6-1-2006

Digital Collection Management through the Library Catalog

Michaela Brenner
Portland State University

Thomas W. Larsen
Portland State University, larsent@pdx.edu

Claudia Weston
Portland State University, westonc@pdx.edu

"The librarian and the archivist, for example, both collect, preserve, and make accessible

materials for research; but significant differences exist in the way these materials are arranged,

described, and used." (Society of American Archivists 2004) Libraries usually collect published

books and serials, and in more recent times commercially available sound recordings, films and

videos, electronic resources of various types, etc. Archives, on the other hand, tend to collect

unique 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.

Library practice

The first cataloging rules consisted of a set of 91 rules established by Anthony Panizzi (1841).

They have been the foundation upon which all Western cataloging rules have since been based.

Panizzi based his rules on the idea that any person looking for any particular book should be able

to find it through the catalog. The first complete set of rules for a dictionary catalog were

established by Charles Cutter (1876). Cutter based his rules on three "Objects" and six "Means."

(Cutter 1876) The three Objects were (1) to enable a person to find a book of which either the

author, the title, or the subject is known; (2) to show what the library has by a given author, on a

given subject, or in a given kind of literature; and (3) to assist in the choice of a book as to its

edition or as to its character. The six means of accomplishing this were (1) author-entry with the

necessary references; (2) title-entry or title-reference; (3) subject-entry, cross-references, and

classed subject table; (4) form entry; (5) giving edition and imprint, with notes when necessary;

and (6) notes. As can be seen, all of these principles are geared towards locating individual

published items. The rules set out in the *Anglo-American Cataloging Rules* (Joint Steering

Committee for the Revision of AACR 2002) are largely expansions on the principles established by Panizzi and Cutter (Wynar 1992).

Within the United States, bibliographic records in a modern library online public access catalog (OPAC) usually are coded and stored in MARC (Machine Readable Cataloging) format. MARC is a standard for the representation and exchange of bibliographic information that has been widely used by libraries for over thirty years. It is specifically designed to encode information about an individual bibliographic item, or *information product* (Haworth 2001), including author, title, subject, physical description, etc.

Archival practice

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 and/or accumulated and used by a particular person, family, or organization in the course of that creator's activities and functions (Society of American Archivists 2004). Thus, the basic unit of archival description, usually a finding aid, is a much more complex entity than the basic unit of bibliographic description and often involves multiple hierarchical levels of description which may or may not extend down to the level of individual items. Before archival description begins, the archivist identifies the 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 (Haworth 2001). The first explicit statement of the levels of arrangement in an archival collection was by Holmes (1964) and had been elevated to the level of "dogma" in the archival community (Abraham 1991). 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. (Society of American Archivists 2004)

The end result of archival description is usually a finding aid which ideally presents an accurate representation of the items in an archival collection so that users can, as independently as possible, locate them (Haworth 2001).

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 (Society of American Archivists) descriptive standard *Archives, Personal Papers, and Manuscripts* (APPM) (Hensen 1989) and its successor, *Describing Archives: A Content Standard* (DACS) (Society of American Archivists 2004). Its adoption, however, has been somewhat controversial among archivists (Carini and Shepherd 2004; Hensen 2001).

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, in practice few

archivists have actually made use of these linking fields (Abraham 1991). 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. The 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, more frequently, their electronic extension, the EAD, whereas the library community heavily relies on the 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 in specially designed websites. This, of course, means that users searching the OPAC may miss relevant materials which are described only in the archival and/or digital documents database or website. Similarly, users searching the archival and/or digital documents database or website 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 OPAC for individual items in their archival and/or digital document 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 records. To minimize processing costs, however, libraries frequently create brief descriptive records for items thereby limiting their value to patrons (Weisbrod and Duffy 1993). 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/or 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 which describe the collections more fully (Carini and Shepherd 2004). Collection-level records often are used in libraries where library resources may be insufficient for cataloging large collections of materials at the item level (See, e.g., Nichols 1996; Weisbrod and Duffy 1993). The guidelines for collection level records in

APPM and DACS, however, allow for additional fields that are not present in library bibliographic records. These include descriptions of the organization and arrangement of the collection, citations for published descriptions of the collection and/or links to the finding aid, acknowledgment of the donors, etc. 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. (Joint Steering Committee for the Revision of AACR 2002)

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 collection as a whole as a continuing resource.

With this in mind, we wished to examine whether we could adapt our electronic resource management system to accommodate evolving collections of digitized and born digital material.

More specifically, we wanted to examine whether the capabilities of our present system could be expanded to capture the hierarchical structure found in traditional archival finding aides. The electronic resource management system in use by Portland State University (PSU) Library is Innovative Interfaces' Electronic Resource Management (ERM) product.

According to Innovative Interfaces' marketing literature, "Electronic Resource Management effectively controls subscription and licensing information for licensed resources such as e-journals, Abstracting and Indexing (A&I) databases, and full-text databases." (Innovative Interfaces Inc. [2005]) 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, multi-level permissions restrict viewing and editing of data to the appropriate level of staff and/or patron.

The ability of ERM to replicate the relationships between aggregators or publishers and the electronic and/or print resources they provide was of particular interest to the authors. Through ERM and Innovative Interfaces' 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, sub-collection or title level thereby enabling the capture of descriptive information not permitted by standard bibliographic records. In Figure 1, for example, the resource record shows that the PSU Library provides limited access to a number of journal titles through its Springer journal online resource.
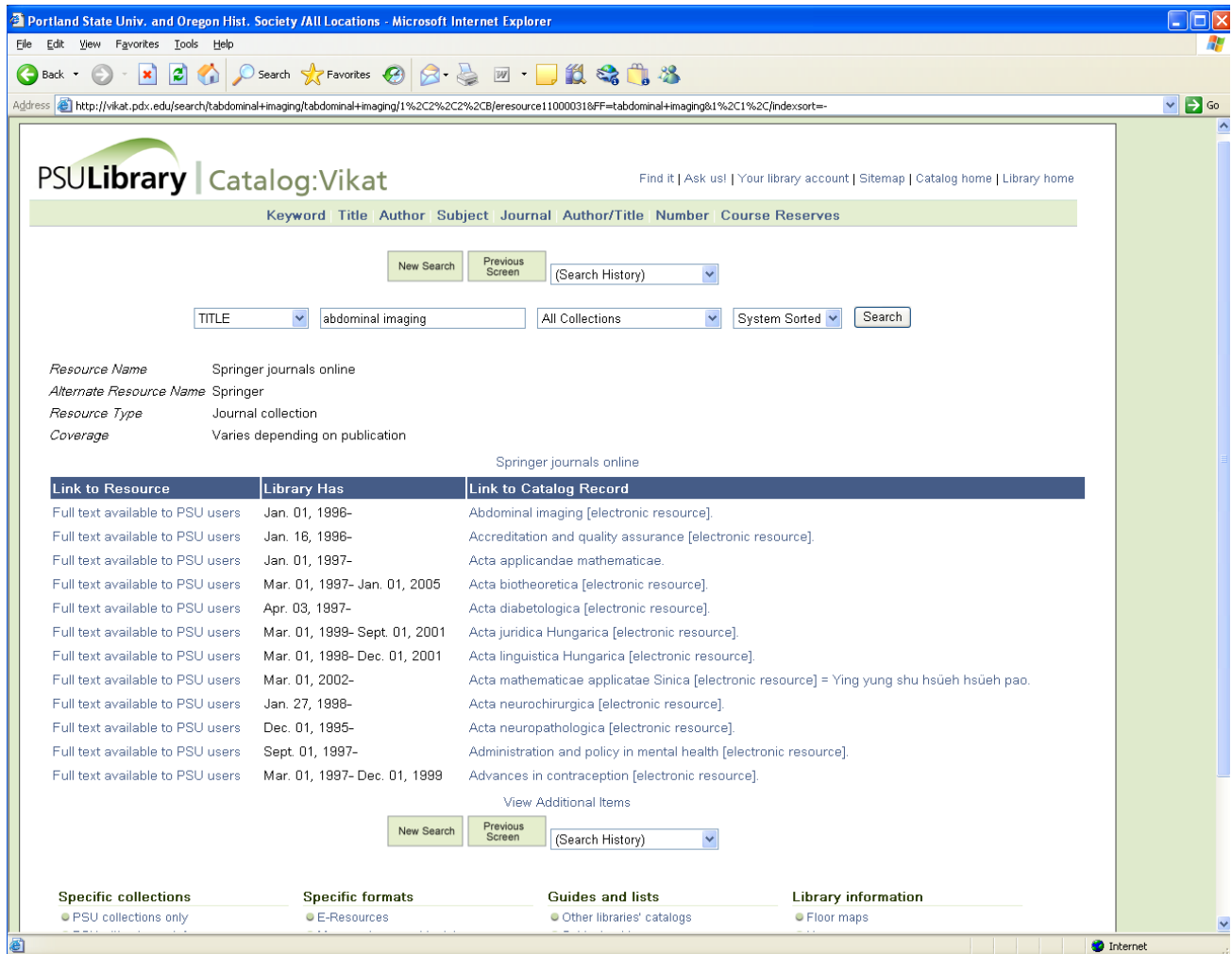
Figure 1 – Example of resource record from the PSU Library catalog (search conducted 11/04/05)
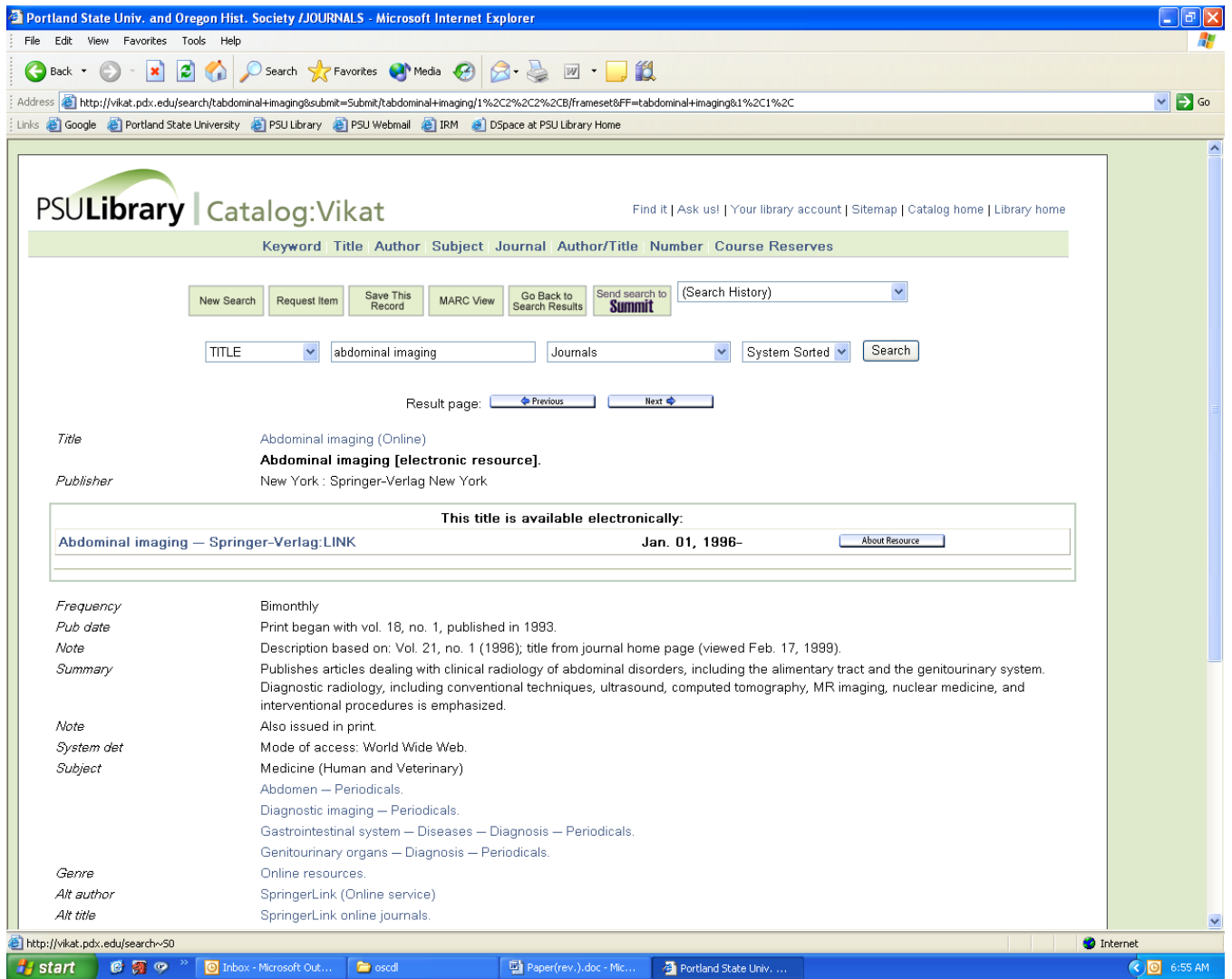
Figure 2 – Example of a bibliographic record for a journal title from the PSU Library catalog

(search conducted 11/04/05)

Conversely, if a search is conducted, for example on the journal title *Abdominal imaging*, as seen

in Figure 2, 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. More

information on the Springer collection can be discovered by clicking on the 'About resource'

button to retrieve the Springer resource record.

To explore the adaptation of ERM for library created digital collections, we took advantage of work being done to fulfill the requirements of a grant the PSU Library received in 2005. 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." (Abbot 2005) The overall collection is called the Oregon Sustainable Community Digital Library (OSCDL).

In addition to having its own Web site, we wanted to make this collection accessible through our library catalog so that patrons could find digitized original documents about the City of Portland together with other library materials. It was our intent to add bibliographic records to our database with hyperlinks to the digitized original documents using existing staff and tools. These bibliographic MARC records would be as complete as possible.

Initially, we focused our attention to documents originating from four different sources: Mr. 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, we received metadata 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 our project was that we wanted to reflect the four hierarchical levels of our collection (Figure 3). We manipulated Innovative Interfaces' system structure in order to accomplish this.
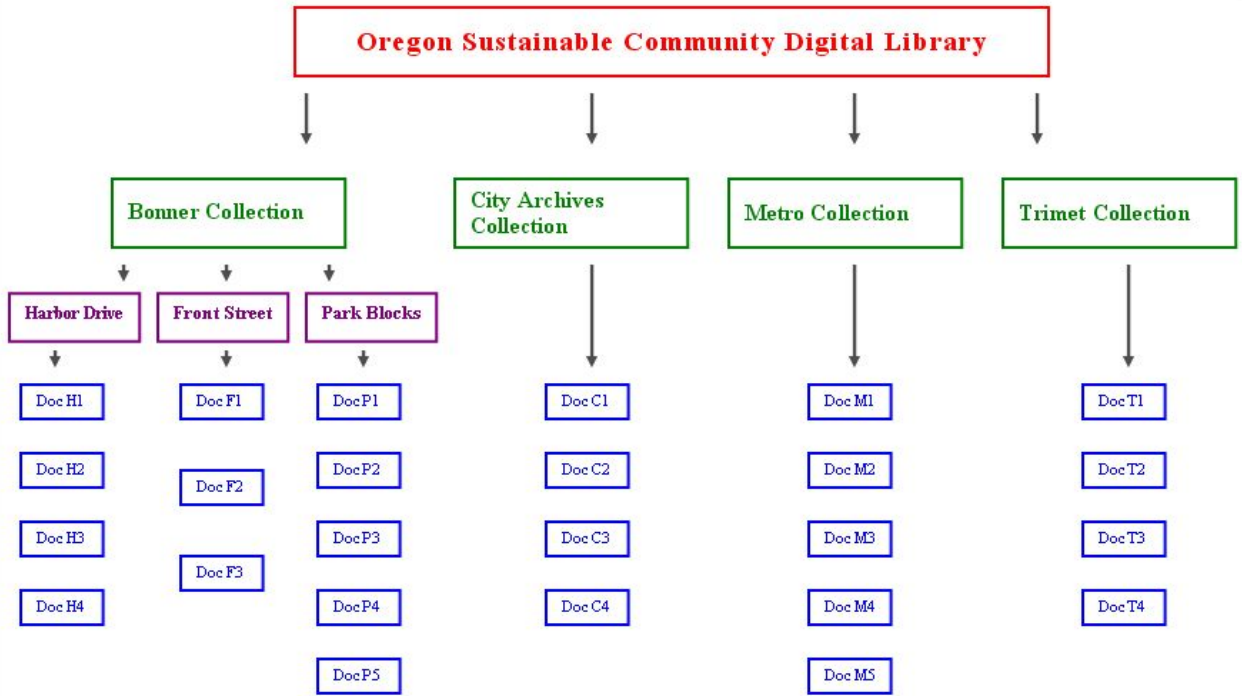
Figure 3 – Hierarchical levels of the collection

At the core of Innovative's ERM module are resource records created to reflect the peculiarities of a particular collection. Linked to these resource records are holdings records containing hyperlinks to the actual digitized documents as well as to their respective bibliographic records containing additional information on the individual items within the collection (Figure 4).
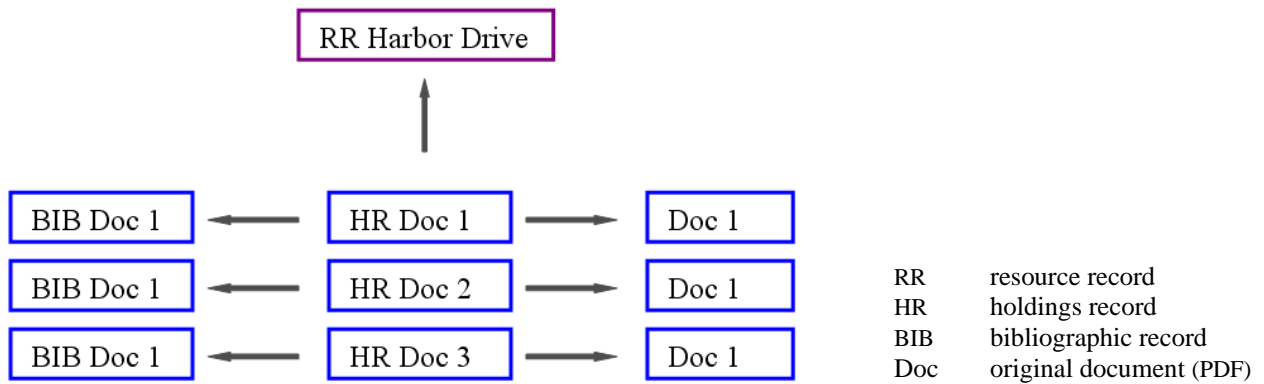

RR	resource record
HR	holdings record
BIB	bibliographic record
Doc	original document (PDF)

Figure 4 – Resource record Harbor Drive with linked holdings records, bibliographic records and original documents

First, we created resource records for three of the sub-collections within the Bonner collection. These sub-collections 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 Website; content and systems contact names; a brief description of the resource; and, most importantly, the Resource ID used to connect holdings records for individual documents to the corresponding resource record.

Taking advantage of tracking data produced during the digitization process (Figure 5), we created spreadsheets for each collection reflecting the data assigned to each individual digitized document. We included document title, the date the document was created, number of pages, and summaries. Since we were dealing with urban planning projects, we also included coordinates for the streets mentioned in the documents. Because Innovative's ERM uses ISSN numbers and titles as match points for record loads, we also manufactured 'ISSN' numbers for each document and included them in the spreadsheet. These homemade numbers were distinguished by using *pdx* as a prefix followed by collection and document numbers or letters, for example pdx0022090 or pdxhdcoll. We were fortunate that Innovative's ERM accepted these dummy ISSNs (Figure 6).

A	Description/Title (245)	Base file name (583 ‡b)	Pages (300)	Height in cm (300 ‡c)	Dimension from transmittal sheet (Ignore for MARC	Notes from transmittal sheets (Ignore for MARC record)	Notes from transmittal sheets (590)	Vendor	Date Sent	Date Rec'vd
1	- Handwritten Notes "Moses Plan" (1943)	pdx0021001_ _ _	1	28 cm.	8.5 x 11 inches. 1 page.	Do not scan reverse/blank side of documents	Black ink handwritten notes on notebook paper. Page has creased top and bottom left corners and a wrinkled bottom right corner.			
2	- Handwritten Notes "Carolyn Philp"	pdx0021002_ _ _	1	28 cm.	8.5 x 11 inches. 1 page.	Do not scan reverse/blank side of documents	Black ink handwritten notes on notebook paper. Page has a blue 5" x 3" sticky note in center of page, and two 3.5" x 2" business cards stapled to page. Please scan as one image in color.			
3	- Handwritten time line, starts 1/9/69 City Club Committee Meeting	pdx0021003_ _	3	28 cm.	8.5 x 11 inches. 3 pages.	Entire document is to be scanned, sequentially -- For example, pdx0021003001.tif is file name for the first page -- pdx0021003002.tif is file name for the next page. -- Pages are unnumbered -- there are 3 pages total. -- DO NOT scan	Green ink handwritten notes on notebook paper. Please scan in color.			
4	- Handwritten time line, starts 19 Dec '69 Oreg p. 44, c.1	pdx0021004_ _ _	8	28 cm.	8.5 x 11 inches. 8 pages.	Entire document is to be scanned, sequentially -- For example, pdx0021004001.tif is file name for the first page -- pdx0021004002.tif is file name for the next page. -- Pages are unnumbered, 8 pages total. -- DO NOT scan reverse/blank sides of	Black ink handwritten notes on notebook paper. Please scan in color.			

Figure 5 – Spreadsheet for tracking data

Journal Title	Notes	Pub date	Summary	ISSN	pages	Provider	URL
Ash Street ramp project		August 11, 1961	Letter to the Portland City Planning Commission attn: Harry Sroufe and Lloyd Keefe from Mayor Schrunk, Aug. 11, 1961	pdx0021031	2 pages	Harbor Drive	http bor'
City Club Committee meetings	Chronology	Juli 1968-January 1971		pdx0021003	3 pages	Harbor Drive	http bor'
City of Portland Council Meeting Minutes		June 18, 1971		pdx0021016	42 pages	Harbor Drive	http bor'
City planning - Clay Street	Note			pdx0021008	1 page	Harbor Drive	http bor'
Correspondence City/County minutes		November 4, 1960	Memo to File from Tom Edwards, Construction Engineer	pdx0021030	2 pages	Harbor Drive	http bor'

Figure 6 – Data spreadsheet

From this data spreadsheet, we also created the system required comma delimited coverage load file (*.csv). 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. We included individual document titles, our 'ISSN' numbers, individual URLs to the documents, and a collection

specific resource ID (Provider) which connects all the documents from a collection to their respective resource record. The resource ID is the same for all documents in one collection (Figure 7).

	A	B	C	D
1	Journal Title	ISSN	URL	Provider
2	Ash Street ramp project	pdx0021031	http://oscdl.research.pdx.edu/docs/harbor%5Fdrive%5F1/pdx0021031.pdf	Harbor Drive
3	City Club Committee meetings	pdx0021003	http://oscdl.research.pdx.edu/docs/harbor%5Fdrive%5F1/pdx0021003.pdf	Harbor Drive
4	City of Portland Council Meeting Minutes	pdx0021016	http://oscdl.research.pdx.edu/docs/harbor%5Fdrive%5F1/pdx0021016.pdf	Harbor Drive
5	City planning - Clay Street	pdx0021008	http://oscdl.research.pdx.edu/docs/harbor%5Fdrive%5F1/pdx0021008.pdf	Harbor Drive
6	Harbor Drive to Ash or Oak Street	pdx0021030	http://oscdl.research.pdx.edu/docs/harbor%5Fdrive%5F1/pdx0021030.pdf	Harbor Drive

Figure 7 - Comma delimited coverage load file (*.csv)

In our first attempt, we set the system up to automatically produce holdings and bibliographic records 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).
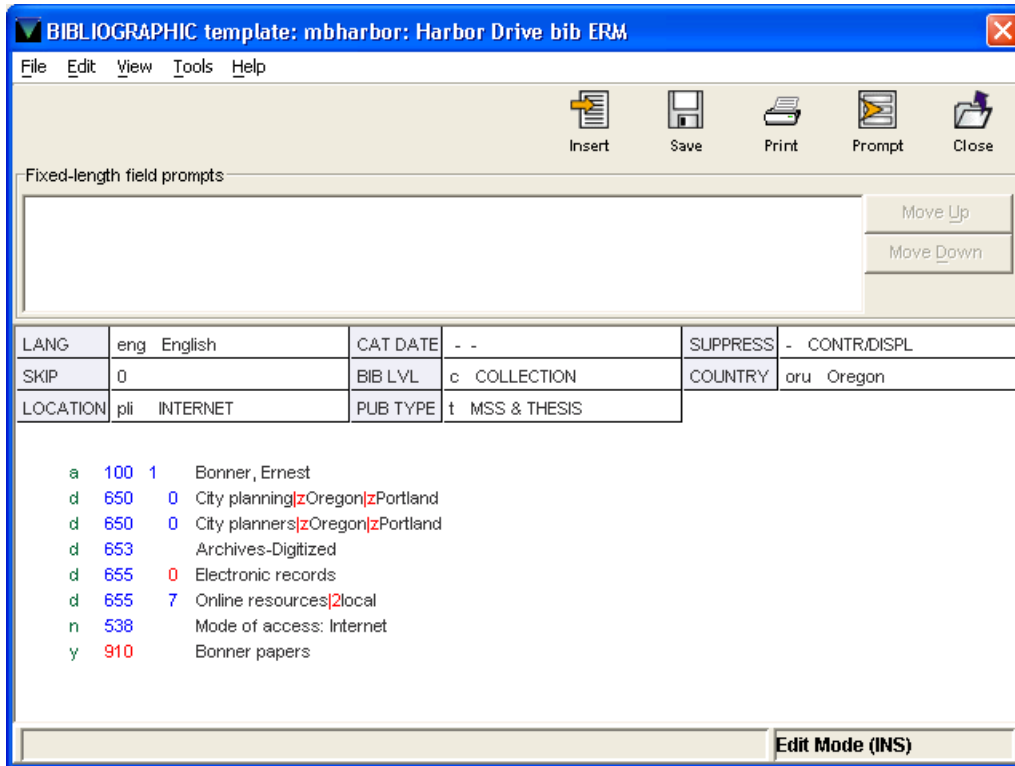
Figure 8 – Bibliographic records template

We loaded the records for the Harbor Drive collection and the system created brief bibliographic and holdings records and linked them to the Harbor Drive resource record. We globally updated the records 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 our online catalog (OPAC) (Figure 9)

Figure 9 – Browse screen in OPAC

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 Website (via the *<street name> - Oregon Sustainable Community 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).

SUBJECT	portland city planners	All Collections	System Sorted	Search

Record 5 of 145

Result page: Previous | Next

Author	Bonner, Ernest.
Title	**Ash Street ramp project [electronic resource].**

This title is available electronically:	
Click here to see manuscript	About Resource

System det	Mode of access: Internet.
Subject	City planning — Oregon — Portland.
	City planners — Oregon — Portland.
Genre	Electronic records.
	Online resources
Subject	Archives–Digitized.
ISSN	pdx0021031

Record 5 of 145

Result page: Previous | Next

Figure 10 – System created brief bibliographic record in OPAC

Resource Name	Harbor Drive – Oregon Sustainable Community Digital Library
Description	Correspondence and notes regarding the closure, renovation engineering and re-opening of Harbor Drive, Portland, Oregon, between City stakeholders, including Ernie Bonner, Neil Goldschmidt, Jack Sollis, James McClure, Robert Bothman, and others of the State Highway Division and the City Club of Portland
Resource Type	Digitized documents
Coverage	Portland City planning 1938 to 2002

Harbor Drive – Oregon Sustainable Community Digital Library

Link to Resource	Library Has	Link to Catalog Record
Click here to see manuscript		Ash Street ramp project [electronic resource].
Click here to see manuscript		City Club Committee meetings [electronic resource].
Click here to see manuscript		City of Portland Council Meeting Minutes [electronic resource].
Click here to see manuscript		City planning – Clay Street [electronic resource].
Click here to see manuscript		Correspondence City/County minutes [electronic resource].
Click here to see manuscript		Correspondence Keefe, et. al [electronic resource].
Click here to see manuscript		Correspondence Schrunk, et.al [electronic resource].
Click here to see manuscript		Department of Transportation [electronic resource].
Click here to see manuscript		Harbor Drive access study [electronic resource].
Click here to see manuscript		Harbor Drive closure – Bonner's notes [electronic resource].
Click here to see manuscript		Harbor Drive closure – editorial [electronic resource].

Figure 11 – Resource record with various links

We were, however, not satisfied with the quality of the brief bibliographic records that had been batch generated through the system-provided template (Figure 8). We wanted to include more document specific data like summaries, number of pages, the dates the documents were created, geographical information and document level local subject headings. These data were already available to us from the original spreadsheets. With limited time and staff resources, we followed a path we had taken before and batch created full bibliographic MARC records 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 us maximum control over the data we wanted to include and the way we wanted to include them. It also eliminated the need to clean up the data following the record load (Figure 12)

Figure 12 – Full bibliographic record in OPAC

Subsequently, we created full bibliographic records for the sub collections Harbor Drive, Front Street, and Park Blocks to connect them to the next higher level, the Bonner Collection (Figure 3). We also contributed these records to WorldCat. Mimicking the process used at the document level, we created a resource record for the Bonner Collection and connected the holdings records for the three sub-collections and their corresponding bibliographic records (Figure 13).
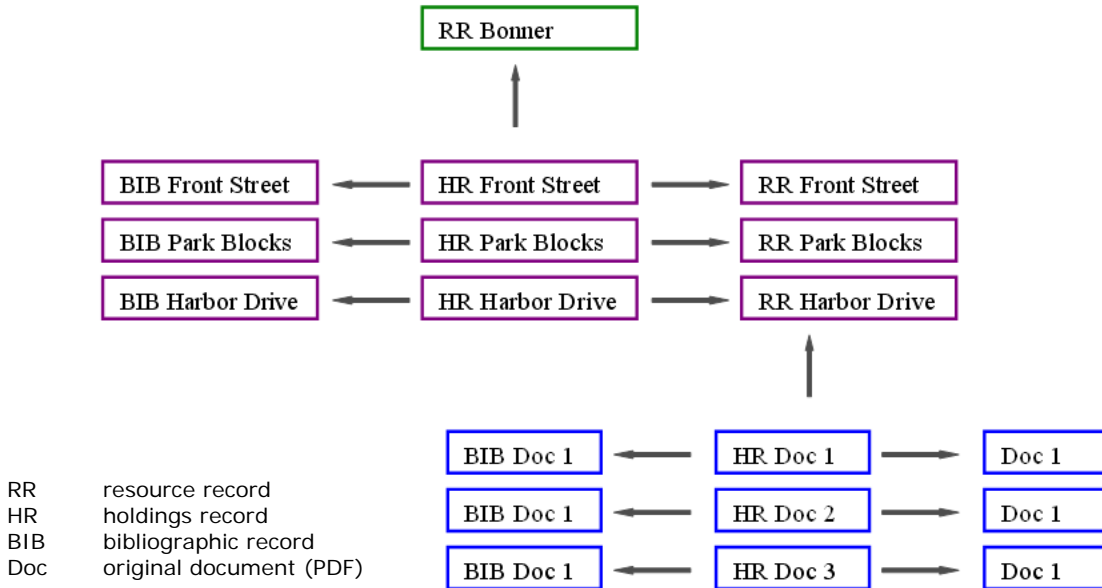
Figure 13 – Bonner resource record with linked holdings records, bibliographic records and

original documents

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 Oregon Sustainable Community Digital Library (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).
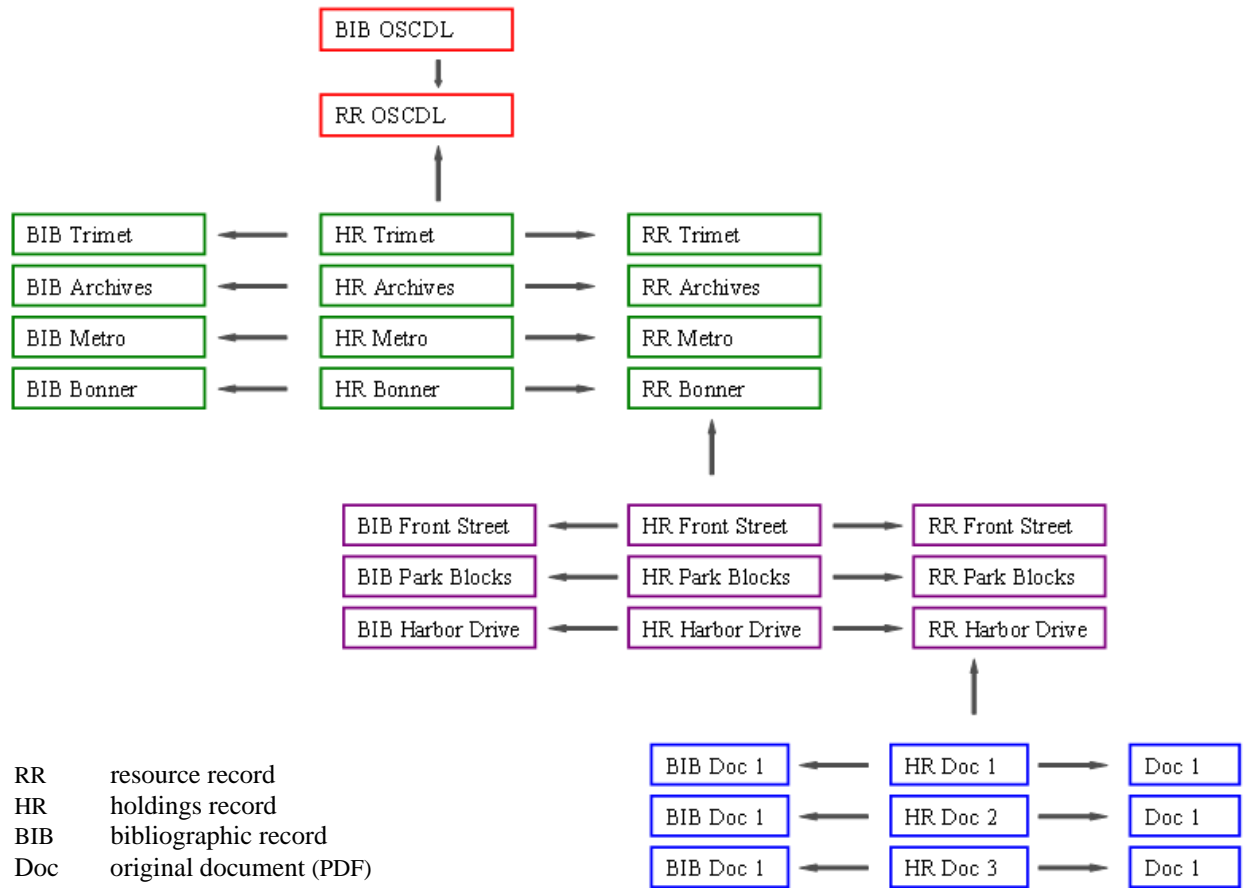
Figure 14 – Outline of linked records in the collection

Figure 15 – Bibliographic record for the OSCDL collection

More sub-collections and their corresponding digital documents are continually being added to

the Oregon Sustainable Community Digital Library. As these collections change, we adjust the

structures in our OPAC.

Conclusion

According to Anne Salter, "Digitizing, the current challenge that straddles the 20th and 21st

centuries, has given archivists and librarians pause to re-consider access to their collections. The

world of digitization is the catalyst for IT people, librarians, and archivists to unify the way they

do things." (Salter 2003) The authors have offered a strategy for adapting a library system to traditional archival practice.

References

Abbot, Carl. 2005. Planning a sustainable Portland: A digital library for local, regional, and state planning and policy documents - Framing paper. http://oscdl.research.pdx.edu/framing.php

Abraham, Terry. 1991. Oliver W. Holmes revisited: levels of arrangement and description of practice. *American Archivist* 54:370-377.

Carini, Peter, and Kelcy Shepherd. 2004. The MARC standard and encoded archival description. *Library Hi Tech* 22 (1):18-27.

Cutter, Charles A. 1876. *Rules for a printed dictionary catalogue*. Washington: Government Printing Office.

Haworth, Kent M. 2001. Archival description: content and context in search of structure. *Journal of Internet Cataloging* 4 (3/4):7-26.

Hensen, Steven L. 2001. Archival cataloging and the internet: the implications and impact of EAD. *Journal of Internet Cataloging* 4 (3/4):75-95.

Hensen, Steven L., comp. 1989. *Archives, personal papers, and manuscripts*. 2nd ed. Chicago: Society of American Archivists.

Holmes, Oliver W. 1964. Archival arrangement: Five different operations at five different levels. *American Archivist* 27 (1):21-41.

Innovative Interfaces Inc. [2005]. Electronic resources management. http://www.iii.com/pdf/lit/eng_erm.pdf.

Joint Steering Committee for the Revision of AACR. 2002. *Anglo-American Cataloging Rules*. 2nd ed., 2002 revision ed. Chicago: American Library Association.

Nichols, Margaret F. 1996. Finding the forest among the trees: the potential of collection-level cataloging. *Cataloging & Classification Quarterly* 23 (1):53-71.

Panizzi, Antonio. 1841. Rules for the compilation of the catalogue. *The Catalog of the British Museum* 1:v-ix.

Salter, Anne A. 2003. 21st-Century Archivist. *Newsletter*. http://www.lisjobs.com/newsletter/archives/sept03asalter.htm.

Society of American Archivists. 2004. *Describing archives: a content standard*. Chicago: Society of American Archivists.

———. 2004. So you want to be an archivist: an overview of the archival profession. http://www.archivists.org/prof-education/arprof.asp.

Weisbrod, Elizabeth J., and Paula Duffy. 1993. Keeping your online catalog from degenerating

      into a finding aid: considerations for loading microformat records into the online catalog.

      Technical Services Quarterly 11 (1):29-42.

Wynar, Bohdan S. 1992. *Introduction to cataloging and classification*. Translated by A. G.

      Taylor. 8th ed. Englewood, Colo.: Libraries Unlimited.