Digitization has bestowed upon librarians and archivists of the late 20th and early 21st centuries the opportunity to reexamine how they access their collections. It draws these two traditional groups together with IT specialists in order to collaborate on this new great challenge. In this paper, the authors offer a strategy for adapting a library system to traditional archival practice.

The librarian and the archivist . . . both collect, preserve, and make accessible materials for research; but significant differences exist in the way these materials are arranged, described, and used. Among the items usually collected by libraries are: published books and serials, and in more recent times, commercially available sound recordings, films, videos, and electronic resources of various types. Archives, on the other hand, tend to collect original records of an organization, unique personal papers, as well as other effects of individuals and families. Each type of institution, given its particular emphasis, has its own traditions and its own methods of dealing with its collections.

Most mid- to large-sized automated libraries in the United States and abroad use Machine Readable Cataloging (MARC) records to form the basis of their online catalogs. Bibliographic records, including those in the MARC format, generally represent an individually published item, or “information product,” and describe the physical characteristics of the item itself. The basic unit of archival description, however, is a much more complex entity than the basic unit of bibliographic description and often involves multiple hierarchical levels that may or may not extend down to the level of individual items. At Portland State University (PSU) the authors examined whether the capabilities of their present integrated library system could be expanded to capture the hierarchical structure of traditional archival finding aids.

Background

As early as 1841, the cataloging rules established by Panizzi were geared toward locating individual published items. Panizzi based his rules on the idea that any person looking for any particular book should be able to find it through the catalog. This tradition has continued over time up through current standards such as the Anglo-American Cataloguing Rules and reaffirmed in MARC, the standard for the representation and exchange of bibliographic information that has been widely used by libraries for over thirty years.

Archival description, on the other hand, is generally based on the fonds, that is, the entire collection of materials in any medium that were created, accumulated, and used by a particular person, family, or organization in the course of that creator’s activities and functions. Thus, the basic unit of archival description, usually a finding aid, is a much more complex entity than the basic unit of bibliographic description, often involving multiple hierarchical levels of description that may or may not extend down to the level of individual items.

Before archival description begins, the archivist identifies related groups of materials and determines their proper arrangement. Once the arrangement is determined, then the description of the materials reflects both their provenance and their original order. The first explicit statement of the levels of arrangement in an archival collection was by Holmes and has since been elevated to the level of dogma in the archival community. A more recent statement in Describing Archives: A Content Standard (DACS) indicates that the actual levels of arrangement may differ for each collection.

By custom, archivists have assigned names to some, but not all, levels of arrangement. The most commonly identified are collection, record group, series, file (or filing unit), and item. A large or complex body of material may have many more levels. The archivist must determine for practical reasons which groupings will be treated as a unit for purposes of description.

Rephrasing Holmes, the five levels of arrangement can be defined as:

1. The collection level which Holmes called the depository level—the breakdown of the depository’s complete holdings into a few major divisions based on the broadest common denominator
2. The record group level—the fonds or complete collection of the papers of a particular administrative division or branch of an organization or of a particular individual or family
3. The series level—the breakdown of the record group into natural series and the arrangement of each series with respect to the others
4. The filing unit level—the breakdown of each series into unit components, which are usually fairly obvious if the documents are kept in file folders
5. The document level—the level of individual items

Michaela Brenner, Tom Larsen, and Claudia Weston

Michaela Brenner (brennerm@pdx.edu) and Tom Larsen (larsent@pdx.edu) are Database Maintenance and Catalog Librarians, and Claudia Weston (westonc@pdx.edu) is Assistant University Librarian for Technical Services, Portland State University.
The end result of archival description is usually a finding aid that ideally presents an accurate representation of the items in an archival collection so that users can, as independently as possible, locate them. Building on the print finding aid, the archival community has explored a number of mechanisms for disseminating information on the availability of items in their collections. In 1983, the USMARC Format for Archival and Manuscript Control (MARC-AMC) was released and subsequently sanctioned for use as one possible standard data structure and communication protocol in the SAA descriptive standard Archives, Personal Papers, and Manuscripts (APPM) and its successor, DACS. Its adoption, however, has been somewhat controversial among archivists.

The difficulty in capturing the hierarchical nature of collections through the MARC format is one factor that has limited the use of MARC by the archival community. While it is possible to encode this hierarchical description in MARC using notes and linking fields, few archivists in practice have actually made use of these linking fields. Thus, in archival cataloging, MARC records have been used primarily for collection-level description, allowing users to search and discover only general information about archival collections in online catalogs while the finding aid has remained the primary tool for detailed data at all levels of description.

In 1995, the Encoded Archival Description (EAD) emerged as a new standard for encoding descriptions of archival collections. The EAD standard, like the MARC standard, allows for the electronic storage and exchange of archival information; but unlike MARC, it is based on the finding aid. EAD is well suited for encoding the hierarchical relationships between the different parts of the collection and displaying them to the user, and it has become more widely adopted by the archival community.

As outlined, the standards and systems chosen by an institution are dictated by the needs and traditions of that institution. The archival community relies heavily on finding aids and, with increasing frequency, on EAD, their electronic extension; whereas the library community heavily relies on the Online Public Access Catalog (OPAC) and MARC records. New trends capitalizing on the strengths of both traditions are evolving as libraries and archives seek ways to improve access to their archival and digital collections.

### Access to digital archival collections in libraries

When searching the Web for collections of information, one frequently encounters separate interfaces for traditional library, archival, and digital collections even though these collections may be owned, sponsored, hosted, or licensed by a single institution. Descriptive records for traditional library materials reside in the OPAC and are constructed according to standard library practice, while finding aids for the archival and digital collections increasingly appear on specially designed Web sites. This, of course, means that users searching the OPAC may miss relevant materials that are described only in the archival and digital documents database or Web site. Similarly, users searching the archival and digital documents database or Web site may miss relevant materials that are described only in the OPAC.

In other instances, libraries, such as the Library of Congress, selectively add records to their OPACs for individual items in their archival and digital collections. This incorporation allows users more complete access to items within the library's collections. Authority control and the assignment of descriptors further enhance access to the item-level records. To minimize processing costs, however, libraries frequently create brief descriptive records for items, thereby limiting their value to patrons. By creating descriptive records for the items only, libraries also obscure the hierarchical relationships among the items and the collections in which they reside. These relationships can provide the user with a useful context for the individual items and are an essential part of archival description.

Still other libraries, such as the University of Washington, include collection-level MARC records in the OPAC for their archival and digital document collections. These are searchable in the OPAC in the same way as bibliographic records for other materials. These collection-level records can then in turn be linked to finding aids that describe the collections more fully. Collection-level records often are used in libraries where library resources may be insufficient for cataloging large collections of materials at the item level. The guidelines for collection-level records in APPM and DACS, however, allow for additional fields that are not ordinarily used in library bibliographic records. These include such things as descriptions of the organization and arrangement of the collection, citations for published descriptions of the collection and links to the finding aid, and acknowledgment of the donors, as well as ample subject access to the collection. Despite their potential for detail, collection-level records cannot provide the same degree of access to individual items as full item-level records.

### An approach taken at Portland State University Library

In many ways, archival and digital-document collections are continuing resources. A continuing resource is defined as “...a bibliographic resource that is issued over time...
with no predetermined conclusion. Continuing resources include serials and ongoing integrating resources.16

Like published continuing resources, archival and digital collections generally are created over time with no predetermined conclusion. In fact, some archival collections continue to grow even after part of the collection has been accessioned by a library or archive. Thus, even though many of the individual items in the collection might be properly treated as monographic (not unlike serial analytics), it would not be unreasonable to treat the entire collection as a continuing resource.

With this in mind, the authors examined whether their electronic-resource management system could be adapted to accommodate evolving collections of digitized and born-digital material. More specifically, the present system was examined to determine whether its capabilities could be expanded to capture the hierarchical structure found in traditional archival finding aids.

The electronic resource management system in use by PSU Library is Innovative Interfaces’ Electronic Resource Management (ERM) product. According to Innovative Interfaces Inc.’s (III) marketing literature, “[ERM] effectively controls subscription and licensing information for licensed resources such as e-journals, Abstracting and Indexing (A&I) databases, and full-text databases.”7 To control and provide improved access to these resources, ERM stores details about purchase orders, aggregators and publishers, subscription terms, licensing conditions, breadth of holdings, internal and external contact information, and other aspects of these resources that individual libraries consider relevant. For increased security and data integrity, multilevel permissions restrict viewing and editing of data to the appropriate level of staff or patron.

The ability of ERM to replicate the two-level hierarchical relationships between aggregators or publishers and the electronic and print resources they provide was of particular interest to the authors. Through ERM and III’s batch record load capabilities, bibliographic and resource records can be loaded into the III system using delimited source files such as those provided by Serials Solutions. Resource records are the mechanisms used by III to describe digital resources at a collection, subcollection, or title level, thereby enabling the capture of descriptive information not permitted by standard bibliographic records. III uses holdings records to document serial holdings statements. According to the MARC 21 Formats for Holdings Data, a holdings statement is the “record of the location(s) and bibliographic units of a specific bibliographic item held at one or more locations.”18 III holdings records may also contain a URL for connecting to an electronic resource. In figure 1, for example, the resource record shows that PSU Library provides limited access to a number of journal titles through its Springer Journals Online resource.

As seen in figure 2, the display of a holdings record embedded in a bibliographic record provides more specific information on the availability of a title through the library’s collection. In this particular example, the information display reveals that print volumes are available for this title but that PSU only has this title available as a part of the Springer-Verlag electronic collection accessible by clicking on the hotlink. More information on the Springer collection can be discovered by clicking on the About Resource button to retrieve the Springer Journals Online resource record. This example, then, represents a two-level hierarchy where the resource Springer Journals Online is analogous to an archival collection and Abdominal Imaging is analogous to an archival series.

Adaptation of ERM for library-created digital collections was explored through work being done to fulfill the requirements of a grant received in 2005 by PSU Library. The goal of this grant was “to develop a digital library under the sponsorship of the Portland State University Library to serve as a central repository for the collection, accession, and dissemination of key planning documents and reports, maps, and other ephemeral materials that have high value for Oregon citizens and for scholars around the world.”19 The overall collection is called the Oregon Sustainable Community Digital Library (OSCDL).

In addition to having its own Web site, it was decided to make this collection accessible through the PSU Library catalog so that patrons could find digitized original documents about the city of Portland together with other library materials. Bibliographic records would be added to the database with hyperlinks to the digitized original documents using existing staff and tools. These bibliographic MARC records would be as complete as possible.

Initially, attention was focused on documents originating from four different sources: Ernest Bonner, a former Portland city planner; the city of Portland archives; Metro (the regional government for the Portland, Oregon, metropolitan area); and Trimet (the Portland metropolitan public transportation system). Along with the documents, metadata was received from various databases. These descriptions ranged from almost nothing to detailed archival descriptions.

Unlike the challenge of shifting titles and holdings with typical serials collections, the challenge of this project was to reflect the four hierarchical levels of PSU Library’s collection (figure 3). Innovative’s system structure was manipulated in order to accomplish this.

At the core of III’s ERM module are resource records (RR) created to reflect the peculiarities of a particular collection. Linked to these resource records are holdings records (HR) containing hyperlinks to the actual digitized documents (Doc H1 – Doc H3) as well as to their respective bibliographic records (BIB Doc H1 – BIB Doc H3) containing additional information on the individual items within the collection (figure 4).
First, resource records were manually created for three of the subcollections within the Bonner collection. These subcollections contained documents reflecting the development of Harbor Drive, Front Street, and the Park Blocks. The fields defined for the resource records include the resource title; type (digitized documents) and format (PDF) of the resource; a hyperlink to the new OSCDL Web site; content and systems contact names; a brief description of the resource; and, most importantly, the Resource ID used to connect holding records for individual documents to the corresponding resource record.

Next, the batch-loading function in ERM was used to create bibliographic and holding records and associate them with the resource records. Taking advantage of tracking data produced during the digitization process (figure 5), spreadsheets were created for each collection reflecting the data assigned to each individual digitized document. The document title, the date the document was created, number of pages, and summaries were included. Coordinates for the streets mentioned in the documents were also included. Because ERM uses ISSN numbers and titles as match points for record loads, “ISSN” numbers were also manufactured for each document and included in the spreadsheet. These homemade numbers were distinguished by using pdx as a prefix followed by collection and document numbers or letters, for example, pdx002090 or pdxhdcoll. Fortunately, ERM accepted these dummy ISSNs (figure 6).

From this data spreadsheet, the system-required comma delimited coverage load file (*.csv) was also created. For this file, the system only allows a limited number of fields, and is very particular about the right terms, including correct capitalization, for the header row. Individual document titles, the made-up ISSN numbers, individual URLs to the documents, and a collection-specific resource ID (Provider) that connects all the documents from a collection to their respective resource record were included. The resource ID is the same for all documents in one collection (figure 7).

In the first attempt, the system was set up to produce holdings and bibliographic records automatically, using the data from the spreadsheets. For the bibliographic records, a system-provided template was created that included some general subject headings, genre headings, an author field, and selected fixed fields, such as language, bibliographic level, and material type (figure 8).

Records for the Harbor Drive collection were loaded, and the system created brief bibliographic and holdings records and linked them to the Harbor Drive resource record. The records were globally updated to add the General Material Designator (GMD) “electronic resource” to the title as well as the phrase “digitized document” as a local “call number” to make these documents more visible in the browse screen of the online catalog (OPAC) (figure 9).

The digitized documents now could be found in the library catalog by author, subject, or keyword. The brief bibliographic records (figure 10) allow the user to go either to the digitized document via URL or to the resource record with more information on the resource itself and links to other items in the same collection. The resource record then provides links either to the new OSCDL Web site (via the <street name> - Oregon Sustainable Community Digital Digital Library link at the bottom of the resource record), to the bibliographic description of the individual document, or to the digitized document (figure 11).

However, the quality of the brief bibliographic records that had been batch generated through the system-provided template was not satisfactory (figure 8). It was decided that more document-specific data like summaries, number of pages, the dates the documents were created, geographical information, and document-level local subject headings should be included. These data were already available from the original spreadsheets. With limited time and staff resources, full bibliographic MARC records were batch created using the spreadsheets, detailed templates adjusted slightly to each collection, Microsoft Mail Merge, and finally, the MarcEdit program created by Terry Reese of Oregon State University (http://oregonstate.edu/~reeset/marcedit/html/index.html). This gave maximum control over the data to be included and the way they would be included. It also eliminated the need to clean up the data following the record load (figure 12).

Subsequently, full bibliographic records were created for the subcollections Harbor Drive, Front Street, and Park Blocks, to connect them to the next higher level, the Bonner Collection (figure 3). These records were also contributed to WorldCat. Mimicking the process used at the document level, a resource record was created for the Bonner Collection and the holdings records for the three subcollections were connected with their corresponding bibliographic records (figure 13).

Resource records with their corresponding item-level records for Trimet, the City Archives, and Metro followed. The final step was then to add the resource record and the bibliographic record for the whole OSCDL collection (figure 14). Since this last bibliographic record is not connected to a collection above it, there is only a hyperlink to the OSCDL resource record (figure 15).

More subcollections and their corresponding digital documents are continually being added to OSCDL. Structures in PSU Library’s OPAC are adjusted as these collections change.
Conclusion

According to Salter, “Digitizing, the current challenge that straddles the 20th and 21st centuries, has given archivists and librarians pause to reconsider access to their collections. The world of digitization is the catalyst for IT people, librarians, and archivists to unify the way they do things.” In this paper, a strategy has been offered for adapting a library system to traditional archival practice. By making use of some of the capabilities of the module in PSU Library’s Integrated Library System that was originally designed for managing electronic resources, a method was developed for managing digital archival collections in a way that incorporates some of the features of a traditional finding aid. The contents of the various hierarchical levels of the collection are fully represented through the manipulation of the record structures available through PSU’s system. This technique provides for enhanced access to the individual items of a collection by giving the context of the item within the collection. Links between the hierarchical levels facilitate navigation between the levels.

Although the records created for traditional library systems are not as rich as those found in traditional finding aids, or in EAD, their electronic equivalent; and the visual arrangements are not as intriguing as a well-planned Web site, the ability to show how items fit within the greater context of their respective collection(s) is a step toward reconciling traditional library and archival practices. Enabling the library user to virtually browse through the overall resources offered by the library and then, if desired, through the various levels of a collection for relevant resources enhances the opportunities presented to the user for finding relevant information.

References and notes

Appendix. Figures

Figure 1. Example of resource record from the PSU Library catalog (search conducted Nov. 4, 2005)
Figure 2. Example of a bibliographic record for a journal title from the PSU Library catalog (search conducted Nov. 4, 2005)
Figure 3. Partial diagram of the hierarchical levels of the collection

Figure 4. Resource record Harbor Drive with linked holdings records, bibliographic records, and original documents
Figure 5. Spreadsheet for tracking data

Figure 6. Data spreadsheet

Figure 7. Comma delimited coverage load file (*.csv)
Figure 8. Bibliographic records template

Figure 9. Browse screen in OPAC
Figure 10. System-created brief bibliographic record in OPAC

Figure 11. Resource record with various links
Figure 12. Full bibliographic record in OPAC

Figure 13. Bonner resource record with linked holdings records, bibliographic records, and original documents
Figure 14. Outline of linked records in the collection

Figure 15. Bibliographic record for the OSCDL collection