Archival description in OAI-ORE

Deborah Kaplan, Anne Sauer, Eliot Wilczek

Introduction

This paper seeks to define a new method for representing and managing description of archival collections using OAI-ORE. This new method has two advantages. Firstly, it adapts traditional archival description methods for the contemporary reality that digital collections, unlike collections of physical materials, are not best described by physical location. Secondly, it takes advantage of the power of OAI-ORE to allow for a multitude of non-linear relationships, providing richer and more powerful access and description.

Archives and Finding Aids

Archival collections are composed of aggregations of interrelated materials. These materials document the activities of individuals or organizations. Unlike library collections, where holdings are generally of discrete items that can stand on their own, the significance of archival collections lies in their aggregate, or collective, nature. Archivists work to preserve *original order* of the documents within a collection because the physical arrangement and context of the documents before archival deposit are as valuable as the information they contain. The archival finding aid is the tool that was developed to provide archivists with intellectual and physical control of their holdings, and users with the means to discover documents within collections and understand the way documents relate to each other.

Traditionally, the finding aid is a hierarchical and linear narrative document. It begins at a high level, describing the collection as a whole, its creator(s), and how the collection is organized. From there, subgroupings of records, or series, may be described, followed by a container list of boxes and perhaps the folders or items contained within a box. The linear flow of the traditional finding aid closely mirrors the physical arrangement of the documents in hand, serving as a description of the collection and a map to where records are physically located within the actual shelves, boxes, and files. Although the EAD (Encoded Archival Description) standard enables archivists to encode and electronically share finding aids, it does not fundamentally transform archival description. It provides a standardized data structure for the same linear narrative of the traditional paper-based finding aid.

The Limits of Finding Aids

Contemporary recordkeeping practices, however, challenge archivists' ability to maintain original order as the basis for describing archival collections. Although original order is still vitally important, it is often complex and multifaceted in institutions with dynamic organizational structures, business functions, and recordkeeping systems that operate in digital environments. The difficulty of archives trying to document undergraduate courses illustrates this challenge.

For example, imagine a course called "Senior Seminar in Widgets, Special Topics, EAS 421" taught by Professor Irene Adler, in the Department of Widgets. A variety of records document this course. These records, however, are distributed across several archival collections. The course syllabus and assignments are managed in the archival

collection that holds the records of the university's Learning Management System. The course description is managed in the archival collection that holds departmental records for the Department of Widgets. The lecture notes for the class are held in the manuscript collection comprising Professor Adler's papers, which she has donated to the university. (Professor Emily Pollifax teaches the same class in alternate years. Her lecture notes will be in the manuscript collection holding Professor Pollifax's papers.) Student theses for this course—20-35 page papers—are held in an archival collection of student theses from across university departments.

Using traditional archival description modes, there is no way to describe the class EAS 421. Indeed, teaching is a notoriously under-documented function in university archives, in part because courses cannot be properly documented in a single archival collection, as the records that document a single course are usually created by multiple record creators. In order to produce a coherent documentary representation of EAS 421, the archives needs a mechanism that can string together items from five collections. Using OAI-ORE, individual elements of each of these five traditional archival and manuscript collections can be linked by their relationships to one another. A user-friendly finding aid could then easily be produced which would depict EAS 421 as if it were a coherent collection—which, as far as the end user is concerned, it is.

Provenance, Relationships, and Digital Realities

This is not merely a matter of producing on-the-fly theme-based collections for the convenience of researchers. This goes to the heart of addressing the limitations of provenance-based arrangement and description. Make no mistake; provenance is vital. The *Glossary of Archival and Records Terminology* notes that the "principle of provenance or the *respect des fonds* dictates that records of different origins (provenance) be kept separate to preserve their context." This is a sound principle; archives preserve records to document the activities of records creators and the individuals, organizations, and societies that interact with the records creators. The context of who originally created these records and how they managed them is just as critical and revealing as the content contained by these records.

The weakness comes not in the principle of provenance, but in its literal application. Archives arrange records into a record group based on the records' creator. The finding aid serves as an internal tool for the archivist and as a research tool for the user. However, the success of this strategy depends on the archival records having a rigid one-to-one relationship with a creator that can only be understood in a single way. In reality, this is rarely the case. The traditional finding aid is not equipped to describe complex relationships between records and a multitude of creators. In addition, traditional archival arrangement and description strategy is not good at documenting activities whose records are spread across multiple record groups in the archives. Professor Adler's lecture notes document the activities of Professor Adler, but they also document the EAS 421 course. A traditional finding aid, even one encoded in EAD, is not equipped to represent this second documentary reality of Professor Adler's lecture notes.

In addition, the nature of electronic records and digital objects adds another layer of complexity to archival description and the notion of original order. "Filing" in the digital world is not an electronic equivalent of paper filing. A single digital document may exist in multiple contexts: on the network, a draft on the desktop, attached to an email, posted to a wiki, as well as being printed and filed. The flexibility of the digital environment allows

people to manage their files using search tools and tagging rather than organizing their files into a particular arrangement.

Moreover, this exposes the significant shortcomings of using the physical arrangement of documents as the basis for describing the structure and organization of digitized and born-digital documents. What "location" means for digital records is difficult to define. The location of servers in a data center, bits on a disk platter, or files on a filesystem is not relevant for discovery and retrieval; a persistent unique identifier, such as a handle, provides all of the organizational information necessary for retrieval. Instead, a meaningful original order should represent how that document functioned intellectually in the creator's world. This requires detaching original order from physical location.

New Ways of Thinking

New approaches should add to, but not undermine, the fundamentals of archival theory. Provenance and original order continue to have primary importance to understanding records. However, the reality of modern records creation is that even in their active use records may exist in multiple contexts and have multiple relationships that describe their significance and value. Our descriptive tools should have the flexibility and power to reflect this instead of forcing us to present multifaceted records in a single hierarchical arrangement.

More radically, for many users, provenance and original order will not have particular significance. Secondary or even tertiary relationships may be where their interest lies, and our access tools should enable multiple avenues of discovery.

In Practice: Display of OAI-ORE finding aids in Fedora Commons

The Tufts University Digital Library currently displays HTML "collection guides" which are transformed from EAD finding aids encoded in XML. Our Fedora content model has a disseminator that displays the metadata for these collection guides, and offers the finding aid transformed by XSLT into chunks.

We would continue to provide this functionality if we encoded our finding aids using OAI-ORE. Since it would be trivial to write a crosswalk which would convert an OAI-ORE collection description into a less informative EAD finding aid, we could produce instances of our finding aid in EAD and use our existing content model to continue to display those finding aids in our digital library. Additionally, we could produce a multitude of traditional-seeming finding aids in EAD, simply by slicing and dicing the OAI-ORE encoded information along different lines. Although encoded as if they were traditional EAD documents, each of these finding aids would presentmeaningful arrangements rather than reproduce the physical arrangement of archival records.

However, we could also provide additional rich and flexible visualizations of our collections using OAI-ORE finding aids. We would not be restricted to the old linear view inspired by the paper finding aid. For example, by leveraging the granular flexibility of OAI-ORE, we could provide a visualization of the portions of the five archival collections that document EAS 421 thereby producing a virtual collection centered on the course. We could deliver this flexibly assembled archival description by having a Fedora disseminator automatically build a Visual Understanding Environment (VUE) map displaying the rich set of relationships in a OAI-ORE finding aid, allowing the end-user to see the relationships between different collections, record creators, records, business functions, and

recordkeeping systems.

Conclusion

A schema for representing finding aids in OAI-ORE would allow richer methods for modeling archival collection descriptions. In conjunction with an XSLT to create EAD from the OAI-ORE, archives could switch to modeling their collections in OAI-ORE immediately. For all existing tools, a transformation to EAD would represent the data in a way that pre-existing collection guide tools expect. Augmenting existing collection description methods with OAI-ORE would revolutionize the capabilities of archives. Tufts DCA is applying for an NHPRC grant to investigate this process further. Working with an advisory board of archival description experts, we will scope out the detailed use cases for the modified collection description.