

January 2011

Making Digital Preservation Practical: A Personal Odyssey

Christopher J. Prom

University of Illinois at Urbana-Champaign

Follow this and additional works at: <https://digitalcommons.kennesaw.edu/provenance>



Part of the [Archival Science Commons](#)

Recommended Citation

Prom, Christopher J., "Making Digital Preservation Practical: A Personal Odyssey," *Provenance, Journal of the Society of Georgia Archivists* 29 no. 1 (2011) .

Available at: <https://digitalcommons.kennesaw.edu/provenance/vol29/iss1/2>

This Article is brought to you for free and open access by DigitalCommons@Kennesaw State University. It has been accepted for inclusion in Provenance, Journal of the Society of Georgia Archivists by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.

Making Digital Preservation Practical: A Personal Odyssey¹

Christopher J. Prom

Thank you for the kind introduction. And thank you all for having me here to speak with you and to learn from you. I am so pleased to be doing so, since the theme of your conference — real world solutions — is near and dear to me. As noted, the theme of my remarks is “Making Digital Preservation Practical.” I will highlight some ways that archives can begin a systematic program to acquire, preserve, and provide access to born-digital materials, by reflecting on my own experiences over the past few years.

Before I begin, I’d like to stress that I am not a digital preservation expert, whatever the term ‘expert’ might mean in this context. That may seem like a strange thing to say given the title of my remarks, but I would like to emphasize that I have no formal training in computer science, digital curation, or a related area. I cared little for computers when I was undertaking my undergraduate work as a philosophy and history major. While completing a history dissertation, I tried to automate my note taking and sorting process, with very limited success. Even though

¹ Keynote address at the 2011 Society of Georgia Archivists annual meeting, held on November 3, 2011 in Morrow, Georgia.

I've done a lot with computers since then, I can say that the more I learn about digital technology, the less I feel like I truly understand it. The ground seems to shift so rapidly beneath our feet.

An incident from my early days as a budding archivist illustrates the limits of my skill. In the summer of 1998, I had just returned to Illinois after completing dissertation research in the United Kingdom. It was a nice trip, and I had gathered quite a bit of research material. I had spent a considerable amount of time tracking down sources from the closets and garden sheds of pensioners, then attempting to convince them to donate them to county record offices. Although I did not know it at the time, I was on my way to becoming an archivist.

Right after my wife and I returned to the States, I began writing up my dissertation, and I soon accepted hourly work in the University of Illinois Archives. Over the years, my part time work led to a full-time position. Knowing nothing about computers, I was given the task of putting our descriptive information online. I charged in where angels dared not fear to tread — and promptly deleted the entire descriptive record of the ALA Archives, representing over 25 years of work! After a half hour of panic, I sheepishly turned to the University Archivist, William Maher and explained the situation. Luckily, we recovered the database, since our Library had a forward-thinking IT manager who guarded against such operator error. I spent the next several weeks putting our other records online. Over the years, that simple project and others like it led Scott Schwartz and me to develop the Archon descriptive software, a product that is now moving toward new life in the ArchivesSpace project.

Why tell this story? Because it cuts close to the theme of this talk, 'Making Digital Preservation Practical.' If someone as error prone as me can learn enough about digital preservation to be make a go of it, anyone can.

Over the years, my dual interest in history and digital technology led me to think that the University of Illinois Archives faced a big problem, born out of our past successes. Over the years, we had developed excellent working relationships around campus, in the process of acquiring traditional paper based archives. As part of this work, we also came to possess of a wide range of digital files. Not knowing what to do with them, our solution was simple: to retain them on their original media and to note the existence of the disk in the finding aid. This resulted in what Ben Goldman has called the ‘disk in a box’ problem, one that I am sure is familiar to many of you.²

About five years ago, we became a bit concerned about this state of affairs. We began to copy the contents of newly accessioned media to a shared drive on our library’s server network. However, we were well aware that we were simply copying the files. They went into a folder labeled ‘Electronic Records,’ and remained inaccessible to our users. Over time, we managed to accession — and I use that term loosely — over one terabyte of born-digital materials, with no real intellectual or physical control over the items.³ We did not know precisely what we were keeping, and we were not managing it for long-term preservation and access. What we needed was a quick and easy way to get these files under control, while building capacity to systematically acquire, describe, and preserve born digital records. Unfortunately, I found few solutions in

² “Guest Post: Ben Goldman,” <http://e-records.chrisprom.com/?p=1993> (checked 29 December 2011)

³ In this situation, we were hardly unique. Forty-five percent of academic research libraries have not even assigned responsibility for the preservation of born digital content to one or more parties in the institution. Jackie Dooley and Katherine Luce, *Taking Our Pulse: the OCLC Research Survey of Special Collections and Archives* (Dublin, OH: OCLC Research, 2010), 57. <http://www.oclc.org/research/publications/library/2010/2010-11.pdf> (Checked 29 December 2011).

my cursory examination of the literature related to electronic records.

In 2008, I was thinking about this problem when presented with a rare opportunity: the chance to take a sabbatical. Noticing that the US-UK Fulbright Program would be open to the type of research I wanted to do, I applied for a fellowship that would support research at the Centre for Archive and Information Studies at the University of Dundee. I was thrilled to find out in April of 2009 that the proposal had been accepted. Today, I'd like to describe two things: What I learned from my research, and how I learned it.

My project began with the goal of developing a method that I, and hopefully others, could use to develop digital preservation capacity, competence, and trust. For me, learning how to 'do' digital preservation has truly been an odyssey, a mixture of the personal and the professional. According to the psychologist Erich Fromm, "The process of learning an art can be divided conveniently into two parts; one, the mastery of theory; the other, the mastery of practice."⁴ Both steps were necessary as I tried to master the art of digital preservation. First, I read digital preservation literature — something I had little time to do as a working archivist. Next, I spent time getting my hands dirty: assessing software tools that could be used to appraise, process, preserve, and provide access to born digital records.

Based on this work, I developed policy templates and software recommendations. These resources are intended to help 'small' archives begin a digital preservation program, using whatever resources they have at hand or can acquire with minimal outlay. They comprise the heart of my practical e-records project, and while I would never suggest that I have mastered the art of digital

⁴ Erich Fromm, *The Art of Loving* (Harper and Row, 1956), 5.

preservation, I do feel as if the project at least helped me be competent in it.

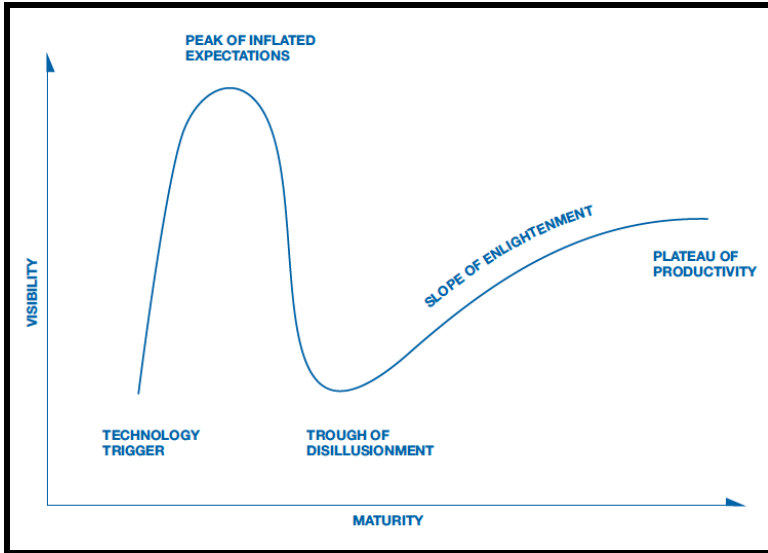


Figure One: Gartner Hype Cycle

In hindsight, I can see that my experience in pursuing this project roughly reflects the typical digital technology adoption process, which is perhaps best represented in the Garnter Hype Cycle (see figure one).⁵ For those of you who are not familiar with it, the Hype Cycle provides a way to understand the lifecycle of transformative technologies. Garnter Research uses it as part of their consulting business, which is to provide technology implementation advice. Today I am using the term in a slightly different way: as a structuring metaphor

⁵ A visual representation of the hype cycles is available at <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp> (Checked 12 December 2011).

to illustrate one way that we might engage with digital preservation activities, both personally and professionally.⁶

The hype cycle model proposes that big changes in an area of practice are initiated by a ‘technology trigger.’ After the initial excitement, problems set in and interest wanes. If, through hard work and some luck, the people developing the technologies begin to climb the slope of enlightenment, the field may develop into a set of mature, productive services. The hype cycle reflects my experience over the course of my sabbatical project. It also, in my opinion, represents the history of the archival profession’s engagement with digital preservation theory and practice.

In my personal case, the opportunity to spend 10 months in Scotland learning from British colleagues served as an effective technology trigger, in helping the University of Illinois Archives to systematically grapple with digital preservation. I could read the digital preservation literature and test software with a level of concentration that would have been impossible to achieve during my usual work schedule.

As I began the Practical E-Records Project, my excitement climbed rapidly. Naturally, I set up a blog to document my experiences. I did not think I had anything all that interesting to say, but I set it up simply to keep myself on track and to organize my thoughts. In the end, I’m glad that I did so. By blogging, I forced myself to actually understand and apply the concepts and tools I was reading about. Without that motivating factor, I’m sure I’d still be spinning my wheels.

My initial activities led rather quickly to what the Garnter Hype Cycle calls the ‘Peak of Inflated Expectations.’ From the lofty heights, I saw the many digital preservation tools, services, and approaches that had been developed over the past 15 years; the possibilities for

⁶ Project recommendations can be found at http://e-records.chrisprom.com/?page_id=508 (Checked 16 December 2011).

preserving digital information seemed endless. It reminded me a bit of the landscape I saw after our family spent the better part of a day climbing in the Scottish highlands: expansive, if a bit remote from my normal experience.

Unfortunately, there was a very dark cloud looming over this pretty landscape, in the form of the seeming technical complexity underlying most approaches to digital preservation. Specifically, the more I looked at *the Reference Model for an Open Archival Information System (OAIS)*, the more confused I seemed to grow.⁷ As I found out later, I was far from the only person to feel this way. As William Kilbride, the Director of the UK's Digital Preservation Coalition likes to joke, the OAIS Reference Model was meant to solve a problem so complicated that NASA had to call in their European buddies for assistance!

In essence, the OAIS Reference Model describes a set of information technology systems, services, and policies that an institution must adopt in order to ensure that the archives is acting as a trusted agent. This means three things: acquiring records in way that preserves their context, storing them in a way makes them authentic, and rendering them in a way that makes them useful. As I puzzled over how the details of the model could be implemented in practice, I came to realize that different parts of an OAIS could be implemented by using some relatively easy-to-use tools and services. However, I found relatively little non-technical guidance as to how these tools could be fitted into a cohesive whole, at least with the types of budget resources available to the typical archive that has cared mainly for paper-based materials. How could the tools be implemented in a reproducible workflow,

⁷ Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS)*, January 2002, <http://public.ccsds.org/publications/archive/650x0b1.pdf> (Checked 29 December 2011).

particularly if one's staff had relatively little advanced training or experience with digital curation technologies?

These are the critical issues facing many repositories. We needed to transform our mission so that we can acquire and manage born-digital resources, even as resources contract. Turning to the profuse digital preservation literature, I perceived a set of complex projects, resources, advice documents, and peer reviewed articles. These sources — each of which was excellent on its own — emanate from such respected sources as the Library of Congress's National Digital Information Infrastructure and Preservation Program, the InterPARES Project, the European Union's PLANETS Project, and those affiliated with those projects.⁸ After reading this literature for much November 2009, I found myself falling deeper and deeper into what the Gartner Hype cycle calls the 'Trough of Disillusionment.' It probably did not help matters that I was finishing up during the dead of the Dundee winter, when the sun rose around 9 am and set about 3:30 pm!

I began to climb out of the trough, into the next part of the Gartner Hype Cycle — the so-called 'Slope of Enlightenment.' Although enlightenment is a good thing, climbing a slope required hard work, which in my case meant practicing digital preservation activities by testing and evaluating software. This exercise was most useful. As Erich Fromm puts it much more eloquently than I: "Thought can lead us only to the knowledge that it cannot give us the ultimate answer. ... The only way in which the

⁸ Library of Congress, "National Digital Information Infrastructure and Preservation Program Website", n.d., <http://www.digitalpreservation.gov/>; InterPARES Project, "Project Website", n.d., <http://www.interpares.org/>; Planets Project, "Planets Project Website", 2010, <http://www.planets-project.eu/> (Checked 29 December 2011).

world can ultimately be grasped lies not in thought, but in the act.”⁹

It is in the actions of retaining evidence, rendering files, and proving authenticity that we understand digital preservation. Completing these actions requires less effort than you might think, in spite of the complexity of the OAIS Reference Model or the diagram that has been developed to represent it.

It took me a long time to figure out that I didn't need to understand or implement the OAIS diagram all at once. At substantial risk of oversimplification, I would even go so far as to say that preserving digital materials really is not that much different than preserving print materials. The trick lies in understanding which tools and services can be used in complete traditional archival functions such as appraisal, identification, arrangement, description, and storage. Once you align sound policies with skilled people and good systems, digital preservation becomes business as usual.¹⁰

Am I making this sound too easy? Perhaps, but I do think that any archivist can undertake a series of relatively simple actions to build digital preservation skills. Let me walk through the process that I used and that I recommend to others.

First, put your own house in order. By gaining control over your own digital files, you will inevitably learn what it takes to manage bigger buckets. In my case, I was forced to clean up my act when I received a notice that our email system was being migrated in several days. As a

⁹ Fromm, *The Art of Loving*, 78–79.

¹⁰ See William Kilbride, “Digital Preservation: What I Wish Someone Had Told Me Before I Started” (presented at the Practical Approaches to Electronic Records: the Academy and Beyond, University of Dundee, May 21, 2010), <http://www.dundee.ac.uk/cais/erm/WilliamKilbride.pdf> (checked 29 December 2011) for a similar perspective.

result, I began researching email preservation options, learning more than I ever wanted to know about how email systems work. Some of my email had been stored in local folders created by Alpine — an old terminal style email application originally developed for a Unix environment. After reading and blogging about all of this, I was able to develop a relatively simple process to move my own email to a more current, preservation-ready format.

The second step up the slope lies in lending a hand to others. This can take several forms. You can help people manage their own records more appropriately, develop guidance documents, set up technologies, or even provide digital legacy planning advice. By taking any of these steps, you will begin to expand the set of tools and services with which you are familiar, building your digital preservation capacity. Helping others leaves you in a position to take the third step: Developing a digital program statement.

By writing such a statement or by adapting an existing one, you will lay a sound foundation for the development of services that acquire and care for electronic materials.¹¹ Developing such a statement will serve several goals. At the most basic level, it will provide you a roadmap, setting out a series of policy and implementation steps that you will undertake over the next few years. Even if you cannot immediately provide all of the services that you specify, the existence of the statement will serve to engender trust among potential donors or other constituents. They will note with pleasure that you are seeking to expand your program by building born-digital collections. In other words, the statement will provide a framework around which you can develop and promote what you do. At a minimum, the statement should include

¹¹ A template statement is available at http://e-records.chrisprom.com/?page_id=540 (checked 29 December 2011).

the following elements: (1) a program mandate; (2) a list of partners; (3) a description of the scope of records to be preserved; and (4) a statement of guiding Values and commitments. Subsequent sections of the program statement (or related documents) can cover additional topics, such as pre-deposit services; acquisition procedures; and methods for processing, describing, storing, and providing access to preserved records.

Once such a policy is in place, you should move to acquire born-digital records, if you haven't already. This is step four in your plan to lead a dynamic, expanding program to digitally document the areas covered by your repository's mission. If you do have records, you should begin working with appropriate tools to undertake some of the preservation actions associated with traditional archival functions, such as processing and storage. Sure, you'll make some mistakes, but if you work with a copy of the original files, you'll save yourself from committing any unpardonable sins.

Probably the most important element in moving up the slope of enlightenment is setting out to become a trusted digital repository (TDR). As you may be aware, those in the digital preservation community have formulated a yardstick by which a repository's trustworthiness can be measured.¹² While your repository may not be able to immediately fulfill the formal criteria, you can work in that direction, using whatever technologies you have at hand.

I am a big believer in using the tools that are available to you. Most repositories already have what they need to set up what I call the Do-it-Yourself Trusted Digital

¹² RLG/OCLC Working Group on Digital Archive Attributes, *Trusted Digital Repositories: Attributes and Responsibilities (pdf)* (Mountain View, CA: Research Libraries Group, 2002), <http://www.oclc.org/research/activities/past/rlg/trustedrep/default.htm> (checked 29 December 2011).

Repository. The concept is described in detail on my blog, but the essential elements are simple to explain.¹³ In essence, by defining a set of local policies and procedures, you can build a method to accession, process, describe, and store records in an archival information packet.

In one of the best descriptions of the OAIS Reference Model, Brian Lavoie offered a graphical representation of the Archival Information Packet (AIP). His schematic is shown in Figure Two.

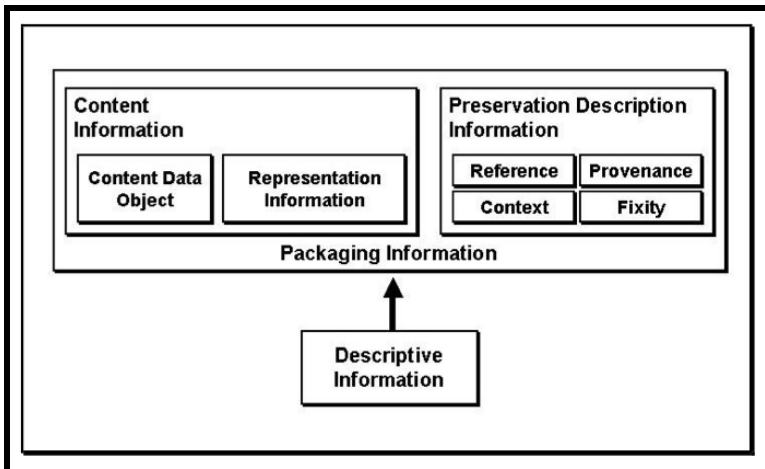


Figure Two: Elements of an Archival Information Packet¹⁴

This diagram tells us that we must keep three buckets of data, if we wish to effectively preserve records. It is not good enough to keep the files themselves (“content

¹³ See http://e-records.chrisprom.com/?page_id=754 (checked 29 December 2011).

¹⁴ Brian Lavoie, *The Open Archival Information System Reference Model: Introductory Guide*, DPC Technology Watch Report 04-01 (London: Digital Preservation Coalition, 2004), 12.

information”), although that is a start. You must also generate and preserve technical information about the files (“preservation description information”) and information explaining the scope and contents of the files (“descriptive information”). When all of this is wrapped together, an Archival Information Packet has been born. The rest of digital preservation work consists simply in keeping that packet alive.

For a long time I puzzled over the OAIS Reference Model diagram, thinking that it would be difficult if not impossible track to the required data for each individual file in a digital collection. One of the objects of my testing work, the files of the American Library Association’s Office of Intellectual Freedom, held over 34,127 files. Thinking archivally, one way to control such a large number of records is to treat them as an accumulation. This is the way we treat the individual documents, photographs, and other records that we keep in record center boxes. Accumulated records are simply groups of records sharing a common relationship to a records creator or a function (a record series), and they can be held in a single archival packet. Treating large groups of records as aggregations makes particular sense for small archives, particularly those wishing to follow a more product, less process model for digital archives.¹⁵

By viewing aggregates as the object of digital preservation, we can overlay our existing tools and services onto the AIP diagram, filling in the framework for a do-it-yourself repository. My attempt to do this is shown in Figure Three.

¹⁵ Mark Greene, “MPLP: It’s Not Just for Processing Anymore,” *The American Archivist* 73, no. 1 (Spring/Summer 2010): 175-203.

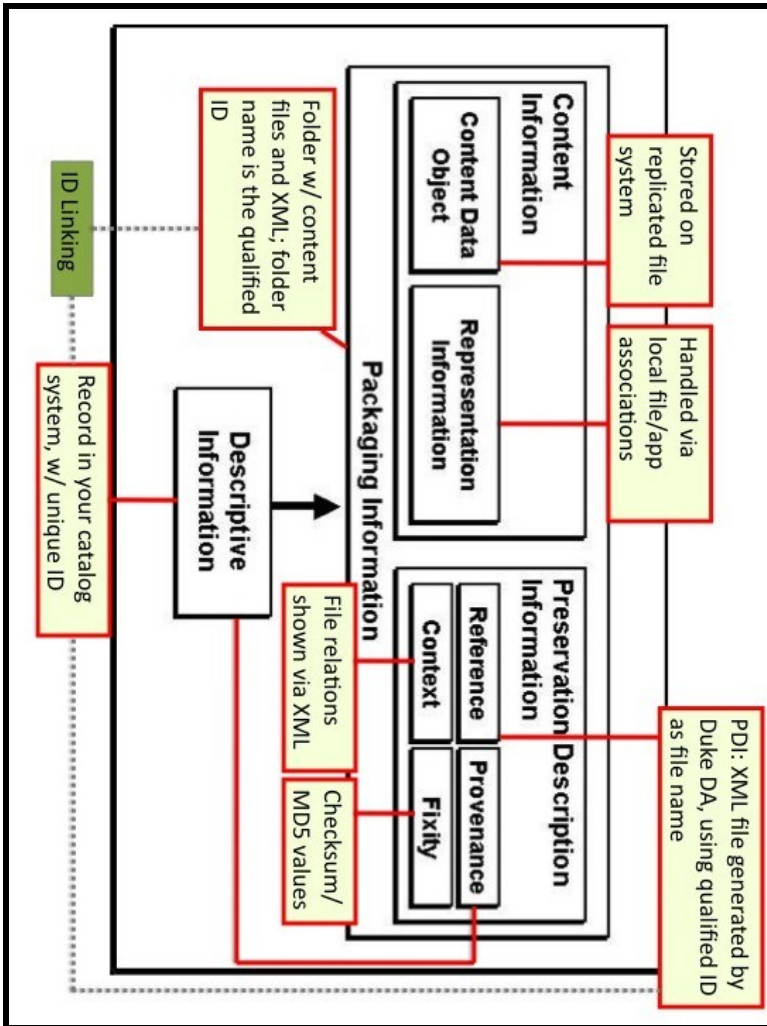


Figure Three: Elements of the AIP in a Do-it-Yourself Repository

Without descending too far into details, I would like to point out several things:

1. The system is in the process of implementation at the University of Illinois, led by my colleague Angela Jordan. While our Library is developing an application for long-term storage of digital objects from the University Archives, the system—code named Medusa—is not yet ready to accept content. In the meantime, the University Archives is able to store all the digital files that we have accessioned in a way that makes them ready for easy transfer into the new system, when it is available.
2. Each element in our system is a software or hardware application that we were already using or which we could implement without any direct help from an IT professional. To track descriptive information, we simply create a record for the Archival Information Packet within our catalog system, Archon. (One could just as easily use the Archivist's Toolkit or another application for this function.) The packet itself is provided a folder name that is the same as the ID of the descriptive record to which it is linked. The packet holds the files we have accessioned and an XML file that is generated by a program developed by Seth Shaw at Duke University, the Data Accessioner.¹⁶ The files themselves, as well as the preservation description information ("PDI") generated by the Data Accessioner, are stored on a replicated file server. Since we do not modify or rearrange the files in the archival packet, their provenance and original order is preserved for posterity.
3. We track file types, making sure that we have software to view or display them in a current

¹⁶ <http://library.duke.edu/uarchives/about/tools/data-accessioner.html> (checked 29 December 2011).

operating system. Where we do not have such software, we readily admit that fact in the descriptive record, providing some indication as to how people might render the files.¹⁷

4. We generate ‘online’ and ‘nearline’ access copies for each archival information packet. We also link these access copies to the descriptive record. They function as what the OAIS reference model calls a Dissemination Information Packet. The online copies are provided in our “E-Records Repository,” via a simple directory-browsing application that we customized for local use.¹⁸ Nearline copies are available by contacting the archives; they can also be provided on USB stick, CD, or other media.

Figure four provides a schematic view of our end-to-end processing, storage, and access workflow that we use under the do-it-yourself repository model; additional details are available on our staff website.¹⁹

While the Gartner Hype Cycle illustrates my personal attempts to grapple with digital preservation literature and methods, I have also come to believe that it is a good metaphor for describing the development of digital preservation as subfield in the archival profession. For example, it is easy to find evidence that members of our profession celebrated the possibilities of digital

¹⁷ This strategy may not be perfect, but it provides what we feel is good enough preservation, relying on the fact that most files have been created or used in readily accessible applications. For the rest, we assume that humans are clever. If we need to get access to an obsolete file, we will locate software from the growing digital preservation community. As needed, we can migrate content to new formats over time.

¹⁸ <http://www.library.illinois.edu/archives/Electronic%20Records/> (checked 29 December 2011).

¹⁹ <http://www.library.illinois.edu/archives/staff/digital/index.php> (checked 29 December 2011).

preservation; those on both ends of the battles over the nature of electronic records work showed such excitement.²⁰ But this high did not last long. Soon, most archivists were plunged into the depths of despair, which lasted for a good part of the first decades of the new millennium. Tools to do effective electronic records were simply not available, leaving most archivists unable to effectively pursue practical approaches to preserving electronic records. It has been a hard climb up the slope of enlightenment, but I do feel as if, professionally, we are now seeing glimpses of what Garnter terms the ‘plateau of productivity.’

Emerging to this location will require teamwork and collaborative leadership. It will require us to nurture partnerships not only in our own institutions, but within the broader digital preservation community. It will require that we experiment with new technologies and services, but in a coordinated way, so that those that truly prove their worth made available to the whole community, in a sustainable fashion.

I hope you do not infer from my somewhat breezy talk today that all of the problems of digital preservation have been solved, or that identifying, preserving, and providing access to electronic records is easy. Recent work that I have been doing with email has convinced me preserving digital information is hard work, but it is possible.²¹

²⁰ David Bearman and Margaret Hedstrom, “Reinventing Archives for Electronic Records: Alternative Service Delivery Options,” in *American Archival Studies: Readings in Theory and Practice* (Society of American Archivists, 2000), 549-567; Linda Henry, “Schellenberg in Cyberspace,” *American Archivist* 61, no. 2 (January 1, 1998): 309-327.

²¹ “Email Preservation Options,” Practical E-Records, November 17, 2011. <http://e-records.chrisprom.com/?p=2351> (checked 29 December 2011).

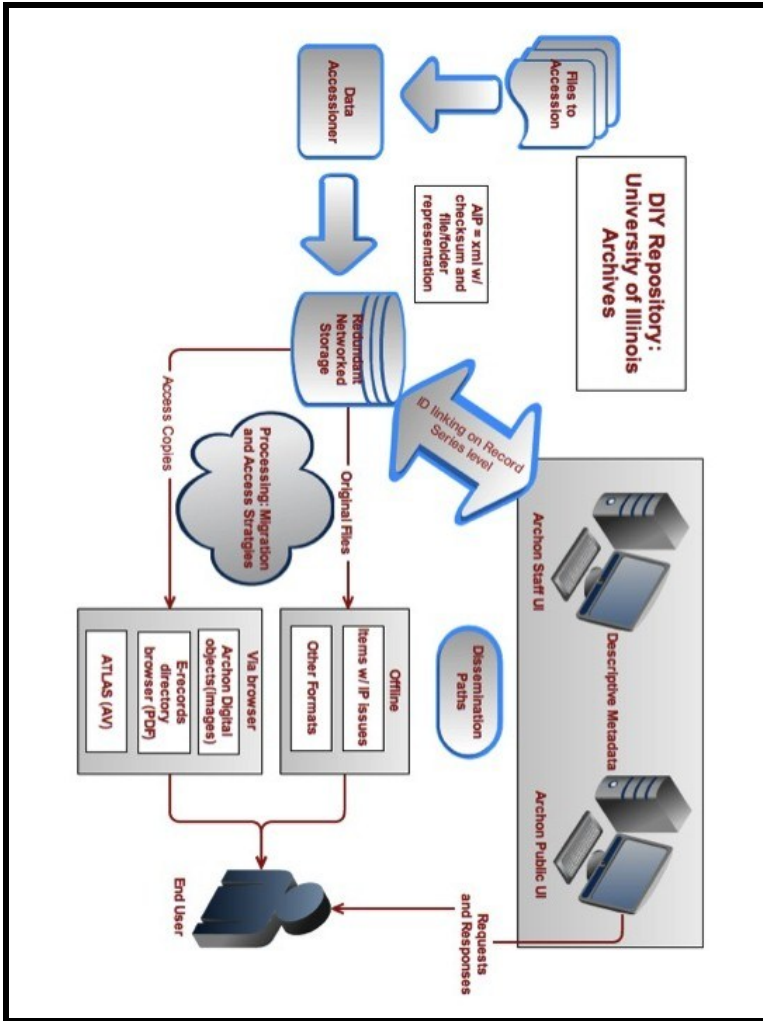


Figure Four: University of Illinois Do-It-Yourself Repository

One final point: I would encourage you to do whatever work you undertake in a way that allows you to experience the Tao of Digital Preservation. The Tao of Digital Preservation is that the nameless state that can only be experienced as a path. It can never be fully grasped; it merges all conflicts and contradictions into its ineffable wholeness. It will require you to be comfortable with the fact that digital objects both exist and don't exist. It will require you to contemplate the problems posed by that issue. It will require you to actively live out solutions, as you cultivate the way. Trust that many others are walking similar paths, and, above all else, know that the work you complete as a digital archivist will touch the lives of many people in the past, in the present, and in the future.

Christopher J. Prom is Assistant University Archivist and Associate Professor of Library Administration at the University of Illinois at Urbana-Champaign. He holds a PhD in history from the University of Illinois and also studied at the University of York (United Kingdom). Chris is a Distinguished Fellow of the Society of American Archivists and has received several other research fellowships including, most recently, a 2009 - 10 Fulbright Distinguished Scholar Award. He maintains the Practical E-Records Blog and is most recently the author of a Digital Preservation Coalition (UK) Technical Watch Report, "Preserving Email."