For the past thirty years, many authorities in the archival community have called for archivists to transform themselves so that they can meet the challenges posed by electronic records. At the same time, many individuals argue that we must remain true to the values and techniques that make us unique among information professionals. For example, the authors of the A*CENSUS survey report—which was a fairly comprehensive survey of US archivists, concluded that two of the biggest tasks that challenged practicing archivists were strengthening technical skills and “[i]dentifying effective methods for transferring the knowledge and values acquired through decades of experience to members of the next generation of archivists.”¹

In practice, it has proven extremely difficult for most archives—aside from well-funded national archives or research libraries, to develop a comprehensive program to identify, preserve, and provide access to electronic records. A forthcoming study of archivists at United States colleges and universities concludes bluntly: “No comprehensive programs exist for managing e-records to use as models for the field.” The authors continue by noting that there is a “[g]eneral lack of interest on the part of

administrators to invest a significant commitment into managing institutional records of any kind,\(^2\) indicating a shocking failure of efforts to advocate for an archives’ basic function.

One can certainly cite many examples of successful pilot programs to manage born-digital information, and some of them even capture the attention of the media, as Emory University recently did with its successful effort to emulate the complete Macintosh desktop environment from one of Salman Rushdie’s computers.\(^3\) Yet, if we are honest with ourselves, many of us would have to quietly admit that too many institutions have followed policies similar to those described by a curator from Harvard University’s Houghton Library, who was quoted in the same article: “We don’t really have any methodology as of yet . . . . We just store the disks in our climate-controlled stacks, and we’re hoping for some kind of universal Harvard guidelines.”

Even though this quote comes from a curator at a decidedly well-off institution, it represents the problem faced by archivists working at what might be termed ‘under-resourced’ institutions. One the one had, many high-profile research projects, such as InterPARES, PLANETS, and the United States Library of Congress National Digital Information Infrastructure Program, have developed frameworks, standards, and tools that can be used to preserve and render records in a way that users can judge their authenticity and integrity. A quick skim of the program for this conference, as well as many others on similar themes, shows the fruits of this labor, in the form of reports.

---


regarding standards development; policy recommendations; local, national or regional software development and implementation projects. On the other hand, relatively little sustained implementation headway has been made in actual collecting and preserving born-digital materials at the vast majority of ‘smaller’ archives, by which I mean those without access to dedicated IT support, and who are represented sparsely at conferences like this one. In the United States, for example, this is true even in cases where much archival material has been identified in college and university institutional repositories applications.4

Why have relatively few smaller archival programs made headway in implementing programs to capture, process, preserve and provide access to digital records? In most cases, archivists lack access to the requisite computing resources (such as dedicated system support budget or a programmer) that are common at large universities or government archives. The resources that they do have available have often been targeted toward attractive digitization projects for valuable analog materials.

It is unrealistic to expect that our institutions will garner significant new funding to address electronic records issues in the period of austerity that we are living through. For these reason, there is a strong need for the community to develop synthesized, relatively easy-to-apply methods that can be used within existing budget lines and funding levels.

---

4 It has been argued, without supporting statistical evidence, that much of the material included in institutional repositories partially replicates the traditional documentary areas covered by archives, but that much of the materials actually included consists of digitized versions of paper-based records, rather the born-digital materials. Elizabeth Yakel et al., “Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change,” American Archivist 71, no. 2 (2008): 346-47.
This means, first and foremost, assessing existing projects and tools developed by and for larger repositories. How fit are they for use in a small archives? Which tools are easy to install, use and implement? Do they support all aspects of work that needs to be completed in a small archives?

Curiously, the published literature is relatively silent on these topics, at least if one is looking for comparative, systematic research studies. Many projects concerning electronic records include a formal assessment, in which the results of the project are measured against an internal set of goals. Such evaluations are useful, but they generally tell us relatively little outside of the context of the local project. In addition, a few projects have been undertaken to formally evaluate a specific genre of software, such as a CLIR-sponsored project to evaluate archival management software. And, there are numerous examples of institutions (usually not archives) evaluating themselves or a piece of software against a set of external criteria, such as the TRAC guidelines or the OAIS requirements. Finally, we have a few examples, such as the excellent but preliminary PLANETS guide to preservation action tools, which assesses specific software products that might be useful for the migration of a few uncommon formats.

---

5 However, the most insightful of these self-assessments, such as Sibyl Schaefer’s evaluation of the Archivist’s Toolkit project, which was recently published in the code{4}lib journal, show that lessons learned in one project can be applied elsewhere. Sibyl Schaefer, “Challenges in Sustainable Open Source: A Case Study,” code{4}lib Journal, no. 9 (March 22, 2010), http://journal.code4lib.org/articles/2493.
7 For an attempt to assess Fedora’s ability to meet the OAIS requirements see “Tufts and Yale: Fedora and the Preservation of University Records Project Reports,” 2006, http://dca.lib.tufts.edu/features/nhprc/reports/index.html As the project proceeded, the principle investigators reoriented the project toward developing a set of requirements for a trustworthy recordkeeping system in the university environment.
What is missing? A comprehensive, high-level assessment of the entire range of approaches, tools, and services that an archivist might be required to use, in order to establish, implement, and sustain a comprehensive program to identify, preserve, and provide access to electronic records.

It is tempting to think that any attempt to develop such a grand assessment would quickly degenerate into a fool’s errand. One might object that there is simply too much work to be done, that no individual is capable of undertaking the task.9 Or you may argue that that any attempt to compare tools or projects side by side would be flawed, since each tool was developed for a specific purpose, and one cannot compare apples to oranges. Or that the evaluation of other people’s software—which they are in many cases making available under an open source license—is ungracious.

While I do not doubt the validity of these objections, I do think there is a role for an individual of moderate technical skills to play in assessing the lay of the land. As part of my project, I informally assessed a variety of software tools that can used to appraise, process, and preserve electronic records. I undertook the assessments from a very specific perspective: that of a lone archivist who has limited access to information technology and/or budgetary support, but who needs to manage records that are about to be deposited at the end of their period of active use.10 During the

---

9 Perhaps such a project would best be conducted formally by many people, as part of a sponsored by a research grant. See “RM Capacity and Compliance toolkits Project Website,” Northumbria University, April 19, 2010, http://www.northumbria.ac.uk/sd/academic/ceis/re/isrc/themes/rmarea/tkit/ for one example. This project may offer a good framework for a formal project to assess electronic records tools.

10 In this sense, I am taking as an assumption the fact that the records being accessioned have not benefitted from much, if any pre-custodial intervention. In essence, the approach I took was to assess methods, tools and procedures that could be used vis-à-vis one simple question: “What would you do if a donor drops some electronic records off at your repository?” By approaching the question in this way, which assumes a traditional lifecycle approach toward records management functions with which many archivists are charged, I hoped that lessons learned could also be applied in cases where records
test, I used two sets of records as working files: the office records of the American Library Association’s Office of Intellectual Freedom and the email correspondence of Nobel prize-winning chemist Paul Lauterbur. As a metric, I used a simple set of software evaluation criteria, measuring the useful of the software’s ease of installation, as well as the quality of its functionality, scalability, sustainability, metadata, documentation, and support.\footnote{Chris Prom, “Software/Tools, Practical E-Records Blog, \url{http://e-records.chrisprom.com/?page_id=175} }

While my assessments certainly are not intended to be the last word regarding any of these projects, it is my hope that they will offer archivists at smaller institutions a rough and ready guide for comparison, using a real world set of problems, rather than a long list of ideal functional requirements.

Over the course of several months, I read project documentation, installed and uninstalled software, and used it to complete tasks like appraising, identifying, sorting and weeding records; migrating files to other formats; characterizing files; extracting metadata; or storing records in repositories. Without divulging full results from the study (which would not be possible in a brief talk, but which will be forthcoming in published form), what have I learned? I’d like to make several points, discussing only a small number of the tools I actually used, as illustrations.

First, and foremost, that there are many, many very useful tools that can be used to establish a program to manage electronic records that are deposited in an archives. Whether an archivist needs to develop a preservation plan, select tools to migrate files

\footnote{Such involvement would, in theory, make the challenges of accessioning, processing, and preserving the records easier than if no such involvement had taken place.}
from one format to another, characterize the nature of files, or store them in a repository, there are many free or paid tools available. In most cases, these tools can be installed with a minimal amount of effort.

However, just because these tools are available does not mean that I found them easy to use. Tools that can be used to appraise records in order to decide which files to deposit in an archives are a good case in point. I found the task of paring my test records down to something that might be considered a Submission Information Packet to be extremely frustrating. In the end, I cobble together an approach to records appraisal using a variety of open source or paid tools, such as specialized file managers and the Danish National Archives SABA Copying program. But none of the tools really allow you to go through and identify records quickly and easily, and in the end I was unable to use them to quickly make appraisal decisions regarding the 31,927 mixed files in 2,172 subfolders that comprised the 25 gigabytes of the Office of Intellectual Freedom ‘accession.’

My experience points to a gaping hole in current software development efforts and to the OAIS conceptual model, which is silent on what Adrian Cunningham has called ‘pre-ingest activities.’ Archivists and records producers need a purpose-built tool for conducting records appraisal and processing actions related to selection. Such tools must ensure that, at the end of the process, we have in hand not just a set of information objects, but a record of events and decisions related to them, so that users can tell why we decided to include or exclude certain records. Until this gap is

rectified, it will be very difficult for archivists to make any headway in dealing with
the reality of the digital lives that records producers are leading, and I have posted a
few thoughts regarding the major features such a tool would include.14

Tools that can be used to identify, characterize, and validate files illustrate a slightly
different issue than those for assembling a submission information packet. Although
they implement the functions that OAIS mandates and that archivists require, they are
difficult to use without extensive tech support. They are meant for integration into
other services, but are very difficult for archivists to use in a stand-alone basis.
JHOVE, for instance, can only be installed and used by someone who has significant
systems administration experience. In addition, many of these tools are very difficult
to use in a batch or bulk fashion. For example, the FITS toolset is a very useful tool
for extracting identification, validation, and characterization metadata from many
common file formats, tying together it together into an xml wrapper. Properly stored,
the FITS output for each file in an Archival Information Packet provides many of the
critical pieces of representation information, such as a checksum value, and
preservation description information that are required to maintain the authenticity of
the file over time. However, FITS is a command line tool and cannot be run in batch
mode without the development of a separate script or user interface.

Regarding tools to undertake file migration: again, many excellent ones exist, but it is
very difficult to use most of them as standalone applications. Imagemagick and Open
Office, for instance, can be used to do batch conversions of image and document files,
respectively. But calling them in batch mode requires use of the command line, and

14 http://e-records.chrisprom.com/?p=1029
both can deliver inconsistent results unless all parameters are correctly configured during the conversion operation.

Xena is a very useful and easy to install tool, developed by the National Archives of Australia, that bundles many of these migration tools into one convenient package. In my experience, it can be easily installed and used on Windows, Mac, and Linux platforms. For most archives, it can be used to convert a wide range of formats to a normalized file type, quickly and easily. However, those using it should be aware that there are many file types it cannot deal with; in particular its support for email, database, and video conversion is not extensive, so additional action will need to be taken in order to convert those file types. Although the developers have supplied an excellent set of instructions for developing plugs into support other files, writing a plug-in would be a job for only the most technically skilled archivists or, more likely, a developer. Xena also encloses each of the converted files in a small xml wrapper, which means that the file can only be viewed in the Xena Viewer application. Although the application works well, and allows for the export of the files outside of the wrapper, the process of using files via the viewer is cumbersome. In addition, some repositories may question the normalization actions that XENA undertakes, and wish to convert the files to another format (for example, PDF-A instead of Open Office Document format.) It should also be noted that, in my testing, Xena occasionally locked up, scuttling the conversion process.15

Regarding storage of Archival Information Packets, the installation and maintenance of a digital repository software, such as that based on Fedora or DSapce, is beyond the

---

capability of most archivists. While such repositories may offer a good place to store records, most need extensive configuration and integration with other tools or services, such as archival descriptive tools. Ultimately, after much trial and effort, I was unable to install working copies of repository software on my laptop. And, of course, such tools would need to be installed in a networked environment, with extensive tech support.

I would like to conclude with what my session title promised: some practical approaches that archivists can and should take to ‘deal with’ electronic materials, based on my experience working with the Office of Intellectual Freedom and Lauterbur Files. These are a set of specific recommendation that, hopefully, an archivist at a repository of any size can use to begin implementing a program for the electronic records preservation.

First and foremost, an archivist should get involved in open source projects which aim to wrap together many of the software tools that are necessary to undertake a range of actions related to electronic records appraisal, migration, storage, and access. This is a new class of applications which is currently becoming available, and the three that are the most likely to be of use to archivists at smaller repositories are the Planets Interoperability Framework, the University of Mihno’s RODA (Repository of Authentic Digital Objects), and the Archivematica project.

Of the three, I have personally chosen to become directly involved with the Archivematica project, since it appears to offer the best chance of meeting the needs of small archives. Essentially, Archivematica is a Linux Ubuntu-based virtual
machine, which the developers eventually hope to make available as a Linux
distribution. As such, it can be installed easily inside an existing operating system or
as the base operating system on a computer. It is being designed so that it can be
connected to external repository applications, simple file-based storage mechanisms,
and a variety of archival description tools, such as ICA-AtoM or any application
which can accept files via an http POST operation. Although the software is currently
available only in proof of concept form, it is being developed in a very inclusive, open
fashion, and it is easy to install, use, and modify. Software like this can benefit from
extensive use and testing during its development cycle, and Artefactual Systems,
which developing is developing the software for a group of clients, has provided the
right mix of tools to allow for such contributions from the community.\textsuperscript{16}

In the meantime, I have provided a series of recommendation and policy guidelines
that archivists can use to gradually implement an electronic records policy using
whatever resources they have available in their local environment. While any system
set up based on this advice would be far from perfect in fulfilling all of the
requirements of the OAIS reference model or the Trusted Repositories Audit
Certification Guidelines, it is my feeling that it will allow under-resourced
repositories to take a significant step in the right direction, while doing no harm to the
files that they are accessioning, processing, storing, and making available.

My recommendations involve seven steps, for which I provide templates and
guidance, with links to external resources:

\textsuperscript{16}The project documentation and communication mechanisms are some of the most effective I have
seen for a project in such an early phase of its development. Interested parties can participate at
• Assess available resources (e.g. staff, technology, IT support, budget).
• Develop an electronic records program statement.
• Develop policies and procedures concerning submission of content.
• Begin implementation of a trustworthy digital repository.
• Develop preservation and access action plans.
• Develop processing and preservation, and storage, workflows.
• Develop an access method.

For most of these, I list specific steps that can be taken to achieve the goal. For example, I provide a method by which a small repository can be building a trusted digital repository using locally available tools, that conforms in general outline to OAIS Reference Model requirements.\(^{17}\)

In the end, my Fulbright project demonstrated how difficult it is to actually implement an electronic records program without the assistance of many individuals and groups. Yet I’ve learned that one of the best aspects of our profession is that so many members of it freely share ideas, software, and methods in an open forum. As a result, I am confident that nearly any archival repository can begin making significant headway in implementing a program to manage electronic records, using existing resources—**if they make the development of programs a priority**. It is easy to say that it is incumbent upon software developers to ensure that tools are available, accessible, and extensively tested by archivists at repositories of all shapes, sizes, and budget. But it is equally incumbent upon archivists at smaller repositories to involve themselves in the nitty-gritty of product design, implementation, and testing. If they do, that will find that there are many practical ways to begin effectively managing and providing access to born-digital materials.