow best practices related to both privacy and records preservation. A requirement such as this would allow wealthier social networking companies to devote resources to collaborative research and development in managing software and policy to protect both privacy and the preservation of records created using the software. This research and development could result in a set of policies, procedures, and standards that smaller social networking service providers could take advantage of.

To encourage this type of partnership using a market mechanism the federal government could offer a tax incentive to corporations that have the required number of subscribers. This tax incentive would offset the cost of maintaining a relationship with the third party TDR, and the investment would benefit consumers by making the research results mentioned above available to other social networking companies, improving the policy and technology management practices at both large and small firms.

Another market incentive could be to require subscribers to opt in to this program. If a subscriber agrees to allow her records to be transferred automatically to the TDR, that transfer could come at no cost. If a subscriber chooses to upload records or to opt in at a later date, that transfer could come at a cost. This pricing structure would work only if the education component, discussed below, underscored the importance and cultural value in preserving these records.

The social norms that would encourage this type of behavior would again need to be influenced through education programs. Facebook subscribers would need to be aware of when records were being transferred to the TDR, especially if their only opportunity to manage permission and access to their records was while those records still existed in Facebook. Subscribers would need to be aware of the use of these records for research and when these records would be made publicly available.

The problem with this arrangement is that, unlike Twitter, whose subscribers primarily have unprotected profiles, meaning all of their Tweets are public, Facebook subscribers have a much more nuanced set of privacy protections. Messages can be exchanged between just two or three individuals, subscribers can use a chat feature to converse in "real time," subscribers can belong to a group and communicate only to individuals in a group, subscribers can limit access to any material they post, and these privacy protections can be combined to create a complex set of rules determining whether an individual comment or conversation is "public." This complex set of rules would make it extremely difficult for Facebook and the Library of Congress to agree on a set of records that would be acceptable to Facebook subscribers, Facebook, and the Library of Congress.

Additionally, if Facebook partnered with LOC, subscribers would lose the ability to delete records once records were transferred to LOC, forfeiting the privacy protection strategy called "the right to delete" discussed in chapter two. In addition, moving the corpus of Facebook records and separating "private" from "public" messages, however that separation is determined, destroys the original order of the records and limits a researcher's ability to conduct research, as well as an archivist's ability to gain intellectual control over these records.

A THIRD PARTY "SCRAPES" RECORDS

"Scraping" or "screen scraping" is a technique that uses code to crawl the Web and download available files, similar to the way a search engine crawls the Web and creates a database of keywords and linking information. This technique can be used within just one domain, such as Facebook.com, to download information that is linked to publicly available information.

If a third party, such as the Internet Archive, were to conduct scraping of Facebook profiles, it would be able to download publicly available information to a repository and continue to make this information available, regardless of whether Facebook stays in business or maintains older records. This technique protects privacy to the extent that only records that are publicly available at the time of collection would be made available by the archive at a later date.

I will start by discussing the problems with third parties' downloading and preserving Facebook records because, in the context of Facebook, scraping would yield a distorted view of Facebook, the type of distortion Howard Zinn and Gerald Ham complained of in the 1970's. By downloading publicly available Facebook records,

mostly corporate and governmental marketing material would be kept. Most correspondence between Facebook friends is not available to a screen scraper, it is protected by code that essentially tells the Web crawling code that it does not have permission to access these particular database records. Privacy would also be affected if the collecting organization did not create a process for record owners to modify or delete the records they collected, limiting the subscriber's right to delete.

The architecture necessary for a third party to download Facebook records is currently in place and in wide use by organizations such as the Internet Archive. The code simply follows links from one Web site to another, downloading the html of each available page to a local server, maintaining the links that exist on each page. To address the problem of accessing protected records discussed above, Facebook would have to agree to allow the organization access to protected files, which essentially changes screen scraping to a partnership between Facebook and a third party.

Laws that would encourage a third party to download Facebook records are also relatively established. Copyright law on how and when publicly available Web pages can be downloaded and made accessible to a different audience has been tested by organizations such as the Library of Congress and the Internet Archive and is relatively clear on what is considered legal. The Internet Archive has been downloading Web pages since 1996 and has developed a set of policies and procedures that limit the liability of a not-for profit organization engaged in this type of activity.

The market incentive for third party collection of Facebook records would most likely be tax incentives through 501(c)(3) not-for-profit status. The Internet Archive also has a tested business model of offering an archiving service to organizations that conduct

business using Facebook and would like the ability to preserve their public Facebook profiles over time.

The norms that would support third party collection and preservation of publicly available Facebook records are awareness of the collection and the ability to solicit the collecting company to make the collected records unavailable if the owner of the record can prove ownership and has a "legitimate" reason for removal of the record.

The Internet Archive is "building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public" (Internet Archive, n.d.). Offering a Web-archiving service that "helps organizations to harvest, build, and preserve collections of digital content", The Internet Archive has also been contracted by the Library of Congress to preserve various Web sites and digital collections (Archive-It, n.d.). The Internet Archive has developed a set of reasons it will remove an existing file from the Way Back Machine, the collection of records it has developed over time from crawling the Web. These reasons include copyright violations and defamation, among others, and the Internet Archive has established a process to request that records be removed from the archive.

I have discussed four general approaches to the preservation of Facebook records: subscribers downloading their own records, Facebook acting as a TDR, Facebook partnering with a third-party company, and a third-party company acting independently. I used Lessig's four constraints of regulation (architecture, law, norms, and market), to discuss a variety of ways to encourage Facebook to preserve records and protect privacy at the same time, and I have discussed the problems with each approach to preservation.

Hopefully this brief overview can act as a starting point for policy entrepreneurs; for individuals who are interested in finding a way to encourage large social networking service providers, such as Facebook, to take steps to protect the cultural heritage records that billions of subscribers are creating each day. Without a change in Facebook's current practices, the preservation of the corpus of these records over a long period of time is uncertain, as are the privacy implications of records that will continue to exist for an indefinite period of time. In the next chapter I will discuss whether the U.S. federal government has cause to regulate Facebook using the market imperfection theory of regulation. I will then discuss whether the federal government is likely to regulate the preservation of Facebook records using Kingdon's garbage can model of public policy.

Chapter 5: Federal Information Policy and the Preservation of Facebook Records

In this chapter I will discuss whether the U.S. federal government should, and is likely to, play a role in preserving valuable Facebook records and protecting subscribers' privacy. I will first explore whether the federal government should intervene to preserve Facebook records and protect subscribers' privacy by discussing the market imperfections theory of regulation. This theory asserts five economic reasons for government regulation, which I will discuss in turn. I find that information asymmetry and positive externalities exist and offer some justification for government regulation. Both the online survey and the focus group data, which I discuss in detail in the next chapter, confirm the existence of information asymmetry between Facebook subscribers and the Facebook Corporation regarding how Facebook manages subscribers' records.

Having determined that there is some standing for the federal government to regulate Facebook, I look to John Kingdon's study *Agendas, Alternatives, and Public Policies* (2003) to frame the discussion of how a potential solution to the threat facing both the preservation of Facebook records and personal privacy might become an item for consideration on the policy "agenda."

GOVERNMENT REGULATION

According to Baron (2006), regulation is government intervention in economic activity using commands, controls, and incentives (p. 323). Government regulation as we understand it today began when the English monarchy granted stagecoach operators certain road privileges while retaining the right to regulate prices and services. This

exchange enabled private companies to provide a for-profit public service while putting the government in the position to control the price and the type of service.

The precedent set by this relationship between government and private companies carried through to the United States and was confirmed by the U.S. Supreme Court decision in *Munn vs. the State of Illinois* (94 U.S. 113, 1876). In this case, the Illinois constitution had designated all grain elevators as "public warehouses." The state then established regulations governing the prices these warehouses could charge, requiring the warehouses to submit weekly public statements as to the amount and quality of the grain in the warehouse, and requiring warehouse owners to allow the owners of the grain stored in the warehouse to be "at liberty to examine such property stored, and all the books and records of the warehouse in regard to such property" (Article XIII, Section 3, 1876).

When Munn and Scott, the owners of the North-Western Elevator, were fined because they did not file for the proper county license, they sued the state of Illinois for violation of the due process clause of the Fourteenth Amendment of the U.S. Constitution. Chief Justice Waite delivered the opinion of the Illinois court, looking back 200 years to English common law to quote Lord Chief Justice Hale stating that when private property is "affected with a public interest, it ceases to be *juris privati* only" (1876). Justice Waite explains "[w]hen, therefore, one devotes his property to a use in which the public has an interest, he, in effect, grants to the public an interest in that use, and must submit to be controlled by the public for the common good, to the extent of the interest he has thus created" (1876).

This state Supreme Court opinion set a precedent for the federal regulation of interstate commerce and private industry in the United States when the "public interest" is implicated by the service being regulated. As we see from *Munn*, the regulation of a

private company, such as Facebook, by the federal government has a long history. In the last part of the nineteenth century, the Sherman Act (1890) and other antitrust laws increased federal regulation of private industries including oil, steel, and railroads. The Mann Elkins Act (1910) specifically extended federal regulation to telecommunications services such as the telephone, telegraph, and cable companies, extending the precedent of *Munn* to the communications industry.

The 1934 Communications Act (1934) had a number of different provisions, including the establishment of the Federal Communications Committee (FCC) as a regulatory body and the designation of telephone, telegraph, and cable providers as common carriers. A common carrier, according to the act, is:

> [A]ny person engaged as a common carrier for hire, in interstate or foreign communication by wire or radio or in interstate or foreign radio transmission of energy, except where reference is made to common carriers not subject to this Act; but a person engaged in radio broadcasting shall not, insofar as such person is so engaged, be deemed a common carrier.

This definition helps to separate the economic term "a public good" from the mostly political term "the public interest." *Munn* established a precedent wherein the federal government regulates private industry in an attempt to protect "the public interest," or a commons in which the public has an interest. "A public good," on the other hand, is a term I will explore in more detail later in the chapter when I discuss the market imperfections theory of regulation.

While a comprehensive history of regulation in the United States is beyond the scope of this study, I will use the right of the government to regulate as established by *Munn* and a common theory of regulation, the theory of market imperfections, to explore whether the federal government should regulate Facebook. Later in the chapter I briefly explore whether Facebook provides a service that should be considered a "public good"

even though my aim in this chapter is much narrower; I am interested specifically in ways we might use public policy to encourage the preservation of records created using Facebook while protecting privacy.

MARKET IMPERFECTIONS THEORY OF REGULATION

There are many theories that explain when government regulates private companies. One of the most widely used theories that helps explain government regulation in a capitalist system is the theory of market imperfections. This theory recognizes five types of market imperfections: natural monopoly, public goods, externalities, asymmetric information, and moral hazard (Baron, 2006, p. 332). Each of these market imperfections may be relevant to the preservation of Facebook records, and I will discuss each in turn.

The first market imperfection to discuss is natural monopoly. The theory of natural monopoly predicts that a monopolist will restrict its output to increase prices above marginal cost (p. 332). Neither scholarly writing nor legal decisions have designated Facebook a monopoly. We can look to the recent designation of Microsoft as a monopoly to help us understand whether Facebook might be considered the same.

U.S. District Judge Thomas Penfield Jackson designated Microsoft a Monopoly in *U.S. vs. Microsoft* in his finding of fact:

> Viewed together, three main facts indicate that Microsoft enjoys monopoly power. First, Microsoft's share of the market for Intel-compatible PC operating systems is extremely large and stable. Second, Microsoft's dominant market share is protected by a high barrier to entry. Third, and largely as a result of that barrier, Microsoft's customers lack a commercially viable alternative to Windows (Civil Action No. 98-1232, section III #34).

Facebook has yet to receive a similar designation in any federal court decision. Despite the large numbers of Facebook subscribers, Facebook does have commercial competitors that offer very similar services such as Google+ and MySpace, and many other social media competitors such as LinkedIn, Twitter, and Pinterest. In addition, the barrier to entry to the competitive social networking space is very low. Open source software is available for a number of platforms, and hosting for a new social networking site start at about $10 per month. The popular blogging platform Wordpress offers a free social networking plug-in called Buddypress, making the cost to start up a social networking service negligible.

The discussion of price and cost that is used to describe a natural monopoly in this theory does not directly apply to social networking records and subscribers. Subscribers do not pay a "fee" to use Facebook. However, the service does have a "price" because subscribers face a cost in the revelation of personal information that is then used by Facebook to sell advertising services. I have discussed the relationship among advertisers, Facebook, and subscribers in earlier chapters, and I will consider these relationships in more detail later in the chapter.

For the purpose of determining whether the federal government should regulate the preservation of Facebook records, a consensus does not exist that Facebook is a monopoly; therefore natural monopoly is not one of the market imperfections to consider as reason for regulation.

The second market imperfection to consider in the context of the regulation of the preservation of Facebook records is when a service is considered a common or public good. The term "public good" has a long history that includes many different concepts and definitions over time. For the purposes of this chapter, I will focus on a commonly

accepted economic definition and two terms that help us understand this definition more clearly.

When use of a product or service by one person does not reduce its availability for other people, the theory of market imperfections recognizes the product as a public good (Baron, p. 335). This concept is commonly referred to as nonrivalry in consumption (Kaul, Grunberg, & Stern, 1999, p. 3). The term "nonexcludability" is often used in conjunction with nonrivalry (Kaul et al., p. 3). Nonexcludability refers to the inability to exclude individuals from using the product or service. The definition of a product or service as a public good is not dichotomous; there is a range of potential designations of a product between private good and public good.

Facebook is nonrivalrous in consumption even though there is an incremental cost to Facebook for each new subscriber in terms of data storage, bandwidth usage, and general account management costs. Facebook has over one billion subscribers, yet each new subscriber does not decrease the availability of the service for the other subscribers. Facebook has added over one billion users in eight years, which represents a rate of close to 300,000 new subscribers per day. At this scale, the addition of a single subscriber does not reduce the availability of the service.

Facebook's ability to provide the service depends on the revenue gained by the firm from advertising. If this revenue were limited or began to decrease, the service might well become rivalrous in order to control the incremental cost of each new subscriber.

Indefinitely preserving the records produced by over one billion people using the service poses a substantial additional cost to Facebook unless, as we have discussed earlier, Facebook considers those records to continue to have primary value for an indefinite period of time. Facebook uses subscribers' historical records to create

consumers' profiles that will allow for more accurate predictions of behavior both for that individual and for other subscribers with similar characteristics. These profiles increase the value of the advertising space Facebook provides to advertisers. Whether that value would continue to accrue over the course of thirty to fifty years is difficult to determine.

Access to Facebook cannot be said to demonstrate the characteristic of nonexcludability. It is trivial for Facebook to exclude particular individuals from access to the Facebook software by either suspending or deleting that individual's Facebook account. Facebook routinely denies access to the software when a subscriber violates Facebook's terms of service, for example.

Given that Facebook can and routinely does exclude subscribers, but that the service may be considered nonrivalrous, Facebook does not clearly fit the definition for the market imperfection term of "public goods."

I have mentioned that Facebook has gained close to 300,000 subscribers per day over the course of the last eight years. This increase in the number of subscribers increases the perceived value of the service to existing subscribers because there are more people to be "friends" with. The increase in the perceived value of the service with each additional subscriber is what Baron would consider a positive externality. Baron defines two broad types of externalities: pecuniary and non-pecuniary (p. 334). A pecuniary externality exists when the choices of one economic agent are affected by the actions of other economic agents through changes in the price of a good or service. Facebook clearly is not an example of a pecuniary externality because subscribers' access to the service comes at a consistent price; the exchange of personal information.

A non-pecuniary externality exists when the choices of one economic agent are affected by the actions of another economic agent through factors other than price. These

externalities can be negative, such as pollution created by a factory, or positive, which is the case for Facebook. The actions of independent actors add value to the service provided by Facebook without significantly changing the cost to provide that service. As more people join Facebook, each individual subscriber can depend on using Facebook to reach a growing audience, whether the audience consists of family members, real-world friends, or potential business customers. Rohlfs discussed the idea that "the utility that a subscriber derives from a communication service increases as others join the system," (1974, p. 16) in relation to telephone communication services, an idea commonly referred to as both a network externality and a network effect.

I have established that Facebook is not a natural monopoly and cannot easily be categorized as a public good, but that the service does create positive externalities for service subscribers. The last two market imperfections I will consider are asymmetric information and moral hazard.

If actors have different information from each other when they choose to act, an asymmetric information market imperfection exists (Baron, 2006, p. 335). The data from both my online survey and the focus group, discussed in detail in chapter six, indicate that most respondents (64%) did not know whether Facebook used adequate backup procedures, and most respondents (65%) did not think that Facebook would delete records when a subscriber requested those records to be deleted. Focus group respondents complained that there was no way for them to "know what Facebook actually does with our records."

This assertion and these data from the focus group and online survey indicate an information asymmetry exists, which could be used as justification for government regulation of the preservation of Facebook records. The U.S. has a history of establishing

regulation to address information asymmetry; the Food and Drug Administration and the Securities and Exchange Commission are only two examples of government organizations established to limit information asymmetry between citizens and corporations. Information asymmetry can affect an individual's ability to choose a service provider that provides services consistent with her needs. In the context of Facebook, if subscribers are not aware of Facebook's plans and procedures to preserve records, subscribers are not able to choose a service that is consistent with their desire to both protect their privacy and preserve their records.

Moral hazard is the last market imperfection I will consider. Moral hazard, according to Baron, occurs when "inefficient actions induced by policy instruments cause people not to bear the full consequences of their actions" (p. 336). In the context of the preservation of Facebook records, moral hazard is not a reason for government regulation. Policy instruments do not exist that require the preservation of Facebook records beyond a time period that Facebook deems profitable. The lack of a policy instrument means that Facebook, the market, Facebook subscribers, or other actors including advertisers and law enforcement officials, not policy instruments, introduce any inefficiency that may exist. The federal government would first need to determine that regulation of the preservation of Facebook records to preserve cultural heritage artifacts was necessary, and then introduce, approve, and implement a policy, for a policy response to introduce inefficiency. Without an existing policy response, moral hazard does not exist.

Through this brief discussion of the market imperfections theory of regulation we have determined that an asymmetry of information and positive externalities exist in the context of the preservation of Facebook records. Based on these conclusions we

understand that the government *might* regulate Facebook, but in order to understand how federal regulation to encourage the preservation of Facebook records might come about, I turn to John Kingdon's study *Agendas, Alternatives, and Public Policy*.

KINGDON AND PUBLIC POLICY

The market imperfection theory of regulation relies on economic theory to understand whether there is an economic reason for the government to regulate an industry. Widely recognized as one of the most important studies of public policy formation, Kingdon looks more broadly at how public policy is determined rather than the economic reasons government regulates business. Kingdon conducted four separate sets of interviews from 1976 – 1979, interviewing 247 individuals during that time (Kingdon, 2003, p. 237). Kingdon's respondents included congressional staff members, members of the executive branch, and people outside of government including lobbyists and researchers (p. 238).

Kingdon defines public policy making as a set of processes including: setting the agenda, determining alternatives, choosing an alternative, and implementing the policy (p. 3). Kingdon's study looks at the process of setting the agenda and defining and selecting alternatives. Kingdon defines the public policy agenda as "the list of subjects or problems to which government officials, and people outside of government closely associated with those officials, are paying some serious attention at any given time" (p. 3).

The data Kingdon collected over the course of four years of interviews revealed three separate sets of processes related to agenda and alternative setting: problems, policies, and politics (p. 16). These three separate processes influence agendas in unique

and sometimes interrelated ways. A problem includes any "crises or prominent event" (p.16), the terrorist attacks of 9/11 are one obvious example of a "problem." Kingdon differentiates between a condition and a problem, "conditions become defined as problems when we come to believe that we should do something about them" (p. 109). An asymmetry of information related to the preservation of Facebook records, for example, is a condition unless an elected official can be convinced that the federal government should do something about that condition. The policy process influences agenda and alternative setting through research and the accumulation of knowledge, as well as development of new policy alternatives. Politics has an effect on agenda and alternative setting both through the regular cycle of electing public officials, the appointment of public officials, and "swings of national mood" (p. 17).

Kingdon is careful to point out that the development of public policy does not follow a rational, comprehensive path. "For various reasons already developed by other writers, such a model does not very accurately describe reality" (p. 78). Instead he builds on what Cohen, March, and Olsen (1972) call "the garbage can model" of organizational decision-making.

Cohen et al. use the term "choice opportunity" to describe a set of circumstances that allows actors to make a decision and effect change, such as the selection of a dean for a university school. These "choice opportunities," according to Cohen et al., are the:

> [G]arbage can into which various kinds of problems and solutions are dumped by participants as they are generated. The mix of garbage in a single can depends on the mix of cans available, on the labels attached to the alternative cans, on what garbage is currently being produced, and on the speed with which garbage is collected and removed from the scene (p. 2).

Kingdon adapts the garbage can model proposed by Cohen et al. to the development of public policy within the federal government by including the three streams we have

discussed earlier: problems, policies, and politics. Each of these streams is separate and develops "garbage" in the form of problems that need to be addressed, policy solutions, and changes in national mood or elected officials without necessarily being influenced by the other streams. A "policy window" is a choice opportunity related to agenda or alternative setting that occurs when "a problem is recognized, a solution is available, the political climate makes the time right for change, and the constraints do not prohibit action" (p. 88).

The policy window has also been discussed as the "issue attention cycle" (Downs, 1972, p. 38). The issue attention cycle has five stages: the pre-problem stage, the alarmed discovery and euphoric enthusiasm stage, the realizing the cost of significant progress stage, the gradual decline of intense public interest stage, and the post-problem stage (pp. 39 – 40). Downs describes three characteristics of problems that are likely to go through the issue attention cycle: problems that affect a minority of people, the suffering of that minority is caused by an arrangement that benefits some majority or powerful minority, and the problem has no intrinsically exciting characteristics.

The issue attention cycle is useful as we discuss possible solutions to protecting privacy and preserving Facebook records. If we consider archivists and researchers to be the group of people affected by the absence of preservation efforts of Facebook records, then the problem does affect a minority. However, the disappearance of Facebook records is a long-term cost to society in general because these records are valuable cultural heritage records, but only archivists and historians currently recognize the loss as a problem. The Facebook Corporation is a strong minority that benefits from avoiding the cost of preserving Facebook records, and the potential disappearance of Facebook records does not have intrinsically exciting characteristics. The issue attention cycle is useful in

helping us understand why archivists and privacy experts should work together now to resolve their seemingly contradictory understanding of the preservation of Facebook records. The preservation of Facebook records as a policy issue is in the pre-problem stage, a good time for privacy experts and archivists to discuss their differences and hopefully agree on a recommended course of action if and when the issue attention cycle reaches the alarmed discovery and euphoric action stage.

In order to address the question of how the federal government might regulate Facebook in order to preserve valuable historic records and protect privacy we need to better understand how the preservation of Facebook records might be included on the federal policy agenda. The three streams (problems, policies, and politics) would need to be coupled as part of a "policy window."

The politics stream is certainly ready for a change in agenda. When a new president is elected or an existing president is re-elected, he is expected to set the agenda. Kingdon finds that the president himself has more power than any other single actor to set agendas (p. 23).

The problems stream offers less of an opportunity for change. While there have been instances of smaller services being shut down overnight, the loss of valuable records rarely occurs as a single dramatic event. One example of loss of access to records hosted on a third party service is the April 2012 seizure of New Zealand based file sharing service MegaUpload's servers by the U.S. Department of Justice (DOJ). Once the servers were ordered offline by the DOJ, the hosting service responsible for keeping them running wanted to know whether the DOJ or Kim Dotcom, the owner of MegaUpload who also had his assets frozen, would continue to pay the $37,000 per month (Pelofsky, 2012) to keep the servers running.

The service had over 150 million customers who lost access to their files overnight. While much of the material may have infringed on copyright, some of the material was legally owned and posted to MegaUpload's servers. The DOJ asserted they had enough evidence to try the MegaUpload case, and that the servers should be shut down and no customers should be granted access to the records they have stored on MegaUpload's servers.

Another example of valuable personal records being threatened by a third party service comes from the Society of American Archivists (SAA), an organization we would expect to err on the side of preserving historical records. In 2007, Nancy Beaumont sent a note to the Archives and Archivists listserve on behalf of the SAA Council, to inform subscribers that the stored copies of messages sent to that list between 1993 and 2006 would be deleted because the "costs of maintaining the listserv archive outweighed the benefits" (Prelinger Library, 2007). The outcry that followed led the SAA to rethink its decision, but this event is an example of how an organization with leadership responsibility for the archives profession can come to a decision to permanently delete subscribers' records.

A single researcher usually notices this type of loss as she looks for records and finds these records are no longer available, rather than a large, attention-worthy event or catastrophe. If Facebook, Twitter, or even MySpace were to suddenly close down and delete all of their records, we may have a problem that would garner national attention. Whether that situation would be seen as a condition or a problem requiring government intervention remains to be seen.

EXISTING AND PROPOSED LEGISLATION

In order to discuss the policy stream, I decided to review existing and proposed legislation related to the preservation of social networking records and the protection of privacy. If legislation exists that is directly related to the preservation of social networking records and the protection of privacy, we can assume that the problem, policy, and politics streams have already been coupled and that government regulation of social networking records is on the agenda. If there exists legislation that addresses similar conditions, we may be able to learn whether the preservation of Facebook records is likely to be included on future agendas. I begin the review of existing legislation by discussing some of the first speech regulation in the United States, the First Amendment to the Constitution, which defines the constraints and protections of the type of personal speech represented by Facebook records.

The regulation of speech has a long history in the United States. The First Amendment to the Constitution mandates "Congress shall make no law… abridging the freedom of speech." (U.S. Const. amend. I) A common understanding of the limit to free speech is from *Schenck v. United States* (1919), in which Justice Oliver Wendell Holmes Jr. supports the conviction of the secretary of the Socialist party for distributing anti-war leaflets by saying:

> The question in every case is whether the words used are used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent (1919).

The concept of free speech has changed over time, but the regulation of speech in the context of the preservation of interpersonal communication records has little, if any, legal precedent. Most speech regulation is reactive; protecting individual rights both to free

speech and freedom from harassment, blackmail, and other infringing activities after the speech has occurred. These regulations pertain to statements made by individuals or groups whether in person, by phone, mail, or text message, and these messages are preserved only as part of court records. Only speech that is recorded by a government employee as part of her official duties or part of a criminal or civil proceeding has been regulated in the past.

The Post Office has historically ensured privacy of personal correspondence. The Continental Congress established the federal postal system in 1775, but Congress did not pass a law protecting the privacy of records sent through the postal system until 1782 (Seipp, 1978). The assurance of the privacy of records sent through the mail is limited, and, as we discussed before, privacy was protected only until the correspondence was delivered by the postal service to the recipient. Once an individual receives her mail, the Post Office has no obligation or ability to protect it. We discussed earlier how Facebook presents a different set of circumstances in that the recipient of a message is granted access to that message. The original record is created using Facebook's service and is stored on Facebook's servers.

The Federal Records Act of 1950 established a framework for federal records management that required the National Archives and Records Administration (NARA) to oversee the management of records in all federal agencies.

Chapter 33 of Title 44, United States Code established the procedures by which an agency could destroy records it has created. NARA established the general records schedule (GRS) that details the procedures for the most common federal records, and each agency was charged with submitting a records schedule that pertains to any records that are not covered by the GRS. So we do have a precedent for the federal government

regulating federal records, including correspondence conducted in the course of federal business.

The records of the Office of the President fall under a more specific set of data retention requirements. The Nixon Watergate scandal lead to the Presidential Records Act (PRA) of 1978 (Barker, 2005). The PRA defined Presidential records in contrast to personal records maintained and created by the President, and allows the President to block the disclosure of Presidential records for specific reasons and according to a defined process. The PRA was later modified by President Reagan's Executive Order 12667 (EO 12667; Barker, 2005, p. 6) and by President George W Bush's EO 13233 (EO 13233; Barker, 2005) which extended the definition of Presidential records to the records of the Vice-President, and expanded the reasons a President or former President can withhold records from a Freedom of Information Act request. President Obama issued EO 13489 (2009), which essentially restored the requirements of EO 12667. But even in the case of these three Executive Orders, the EO separates personal records from business records, with no regulation of the retention of personal records, and limited access to former Presidents' business records.

The E-Government Act of 2002 was passed to "to improve the methods by which Government information, including information on the Internet, is organized, preserved, and made accessible to the public" (2002). The act had a number of stated goals including promoting the use of the Internet to increase citizen participation in government, to make the Federal Government more accountable and transparent, all in a "manner consistent with laws regarding protection of personal privacy, national security, records retention" (2002) The act required the newly established Office of Electronic Government to establish an e-government strategy and to submit an annual progress report to Congress.

The E-Government Act has not resulted in a strategic plan for the preservation of social media records. While the Library of Congress' partnership with Twitter and NARA's list of social media initiatives may have been inspired by the requirements of the E-Government Act, the discussion of the preservation of electronic records is limited. The E-Government Strategy created the Associate Director for Information Technology and E-Government (Forman, 2002), but preservation was mentioned only once, in Appendix C:

> This project will provide guidance on electronic records management applicable government-wide and will provide tools for agencies to transfer electronic records to NARA in a variety of data types and formats so that they may be preserved in for future use by the government and citizens (p. 33).

The fact that preservation was mentioned at all is interesting and demonstrates that NARA has some influence over the policy and strategic plan development process, however limited that influence might be.

The federal government has chosen to regulate Facebook to address privacy concerns. In November of 2011 the Federal Trade Commission settled a lawsuit that accused Facebook of compromising subscribers' privacy. Jon Leibowitz, Chairman of the FTC said, "Facebook is obligated to keep the promises about privacy that it makes to its hundreds of millions of users" (FTC a, 2011). Changes Facebook made to the software revealed the previously private information of subscribers, prompting the lawsuit. This regulation demonstrates that the way Facebook treats subscribers' records is a matter for federal regulation.

To determine whether the preservation of social networking records has been part of the federal legislative agenda over the past three years (2009 – 2012) I conducted a search of proposed legislation using the Library of Congress Thomas search engine. I

searched using the terms "social," "media," "records," "preservation," "network," "networking," "Facebook," "preserve," and "preserving."

The search yielded two interesting results that are not directly related the preservation of social networking records, but that indicate some level of interest by Congress in similar topics. The first proposal is titled The Preserving the American Historical Record Act, S. 3227, submitted by Senator Orrin Hatch (R- Utah) and referred to committee on April 19, 2010. There is a related proposal by the same name in the House, and it was proposed by Representative Maurice Hinchey and referred to subcommittee on July 31, 2009. The act requires the Archivist of the United States to make grants to state Archivists to preserve and make available historical records.

Representative Paul Rhodes from New Hampshire proposed the Electronic Message Preservation Act on March 9, 2009. The act requires the Archivist of the United States to establish and periodically review regulations governing the preservation of electronic messages that are considered federal records. The act was referred to the Committee on Homeland Security and Governmental Affairs on March 18, 2010.

As we can see from these search results, some representatives have recognized that the preservation of historical records and the preservation of electronic records should be addressed by federal policy. The fact that only two results related to my search terms for the 112th and 111th Congress' indicates that there is limited recognition that preservation is a "problem" as Kingdon uses the term. In addition, neither of these results addresses personal correspondence using social networking services.

Without a policy precedent on which to build, we need to look to other contemporary events that may indicate an interest in preserving social networking records. The only nationally recognized event that indicates interest in the preservation of

social networking records at this scale is the agreement between the Library of Congress and Twitter. While this is an important collaboration, it hardly indicates an overall increase in national interest in preservation.

I have discussed the market imperfection theory of government regulation and concluded that information asymmetry and a positive externality exist as reasons for government to regulate Facebook records. I also discussed Kingdon's garbage can model of agenda and alternative setting and used his model to analyze current policies, problems, and politics. From this analysis I learned that the federal government does not have a history of regulating ordinary interpersonal communication, that no relevant federal policies have been considered in either the 111th or 112th Congress, and that there is only one large-scale example of a government entity collaborating with a social networking firm to preserve social networking records; the Library of Congress' collaboration with Twitter. The market imperfection theory leads me to believe the federal government could regulate Facebook records for preservation purposes, but my exploration of current and proposed legislation leads me to believe that a policy window does not currently exist.

Chapter 6: Subscribers' Attitudes: Online Survey and Focus Group

In the preceding chapter I used market imperfection theory to discuss whether the U.S. federal government has reason to regulate the preservation of Facebook records. In addition I referred to Kingdon's version of the garbage can model of public policy to discuss whether the federal government is likely to regulate the preservation of Facebook records. I concluded that positive externalities and information asymmetries exist in the context of the privacy and preservation of Facebook records, giving the federal government reason to regulate. I also concluded that a policy window does not exist; the lack of existing or proposed legislation and the lack of a major public disaster or event make federal regulation of the preservation of Facebook records unlikely.

A close examination of the seemingly opposing viewpoints of privacy scholars and archivists revealed that when privacy scholars say they are concerned that social networking records threaten privacy because they will exist "forever," they are actually referring to access control and persistence, rather than persistence alone as "forever" implies. In addition to a more clearly articulated concern for privacy, I discussed the lack of agreement among archivists that digital records will continue to exist for long periods of time. I discussed possible reasons Facebook records may be considered valuable and worth preserving, including cultural heritage, personal memory, and the implications of the volume of Facebook records for research. I explored some possible ways these records may be preserved.

Facebook is being used by one billion subscribers to create potentially valuable records that we cannot be sure Facebook will choose to preserve, which means the

preservation of these records may depend on the actions of individual Facebook subscribers. In this chapter I have taken a first step at empirically addressing the expectations and behavior of a group of Facebook subscribers to better understand what Facebook subscribers expect to happen to their records and whether subscribers believe they have control over these records by asking two research questions:

1. What are Facebook subscribers' preservation expectations?
2. To what extent are Facebook subscribers' preservation expectations associated with subscribers' disclosure behavior?

The first question is a preliminary step at exploring subscribers' preservation expectations. There is a dearth of previous research on how long individuals expect records created using software to exist. As I discussed in chapter four, how long subscribers expect their records to exist, whether subscribers believe they have control over those records are primary considerations in understanding how subscribers may choose to act, and whether subscribers are thinking about preservation when they use Facebook, might help archivists understand what they would need to do in order to train Facebook subscribers to become "citizen archivists."

The second question has been explored by previous research, as I have discussed in earlier chapters. Privacy scholars have asserted that social networking records are a threat to privacy because they will continue to exist "forever." Given Rogers' (1983) protection motivation theory, we would expect Facebook subscribers to decrease their disclosure behavior if they recognize the persistence of Facebook records as a threat to privacy, if they agree that the prescribed steps they need to take to protect their privacy would be effective, and if they believe they have the technical skill to execute these protection strategies. We would also expect subscribers who do not trust Facebook to

disclose less information than subscribers who do trust Facebook (Dwyer, Hiltz, & Passerini, 2007).

Understanding the extent to which preservation expectations are associated with disclosure behavior contributes to both the archival and the privacy literature. A better understanding of what Facebook subscribers choose to share and how subscribers' disclosure behavior varies among different groups is an important area of study as archivists continue to determine the value of Facebook records. One way for subscribers to protect their privacy is to share less information using Facebook. An improved understanding of the influences that are associated with sharing behavior will improve privacy scholars' ability to understand how to influence Facebook subscribers to better protect their privacy and allow scholars to have a clearer understanding of the threat to privacy posed by the persistence of Facebook records.

DATA COLLECTION

In a first attempt to determine Facebook users' preservation expectations I conducted both an online survey and a focus group with primarily master's level students at the Information School at a major research university. I recruited participants for the online survey using both the Masters and PhD student listservs hosted by the school.

The online survey instrument included 25 questions, took on average ten minutes for respondents to complete, and included three groups of questions. One group of questions measured preservation expectations, or how long participants expect their Facebook records to exist, and whether respondents believed they had control over the existence of and access to these records. The second group of questions identified disclosure behavior, which is the amount and type of information respondents are likely

to share. The third group of questions included various demographic and psychometric questions used both as control variables and to compare results of this survey to those of previous research.

I distributed questions from the three groups throughout the survey instrument. Appendix C includes the full set of questions from the survey.

I conducted the focus group interview at the same Information School and included six graduate students. I recruited participants by asking master's students in a particular class at the school to participate in the one-hour long focus group session. I used a predetermined set of questions (see Appendix D) to structure the interview, but participants were able to change the direction of the conversation and introduce topics or questions that they were interested in.

ANALYTIC SAMPLE

Of the 155 online survey respondents, I use 144 in this analysis. One respondent was under 18 and was not allowed to participate. Ten respondents were not included in the sample because they did not respond to at least one of the important questions used in the analysis of the survey data. The majority of participants were Information School graduate students, who are more likely to be more knowledgeable about social networking, privacy, and archival issues than the average Facebook subscriber. The characteristics of sample respondents are described subsequently in this chapter.

VARIABLES

The following variables were used in my analysis of participants' responses to the online survey.

Preservation Expectations

I asked eight questions related to preservation expectations derived from archival theory and elements of the requirements of a trusted digital repository. These eight questions included a subset of four questions related to the persistence of Facebook records and a subset of four questions related to subscribers' trust that Facebook would act according to subscribers' intentions.

The first three questions related to the persistence of records asked respondents to review a statement and indicate whether they agreed, disagreed, or did not know whether the statement was accurate. The first statement was "Facebook uses adequate backup procedures to protect my posted information from being lost." Confidence in Facebook's backup is an indication that subscribers expect their records to persist over time.

The second statement I included in the survey instrument was "I expect my Facebook information to be available indefinitely into the future even if I never log in again." This statement refers directly to preservation expectations and the persistence of Facebook records by asking for respondents' reaction to both of the major concerns of archivists I discussed earlier; the persistence of records and access to those records.

The third statement "my Facebook information will exist forever" tested preservation expectations in a slightly different way. This question was developed based on the expressed concerns of privacy scholars and the popular press that social

networking records are a threat to privacy in part because they will continue to exist "forever."

I used different terminology in questions two and three to refer to the persistence of records over long periods of time. Archivists typically refer to preservation as a way to maintain records for future generations but rarely use the term "forever." As I have discussed in earlier chapters, privacy scholars and the popular press use the term "forever" when describing the threat to privacy attributed to the persistence of social networking records. By using these two different terms I am able to compare my sample participants' reaction to the terms "forever" and "indefinitely."

The fourth persistence question asked participants to respond to the statement "I expect my Facebook information to be available to me for the next," and included six possible responses: zero to six months, seven months to less than one year, one year to less than three years, three to five years, six to ten years, or more than ten years. This question was asked to gain a more clear understanding of the time period respondents expected records to persist.

The second set of four questions measured whether the subscriber expects Facebook to allow her to have access to and control of her records in the future, a concept I refer to as "trust." I use the term "trust" in part because focus group respondents continually used that term to discuss whether they expected to have control over or access to their Facebook records in the future. I also use the term "trust" because it subsumes the idea of controlling access to the records.

For the subscriber to believe she will have control over access to her records, she needs to "trust" that Facebook will act according to her wishes; that Facebook will delete records when she asks for them to be deleted, that only subscribers who she decides

should have access to her records will be granted access, and that same set of permissions will be consistent over time. In the context of a trusted digital repository, these access and trust concerns are addressed by the Deed of Gift agreement that I discussed earlier. There is no such Deed of Gift for Facebook, only the terms of service agreement that is subject to change over time.

When archivists talk about "trust" in the context of a trusted digital repository, the term usually refers first to providing reliable, long-term access to records, but "trust" also is used to refer to transparency and accountability, including sharing the results of audits with the designated community (CCSDS, 2011, p. 3-8). Respondents clearly did not believe Facebook was acting in a transparent manner that would allow for accountability, and at least one comparison of Facebook and MySpace found that Facebook subscribers trusted Facebook more, and tended to disclose more personal information, than MySpace subscribers (Dwyer at al., 2007).

For this set of four questions, I asked respondents to indicate whether they agreed, disagreed, or did not know whether the statement was true. The first of the trust statements was "If I deactivate my account, Facebook will immediately delete all of my records." This statement does not accurately describe how the Facebook software acts and is intended as a proxy for respondents' knowledge about Facebook. The second statement was "If I delete information from my profile, or delete my account completely, information in backup systems will not be available to anyone," and was intended to measure respondents' trust that deleted Facebook records would not be available even if those records were deleted from the "production" Facebook database.

The third and fourth statements were "Facebook deletes posts immediately after I remove them" and "If I delete my account, Facebook will immediately delete all of my

records" respectively. These statements intended to measure whether respondents believed that the Facebook software acted according to subscriber intentions; whether the software did what subscribers told it to do.

Responses to the four trust questions were summed to create a trust index with a possible score ranging from four to twelve.

Disclosure Behavior

The 12 questions I asked related to disclosure behavior were originally used by Gross and Acquisti (2005), and Christofides, Muise, & Desmarais (2009) to describe the disclosure behavior of Facebook users. I included the following disclosure statements from Christofides et al., rated on a five-point Likert scale from very unlikely to very likely:

1. How likely are you to post profile pictures of yourself?
2. How likely are you to post pictures of yourself with friends?
3. How likely are you to post pictures of yourself with friends doing something illegal?
4. How likely are you to post pictures of yourself naked or partially naked?

I combined responses to these questions to create an index that ranged from a score of four if the participant responded "very unlikely' to each question, through 20 if the participant responded "very likely" to each question. I refer to this scale throughout the rest of the study as "future disclosure" even though the respondents may indicate "very likely" as a response if they are currently disclosing the type of information discussed in the question.

I also included the following statement from Christofides et al. related to the type of information subscribers post on Facebook:

Choose all of the following items you have posted on Facebook:

1. Birthday
2. Email
3. Hometown
4. Relationship status (single, married, divorced)
5. School you have attended
6. Program you have attended
7. Phone number
8. Home address.

These questions were asked to help explore disclosure behavior and the type of information subscribers choose to share. I used participants' responses to these eight disclosure behavior questions to create a scale that measured current disclosure behavior, or items respondents report they have currently disclosed in Facebook. Each item a particular respondent disclosed was given a value of one. The disclosure items were then summed for each respondent creating an index range of zero to eight for current disclosure.

Gross and Acquisti articulate a number of privacy threats that are based on the disclosure of these data elements. These threats include stalking, re-identification or the ability to use these data to identify subscribers across multiple different data sets, and building a digital dossier (Gross & Acquisti, 2005, sections 4.1 – 4.3). One very significant privacy threat Gross and Acquisti point out is the use of hometown and birthday information to "re-identify" a subscriber's social security number (section 4.2), facilitating identity theft.

Control Variables

To understand whether preservation expectations and disclosure behavior vary based on demographic and behavioral characteristics, I asked questions related to age,

race, level of education, whether the respondent was currently pursuing a degree, how often she uses Facebook, and how long she has been using Facebook. The frequency of Facebook use question was a five point Likert scale ranging from rarely to 2 hours per day or more. The duration of Facebook use question used a four point Likert scale ranging from under six months to more than two years.

ANALYTIC STRATEGY

The data I gathered from the focus group were used to confirm that the questions I asked in the online survey made sense to the population of interest for this study, the majority of whom are graduate students at the Information School. I took detailed notes of the focus group responses as well as an audio recording. I transcribed the audio recording, removing any information that might identify any respondents, and used the transcript to look for themes in our discussion and for consensus and disagreement among the participants. These data were used to confirm that the survey respondents understood the questions I asked on the online survey, and the responses are discussed throughout the rest of the chapter.

To understand the preservation expectations of Facebook subscribers I conducted a descriptive analysis and examined the frequency with which the 144 respondents agreed or disagreed with the three persistence and four trust questions related to preservation expectations. I then looked for patterns and differences in responses. I also used a frequency comparison to examine the timeframes respondents associated with each response to the "forever" and "indefinite" statements.

To examine the disclosure behavior of my survey respondents I started with a descriptive analysis of their reported patterns of disclosure. I looked at the frequency and

likelihood of disclosing information by comparing the means and response rates of the two different groups of disclosure behavior questions: future and current disclosure behavior. I also briefly compare the current disclosure behavior rates to the rates reported by Gross and Acquitsti (2005).

Preliminary analysis using comparison of means showed there was not enough variation to examine whether the future disclosure responses were associated with preservation expectations.

I was particularly interested in understanding the extent to which respondents' reaction to the statement "I expect my Facebook information to be available indefinitely into the future" is associated with respondents' current disclosure behavior. I began this part of the examination by estimating the bivariate relationship between current disclosure behaviors and participants response to the "indefinite" statement. I used this particular statement to represent preservation expectations because the statement includes both the idea of the persistence of the record over a long period of time and access to the record, two of the primary concerns of archivists. I used ordinary least squares (OLS) regression with the eight questions representing current disclosure as the dependent variable with a range of zero to eight. The primary independent variable was the respondents' reaction to the "indefinite" statement based on the responses of yes, no, and don't know. "Don't Know" was the referent category for this analysis.

The OLS analysis allows me to understand the association between the dependent and independent variable accounting for differences in the characteristics of the respondents. In the second OLS model I included the age, race, education level, whether the respondent was currently pursuing a degree, how often the respondent used Facebook,

and how long the respondent had been using Facebook, as control variables to account for potential differences in respondents' reaction to the "indefinite" statement.

RESULTS

The primary focus of this chapter is to have a first look at what these 144 Facebook subscribers expect to happen to their records and the extent to which these expectations are associated with the information they choose to share. Privacy scholars have expressed concern that these records will be available "forever" as a way to express the lack of access control over time and the threat to privacy implicit in this concern. There are important reasons for archivists to preserve both individual Facebook records and the corpus of Facebook records, but our ability to preserve these records over long periods of time is uncertain and may depend on the actions of individual subscribers.

The results reported below help to clarify whether individual subscribers are thinking about preservation when they use Facebook, whether they believe they have control over their Facebook records, how long they expect these records to exist, and whether these expectations are associated with the type of information they choose to share.

Research Question One: What are Facebook subscribers' preservation expectations?

Table 2: *Response Rates to the Trust and Persistence Questions*

Trust	**Agree**	**Disagree**	**Don't Know**
If I deactivate my account, Facebook will immediately delete all of my records.	1.39% n=2	84.72% n=122	13.89% n=20
If I delete information from my profile, or delete my account completely, information in backup systems will not be available to anyone.	8.33% n=12	64.58% n=93	27.08% n=39
Facebook deletes posts immediately after I remove them.	12.50% n=18	56.94% n=82	30.56% n=44
If I delete my account, Facebook will immediately delete all of my records.	2.08% n=3	80.56% n=116	17.36% n=25
Persistence	**Agree**	**Disagree**	**Don't Know**
Facebook uses adequate backup procedures to protect my posted information from being lost.	13.19% n=19	12.50% n=18	74.31% n=107
I expect my Facebook information to be available indefinitely into the future even if I never log in again.	47.22% n=68	28.47% n=41	24.31% n=35
My Facebook information will exist forever.	31.25% n=45	25.00% n=36	43.75% n=63

Table 2 shows the difference in the extent to which this group of Facebook subscribers expect the Facebook software to do what they tell it to do and how long they expect these records will exist.

The majority of respondents disagreed with the four trust statements, indicating that respondents to my online survey did not believe they had control over their Facebook records. This finding was supported by the data collected from participants in the focus group, who agreed they did not trust Facebook. The expectations of respondents are supported by recent events: Facebook has been accused of not deleting records when a subscriber requests those records to be deleted (Solon, 2012), and separately accused of

not deleting photographs up to three years after a subscriber had asked that the photographs be deleted (Cheng, 2012).

Table 2 also summarizes responses to the persistence questions from the online survey. Most respondents (74%, n = 107) do not know whether Facebook uses adequate backup procedures, the largest group of respondents expect their records to be available indefinitely and do not know whether their records will exist forever.

Respondents who answer the first three questions "don't know" are probably the most realistic; as I have discussed throughout this study, we don't know how long Facebook records will continue to exist. Some focus group respondents said they answered "don't know" not only because they did not have access to information that would allow them to answer accurately, but also because they did not think about preservation when using Facebook. In the context of the preservation of Facebook records, "don't know" may also mean the respondent does not care, or does not recognize the value of Facebook records and why those records should be preserved.

Table 3: *Response Rates to the Facebook Records Availability Question*

I expect my Facebook information to be available to me for the next:	
0 - 6 months	6.25% n=9
7 - 11 months	4.86% n=7
1 - 3 years	22.92% n=33
3 - 5 years	28.47% n=41
6 - 10 years	14.58% n=21
More than 10 years	22.92% n=33

Table 3 summarizes responses to the statement "I expect my Facebook records to be available to me for the next." Respondents generally expect their records to be available for one to five years (51%, n = 74), a large group (23%, n = 33) believe their records will be available for more than ten years, and very few (11%, n = 16) expect their records to exist for less than one year. We see that 66% (n= 95) of respondents think records will be around for more than three years.

Table 4: *Respondents' Estimate of the Duration of Facebook Records Compared to Response Rates of the Indefinitely and Forever Questions*

	Indefinitely			**Forever**		
	Agree	**Disagree**	**Don't Know**	**Agree**	**Disagree**	**Don't Know**
0 - 6 months	1.47% n = 1	17.07% n = 7	2.78% n = 1	2.22% n = 1	2.78% n = 1	10.94% n = 7
7 - 11 months	1.47% n = 1	9.76% n = 4	5.56% n = 2	2.22% n = 1	8.33% n = 3	4.69% n = 3
1 - 3 years	16.18% n = 11	31.71% n = 13	25.00% n = 9	22.22% n = 10	30.56% n = 11	18.75% n = 12
3 - 5 years	19.12% n = 13	34.15% n = 14	38.89% n = 14	11.11% n = 5	44.44% n = 16	31.25% n = 20
6 - 10 years	20.59% n = 14	4.88% n = 2	13.89% n = 5	17.78% n = 8	8.33% n = 3	15.63% n = 10
More than 10 years	41.18% n = 28	2.44% n = 1	13.89% n = 5	44.44% n = 20	5.56% n = 2	18.75% n = 12

Table 4 above demonstrates respondents who agreed with the statement "my Facebook information will be available indefinitely" also believed their records would continue to exist for three years or more, with 41% (n = 28) reporting they thought their records would exist for more than ten years and close to 62% (n = 42) expecting their records to persist for six years or more. Respondents who disagreed with the "indefinite" statement reported they expected Facebook records to exist for one to five years.

Respondents who reported they did not know whether Facebook records would continue to exist indefinitely were also clustered in the one to five year range. We can see from these data that respondents who agreed with the "indefinite" statement expected records to exist for a longer period of time, but even respondents who disagreed with this statement expected records to last for 3 to 5 years (39%, n = 14).

Respondents who agreed with the statement "My Facebook information will exist forever" had a large group (44%, n = 20) report they believed their records would exist for ten years or more. However, 51% (n = 23) indicated they believed their Facebook records would exist for more than one year but less than ten. Respondents who disagreed with the "forever" statement indicated they expected their Facebook records to exist for one to five years. Respondents who did not know whether the "forever" statement was true generally expected records to persist for more than one year, with the largest group (31%, n = 20) selecting the three to five year range.

These responses indicate that many participants are unsure about Facebook's backup procedures, they expect their Facebook records to persist indefinitely, but they are unsure whether their Facebook records will exist "forever."

Research Question Two: To what extent are Facebook subscribers' preservation expectations associated with subscribers' disclosure behavior?

The second research question asks to what extent preservation expectations are associated with disclosure behavior. Table 5 summarizes participants' responses to the future disclosure set of questions.

Table 5: *Response Rates to the Future Disclosure Questions*

Future Disclosure	Very Unlikely	Unlikely	Neither	Likely	Very Likely
Naked	95.14% n = 137	4.86% n = 7	0.00% n = 0	0.00% n = 0	0.00% n = 0
Illegal	88.89% n = 128	10.42% n = 15	0.00% n = 0	0.00% n = 0	0.69% n = 1
Profile Pictures	8.33% n = 12	7.64% n = 11	4.86% n = 7	29.17% n = 42	50.00% n = 72
Pictures with Friends	9.03% n = 13	10.42% n = 15	6.94% n = 10	36.11% n = 52	37.50% n = 54

Table 5 illustrates that most respondents say they are very unlikely to share pictures of themselves naked or partially naked, or doing something illegal. This finding is not surprising and is consistent with previous research (Christofides et al., 2009). Most respondents are likely or very likely to share profile pictures or pictures of themselves with friends. There is very little variation to these answers and no correlation with preservation expectations based on preliminary analysis.

Table 6: *Response Rates of Current Disclosure Questions and Comparison of Responses to Gross and Acquisti Reported Responses*

Current Disclosure	Yes	Gross & Acquisti
School	91%	72%
Birthday	74 %	96%
Relationship Status	72%	81%
Email	69%	85%
Hometown	69%	85%
Program of Study	49%	72%
Phone	13%	24%
Address	2%	4%

Table 6 shows that most respondents share their school, birthday, relationship status, email, and hometown. Very few respondents share their phone number and address. I include the disclosure rates reported by Gross and Acquisti as a basis of comparison. More students in my sample disclose their school than those included in Gross and Acquisti's sample, a difference that may be attributable to the fact that in 2005 Facebook limited access to higher education institution email addresses, but that a lower percentage of respondents disclose information in each category in my sample.

Preliminary analysis comparing current and future disclosure to trust indicated that a statistically significant relationship did not exist. This was somewhat surprising given that Dwyer (2007) found a correlation between the level of trust in social networking services and subscribers' level of disclosure. This study differs from Dwyer in that I do not compare disclosure across two different social networking services: MySpace and Facebook. This study also looks only at subscribers' trust in the service provider; Dwyer included subscribers' trust in fellow subscribers to that same service. In other words, Dwyer asked whether subscribers trusted Facebook and the community of Facebook subscribers.

These two differences may explain why the finding of low trust in the present study is not correlated with a change in disclosure behavior as we would expect. This finding is important; subscribers in my sample do not trust Facebook, yet this lack of trust does not seem to correspond with a change in disclosure behavior.

To understand the extent to which preservation expectations are associated with disclosure behavior I began by comparing the means of the responses to the current disclosure scale with the means of the responses to the "indefinite" statement.

Table 7: *Current Disclosure and "Indefinite" Statement Comparison of Means*

	I Expect My Facebook Information to be Available Indefinitely		
Current Disclosure	Yes	No	Don't Know
Number of Responses	68	41	35
Mean	4.16 σ = 1.65	4.46 σ = 1.72	4.77 σ = 1.24

Table 7 illustrates that the mean current disclosure score of respondents who agreed with the "indefinite" statement was lower than the other two responses. The highest mean disclosure score belonged to participants who did not know whether to expect their Facebook records to be available indefinitely. Respondents who answered "don't know" share 0.61 more items on average with Facebook than respondents who believed their records would be available indefinitely.

To examine the relationship between current disclosure and the "indefinite" statement more closely, I used a bivariate analysis including the current disclosure scale previously described as the dependent variable and responses to the statement "I expect my Facebook information to be available indefinitely into the future even if I never log in again" as the independent variable. The bivariate analysis suggests a significant correlation at the 0.10 level (1, 144) $p < 0.07$, respondents who agreed with the "indefinite" statement were more likely to disclose information.

The second model accounts for some of the demographic characteristics of respondents as well as how often respondents use Facebook and how long respondents have been using Facebook. The association is even stronger between not knowing and disclosure (1, 144) $p = 0.02$. Those who agree with the "indefinite" statement disclose 0.74 fewer items than those who respond "don't know." There is no statistically significant difference between those who respond "no" and those who respond "don't know." We also see that white respondents (n = 121, 84%) and subscribers who report

using Facebook 1 hour or more per day (n = 40, 28%) are likely to disclose more information. Respondents over age twenty-nine (n = 58, 40%) are likely to disclose less information than respondents under age twenty-nine. Table 8 illustrates this second model.

Table 8: *Results of OLS regression Estimating the Effect of Preservation Expectations on Current Disclosure*

	Current Disclosure	
	Model 1	Model 2
Indefinite Statement		
(Don't Know)		
Agree	**-0.610†**	**-0.786***
Disagree	-0.308	-0.320
Over 29		**-0.651***
White		**0.724†**
Master's Degree or Greater		0.287
Currently Pursuing a Degree		-0.475
Uses Facebook 1 hour or more per day		**1.270***
Has Used Facebook for More Than Two Years		-0.003

†p ≤ .10. *p ≤ .05. **p ≤ .01. Reference groups are in parenthesis.

These results indicate that Facebook subscribers who expect their records to be available indefinitely are less likely to disclose information than respondents who do not expect, or don't know whether to expect their records to be available indefinitely. This finding is important because preservation expectations are an additional reason privacy scholars need to consider when discussing disclosure behavior, not just the respondents' desire for popularity, lack of technical knowledge of the software, or relative level of trust in the software. These results are important to archivists because they are a first step

toward understanding the likelihood that individual subscribers will act to preserve Facebook records, as well as the mediating effect of Facebook on interpersonal correspondence.

The responses from the online survey and the focus group indicate that the participants in my study demonstrated indifference, mistrust, and confusion about the preservation of their Facebook records. Respondents did not recognize, in the context of Facebook use, a general concept of "preservation expectations" that includes both the persistence of records and access to records over time.

In chapter 4 I demonstrated that subscriber expectations result in a clear threat to the cultural record in only one case, when a subscriber expects Facebook to preserve records and Facebook does not. In this case, the subscriber is unlikely to take action to preserve records on their own and it is unclear whether Facebook will take action to preserve records. If a subscriber does not expect Facebook to preserve records, archivists may be able to influence that subscriber to preserve her own records. We see from the survey results that most respondents do not expect Facebook to preserve records, but this is not necessarily good news for the preservation of Facebook records.

A large group of respondents expect their records to be available indefinitely (47%, n = 68), and, of that group, 62% believe their records will continue to exist for more than six years (n = 42). Even with this large group of respondents who expect their records to exist indefinitely, there still exists uncertainty and doubt about the persistence of Facebook records. 66% (n = 95) of respondents expect their records to exist for more than one year but less than ten, 69% (n = 99) disagree with, or are unsure whether their Facebook records will exist "forever," and 53% (n = 76) disagree with, or are unsure whether their records will be available "indefinitely."

The majority of survey respondents expect their records to be available indefinitely, and at the same time a majority is not sure that their records will be available forever. This mix of reactions to the preservation of records is due to respondents' understanding of the word "forever," their reasons for using social networking, and their perception of Facebook's motivation and capacity for preservation.

Focus group participants explained that they were not sure that Facebook would exist as a company for the next ten years, given what they perceive as the rapid pace of change in information and communication technology. They also emphasized that the word "forever" had a different meaning to them than "indefinitely." Indefinitely implied, on the one hand, that there will come a time when these records are not available, but that time is not easily foreseeable. Forever, on the other hand, implied a permanent state of affairs with no foreseeable end. Keeping records indefinitely would require a concerted effort, expertise, and a trustworthy entity charged with maintaining those records.

The reason respondents used social media was an important aspect of their indifference toward the persistence of records. According to focus group participants, Facebook subscribers use the service to communicate with friends and family, to keep in touch with other subscribers, to share their daily activities, and stay up to date with events in the lives of their Facebook friends. Participants indicated that, while there are valuable records being created using Facebook, their own records were not valuable for historic or research purposes.

One participant asserted that the ability to download the records she had created using Facebook was adequate to preserve her records. The group had not thought about whether and how records of historic and research value might be preserved if a creator passes away or loses access to her Facebook account, as might happen when a subscriber

is a citizen of a repressive regime and participates in a revolution. Nor had participants thought about the inconsistent preservation practices Facebook subscribers would follow if they preserve their own records, or that the resulting archives would not be generally accessible, or that the corpus of records as it exists in Facebook would be destroyed by separate, disconnected, individual preservation efforts.

Focus group participants agreed with the assertion that Facebook subscribers created records of historic and cultural value, and that these records should be preserved. Focus group participants also agreed with one respondent when she said "I don't think about preservation when I use Facebook."

Similarly, focus group participants felt that Facebook would preserve records for business purposes, not for the protection of the historic record or for academic research. Participants pointed out that these records are valuable to Facebook for advertising and marketing purposes, and respondents expected the records to be protected only as long as the records provided revenue opportunities.

These responses indicate that the Facebook subscribers in my sample lack adequate information to assess whether Facebook should be considered a trusted digital repository and that respondents' preservation expectations can be described as doubtful, or at best confused.

Chapter 7: Conclusion

This research makes four contributions to existing literature: a discussion of the value of social networking records and whether and how they should be preserved, a close examination of the differing opinions of archivists and privacy scholars about these records, a discussion of the role public policy might play in the preservation of Facebook records and privacy in the United States, and an empirical exploration of the attitudes and behaviors of a small group of Facebook subscribers related to preservation and privacy.

I find that Facebook records are valuable because the volume of Facebook records represents new research challenges that are changing traditional academic disciplines, that Facebook records might include important cultural heritage records, and that Facebook records may be of value to individuals. For these and other reasons, Facebook records are worth preserving.

I discussed four options for preserving Facebook records. Facebook can preserve the records, individual subscribers can preserve their own records, Facebook can partner with a third party for preservation, or a third party can act independently to preserve Facebook records. Only Facebook can act to preserve the corpus of Facebook records, and only Facebook and individual subscribers have the most direct access to, and control of Facebook records, so I examined these two preservation options more closely.

Given that Facebook records are valuable and worth preserving, I explored the differing opinions of privacy scholars and archivists about such preservation. Privacy scholars are concerned that, once a Facebook record is created, that record can be copied, stored, preserved, and published by other subscribers or actors unknown to the record

creator. Because of the digital nature of Facebook records, records creators can be completely unaware of the duplication and re-use of these records. “Friends” of the record creator can easily copy sensitive information shared in confidence with one or two others and inexpensively publish this information to billions of other subscribers or to anyone with Internet access. Unknown actors, such as law enforcement agents, journalists, and stalkers, can copy information shared on Facebook and use that information to create a digital dossier or share that information at a later date. Once a copy is made of a Facebook record, the original record creator loses control over the persistence, destruction, use, and publication of the copy of that record. This loss of access control creates a persistent privacy threat.

Archivists are also concerned with the lack of control over Facebook records, but their concern is that the lack of control threatens the persistence of valuable records. The rapid change in hardware and software versions, the potentially short lifespan of social networking companies, and the use of proprietary software formats all threaten archivists’ ability to ensure the persistence of, and access to, valuable cultural artifacts.

Despite the privacy risks inherent in the networked architecture of Facebook, the archives profession can offer an example of how privacy may be protected while records are preserved. Archives have a long history of protecting privacy and preserving records. The ability of an archives to solicit donations of both records and financial support depends on the reputation that institution has developed over time by providing responsible stewardship for the records, service to the community served by the archives, and protecting the interests of records donors. I discussed the recently passed ISO standards for a trusted digital repository, and how these standards offer a roadmap to

improve the transparency and consistency of Facebook and consequently improve the potential to preserve records.

The concept of a trusted digital repository and the deed of gift do not address all of the concerns of privacy scholars. Facebook and other social networking services require the use of the Internet to communicate and allow subscribers to read, comment on, and copy the status updates, photographs, and videos of other subscribers. As long as this requirement is in place, privacy will continue to be threatened by (1) the lack of control subscribers have over the persistence of records and (2) the ability of individuals, corporate, and government actors to access and copy records without subscribers' awareness or permission. Privacy scholars' concerns with the lack of access control over Facebook records will continue to exist regardless of how well archivists protect the privacy of records donors.

If archivists decide that not only are individual and small groups of Facebook records valuable for cultural heritage and personal reasons, but the entire corpus of Facebook records is valuable for research purposes, they face more than just the challenge of protecting individual donors' privacy. Archivists face a serious challenge gaining access to the corpus of records if they choose to preserve all Facebook records. Without government intervention, it is impossible to know whether Facebook will choose to preserve the corpus of records. Facebook would need to either create a self-sustaining digital archive of its own, or to partner with a third party such as the Library of Congress to ensure the long-term preservation of these records.

To examine whether the U.S. federal government had cause and was likely to impose regulation to encourage Facebook to preserve valuable cultural heritage records I used the market imperfections theory of regulation and Kingdon's garbage can model of

public policy. I find that the U.S. federal government has cause to regulate but is unlikely to do so, which means the preservation of the corpus of Facebook records depends entirely on the decisions made by the Facebook corporation. Archivists may be able to influence these decisions, but Facebook will not be compelled by the U.S. federal government to take action to preserve these records.

Given that Facebook is unlikely to be required to preserve Facebook records by federal law, I investigate individual record creators' attitudes about the preservation of Facebook records. I conducted both an online survey and a focus group to examine Facebook subscribers' expectations related to preservation.

I find that Facebook subscribers in my sample differentiate between the ideas "indefinitely" and "forever." Most respondents believe their records will continue to exist indefinitely, but they are not sure about "forever." I also find that Facebook subscribers in my focus group claim that users do not use Facebook with the idea that their records should be preserved, nor do they "trust" that Facebook will act on subscribers' behalf when Facebook decides to preserve or destroy records. Finally, I find that disclosure behavior is associated with preservation expectations; subscribers disclose less information if they believe their records will continue to exist indefinitely.

As we have learned from this study, it is unlikely that individual Facebook subscribers will take action to preserve their own records. Individual subscribers have little information about whether Facebook acts according to a subscriber's desires; when a subscriber deletes information from her profile, she has no way of knowing whether that information has been deleted. This asymmetry of information makes archivists, privacy scholars, and subscribers uncertain about whether the actions they take to protect subscribers' privacy or preserve subscribers' records are effective.

More research is necessary to understand why subscribers do not think about preservation when they use Facebook and whether the attitudes expressed by my focus group and online survey respondents reflect those of a larger portion of Facebook users. This first step in understanding subscribers' attitudes and expectations demonstrates that archivists should not count on individual subscribers to preserve the corpus of Facebook records, despite Cox's claims about the "citizen archivist" and others' enthusiasm about personal information management as the key to preserving valuable individual records.

WEAKNESSES AND LIMITATIONS OF THIS STUDY

Some of the weaknesses and limitations of my study include the use of a homogenous, non-representative sample, and the possibility that the questions that I created regarding preservation expectations include inaccurate assumptions about digital preservation.

The respondents to the online survey were mostly a homogenous group. Most were "white" (84%) and highly educated, reporting a postgraduate degree (84%) and currently pursuing a degree (68%). These results are consistent with the graduate student population at the university.

The respondents to my online survey are not representative of any larger group, which means we cannot generalize the results of my survey. While lack of generalizability is a limitation of my study, the sample of respondents to the online survey and the focus group came from a population of individuals likely to be more concerned about and aware of both preservation and privacy than the general public. The Information School is home to a highly ranked preservation program, and all of the

students at the school are engaged in course work that includes contemporary Web-based technology.

This increased likelihood of awareness and lack of representativeness allow me to speculate that the general population would likely be less informed about preservation than the respondents to my survey. It would be surprising if the average Facebook subscriber were more likely to be thinking about preservation as she uses Facebook than were students in information studies.

I conducted a focus group, with six participants, in part to address the validity limitations of the survey instrument and to better understand why survey respondents responded as they did. The qualitative data gained from this group increased confidence that the survey participants understood the questions on the survey and that the language I had chosen for questions about preservation was consistent with the language used by the survey respondents. Both outcomes indicated that it was unlikely that the language I used in the questions introduced bias or led respondents to answer in too direct a way. The focus group respondents also helped me understand the preservation and privacy concerns of the survey respondents in more detail by both affirming the choice of the questions I included in the survey and offering an explanation of their understanding of these questions.

The possibility of bias and assumptions in the creation of survey questions is possible in any study. I was fortunate to have a dissertation committee to review my questions and to have over thirty faculty members and students attend my dissertation proposal presentation to consider the proposed study and questions I planned to address. Both of these discussion opportunities helped me to reduce the likelihood of hidden bias.

FUTURE RESEARCH

In the current study I contribute to the archives, privacy, and social media literature by discussing the value of Facebook records as cultural heritage objects, by more clearly defining the differing opinions of privacy scholars and archivists about Facebook records, and by examining who might best preserve these records. This study is a first step in a relatively unexplored area of study; many questions remain.

Archival research would benefit from a broader and more direct investigation or series of investigations of subscribers' attitudes related to the preservation of social networking records. These investigations would include a larger sample that may be more representative of some subset of Facebook subscribers, such as U.S. subscribers of a particular age group. These investigations would address whether Facebook subscribers think Facebook records in general are worth preserving, whether subscribers trust the Facebook software to act according to subscriber commands, whether subscribers would consider preservation while using Facebook if they did trust Facebook, and whether subscribers would allow a trusted digital repository to preserve all of their Facebook records.

Despite the use of Facebook to organize resistance to oppressive regimes, conduct illicit affairs, solve crime, recruit terrorists, and share baby's first steps, Facebook subscribers do not think about whether the records they create using the service will exist in the distant future. A better understanding of why Facebook subscribers do not think about preservation when using Facebook would be an important contribution to the archives literature. Do respondents work to preserve other personal digital records, and do they think their own Facebook records are valuable and worth preserving? Addressing these questions would require trust to be defined more clearly and to determine whether

subscribers do not use Facebook because they do not trust Facebook, because they do not see these records as worth preserving, or for some other reason.

The archives literature would also benefit from a study of whether archivists believe all social networking records should be preserved, or only a selection of these records should be preserved. In this study I have offered four general approaches to preserving Facebook records, but a survey of archivists' opinions on the range of approaches we should take to preserve social networking records and protect privacy would also be a valuable contribution to the discussion of social networking and preservation. In addition, a more detailed survey of contemporary approaches to preserving social networking and personal digital records would be valuable to scholars interested in the privacy implications of the long-term persistence of personal digital records.

Both the privacy and archives literature would benefit from a broader investigation of archivists' and privacy scholars' attitudes about preservation. We should ask privacy scholars what specifically concerns them about the long-term persistence of social networking records. I have discussed the implications of the Internet-based architecture of Facebook and the lack of control this architecture entails, but a clearer description of the concerns of privacy scholars may allow privacy scholars and archivists to work together to improve the protection of privacy and the preservation of Facebook records.

A deeper investigation of the concepts of persistence and control as they relate to preservation would also contribute to improving both the protection of privacy and the preservation of valuable records. This study uncovered different attitudes about trust in

the service provider and control of the record, and how these attitudes affect subscribers' perception of the persistence of the record and the protection of privacy.

The goal of protecting privacy and preserving records would also be served by exploring whether privacy scholars believe social networking records are valuable and should be preserved and whether privacy scholars believe archives are capable of preserving social networking records and protecting privacy.

A closer examination of the role of open standards, especially in the context of the history of email systems and how the change from closed, proprietary systems to open, protocol-based systems allowed for growth, standardization, and communication across service providers would be a valuable addition to the history of technology, privacy, and social networking.

This study began in January of 2011 with the dissertation proposal defense. Between that time and May of 2013, Facebook has held its initial public offering; has been ordered by the FTC to submit to 20 years of privacy audits; has introduced, and compelled subscribers to use, the "timeline" which promises to display all of one's important life events in Facebook in chronological order; and has developed Facebook "Home," a group of mobile applications that change the way Android users interact with their phones by making Facebook the default, primary communication software on the Android phone.

Clearly Facebook is both a very complex and rapidly changing company and software application. While Facebook continues to change, archivists are focused on how best to preserve html Web pages, email, and blogs, and are only beginning to discuss the value of social networking records, the definition of a social networking "record" for

preservation purposes, and the complexities of preserving Facebook as more than just a collection of database entries, but as a changing software environment used by many millions to communicate with each other.

Privacy scholars have recognized and are discussing the potential threat to subscribers' privacy represented by the persistence and loss of control of the records created by over one billion Facebook subscribers on a daily basis. The research described here hopefully offers a step forward in the archival discussion by clarifying the concerns, assumptions, and vocabulary of both archivists and privacy scholars. The right to delete is a clear call to destroy records in the interest of protecting privacy. Archivists need to offer just as clear a call to preserve valuable cultural heritage, personal memory, and research records while still protecting the privacy of records creators.

Appendix A: Facebook Profiles from 2005 – 2011 (Buck, 2011)

Illustration 6: *2005 – Users browsed profiles to see what people were doing and who they were connected to.*

Illustration 7: *2006 – News Feeds were launched, raising privacy concerns by making subscriber activity easier to see.*

Illustration 8: *2007 – No major changes from 2006*

Illustration 9: *2008 – Facebook allows subscribers to add application tabs to the top of their profile and a "publisher tool bar" making it easier to publish status updates, photos, and videos.*

Illustration 10: *2009 – Pages were introduced, allowing individual subscribers to create business or organization pages and to post using that business or organization name.*

Illustration 11: *2010 – subscribers are able to alter the photo banner at the top of the page and view the friends you have in common with the profile you are viewing.*

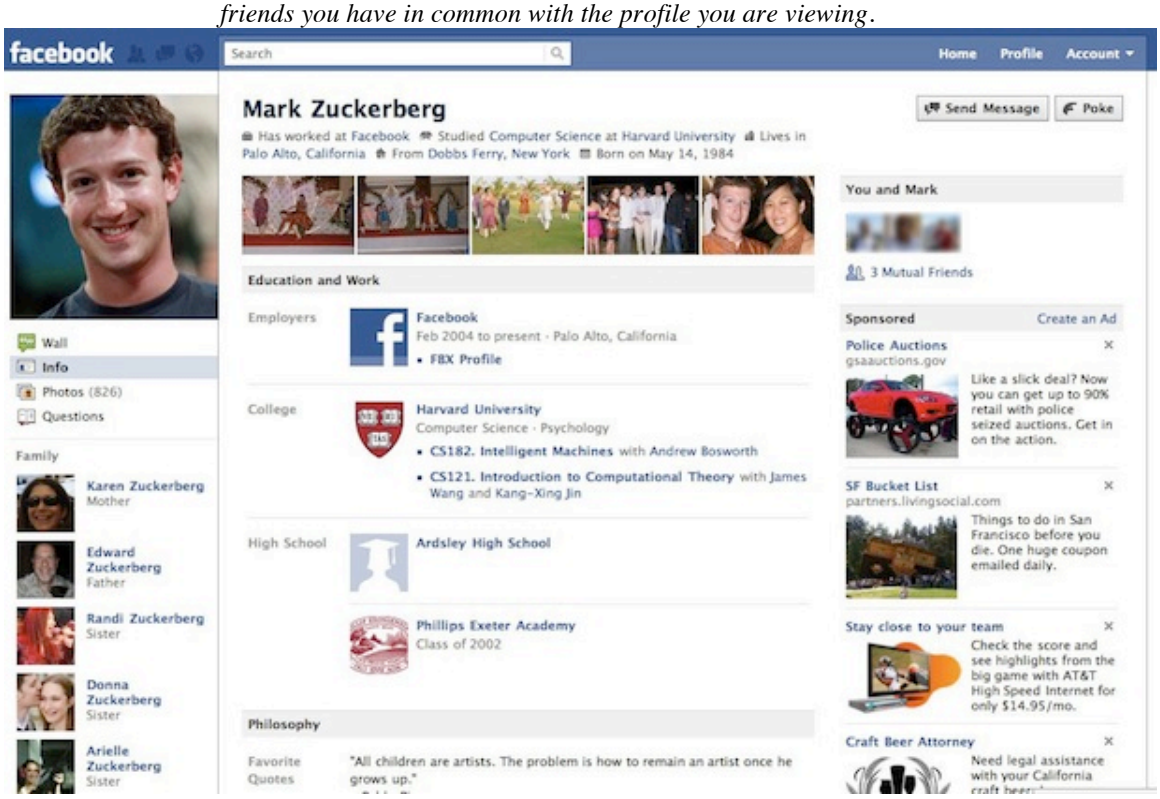
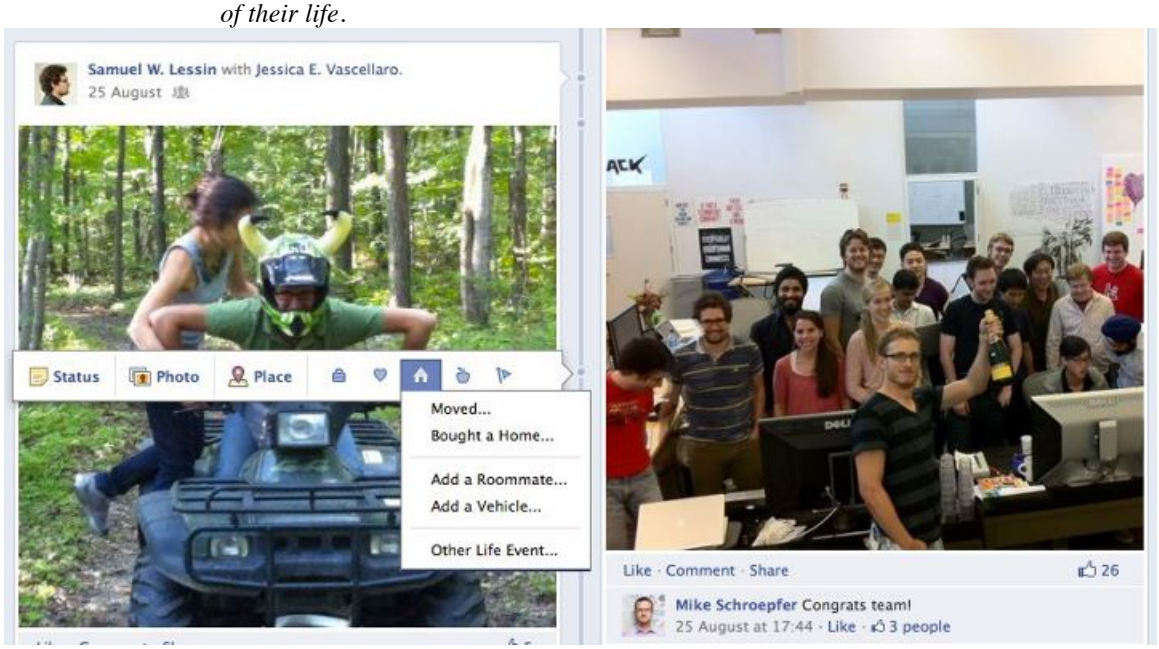


Illustration 12: *2011 – The "ticker" – a summary of the newsfeed option - follows the subscriber across pages and subscribers are able to "view as" – to see their profile as other subscribers might see it.*

Illustration 13: *2011 – A Timeline feature was added allowing subscribers to create a visual representation of their life.*



Appendix B: Facebook Information Available for Download

From: https://www.facebook.com/help/326826564067688 on10/18/2012

What info is available?	What is it?	Where can I find it?
About Me	Information you added to the **About** section of your timeline like relationships, work, education, where you live and more. It includes any updates or changes you made in the past and what's currently in the **About**section of your timeline.	Activity Log
Account Status History	The dates when your account was reactivated, deactivated, disabled or deleted.	Expanded Archive
Active Sessions	All stored active sessions, including date, time, device, IP address, machine cookie and browser information.	Expanded Archive
Ads Clicked	Dates, times and titles of ads clicked.	Expanded Archive
Address	Your current address or any past addresses you had on your account.	Expanded Archive
Ad Topics	A list of topics that you may be targeted against based on your stated likes, interests and other data you put in your timeline.	Expanded Archive
Alternate Name	Any alternate names you have on your account (ex: a maiden name or a nickname).	Expanded Archive

Apps	All of the apps you subscribe to.	Expanded Archive
Birthday Visibility	How your birthday appears on your timeline.	Expanded Archive
Chat	A history of the conversations you've had on Facebook Chat.	Downloaded Info
Check-ins	All of the places you've checked into.	Downloaded Info Activity Log
Connections	The people who have liked your Page or Place, RSVPed to your event, installed your app or checked in to your advertised place within 24 hours of viewing or clicking on an ad or Sponsored Story.	Activity Log
Credit Cards	If you make purchases on Facebook (ex: in apps) and have given Facebook your credit card number.	Account Settings
Currency	Your preferred currency on Facebook. If you use Facebook Payments, this will be used to display prices and charge your credit cards.	Expanded Archive
Current City	The city you added to the **About** section of your timeline.	Downloaded Info
Date of Birth	The date you added to Birthday in the **About**section of your timeline.	Downloaded Info

Deleted Friends	The people you've unfriended.	Expanded Archive
Education	Any information you added to Education in the About section of your timeline.	Downloaded Info
Emails	Email addresses added to your account (even those you may have removed).	Expanded Archive
Events	Events you've joined or been invited to.	Activity Log
Family	Friends you've indicated are family members.	Expanded Archive
Favorite Quotes	Information you've added to the Favorite Quotes section of the **About** section of your timeline.	Downloaded Info
Friend Requests	Pending sent and received friend requests.	Expanded Archive
Friends	A list of your friends.	Downloaded Info
Gender	The gender you added to the **About** section of your timeline.	Downloaded Info
Groups	A list of groups you belong to on Facebook.	Downloaded Info
Hidden from News Feed	Any friends, apps or pages you've hidden from your News Feed.	Expanded Archive

Hometown	The place you added to hometown in the **About** section of your timeline (profile).	Downloaded Info
IP Addresses	A list of addresses where you've logged into your Facebook account (won't include all IP addresses as they are deleted according to a retention schedule).	Expanded Archive
Last Location	The last location associated with an update.	Activity Log
Likes on Other's Posts	Posts, photos or other content you've liked.	Activity Log
Likes on Your Posts from others	Likes on your own posts, photos or other content.	Activity Log
Likes on Other Sites	Likes you've made on other sites off of Facebook.	Activity Log
Linked Accounts	A list of the accounts you've linked to your Facebook account	Account Settings
Locale	The language you see on Facebook is based on where you're located.	Expanded Archive
Logins	IP address, date and time associated with logins to your Facebook account.	Expanded Archive
Logouts	IP address, date and time associated with logouts from your Facebook account.	Expanded Archive

Messages	Archive of messages you've sent and received on Facebook. Note, if you've deleted a message it won't be included in your download.	Downloaded Info
Name	The name on your Facebook account.	Downloaded Info
Name Changes	Any changes you've made to the original name you used when you signed up for Facebook.	Expanded Archive
Networks	Networks (affiliations with schools or workplaces) that you belong to on Facebook.	Expanded Archive
Notes	Any notes you've written and published to your account.	Activity Log
Notification Settings	A list of all your notifications and whether you have email and text enabled or disabled for each.	Expanded Archive
Pages You Admin	A list of pages you admin.	Expanded Archive
Pending Friend Requests	Pending sent and received friend requests.	Expanded Archive
Phone Numbers	Mobile phone numbers you've added to your account, including verified mobile numbers you've added for security purposes.	Expanded Archive
Photos	Any photos you've uploaded to your account.	Downloaded Info
Photos	Any metadata that is transmitted with your	Pending, will be added to

Metadata	uploaded photos.	Expanded Archive
Physical Tokens	Badges you've added to your account.	Expanded Archive
Pokes	A list of who's poked you and who you've poked.	Expanded Archive
Political Views	Any information you added to Political Views in the About section of timeline.	Downloaded Info
Your Posts	Anything you posted to your own timeline, like photos, videos and status updates.	Activity Log
Posts by Others	Anything you posted to someone else's timeline (profile), like photos, videos and status updates.	Activity Log
Privacy Settings	Only current settings are stored.	Privacy Settings
Recent Activities	Actions you've taken and interactions you've recently had.	Activity Log
Registration Date	The date you joined Facebook.	Activity Log
Religious Views	The information you added to Religious Views in the **About** section of your timeline.	Downloaded Info
Removed Friends	People you've removed as friends.	Activity Log

Screen Names	The screen names you've added to your account, and the service they're associated with. You can also see if they're hidden or visible on your account.	Expanded Archive
Searches	Searches you've made on Facebook.	Activity Log
Shares	Content (ex: a news article) you've shared with others on Facebook using the **Share** button or link.	Activity Log
Spoken Languages	The languages you added to Spoken Languages in the **About** section of your timeline.	Expanded Archive
Status Updates	Any status updates you've posted.	Activity Log
Subscribers	A list of people who are subscribed to you.	Expanded Archive
Subscriptions	A list of people you subscribe to.	Activity Log
Tag Suggestions Template	A unique number based on a comparison of the photos you're tagged in. We use this template to help your friends tag you in the photos they upload.	Expanded Archive
Work	Any information you've added to Work in the About section of your timeline.	Downloaded Info
Vanity URL	Your Facebook URL (ex: username or vanity for your account).	Visible in your timeline URL

Videos	Videos you've posted.	Activity Log

Appendix C: Online Survey

1. Which questions indicate preservation expectations?
    a. Persistence
        i. Facebook uses adequate backup procedures to protect my posted information from being lost.
            1. 1 = agree 2 = disagree 3 = I don't know
        ii. I expect my Facebook information to be available indefinitely into the future even if I never log in again.
            1. 1 = agree 2 = disagree 3 = I don't know
        iii. My Facebook information will exist forever.
            1. 1 = agree 2 = disagree 3 = I don't know
        iv. I expect my Facebook information to be available to me for the next:
            1. 1 = 0 - 6 months
            2. 2 = 7 - 11 months
            3. 3 = 1 - 3 years
            4. 4 = 3 - 5 years
            5. 5 = 6 - 10 years
            6. 6 = more than 10 years
    b. Control
        i. If I deactivate my account, Facebook will immediately delete all of my records.
            1. 1 = agree 2 = disagree 3 = I don't know
        ii. If I delete information from my profile, or delete my account completely, information in backup systems will not be available to anyone.
            1. 1 = agree 2 = disagree 3 = I don't know
        iii. Facebook deletes posts immediately after I remove them.
            1. 1 = agree 2 = disagree 3 = I don't know
        iv. If I delete my account, Facebook will immediately delete all of my records.
            1. 1 = agree 2 = disagree 3 = I don't know
2. Facebook use
    a. How often do you use Facebook?
        i. 1= rarely 2 = a few times per week, 3 = less than 1 hour per day 4 = 1 -2 hours per day 5 = 2 hours per day or more
    b. How long have you been using Facebook?
        i. 1 = 0-6, 2 = 7 - 11, 3 = 1 - 2, 4 = more than 2
3. Privacy concern
    a. Have you ever experienced an event that you would consider a breach of privacy?
        i. 1-= yes 2 = no

b. Do you only register for websites that have a privacy policy?
    i. 1-= yes 2 = no
c. Do you read a website's privacy policy before you register your information?
    i. 1-= yes 2 = no
d. Do you look for a privacy certification on a website before you register your information?
    i. 1-= yes 2 = no
e. Do you read license agreements fully before you agree to them?
    i. 1-= yes 2 = no

4. Demographics
    a. Please indicate the education level you have achieved:
        i. 1 = (none, grades 1 - 8)
        ii. 2 = high school incomplete
        iii. 3 = high school grad
        iv. 4 = technical or trade school
        v. 5 = some college
        vi. 6 = college grad
        vii. 7 = post college
    b. Are you currently pursuing a degree?
        i. Yes/no
    c. Age
        i. open
    d. Race/ethnicity
        i. 1 = White
        ii. 2 = Black or African-American
        iii. 3 = Asian or Pacific Islander
        iv. 4 = Mixed race
        v. 5 = Native American/American Indian
        vi. 6 = Other (please specify below)
        vii. 7 = Choose not to say
    e. Marital status
        i. 1 = Single
        ii. 2 = Married
        iii. 3 = Living with a partner
        iv. 4 = Divorced
        v. 5 = Separated
        vi. 6 = Widowed
        vii. 7 = Choose not to say
5. Disclosure behavior
    a. How likely are you to post profile pictures of yourself?
        i. 5 = very likely 1 = very unlikely
    b. How likely are you to post pictures of yourself with friends?
        i. 5 = very likely 1 = very unlikely

c. How likely are you to post pictures of yourself with friends doing something illegal?
    i. 5 = very likely 1 = very unlikely
d. How likely are you to post pictures of yourself naked or partially naked?
    i. 5 = very likely 1 = very unlikely
e. Choose all of the following items you have posted on Facebook
    i. Birthdate
    ii. Email address
    iii. Hometown
    iv. Relationship status
    v. Schools you have attended
    vi. Programs you have attended
    vii. Phone number
    viii. Home address

Appendix D: Focus Group

A focus group composed of graduate students from an Introduction to Research in Information Studies class was asked the following questions:

1. If you have taken the survey, how long did it take to complete?
2. Consider the preservation expectation questions numbered one through seven in your handout. Do you think these questions allow participants to describe whether they expect Facebook to preserve their records? Do you think these questions describe the idea of "preservation expectations"?
3. There are seven questions related to preservation expectations. Do you feel seven are too many?
4. Did I fail to include anything that you believe is important about preservation expectations of Facebook users?
5. Do you feel the disclosure behavior questions, numbers 9 – 13, will allow me to adequately describe Facebook disclosure behavior?
6. Did I fail to include anything that you believe is important with regard to disclosure behavior? If so, what did I fail to include?
7. Are questions like "Have you ever posted pictures of yourself naked or partially naked" likely to elicit an honest response?
8. Would you be more likely to reveal your disclosure behavior if the questions were asking about the likely disclosure behavior of others? For example, "Do you know someone who has posted naked or partially naked pictures of themselves?"

Any other comments, reactions or suggestions about the survey?

References

Abbate, J. (2000, July 31). *Inventing the Internet*. Cambridge, MA: MIT Press.

Ackerman, M. S., Cranor, L., & Reagle, J. (1999). Privacy in e-commerce: Examining user scenarios and privacy preferences. *In Proceedings of the 1st ACM conference on electronic commerce EC '99* (pp. 1-8). New York: ACM Press. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.38.595

Alexa Internet Inc. (n.d.). Top sites. *Alexa.com*. Retrieved on 2/7/2012 from: http://www.alexa.com/topsites

Allen, E. (January 4, 2013). Update on the Twitter archive at the Library of Congress. Retrieved from: http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/

Archive-It. (n.d.) *About us*. Retrieved 10/29/2012 from: http://archive-it.org/learn-more

Barker, A. N. (2005). Executive Order no. 13,233: A threat to government accountability. *Government Information Quarterly, 22*(1), 4-19.

Baron, D. P. (2006). *Business and Its Environment*. New Jersey: Pearson Education, Inc.

Beagrie, N. (2005). Plenty of room at the bottom?: Personal digital libraries and collections. *D-Lib Magazine, 11*(6). Retrieved 1/7/2013 from: http://www.dlib.org/dlib/june05/beagrie/06beagrie.html

Berk, R. & Bleich, J. (February, 2013). Forecasts of violence to inform sentencing decisions. *Journal of Quantitative Criminology*. Published online, retrieved 3/8/2013 from: http://www-stat.wharton.upenn.edu/~berkr/SentCART%20copy.pdf

Bilton, N. (2010, May 12). Price of Facebook privacy? Start clicking. *The New York Times*. Retrieved 3/8/2013 from http://www.nytimes.com/2010/05/13/technology/personaltech/13basics.html?_r=1&scp=3&sq=facebook%20privacy&st=cse

Bilton, N. (2013, February 24). Disruptions: Data without context tells a misleading story. *The New York Times*. Retrieved 3/8/2013 from: http://bits.blogs.nytimes.com/2013/02/24/disruptions-google-flu-trends-shows-problems-of-big-data-without-context/

boyd, d. (2008). Facebook's privacy trainwreck: Exposure, invasion, and social convergence. *Convergence, 14*(1), pp. 13 – 20.

boyd, d. & Marwick, A. (2011). Social privacy in networked publics: Teens' attitudes, practices, and strategies. *A decade in Internet time: Symposium on the dynamics of the Internet society*. Oxford Internet Institute. Retrieved 2/26/2013 from: http://www.danah.org/papers/2011/SocialPrivacyPLSC-Draft.pdf

Brenner, J. (2012, May 29). Pew Internet: Social networking (full detail). *Pew Internet and American Life Project*. Retrieved 4/24/2012 from:

http://pewinternet.org/Commentary/2012/March/Pew-Internet-Social-Networking-full-detail.aspx

Buchanan, T., Paine, C., Joinson, A. N., & Ulf-Deitrich, R. (2007). Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology, 58*(2), 157-165.

Buck, S. (2011). *The evolution of the Facebook profile*. Mashable. Retrieved 10/29/2012 from: http://mashable.com/2011/09/22/facebook-profile-evolution/#270572011--Timeline

Bureau of Educational and Cultural Affairs. (n.d.). *Special Projects*. Retrieved 1/7/2013 from: http://eca.state.gov/cultural-heritage-center/special-projects

Calabrese, C. (2009, November 18). Technology changes things. Latest example: Student records. *ACLU Blog of Rights: Technology and Liberty*. Retrieved June 22, 2010, from http://www.aclu.org/blog/technology-and-liberty/technology-changes-things-latest-example-student-records

CBS News. (2011). *Wael Ghonim and Egypt's new age revolution*. Retrieved from: http://www.cbsnews.com/stories/2011/02/13/60minutes/main20031701.shtml

Cheng, J. (2012, February 5). Over 3 years later, "deleted" Facebook records are still online. *ArsTechnica*. Retrieved 1/3/2013 from: http://arstechnica.com/business/2012/02/nearly-3-years-later-deleted-facebook-photos-are-still-online/

Christofides, E., Muise, A., & Desmarais, S. (2009). Information disclosure and control on Facebook: Are they two sides of the same coin or two different processes? *CyberPsychology & Behavior, 12*(3), 341-345.

Chronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3), pp 297 – 334.

Cohen, M., March, J., & Olsen, J. (1972). A garbage can model of organizational choice. *Administrative Quarterly*. *17*(1). 1 – 25.

Communications Act of 1934, 47 U.S.C. § 609 (1934).

Consultative Committee for Space Data Systems (CCDS). (2011, September). *Audit and certification of trustworthy digital repositories*. Retrieved 7/20/2012 from: http://public.ccsds.org/publications/archive/652x0m1.pdf

Cook, T. (2005). Macroappraisal in theory and practice: Origins, characteristics, and implementation in Canada, 1950-2000. *Archival Science, 5*(2 – 4), 101 – 161.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and application. *Journal of Applied Psychology, 78* (1), pp. 98 – 104.

Cox, R. J. (2008). *Personal Archives and a New Archival Calling*. Duluth, MN: Litwin Books.

DeCew, J. W. (1997). *In pursuit of privacy*. Ithaca, NY: Cornell University Press.

DGConnect. (11/9/2012). *Mission and Priorities*. Retrieved 11/1/2012 from: http://ec.europa.eu/dgs/connect/mission/index_en.htm

Divorce Online. (December 28, 2011). *Alarming increase in Facebook related divorces in 2011*. Retrieved 5/22/2012 from: http://blog.divorce-online.co.uk/?p=2338

Downs, A. (1972). Up and down with ecology: The issue-attention cycle. *Public Interest, 28* (summer), p. 38 – 51.

E-Government Act of 2002, 44 U.S.C. § 101 (2002).

Eisenstadt v. Baird, 405 U.S. 438 (1972).

Electronic Communication Privacy Act (ECPA), 18 USC § 2510, (1986).

EPIC. (n.d.). *Video Privacy Protection Act*. Retrieved 1/6/2013 from: http://epic.org/privacy/vppa/

U.S. Executive Office of the President. (1989, January 16).Executive Order 12667. 54 FR 3403. Retrieved 4/5/2013 from: http://www.archives.gov/federal-register/codification/executive-order/12667.html

U.S. Executive Office of the President. (2001, November 1). Executive Order 13233 - Further Implementation of the Presidential Records Act. 66 FR 56025. Retrieved 4/5/2013 from: http://www.archives.gov/about/laws/appendix/13233.html

U.S. Executive Office of the President. (2009, January 21). Executive Order 13489. 74 FR 15. Retrieved 4/5/2013 from: http://www.gpo.gov/fdsys/pkg/FR-2009-01-26/pdf/E9-1712.pdf.

Ex Parte Jackson, 96 U.S. 727 (1878).

Facebook. (2011, December 13). New partnership between Facebook and the National Suicide Prevention Lifeline. Retrieved 5/22/2012 from: https://www.facebook.com/note.php?note_id=310287485658707

Facebook (n.d. a) Pressroom. *Facebook.com*. Retrieved 8/12/2011 from: http://www.facebook.com/press/info.php?statistics

Facebook (n.d. b) Information we receive and how it is used. *Facebook.com*. Retrieved 4/30/2012 from: https://www.facebook.com/about/privacy/your-info#inforeceived

Facebook (n.d. c) Key Facts. *Facebook.com*. Retrieved 10/30/2012 from: http://newsroom.fb.com/content/default.aspx?NewsAreaId=22

Facebook (n.d. d) How can I download my information from Facebook? *Facebook.com*. Retrieved 10/30/2012 from: https://www.facebook.com/help/?faq=212802592074644#How-can-I-download-my-information-from-Facebook?

Family Educational and Privacy Rights (FERPA), 20 U.S.C. § 1232 (1965).

Federal Trade Commission (FTC) a. (2011, November 29). Facebook settles FTC charges that it deceived consumers by failing to keep privacy promises. Retrieved 7/20/2012 from: http://www.ftc.gov/opa/2011/11/privacysettlement.shtm

Federal Trade Commission (FTC) b. (2011, October 24). In the matter of Google, Inc. a corporation. Retrieved 11/1/2012 from: http://www.ftc.gov/os/caselist/1023136/index.shtm

Foreign Intelligence Surveillance Act (FISA), 50 USC § 36 (1978).

Forman, M. (2002). Implementing the President's management agenda for e-government. E – Government strategy. Retrieved 1/7/2013 from: http://www.usa.gov/Topics/Includes/Reference/egov_strategy.pdf

Freedom of Information Act (FOIA), 5 U.S.C. § 552 (1966).

Gallagher, R. (February 10, 2013). Program that tracks people on social media created by defence firm. *The Guardian*. Retrieved from: http://www.guardian.co.uk/world/2013/feb/10/software-tracks-social-media-defence

Gillette, F. (2011, June 22). The Rise and Inglorious Fall of MySpace. *Bloomberg BusinessWeek*. Retrieved 7/16/2012 from: http://www.businessweek.com/magazine/content/11_27/b4235053917570.htm

Gormley, K. (1992). One hundred years of privacy. *Wisconsin Law Review, 4*, 1335 - 1441.

Griswold v. Connecticut, 381 U.S. 479. (1965).

Gross, R., & Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society* (pp. 71-80). Alexandria, VA: ACM. doi:10.1145/1102199.1102214

Halpin, P. (September 21, 2012). Facebook tightens privacy to satisfy Irish regulators. *Reuters*. Retrieved 11/1/2012 from: http://www.reuters.com/article/2012/09/21/us-facebook-ireland-data-idUSBRE88K0PM20120921

Ham, G. (1975). The archival edge. *Society of American Archivists, 38* (1), 5 – 13.

Ham, G. (1981). Archival strategies for the post-custodial era. *American Archivist*. *44*(3), 207 – 216.

Helft, M. (2008, November 11). Google uses searches to track flu's spread. *The New York Times*. Retrieved 3/8/2013 from: http://www.nytimes.com/2008/11/12/technology/internet/12flu.html?_r=0

Hinduja, S., & Patchin, J. W. (2008). Personal information of adolescents on the Internet: A quantitative content analysis of MySpace. *Journal of Adolescence, 31*, 125-146.

Hobbs, C. (2001). The character of personal archives: Reflections on the value of records of individuals. *Archivaria, 52*, 126 – 135.

Hoofnagle, C., King, J., Li, S., & Turow, J. (2010). How different are young adults from older adults when it comes to information privacy attitudes and policies? Working paper retrieved 2/26/2013 from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1589864

Hyry, T. & Onuf, R. (1997). The personality of electronic records: The impact of new information technology on personal papers. Archival Issues, 22(1), 37–44.

ITU. (n.d.). *History*. Retrieved 11/1/2012 from: http://www.itu.int/en/about/Pages/history.aspx

ITU. (n.d. a) *What Does ITU Do?* Retrieved 11/1/2012 from: http://www.itu.int/en/about/Pages/whatwedo.aspx

International Standards Organization. (October, 2012). Information and documentation – Trusted third party repository for digital records. Retrieved 12/6/2012 from: http://www.iso.org/iso/catalogue_detail.htm?csnumber=58087

Internet Archive. (n.d.). *About the Internet Archive*. Retrieved 10/23/2012 from: http://archive.org/about/

Jenkinson, H. (1922). *A manual of archive administration. Second edition*. London, England: P. Lund, Humphries & Co., Ltd.

Kang, J. (1997). Information privacy in cyberspace transactions. *Stanford Law Review, 50*, 1193-1297.

Katz v. United States, 389 U.S. 347 (1967).

Kaul, I., Grunberg, I, & Stern, M.A. (1999). Defining global public goods. In Kaul, I., Grunberg, I, & Stern, M.A. (1999). *Global public goods: International Cooperation in the 21st Century* (pp. 2 - 19). New York, NY: Oxford.

Kesan, J. & Shah, R. (2006). Setting software defaults: Perspectives from law, computer science, and behavioral economics. *Notre Dame Law Review, 82*(2), pp. 583 – 634.

King, L. W. (2004). *The code of Hammurabi*. Whitefish, MT: Kessinger Publishing.

Kingdon, J. (2003). *Agendas, alternatives, and public policies*. Boston, MA: Addison-Wesley.

Koetsier, J. (July 20, 2012). *8,000 Facebook members die every day. What happens to their profiles?* VentureBeat. Retrieved 10/23/2012 from: http://venturebeat.com/2012/07/20/8000-facebook-members-die-every-day-what-happens-to-their-profiles/

Kuny, T. (1997). A digital dark ages? Challenges in the preservation of electronic information. Presented at the *63rd IFLA Council and General Conference*. Retrieved from: http://archive.ifla.org/IV/ifla63/63kuny1.pdf

Kyllo v. United States, 533 U.S. 27 (2001).

Lampe, C., Ellison, N. B., & Steinfield, C. (2008). Changes in use and perception of Facebook. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 721-730). San Diego, CA: ACM. doi:10.1145/1460563.1460675

Lenhart, A. (2006). Testimony to the House Committee on Energy and Commerce subcommittee on Telecommunications and the Internet hearing on H.R. 5319, The Deleting On-line Predators Act of 2006. July 11, 2006.

Lessig, L. (2006). *Code: And other laws of cyberspace*, version 2.0. New York, NY: Basic Books.

Library of Congress. (2010, April 15). Twitter donates entire tweet archive to Library of Congress. Retrieved 4/27/2012 from: http://www.loc.gov/today/pr/2010/10-081.html

Livingstone, S. (2008). Taking risky opportunities in youthful content creation: Teenagers' use of social networking sites for intimacy, privacy, and self-expression. *New Media and Society, 10* (3), pp. 393 – 411.

Lloyd, A. (2007). Guarding against collective amnesia? Making significance problematic: An exploration of issues. *Library Trends, 56* (1), pp. 53 – 65.

Lowenthal, D. (1985). *The Past is a Foreign Country*. Cambridge: Cambridge University Press.

Lyle, J. (2004). Sampling the umich.edu domain. Fourth international Web archiving workshop. Retrieved from: http://iwaw.europarchive.org/04/Lyle.pdf

Madden, M. (2010, August). Older adults and social media: Social networking use among those ages 50 and older nearly doubled the past year. *Pew Internet and American Life Project*. Retrieved from: http://www.pewinternet.org/~/media//Files/Reports/2010/Pew%20Internet%20-%20Older%20Adults%20and%20Social%20Media.pdf

Malkmus, D. (2008). Documentation strategy: Mastodon or retro-success? *American Archivist, 71*(2), 384 – 409.

Mann Elkins Act of 1910, Pub. L. No. 111-226, Stat. 539 (1910).

Marsh, George P. (1864). Man and nature. New York, NY: Scribner.

McKemmish, S. (1996). Evidence of me... Archives and Manuscripts 24(1), 28-45.

McKie, L. & Ryan, L. (2012). Exploring trends and challenges in Sociological research. *Sociology, 46* (6), pp. 1 – 7. Retrieved from: http://soc.sagepub.com/content/46/6/1.full.pdf+html

Misak, Cheryl. (Ed.). (2007). *New pragmatists*. Oxford, UK: Clarendon Press.

Moor, J. H. (1997). Towards a Theory of Privacy in the Information Age. *Computers and Society*. *27*(3), 27 – 32.

Moore, R. (2008). Towards a theory of digital preservation. *The International Journal of Digital Curation, 1*(3), 63-75.

Munn Vs. State of Illinois. (1876). 94 US 113. Retrieved on 7/18/2012 from: http://caselaw.lp.findlaw.com/scripts/getcase.pl?court=US&vol=94&invol=113

Murthy, D. (2008). Digital ethonography: An examination of the use of new technologies for social research. *Sociology, 42* (5), pp. 837 – 855. Retrieved from: http://soc.sagepub.com/content/42/5/837.short

OCLC, CRL. (2007). Trustworthy repositories audit and certification: Criteria and checklist. Retrieved from: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Office of Management and Budget. (2002, February 27). E-Government strategy: Implementing the President's management agenda for e-government. Retrieved 8/30/2012 from: http://www.usa.gov/Topics/Includes/Reference/egov_strategy.pdf

Ohm, P. (2005). The Fourth Amendment right to delete. *Harvard Law Review Forum, 119*(10), 10-18.

Olmstead v. United States, 277 U.S. 438. (1928).

Omnibus Crime Control and Safe Streets Act, 42 U.S.C. § 3789. (1968).

Opsahl, K. (2010, April 28). Facebook's eroding privacy policy: A timeline. *Electronic Frontier Foundation*. Retrieved 4/25/2012 from: https://www.eff.org/deeplinks/2010/04/facebook-timeline

O'Toole, J. & Cox, J. (2006). *Understanding archives and manuscripts*. Chicago, IL: The Society of American Archivists.

O'Sullivan, C. (2005). Diaries, on-line diaries, and the future loss to archives; Or, blogs and the blogging bloggers who blog them. *The American Archivist,* 68(Spring/Summer), 53 – 73.

Palfrey, J. (2008). The public and the private at the United States border with cyberspace. *Mississippi Law Journal, 78*, 241-294.

Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York, NY: Basic Books.

Parry, M. (July 10, 2011). Harvard researchers accused of breaching students' privacy. *The Chronicle of Higher Education*. Retrieved 2/25/2013 from: http://chronicle.com/article/Harvards-Privacy-Meltdown/128166/

Pearce-Moses, R. (n.d.). *A glossary of archival records and terminology*. Retrieved 1/7/2013 from: http://www2.archivists.org/glossary/terms/a/archives

Pelofsky, J. (2012, April 13). Fight emerges over fate of Megaupload U.S. servers. *Reuters*. Retrieved 1/7/2013 from: http://www.reuters.com/article/2012/04/13/net-us-megaupload-idUSBRE83C1DK20120413

Pollard, R. (2001). The appraisal of personal papers: A critical literature review. *Archivaria 52*, 136-150.

Pomeroy, W. B., Martin, C. E., & Kinsey, A. C. (1949). *Sexual behavior in the human male*. W. B Saunders Company.

Pope, N. (1997). Envelopes in the machine age. *EnRoute, 6*(2). Retrieved from http://postalmuseum.si.edu/resources/6a2o_envelopes.html

Prelinger Library. (2007). Society of American Archivists decides to nuke its listserv archives. *Prelinger Library Blog*. Retrieved 10/16/2012 from: http://prelingerlibrary.blogspot.com/2007/03/society-of-american-archivists-decides.html

Prosser, W. L. (1960). Privacy. *California Law Review*. *48*(3), 383 – 423.

Quinn, Patrick M. (1977). The Times They Are A-Changin'. *The Midwestern Archivist, 2* (2), 5 – 13.

Relyea, H. C. (2002). Homeland security and information. *Government Information Quarterly, 19*(3), 213-223.

Research Libraries Group. (2002, May). Trusted digital repositories: Attributes and responsibilities. Retrieved 7/20/2012 from: http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf

Reuters. (2011, November 29). *FTC says Facebook must end deceptive privacy practices*. Retrieved 6/7/2012 from: http://www.reuters.com/article/2011/11/30/facebook-privacy-idUSN1E7AS1AA20111130

Right to Financial Privacy Act 12 U.S.C. § 35, (1978).

Roe v. Wade, 410 U.S. 113 (1973).

Rogers, R. (1975). A protection motivation theory of fear appeals and attitude change. *The Journal of Psychology: Interdisciplinary and Applied, 91*(1), 93 – 114.

Rogers, R. (1983). Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. *Journal of Experimental Psychology, 19*(5), 469 – 479.

Rohlfs, J. (1974). A theory of interdependent demand for a communication service. *The Bell Journal of Economics and Mangement Science, 5* (1), 16 – 37.

Rushe, D. (2011, June 30). Myspace sold for $35m in spectacular fall from $12bn heyday. Retrieved 3/8/2013 from: http://www.guardian.co.uk/technology/2011/jun/30/myspace-sold-35-million-news

Samuels, H.W. (1986). Who controls the past. *American Archivist. 49*(2), 109 – 124.

Savage, M. & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology, 41* (5), 885 – 899. Retrieved from: http://soc.sagepub.com/content/41/5/885.full.pdf+html

Schenck vs. United States. 249 U.S. 47. (1919).

Schellenberg, T. (1956). *The appraisal of modern public records*. Bulletins of the National Archives No. 8; Washington: National Archives. Retrieved 1/7/2013 from: http://www.archives.gov/research/alic/reference/archives-resources/appraisal-of-records.html

Schneier, B. (2009, November 19). *A taxonomy of social networking data*. Blog post. Retrieved 4/30/2012 from: http://www.schneier.com/blog/archives/2009/11/a_taxonomy_of_s.html

Seipp, D. J. 1981. *The right to privacy in American history*. Cambridge: Harvard University Press.

Sherman Act, 15 U.S.C. 1 (1890).

Smith, A. (2007). Valuing preservation. *Library Trends, 56* (1), pp. 4 – 25.

Solon, O. (2012, December 28). How much data did Facebook have on one man? 1,200 pages of data in 57 categories. *Wired News*. Retrieved 1/3/2013 from: http://www.wired.co.uk/magazine/archive/2012/12/start/privacy-versus-facebook?page=all

Solove, D. J. (2008). *Understanding privacy*. Cambridge, MA: Harvard University Press.

Solove, D. J. (2011). *Nothing to hide. The false tradeoff between privacy and security*. New Haven, CT: Yale.

Stross, R. (May 1, 2010). When history is compiled 140 characters at a time. The New York Times. Retrieved 2/12/2013 from: http://www.nytimes.com/2010/05/02/business/02digi.html?_r=0

Subrahmanyam, K. & Greenfield, P. (2008). Online communication and adolescent relationships. *The Future of Children, 18*(1), 119-146.

Texas Advanced Computing Center (TACC). (n.d.). A thousand words: Advanced visualization for the humanities. Retrieved 2/21/2013 from: http://www.tacc.utexas.edu/tacc-projects/a-thousand-words

Tavani, H. T. (2007). Philosophical theories of privacy: Implications for an adequate online privacy policy. *Metaphilosophy, 38*(1), 1-22.

Telecommunications Act of 1996, 47 U.S.C. §151. (1996).

Telecommunications Act of 1996, 47 U.S.C. § 609 (1996).

Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in the coming years. *The state of digital preservation: An international*

perspective. *Conference proceedings*. Washington, D.C.: Council on Library and Information Resources.

Thomas, S. (2007). A practical approach to the preservation of personal digital archives. Final report, Joint Information Systems Committee (JISC). Retrieved December 22, 2008, from: http://www.paradigm.ac.uk/projectdocs/jiscreports/ParadigmFinalReportv1.pdf

Twitter. (2012, March 12). Twitter turns six. Retrieved 7/20/2012 from: http://blog.twitter.com/2012/03/twitter-turns-six.html

UNESCO. (n.d.). *The organization's history*. Retrieved 11/1/2012 from: http://www.unesco.org/new/en/unesco/about-us/who-we-are/history/

UNESCO. (n.d. a). *Preservation of documentary heritage*. Retrieved 11/1/2012 from: http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/preservation-of-documentary-heritage/

UNESCO. (2006). *Guidelines for the preservation of digital heritage*. Retrieved 11/1/2012 from: http://unesdoc.unesco.org/images/0013/001300/130071e.pdf

Uniting and Strengthening America by Providing Appropriate Tools Required To Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, Pub. L. No. 107-56. 115 Stat. 272 (2001).

U.S. Constitution amend. I.

U.S. vs. Microsoft. Civil Action No. 98-1232 (TPJ). (1999, July 28). Court's Findings of Fact.

Vertuno, J. (2008, January 3). Norman Mailer archive opens in Texas. *Associated Press*. Retrieved 4/30/2012

The Video Privacy Protection Act (VPPA), 18 USC § 2710, (1988).

Warren, S. D., & Brandeis, L. D. (1890). *Right to privacy*. Harvard Law Review, 4, 193-220.

Weideman, C. (n.d.) A guide to deeds of gift. *Society of American Archivists online brochure*. Retrieved on 4/15/2012 from: http://www.archivists.org/publications/deed_of_gift.asp

Westin, A. F. (1967). *Privacy and freedom*. New York, NY: Atheneum.

Youn, S. (2009). Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents. *The Journal of Consumer Affairs, 43*(3), 389 - 418.

Zinn, Howard. (2001). *Howard Zinn on history*. New York, NY: Seven Stories Press.