Towards an Archival Approach to Digital Identity Management

Fred Stutzman

May 9, 2006

Abstract

Identity, in the context of our communal online space, is a fungible concept. Who we are, how we represent ourselves, how we're identified, how we're verified, and how we're investigated all are critical components of the larger notion of online identity. To that extent, the field is truly multi-faceted; cryptographers and computer scientists debate procedures of verification and authentication, sociologists and those in the cultural studies fields examine our "senses" of identity, and information scientists and systems developers attempt to develop tools that will help us collect, manage, store and retrieve elements of our personal identity at a later date. Viewing all these parts holistically, one can clearly see that the results of our lifetime identity production compose a set of documents (or, a document representation of our life).

If we are to create a constant digital record of our identity production, how are we to manage it? The challenges are both traditional and

emergent; Issues of technology, capacity, migration and freshening - those we see commonly in the digital archives field - are fundamentally challenging. Emergent issues, such as how we contextualize and annotate our identity for later retrieval, and how we publicly present our identity artifacts online are fascinating, requiring study. In this paper, I'll attempt to bring together a number of loosely joined pieces from the disparate areas of identity, with a goal of formulating a thought process for dealing with archival notions of our identity.

1 Our Digital Identity

In Scrolling Forward, David M. Levy [18] explores the changing nature of documents as we migrate to a digital world. What is lost and gained in this transformation, and how does it affect us? In a chapter entitled *Reach out and Touch Someone*, Levy explores how email is changing personal communication. He asks:

> It makes me wonder if the real question being asked isn't about the letter or about e-mail, but about a mode of life. Many of us feel that our lives are speeding up, becoming more fragmented and dislocated. Under such circumstances, what are the possibilities for deep human contact and communion? Will e-mail help us or hurt us? Is it possible that e-mail will enable a true correspondence of souls, or will it prove to be a technology of alienation? This is

> what I hear us asking, at any rate, even when the words are not spoken as such, even when we seem to be speaking most directly about properties of the new technologies.[18]

These questions, and the greater context of Levy's work are informative as we look at archival notions of our digital identity. First, the question of how digitization changes the nature of interaction is rather valid when it comes to identity production. Put simply, if we are to digitize (or virtualize) the production of self, how is it we will come to understand this new notion of identity? Second, how do we deal with the fact that our identity fragments are increasingly digitized? In the past, we generally left behind a written record that could be pieced back together if we were lucky enough to merit the labor costs. These identity artifacts were our credentials (birth certificate, licenses), our production (written works), and the works that referenced us (photographs, articles, reviews). Each of these objects existed inside a formalized genre, and we had a set of practices for the evaluation (in terms of authenticity, archival and artifact value) of these objects. Transforming notions of this artifactual identity to the digital age, we're posed with a number of questions that didn't exist previously - and certainly didn't exist on today's scale.

There are a number of informative scenarios. First, imagine an individual, over the course of a lifetime, placing comments on blogs and wikis. These comments are attached to the individual's real name, and they are available in search engines. These virtual comments, which may have previously existed as written records (correspondence) or spoken words, now are available forever, contributing to a picture of our identity. Assuming an unstructured web, how would one piece together a picture of an indvidual's identity with these comments? It would be difficult, because in many cases date and contextual information wouldn't be present.

Next, imagine an individual, using a memex-type device, recording all aspects of his or her production. Certainly, everything the individual produces (the previous web comments and wiki postings being a subset of production) is recorded in the memex. How would this mass store of information be valuable to archivists? Surely, the information is valuable, but value is contingent on the usability of the information. If an archivist must spend a lifetime examining the life-record of an individual in realtime, we begin to see how this doesn't scale well. Indeed, questions of context and volume are not remotely new in the archival field, and they have been addressed in digital contexts as well [11] [20] [19]. However, the scale of the emergent problem, and the importance, both practical and theoretical, of developing solutions makes this a very relevant topic.

In the following paper, an archival approach to identity will be presented. We'll look at directions our digital identities are taking, explore the problems that are emerging, and look at interesting potential solutions to the problem. In doing so, we'll cross over a number of different areas of literature, exploring senses of identity, means of storage and retrieval, and the implications - both cultural and legal - of our digital identity.

2 Digital Identity in Context

Sherry Turkle, the MIT psychologist and professor, explored notions of digital identity in her pathbreaking work *Life on the Screen*[26]. Identity, says Turkle, "refers to the sameness between two qualities, in this case between a person and his or her persona." [27] Turkle's key distinction, however, is that in an online context, one can have many personas; fundamentally, the notion of identity is redefined. Assuming we live a fairly modern existence, our digital footprints are everywhere. These footprints are in the e-mail messages we send, the websites we join, the fora in which we participate, the list serves we post to, the blogs we comment upon, the chat rooms we enter, the systems our identity information resides upon, and so on. In fact, our digital presence is so vast it is exceedingly hard to wrap a meaningful definition around what it means to "be digital." Even when we're not actively leaving our footprints online, digital agents - auction bots, away messages - are acting on our behalf, leaving traces of our intention in a digital record [29].

Digital technology has had a powerful multiplicative effect on the work we can do. We can simultaneously be present in many simulated spaces, interacting with humans and bots, carrying out transactions, leaving a record. The ability to do more, compounded with the fact that all of our transactions can be recorded and archived have powerful repercussions.

Turkle's notion of identity as the sameness between two qualities is reinforced in Lessig and Abelson's definition of digital identity. According to the authors:

> Basically, the essential and unique characteristics of an entity are what identify it. These characteristics might include, among other things, the unchanging physical traits of the person, his preferences, or other people's perception of the individual's personality. The skills that a person possesses can also become part of one's identity No two identities are the same. Each identity maps to a unique set of characteristics. Two people may share some of the same characteristics, such as being old enough to drive or having the same hair color, but that does not mean they have the same identity. One simply is not looking at enough characteristics.[1]

We see that our identity is the sum of our characteristics; according to Lessig and Abelson's logic chain, we'll never be the "same" as anyone else, since we'll always possess a difference in characteristic, no matter how granular, from the other entity. In the digital realm, there are a number of states that our identity can occupy. Baier et. al. [3], Allison et. al [2], Pato [23], Renear et. al. [24], and Crawford [8] explore various states of digital identity, particularly how we become an identified individual on the net. The processes of identification are the technological implementations of our cultural heritage, and since we identify ourselves in many ways, there's support for these different states.

3 States of Identity

Gary Marx, in *What's in a Name? Some Reflections on the Sociology of Anonymity* explores our seven types of identity knowledge. Identity and identification, he argues, comes in seven states. They are:

1. Legal names
2. Locatability
3. Pseudonyms that can be linked to legal name and/or locatability
4. Pseudonyms that cannot be linked to other forms of identity knowledge
5. Pattern knowledge
6. Social categorization
7. Symbols of eligibility/non-eligibility [21]

To simplify, we can think of identity as verified, pseudononymous or anonymous. In the state of verification, we are credentialed; our legal names are used, government documents provide gateways to the systems that verify our identity. While we often use our legal names online, the process of doing so does not imply verification. Consider two scenarios: In the first, an individual posts a comment to a blog using his or her real name; in the second, an individual conducts an electronic commerce transaction using a credit card. The first transaction, though using legal names, is not verified, while the second transaction involves verification because the credential (in this case being a credit card number), used in concert with a legal name establishes verification.

Pseudononymous identity is quite common on the internet. When one joins a new website, they are generally given a "user name" that will represent their identity. As human namespace is not unique, these unique user names are necessary for the technical operation of the site (record locators must be unique, as any collision in locator address would represent the unwanted enmeshing of two entities). User names, or "handles", step in to represent our actions and intentions in the context of the particular site we're logged in to; we can develop a reputation attached to or separate from our verified identity. The practice of using pseudononymous identity is so prevalent on the net, in a novel study, Millen and Patterson [22] found actors behaved quite differently in a forum where verified identity was required.

There's a distinct difference between pseudononymous identity and verified identity on the net. Marx, in his seven states, points out a number of cases where our pseudononymous identity can be profiled. First, the locations we occupy provide clues as to our identity. Imagine pseudononymous users on a cancer support message board; their location would give us good clues to their identity, and we'd be able to begin making a profile of their identity. Expanding this, Marx's notion of pattern knowledge would inform our identity profile. Someone who uses a cancer support forum may also visit the American Cancer Society's website, or conduct MedLine searches for cancer literature. Even though we may not know who the person is, their behavior and signals create a unique identity - different from any other on the net.

Finally, there is anonymity. Anonymity is pseudonymity without any identifying characteristics or traits. In Marx's view, there's quite a limited space for true anonymity in the real world; our patterns, social categories or symbolism contribute to a picture of our identity. However, online, anonymity is quite possible. Software such as Tor [1] and BugMeNot [2] combine to provide individuals full anonymity in identifiable transactions. User names, IP addresses, browser and system information, links to verified identity - can all be falsified with client software. We can truly be "ghosts" online, with minimal effort.

This exploration of the states of identity is important so we can understand a critical archival approach to digital identity. As diplomatics and the archival appraisal process evaluates certain facets of an item for quality and archival value, the bits and pieces of digital identity we leave will also be subject to this evaluative process. Our digital identity production may be an email, a blog posting, a forum message, an electronic document or a social network profile. Our digital iden-

[1] http://tor.eff.org

[2] http://BugMeNot.com

tity may live on the net, in a memex, in an institutional repository or a dark archive. Our digital identity may have our verified name attached to it, one of our pseudonyms, or we may suspect it to be an anonymous transaction. For all of these transactions, there is a clear state of determinable identity. As we've now explored the states of identity, we'll next look at how we might manage our identity.

4 Personal Identity Management

> Consider a future device for individual use, which is a sort of mechanized private file and library. It needs a name, and to coin one at random, "memex" will do. A memex is a device in which an individual stores all his books, records and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory. *Vannevar Bush, As We May Think* [6]

We've never been closer to the realization of Vannevar Bush's memex. Microsoft has invested significant resources into the MyLifeBits [15] [14] [16] project, with a goal of creating a personal archive of a life's production. All that one does, all that they see, all their interactions would be placed into the relational database backend of MyLifeBits, creating a searchable personal archive of one's life. Indeed, Microsoft isn't alone in their quest. The Lifestreams project [13] was a metadata-intensive precursor to the MyLifeBits project. Beagrie [5], Ahmed et. al. [12], Cohn and Hibbits [7] and Czerwinski et. al. [9] each propose implementations of a memex-like tool for personal archiving. As Gemmell [15] shows, all memex-like projects are built on the projection that disk storage will continue to become cheaper as our needs for storage grow, therefore, the practical impediments to the implementation of a memex lie only in the software and practices. (While devices for audio and video capture can be extremely small, we're not quite accustomed to the idea of walking around with a video recorder on our shoulder at all times. Cultural issues, combined with recorder form-factor issue do impede the realization of a ubiquitous memex-enabled future.)

One of the most challenging issues of personal information management is dealing with the volume and heterogeneity of information we encounter. Jones et. al. [17] describe the problem as information fragmentation. Fragmentation addresses the fact that our personal information, and thereby our digital identity, are spread over many systems and physical locations. Therefore, the re-finding of important personal information at a later date is extremely complex. If we are to assume a memex-like solution, fragmentation ceases to be the major issue it is today (assuming that all fragmented devices would report back to the memex). As we've seen, this, this is not yet the case. According to Barreau [4] and Dumais [10], one of the key challenges of PIM, particularly in re-finding, are losses of context. To have an important fragment of information is not enough; we must also be able to put it in context so that we can understand the information's value.

There are a number of practical approaches to applying context to our personal information. The Microsoft team designed tools for the easy annotation of data. For example, a digital video alone isn't worth much if we can't remember when it was recorded, who was in the video, and in what context the video was recorded. On top of that, current information retrieval technology does not allow complex video searches - that we might be able to enter "birthday" into a search box, and have a video retrieval engine find videos of candles being blown out is only fantasy at this point. Providing the user with simple tools to annotate their personal information with stories, MyLifeBits would process the annotations, and use them for context and search. Another emergent trend in re-finding is the use of tagging. Tags are simple bits of non-hierarchical (folksonomic) metadata applied to an item. For example, birthday video may be tagged *birthday 2006 anne.* These three tags would let us know that the video was from Anne's 2006 birthday. Along those lines, we'd also be able to pivot and find any videos matching anne, birthday or 2006 in our collection. According to Gemmell et. al., these tagging systems "have difficulty coping with scale." [15] The authors fail to produce a reference that establishes this point, however.

Personal information management provides a very fitting model for personal identity management (PIM) or digital identity management (DIM). The central challenges are remarkably similar. In personal information management, the field seeks to find ways to organize all of an individual's digital information, creating structure so that the information can be found and used at a later date (with high precision). Personal identity management seeks to provide a place where an individual can track, manage and maintain his or her digital identity, and all of the traces his or her digital identity have left behind. In a sense, this is a subset of personal information management, but an important subset.

Non-uniqueness in the human namespace is a non-trivial problem. Since we share names, we're forced to adopt pseudonyms. At the same time, others who share our name may post content that appears to be ours. If our digital identity is part of our reputation, we have strong economic and social motivations to manage our digital identity. We seek to extract value from our pseudonyms (hence, tying them to our reputation), while clarifying name collisions (both in our real name and our pseudonyms). What's more, since much of this information is out of our control (fragmented across systems we do not control), we're faced with some serious challenges.

5 An Archival Approach to Personal Identity Management

The management of personal identity information presents a multifaceted challenge. If we are to take an archival approach to our digital identity production, how should we define the problem space? Lynch [20], Duranti [11] and Lyman [19] frame the discussion of archival digital information. Lynch provides an overview of the archival approach:

> Before attempting to define integrity or authenticity, it is worth trying to gain an intuitive sense of how the digital environment differs from the physical world of information-bearing artifacts ("meatspace" as somenow call it). The archetypical situation is this: We have an object and a collection of assertions about it. These assertions may be internal, as in a claim of authorship or date and place of publication of the title page of a book, or external, represented in metadata that accompany the object, perhaps provided by third parties. [20]

Lynch goes on to list an approach for answering archival questions. To trust an archival object, archivists:

1. Examine the provenance of the object.

2. Perform forensic and diplomatic examination of the object.

3. Examine the signatures and seals that come with the object.

4. For mass-produced objects, compare the object to other known versions. [20]

Digital archiving policies exist for all facets of Lynch's described process. Generally, digital archivists are dealing with a known, defined corpus with predictable metadata. An archivist may process digital scans of museum artwork or a collection, or examine the digital record of a creative process. The corpus has boundaries, and the control inherent in the creation of the corpus solves a number of the archivist's problems.

Unfortunately, though, the process an archivist would undertake in examining the digital life record of an individual would be messy, and not particularly analogous to our previous examples. An archival record of digital identity would be comprised of our digital credentials, our digital creations, our digital correspondence, the multimedia we create - and many more things. As an illustrative example, let's look at a blog post.

A blog post is a digital document that speaks to an individual's digital identity. The individual is creating a representation of his identity through his work, and outside consumers are evaluating and creating a picture of the individual's identity as a result. Indeed, a blog posting may be a bit of throwaway content, but it might also be the point that marks the germination of an individual's world-changing research agenda. In that case, it is worthwhile to figure out an approach strategy.

A blog posting, in its original, defined form, is simply a temporal update to a webpage. The individual writes some copy, comes up with a title, and hits submit. Behind the scenes, a software program stamps that content chunk with a unique identifier, associates structured metadata with the post, creates an archival location for permanent reference, and then publishes the post to the site.

However, as the blog post is being processed, the original content is entering a number of different forms. A script takes the content and pushes it into a structured form known as RSS. As a result, RSS readers come and "pick up" a digital copy of the document. At the same time, other scripts package the content into an email, which is then dispatched to a number of the blog's email subscribers. Finally, that blog's content is pulled into a meta-blog that republishes the content in a new context, with a new set of unique identifiers, permanent archival locations, and comment thread. The push of one single button has blanketed the web with a singular item of identity production, though in many different forms. Thirty years later, when the original website no longer remains (assumption for the purpose of the example), that content may live on. How will an archivist (or an individual self-archiving their identity) approach this situation?

If we were to apply Lynch's rules, the first would be an examination of the provenance, forensics and diplomatics of the object. This singular object has taken many forms - be they the original blog post, the meta-blog post, the email or the RSS feed. How would we approach an examination of provenance? The critical question of provenance involves who has owned the content, how well that is documented, where the content has been kept, and how well we can trust the system of records that proves these claims or assertions. Considering our task is to create an archive of a digital post from the ether, it seems to be such a challenge we may be tempted to give up before we begin. However, there is hope. Since the creation of this post was time and date stamped, and a unique identifier was attached to the content object, we can reasonably assume that all content objects will have a copy of this identifier. At the same time, the post's content may reference dates, times, links and other characteristics that may prove useful in identification. Our evaluation of provenance may be supplanted by the integrity of the document, if it is available in many forms.

In its multiple forms, this content may exist in people's RSS archives, their email archives, in a browser cache or on tape backup in a vault somewhere. This content may be spread across a multiplicity of devices, such as a server, laptop, PDA, or mobile phone. If we're attempting to construct an archive of one's digital identity, technological advances force us to abandon the notion of owning content in its original form. As the provenance of a digital identity object gets harder to track and verify, we become able to rely on other means for verification. Thankfully, due to the number of forms and wide dissemination of content, our processes of verification may in fact turn to checking against other known copies. For example, does the RSS version of content match the email version of content, are the unique identifiers the same? Although certainly not perfect, we see possibilities for future identity archivists. In a sense, the provenance and forensic/diplomatic examination of the object merge. In fact, the four criteria Lynch lays out for trusting a digital object are all forced to merge - the verification process will intertwine, forcing us to use a flexible protocol for understanding digital authenticity.

6 The Evolution of Digital Identity Records

Duranti, in *The Impact of Digital Technology on Archival Science*, describes the eight components of a digital record. They are:

1. Medium
2. Content
3. Physical form
4. Intellectual form
5. Action
6. Four persons (author, addressee, writer, creator)
7. Archival bond
8. Context [11]

Does the definition of a digital record change when we place it in the context of digital identity? Digital identity records are like any other record; in this case, we are very concerned with how the particular record represents the identity of the attached individual. While this certainly falls under Duranti's "four persons" rule, it would make sense to expand on this a bit.

In the context of identity, our identity can have many states. In the context of an digital identity record, the part of our identity attached to the record can also have many states. Consider how a "real-world" archivist would approach a personal identity item, such as a driver's license. This license is tied to our identity through our name, our picture, the facets of our identity represented in the license. The license, however, does not get its authority from these particular elements. The license gets its authority because it is state-issued, and it has a particular individual record number (drivers license) that serves as an identifying proxy in the system. Imagining the examination of this record 50 years after its issue date - while the identity information is valid, the authority information may be invalid. For example, the license may be invalid due to expiration, the computer record attached to the driver's license number may no longer exist, or the state itself may no longer exist. This record was a valid identity document at one time, and it vouched for our identity at that time; how do we consider it valuable in the present, where its only purpose is that of historical artifact?

The driver's license example is illustrative. Now imagine our presence on the internet. Our presence on the internet is pseudononymous, verified with a pseudononymous email address, and authenticated via a identity management system located in a data center somewhere. To our digital archivist, who knows the various pseudonyms we use around the net, how would they attempt to attach our verified identity to our pseudononymous postings? This is to say, while the multiplicity of content forms may allow us to "trust" a document in Lynch's sense, can we truly trust an identity document if we're not clearly able to track the document back to a verifiable author?

This is a unique problem of the digital age, particularly for archivists. If MyLifeBits or LifeStreams are going to be our personal digital archive, we can implicitly trust the material we self-archive. However, we're a number of years away from the implementation of such a system, and our digital identity documents pervade the internet. While the author has suggestions for how to approach this problem of verification, without implicit verification (some sort of process that proves we are who we say we are), this will continue to be an ever-increasing challenge. In the real world, we're used to being able to "drop in" to a conversation without credentialing ourselves. While we carry documents, we're not used to proving out identity (until we want to buy alcohol or withdraw money from the bank). There's no reason to believe a sort of model where we are constantly verified would ever work on the internet; Microsoft's Passport attempted to start this process, and failed miserably. We're used to being able to be free - to participate where we want, when we want. David Weinberger, in *Small Pieces Loosely Joined*, describes how one individual may approach identity.

> Buyers and sellers on eBay adopt a name by which they will be known. The eBay name of a woman selling the quilt I was interested in was “firewife30.” Firewife30 is an identity, a self, that lives only within eBay. If she’s a selfish bastard elsewhere but always acts with honor in her eBay transactions, the “elsewhere” is not a part of firewife30 that I can know about or should particularly care about. The real-world person behind firewife30 may have other eBay identities. Perhaps she’s also SexyUndies who has 132 “sexy items” for sale at eBay while firewife30 was auctioning her quilt. Unlike real-world selves, these selves are intermittent and, most important, they are written. For all we know, firewife30 started out as firewife1 and it’s taken her this many drafts to craft a self that feels right to her. [28]

That we can create multiple identities, and use them in context is a hallmark of the net. That these identities can then be used for the accretion of social or economic capital is fairly a new phenomenon. This leaves us with the obvious conclusion that some facets of our online identity are worth archiving, and some are not. The framework we develop for archiving digital identities must be based on existing practice and structures. We must be able to classify the object as Duranti points out, and we must be able to trust it following Lynch’s model. There are unique challenges in that the systems we participate in, while being open to the world, have closed authentication systems. Sciences will emerge, however, that provide us best practices for the archiving of a person's digital creation.

7 ClaimID, a System for Digital Identity Management

ClaimID, as described in Stutzman and Russell [25], is a system that allows people to manage an important facet of their digitial identity: how they are represented in search engine results. On the internet, anyone may talk about us. At the same time, our digital identity, spread thin through the use of pseudonyms, are not necessarily attached to our offline identity. When a person searches us, they are forced to disambiguate our name identity, verify the veracity of a claim made about us, and intuit *who else we might be on the net* to get a complete picture of our identity.

The challenge this presents to an identity searcher is multifaceted. Indeed, someone with an insider knowledge of an individual may be able to guess or factor out things about us or not about us on the internet, but an outsider will never have this advantage. The outsider, however, is often the most important identity searcher - they may be deciding whether to hire or date the individual. Since the identity search process is complex, Stutzman and Russell created a system for the contextual self-archive of one's internet identities.

ClaimID is a simple web application that allows people to track, classify, annotate, prioritize and share their digital identity. Allowing individuals to piece together the mentions of their real name, and the mentions of their pseudonyms, ClaimID allows the self-archiving of identity presentation. The problem space is twofold. First, ClaimID allows people to disambiguate their legal name; they can specify what things online that mention their name are actually about them. Second, ClaimID allows people to associate their pseudonyms with their real identity. To further address archival needs, ClaimID automatically archives a copy of the item the person points to; this allows them to have a permanent copy of everything that references their identity, if they so wish.

As context is important for archiving purposes, ClaimID allows individuals many opportunities to place context around their digital identity objects. They may describe the material, place it in a group hierarchy, date the material, and finally tag the material for cross-hierarchical, semi-structured browsing. As identity material is of many forms, the ability to place many types of context and metadata around on a piece of identity material was immediately valuable. ClaimID then allows individuals to share their identity on the net, providing a more complete picture of their digital identity. In a sense, ClaimID is a singular, targeted implementation of MyLifeBits. Realizing that a self-archiving process must not be complicated, but the necessity for rich metadata is strong, ClaimID allows people an easy way to self-archive their public identity with rich metadata. It is in no way as complex as MyLifeBits, but the popularity of the application has proved the need for personal self-archival tools.

8 Conclusions

As we live online, we leave our digital footprints everywhere. They are in the emails we send, the blogs we post to, the fora in which we participate. All of our production speaks to our identity, and we may have many identities that coexist simultaneously. Indeed, these are interesting times, as we've never before encountered the levels of personal production that are now commonplace, nor have we thought about its particular value. A real-life conversation between two people exists in the instant it happens, and in personal retellings. An internet blog posting, over a topic mundane or important, becomes a public conversation visible to all. This throwaway bit of production speaks to our identity - it may influence the jobs we can get, who we can befriend, or who we might marry. And this is happening on a global scale, with more participants joining the conversation each day. How our society will deal with this is still being determined; in fact, these emergent questions of identity will prove to be some of the greatest challenges for the field of information science going forward.

In this paper, we've brought together and explored a number of disparate topics: Identity, personal information management, archival theory, personal archiving. We've looked at how identity can be represented, and how we come to create our online identity. Using the memex as an example, we've explored a future in which all of our identity production can be collected, and we lean on the archival perspective for management strategies. Indeed, the memex is not a reality, but memex-like applications crop up each day. In a sense, the internet is a collective memex, remembering things good and bad about us, while not doing a particularly great job of representing our identity. The challenges presented, as we attempt to both collect and manage our identity represent significant problem spaces. Drawing on a number of different established areas, we begin to see how solutions can emerge.

ClaimID is an example of a lightweight solution to a personal archiving problem. Letting people collect their online identity, and placing an archival context around the identity parts, we see a human-usable solution informed by the archival perspective. The sheer volume of material available today, and the growing volume of material that will be available tomorrow forces humans to become involved in the archival task. If there is to be economic and social value extracted from the self-archiving process, people will take part in the exercise. Our conception of online identity is variable - the words "online identity" mean different things to different people. However, online identity information is personally valuable; we're all leaving a digital record behind that will speak to our identity. This creates very interesting challenges for technologists, archivists and those in the cultural studies. Through a synthesis of these disciplines, we can better understand our digital identity, and better develop solutions for the management of our digital identity.

References

[1] Hal Abelson and Lawrence Lessig. Digital identity in cyberspace. Technical report, White Paper Submitted for 6.805/Law of Cyberspace: Social Protocols, 1998.

[2] Arthur Allison, James Currall, Michael Moss, and Susan Stuart. Digital identity matters. *JASIST*, 56(4):364–372, 2005.

[3] Toby Baier, Christian Zirpins, and Winfriend Lamersdorf. Digital identity: How to be someone on the net. *e-Society*, 2:815–820, June 2003.

[4] Deborah K. Barreau. Context as a factor in personal information management systems. *JASIST*, 46(5):327–339, 1995.

[5] Neal Beagrie. Plenty of room at the bottom? personal digital libraries and collections. *D-Lib Magazine*, 11(4), 2005.

[6] Vannevar Bush. As we may think. *The Atlantic Monthly*, July 1945.

[7] Ellen Cohn and Bernard Hibbitts. Beyond the electronic portfolio: A lifetime personal web space. *Educause Quarterly*, (4):7–10, 2004.

[8] Susan Crawford. Who's in charge of who i am?: Identity and law online. *New York Law School Law Review*, 49:211–229, 2004.

[9] Mary Czerwinski, Douglas Gage, Jim Gemmell, Catherine Marshall, Manuel Perez-Quinonesis, Meredith M. Skeels, and Tiziana Catarci. Digital memories in an era of ubiquitous computing and abundant storage. *Communications of the ACM*, 49(1):45–52, 2006.

[10] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. Stuff i've seen: A system for personal information retrieval and re-use. In *Proceedings of the SIGIR 2003*, pages 72–79, 2003.

[11] Luciana Duranti. The impact of digital technology on archival science. *Archival Sciecne*, 1:39–55, 2001.

[12] Ahmed et. al. Semanticlife - a framework for managing information of a human lifetime. In *Proceedings of 6th International Conference on Information Integration, Web-Applications and Services*, 2004.

[13] Eric Freeman and David Gelernter. Lifestreams: A storage model for personal data. *SIGMOD Record*, 25(1), 1996.

[14] Jim Gemmell, Gordon Bell, Roger Leuder, Steven Drucker, and Curtis Wong. Mylifebits: Fulfilling the memex vision. In *Proceedings of Multimedia 2002*, 2002.

[15] Jim Gemmell, Gordon Bell, and Roger Lueder. Mylifebits: A personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.

[16] Jim Gemmell, Roger Leuder, and Gordon Bell. The mylifebits lifetime store. In *Proceedings of ETP 2003*, 2003.

[17] William Jones, David Karger, Ofer Bergman, Mike Franklin, Wanda Pratt, and Marcia Bates. Towards a unification and integration of pim support. Technical report, The University of Washington, 2005.

[18] David M. Levy. *Scrolling Forward: Making Sense of Documents in a Digital Age*. Arcade, 2001.

[19] Peter Lyman and Brewster Kahle. Archiving digital cultural artifacts: Organizing an agenda for action. *D-Lib Magazine*, July/August 1998.

[20] Clifford Lynch. *Authenticity and Integrity in the Digital Environment*, chapter Authenticity and Integrity in the Digital Environment: An Exploratory Analysis of the Central Role of Trust. Council on Library and Information Resources, 2000.

[21] Gary Marx. What's in a name? some reflections on the sociology of anonymity. *The Information Society*, Special Issue on Anonymous Computing, 1999.

[22] David Millen and John Patterson. Identity disclosure and the creation of social capital. In *Proceedings of CHI 2003*, 2003.

[23] Joseph Pato. Identity management: Setting context. Technical report, Trusted Systems Laboratory, HP Laboratories Cambridge, HPL-2003-72, 2003.

[24] Allen Renear and David Durbin. Toward identity conditions for digital documents. In S. Sutton, editor, *Proceedings of the 2003 Dublin Core Conference*. University of Washington, 2003.

[25] Frederic Stutzman and Terrell Russell. Claimid: A system for personal identity management. In *Proceedings of JCDL 2006*, 2006.

[26] Sherry Turkle. *Life on The Screen*. Simon and Schuster, 1995.

[27] Sherry Turkle. Who am we? *Wired Magazine*, 4(1), January 1996.

[28] David Weinberger. *Small Pieces Loosely Joined: A Unified Theory of the Web*. Perseus Publishing, 2002.

[29] J. Macgregor Wise. Intelligent agency. *Cultural Studies*, 12(3):410–428, 1998.