Research Prospectus

Fostering Archival Connections: Project Rationale and Goals

Archivists draw upon a rich history of theories, practices, and techniques, which we might (oversimplifying matters considerably) term the archival model. Over the past century, the archival model has been codified and implemented through professional practices, institutional structures, standards development, and even software designs. Generally speaking, this model seeks to identify, preserve, and make accessible the value of records as information and evidence by describing them and by preserving information regarding the contexts under which they were created, managed, used, transmitted, and held. 1

The archival model posits that records are first and foremost traces of human activity. Records are left behind as people play a particular role or function in an organization or social group, and as such they are valuable not only for the information they contain, but for what they say about the activities of those who created, accessed, or used them. Archives allow us to answer questions such as that posed in 1973 by Howard Baker, “What did the President know, and when did he know it?”

While the methods that archivists use to manage these traces of human activity were originally developed as a means to cope with analog materials, they are being adapted to the digital environment. For example, digital preservation standards and practices, such as the Open Archival Information System (OAIS) Reference Model and Chain of Preservation Model, are based on core archival concepts as well as those developed by related communities such as library science, record keeping, computer science, and informatics. 2 Accordingly, the software tools that archivists and librarians use tend to implement these abstract standards in a way that mirrors their embedded assumptions and data models. 3

But are those tools in sync with the real-world needs of archivists, librarians, and users? This is a concern, because most ‘born-digital’ records are created in networked environments and are transmitted with embedded metadata that relates them to other records, people, and activities. And if the tools do not reflect and preserve both the records and the complex relationships reflecting their creation and use, what can be done to help archivists and librarians implement a more cohesive set of archival services? These are questions that should be grappled with in every library and archives, and I propose to use the resources afforded by the Andrew Turyn Professorship to deepen the University of Illinois Library’s involvement in projects that are seeking to answer them. 4

Specifically, I propose to lead a collaborative research and implementation program taking the theme of “Fostering Archival Connections.” The proposed research comprises three specific elements:

Evaluating, testing, and implementing the wide range of methods, data models, and tools that can help archivists and librarians complete the essential tasks of archival acquisition, arrangement, description, and access for born-digital materials;

Fostering collaborative research by archivists, functional experts, subject librarians, and teaching faculty at Illinois, to inform the design and implementation of the emergent tools; and

Increasing collaboration and sharing between the many external standards groups and technology development projects that seek to cope with the advent of ‘cloud-based’ records. 5
A key feature of this project lies in its collaborative nature, as I work closely with faculty and staff at the University of Illinois, as well as with carefully selected external partners. Overall, the project’s goal is to facilitate an improved means of preserving and making available ‘born-digital’ records, both at Illinois and elsewhere, whether those are handled by the University Archives or other Library services, such as the Illinois Data Bank, subject libraries, or IDEALS. One core element will be to develop and implement an evaluative framework to assess and improve the tools and services that archivists and librarians can use to identify, arrange, describe, store, and disseminate born-digital content of long-term research value. Based on the assessment, we will attempt to shape and influence the tools in such a way that they can interact with new archival data models that take better advantage of embedded metadata and provenance information. Not only will completing this work help Illinois to mainstream the preservation and access of born-digital materials in the University of Illinois Library’s Medusa repository but—since the project is highly collaborative in nature—it will enhance the capacity for other faculty and academic professionals to play national and international leadership roles, broadening Illinois’ overall presence in the fields of archives, digital humanities, data curation, scholarly communication, metadata, discovery, and delivery.

The project has a secondary, but equally important, goal: to contribute positive benefits to the development processes for standards, tools, and services that are being built elsewhere, but relied upon by Illinois and other archives/libraries. This will help the institutions building those tools and standards to develop them in a fashion that makes it easier for a wide range of repositories to select and use them in consonance with institutional needs and capacities. By publishing our findings and by traveling to engage directly with external groups mentioned in footnotes four and five (with the support of the Turyn Professorship), I will communicate project findings and influence the development trajectories of these and other projects.

Proposed Outline of Work

Support for the Turyn Professorship will be used to pay for documented leave time (with graduate hourly support dedicated to archives), for travel, and for graduate hourly support to pre-screen literature, to compile bibliographies, and test tools such as those mentioned above; the hourly assistant will typically be an archives GA working with our digital collections, to ensure cross-fertilization of theory and practice. At this time I envision the following timeline and outcomes:

Spring/Summer 2016 Literature review regarding “Arrangement and Description in the Cloud/Current Tools and Standards for Archival Work;” relaunch Practical E-Records Blog.

Fall 2016 Convene collaborators group; develop initial evaluative framework, begin tool testing and local implementation; publish literature review and issues analysis articles.

2017 Apply evaluative framework to range of tools available for description, transformation and preservation; contribute to external tool development and publish article about evaluation framework development and initial testing.
2018-19 Implement selected tools and services in conjunction with Library faculty and Staff; test and refine tools for potential integration with Medusa repository; continue publishing articles.

2020-21 Continue testing, implementation and publication/contribution to standards development; author book related to digital archives (working Title Gathering and Maintaining Digital Archives).

Publication and Dissemination Plans

The work described above would be shared initially via my Practical E-Records blog (http://e-records.chrisprom.com) and formalized into a series of research articles and case studies, consisting mostly of co-authored pieces, placed strategically in a wide range of journals. To ensure that these recommendations are as widely discussed (and hopefully, implemented), these articles would be targeted not just to the standard archival journals, but also to a range of top-tier publications in the library and information science community, as well as a capstone book to be proposed to the Society of American Archivists Press.

Notes:


- For example, the user interface for ArchivesSpace (an archival descriptive tool) mirrors the structural model of the technical standard for encoding archival descriptive information (Encoded Archival Description), with some intermixing of concepts from BIBFRAME (i.e. the instance model). Similarly, the digital preservation tools Archivematica and Preservica presuppose a linear workflow, mirroring the OAIS reference model. On the Library side, the assumptions embedded in access systems like contentDM require the recording of item-level metadata—conflicting with the practices of some digital photographers and archivists seeking to manage large photograph collections as aggregations. ↩

- A very partial list of such projects includes iTrust, the Experts Group on Archival Description, and ArchivesSpace, imProvenance Study Group, ePADD, Social Feed Manager, and Archivematica. ↩

- Specifically, I would focus much work through the Experts Group on Archival Description (EGAD) and an effort to develop implementations of archival description that make use of the WC3 PROV-O recommendation. Prov-O is an ontology for recording information about entities, activities, and people involved in producing a data object, in order to make a judgement of its reliability or authenticity. https://www.w3.org/TR/prov-o/. Collaborators working with the DataONE project (https://www.dataone.org/) have made significant progress in applying this datasets, and this work has been suggested by Vicki Lemieux as a model for the generation of archival descriptions. ↩

- The following Univeristy of Illinois faculty member and academic professionals have agreed to work with me as collaborators, if I am chosen for the Turyn Professorship: Bethany Anderson, Harriet Green,
- The evaluation framework will draw initially on methods suggested in Dave Collins, Designing Object-Oriented User Interfaces (Benjamin/Cummings Publishing: Redwood City, CA: 1995). This classic text provides a cognitive-psychology approach to understanding whether the deep structures and user interfaces underpinning a system match users’ real-world mental models (of objects, relationships, and actions). The framework will be revised as additional methods surface during the literature review.

- Formats like email, social media, online photographs, websites, and reports often contain a wealth of embedded metadata, but few archival systems take advantage of it. As Vicki Lemieux noted in a recent PROV-O working group call, archival description is very labor intensive and current practices are unsustainable. (https://osf.io/kyf5q/, currently restricted to project partners.)