

## Developing and Sustaining the Northwest Digital Archives

Alan Kevin Cornish and Trevor James Bond  
Washington State University Libraries, Pullman, WA USA  
Email: [cornish@wsu.edu](mailto:cornish@wsu.edu), [tjbond@wsu.edu](mailto:tjbond@wsu.edu);  
Web: <http://nwda.wsulibs.wsu.edu/index.html>

### Abstract

The Northwest Digital Archives is a union database of Encoded Archival Description (EAD) finding aids from institutions in Washington, Oregon, Idaho, Alaska, and Montana. [1] The purpose of the NWDA is to make information about collections of primary sources widely accessible to researchers over the Internet. This article will explore the selection and development of the NWDA database, the creation of tools to associate digital objects with EAD/XML documents, and the methods employed by the technical staff at the Washington State University Libraries to expose metadata contained in the NWDA to Google (and aggregators) with Google Sitemaps and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Through the analysis of web site use statistics, we realized that most users are locating individual NWDA finding aids directly from Google and other search engines. While not entirely surprising, the implications of this when combined with three rounds of usability testing led to major revisions of the NWDA website and finding aid stylesheet. The article concludes with a discussion the model developed to sustain the NWDA after the end of National Endowment for the Humanities funding.

### Introduction

Daniel Pitti, the creator of the Encoded Archival Description (EAD) standard, defines a finding aid as a document used to describe “corporate records and personal papers.” Finding aids typically include component lists which describe the contents of boxes, folders, and documents. As Pitti notes, an important motive in creating the EAD standard was enabling “universal, union access to primary resources.” Kristi Kiesling describes the EAD standard as “the SGML/XML based document type definition that archives, libraries, and museums are using to create, store, and distribute descriptions of their collections.” Elizabeth Yakel and Jihyun Kim note the challenges in adoption of the EAD standard, which involves two separate processes: encoding the finding aids into XML (following the EAD rules) and publication. Their 2005 study of EAD adoption concludes that only a minority (39%) of nearly 400 archival institutions surveyed had implemented EAD, with institution size being a key driver in adoption (i.e., larger institutions with larger staffs were better able to adopt the EAD standard). Several articles have appeared on the difficulties in publishing EAD finding aids. James Roth’s noteworthy paper on EAD publishing is informative, but describes a number of technical solutions which are no longer available (e.g., DynaText/DynaWeb, the Panorama client-side viewer). Alan Cornish performed a later review of EAD publishing solutions, noting the limits in available commercial and open source solutions available to libraries and archives.

### Background on the Northwest Digital Archives:

The genesis of the Northwest Digital Archives (NWDA), a web accessible union catalog of more than 4,500 EAD finding aids from institutions in Alaska, Idaho, Oregon, Montana, and Washington took place during an informal meeting at Online Northwest in Portland, Oregon, in January 2001 when a group of eager archivists agreed to collaborate on a proposal to the National Endowment for the Humanities (NEH). This and a second proposal were submitted to the NEH Division of Preservation and Access and resulted in the funding of two major grants: phase I (July 2002 through December 2004 and phase II (July 2005 through September 2007). With both grants, Oregon State University administered the project and Washington State University provided technical support and hosted the NWDA database.

At the outset, several institutions realized that they did not have enough complete finding aids to participate in the NEH grant and that they would need to re-process legacy collections and add new information to their finding aids. Eight institutions who wished to rework their legacy finding aids formed a consortium with the moniker the Northwest Archives Processing Initiative (NWAPI) and successfully received two grants from the National Historical Publications and Records Committee (NHPRC). During the first NHPRC grant, the group hired consultant Brad Cole who assisted in developing a regional finding aid standard that anticipated elements of Describing Archives a Content Standard (DACS). As part of their second proposal, the NWAPI group hired Mark Greene as a consult to apply the “More Product, Less Process” method to a regional processing initiative. [2] According to Hauck (2007) the eight institutions participating in the grant processed 800 collections comprising 1,120 feet of materials (an average of roughly 3 hours of labor were spent to process each linear foot collection). The finding aid standard developed in the first phase was revised to adhere to EAD version 2002 and DACS multilevel description. [3]

### NWDA software

In the early stages of planning for the NWDA, there was not a clear choice of database programs that met our needs. In 2003, after an evaluation of commercial and open source EAD delivery software programs, we selected IXIASOFT TEXTML as the NWDA’s database solution. [4] TEXTML had been successfully used in supporting the search and retrieval of archival finding aids in the United Kingdom’s Access to Archives (A2A) project. A2A served as a model of what the NWDA hoped to accomplish in that the A2A consortium brought together finding aids from a diverse range of organizations including libraries, record offices, universities, and museums. The A2A site also demonstrated the scalability of the TEXTML database with more than 400 repositories participating.

TEXTML has powerful indexing capabilities based upon XML-related specifications such as XPath 1.0. The NWDA licensed the TEXTML database software in late 2003; a basic search site was released in 2004. A more robust search application using Active

Server Pages .NET (ASP.NET) was developed by the technical staff at the Washington State University (WSU) Libraries and made public in spring 2006. With TEXTML, the local development demands were significant, as search and retrieval, use assessment, and database submission modules, along with web templates for EAD encoding and validation checking tools were eventually developed by WSU project staff and implemented for use.

Early in the first phase of the NWDA grant, project directors requested that a system for durable identifiers be created so that EAD encoded finding aids in the NWDA database could be linked to from MARC records, departmental homepages, and other online documents. To meet this request, the technical staff at WSU implemented the Archival Resource Key (ARK) specification for the NWDA repository. The ARK specification was developed by staff at the California Digital Library “to allow persistent, long-term access to information objects.” [5]

This implementation has had two specific benefits, including the creation of a stable identifier that can be referenced to retrieve a document and a durable link to facilitate search engine harvesting. As an example, Figure 1 shows a search results screen after a phrase search was executed in Google. The first hit is an NWDA finding aid, with its URL being the identifier based upon the ARK scheme.

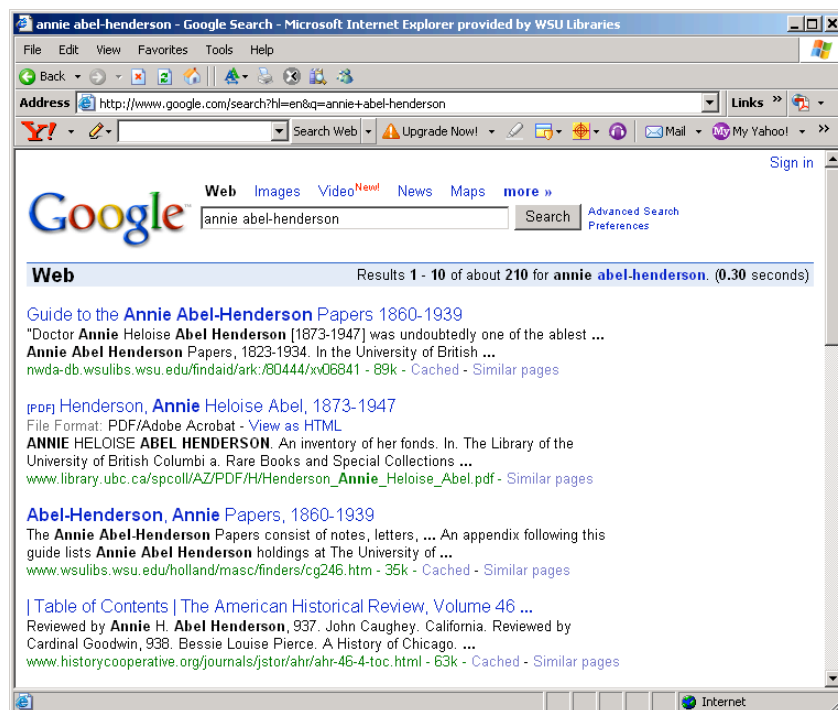


Figure 1. Google search results, NWDA document is the first hit

The second benefit of the ARK implementation concerns sustainability. The ARK specification requires information to be presented when users append parameters to an ARK identifier. For example, these URLs display information to searchers:

<http://nwda-db.wsulibs.wsu.edu/ark:/80444/xv10269>

The digital object (the finding aid)

<http://nwda-db.wsulibs.wsu.edu/ark:/80444/xv10269?>

Metadata about the digital object

<http://nwda-db.wsulibs.wsu.edu/ark:/80444/xv10269??>

A commitment statement for the digital object

The commitment statement provides a good example of sustainability support; according to the California Digital Library's *Preservation Glossary*, it is "a declaration by an organization of its intention to retain and make available a given object or set of objects." This capability enables the consortium to define a preservation policy for digital objects and to have that information transparently available to users of NWDA resources. If the commitment statement URL is entered, information on the institution's intention to preserve the object will be presented to the user, including the details of the preservation policy and, perhaps, the length of time that the identifier is guaranteed to be preserved.

#### NWDA dissemination: Scholarly repositories and search engines

As Tibbo and Meho observed, just because you put finding aids on the Internet does not necessarily mean that researchers will find them. Given that the purpose of the NWDA is to provide access to archival collection to a broad audience of researchers, the technical staff at WSU worked to disseminate information from the NWDA site as widely as possible. To support the integration of NWDA resources into other repositories, an Open Archives Initiative-Protocol for Metadata Harvesting (OAI-PMH) data provider application was developed [6]. In performing the NWDA OAI implementation, sets (the base unit of harvest) were created for both member institutions and for controlled vocabulary terms. This enables, as one example, a discipline-specific repository to harvest a subset of the NWDA content based upon one or more subject terms defined by the project's archivists. Creating an OAI data provider involved the complex process of mapping EAD to the OAI-PMH specification [7]. The current NWDA mapping focuses on collection-level (as opposed to component-level) information.

Over time, it became increasingly desirable to support search access to NWDA documents through Internet search engines, instead of assuming that the NWDA site would be the primary access portal for searchers. To provide this support, project staff members have proactively worked to have NWDA finding aids harvested by Google. This has involved several steps; including serving information to Google using its Sitemaps program and linking to finding aid identifiers from high-use, quality web pages. For Sitemap support, the WSU technical staff wrote and maintains an ASP.NET application that serves up current database identifiers within a well-formed XML document that is coded according to the Sitemap Protocol specification. Linking to NWDA identifiers from high-use, quality web pages, however, appears to be the primary driver for NWDA finding aid indexing and page ranking in Google. To increase the rankings of WSU finding aids in Google search results, we created a basic page listing WSU finding aids with links utilizing the durable ARK identifiers. This page was placed on the WSU web site so that it could be indexed by Google bots.

Document Name	Document Size
A.S.W.S.U. Activities Board Records	5.88 KB
A.S.W.S.U. Activities Center Records	60.71 KB
A.S.W.S.U. Activities Center Records	47.20 KB
A.S.W.S.U. Activities Center Records	81.79 KB
A.S.W.S.U. Annual Budget Reports	5.62 KB
A.S.W.S.U. Board of Control Records	8.64 KB
A.S.W.S.U. President Records	85.84 KB
A.S.W.S.U. Radio Broadcast Committee Records	7.14 KB
A.S.W.S.U. Vice President Records	6.93 KB
Abel-Henderson (Anne) Papers	31.55 KB
Ad Hoc Accreditation Committee Records	53.36 KB
Adams (Glen Cameron) Papers	8.67 KB
Admissions Office Records	27.11 KB
Admissions Office Records	26.91 KB
Agricultural Economic Survey and Studies	94.56 KB
Agricultural Engineering Records	144.79 KB
Agricultural Experiment Station Irrigation Branch, Prosser, Reports	8.50 KB
Agricultural Extension Service County Agents Reports	70.02 KB
Agricultural Extension Service County Leaders Reports	6.02 KB
Agricultural Extension Service Director of Extension Reports	6.63 KB
Agricultural Extension Service Project Summaries	11.11 KB
Agricultural Extension Service Reports	5.73 KB
Agricultural Extension Service Reports	5.09 KB

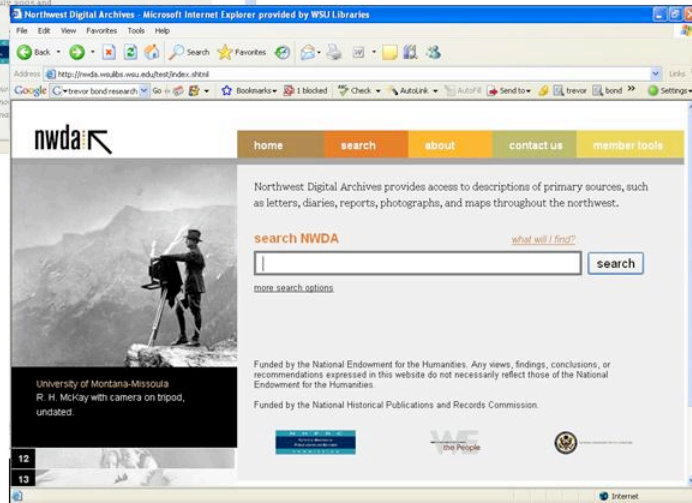
Figure 2. This modest page consisting of finding aid titles with durable links to their corresponding finding aids in the NWDA database has helped drive indexing by Google. Though this page is not directly linked to other pages on the website of Manuscripts, Archives, and Special Collections (MASC) at the Washington State University Libraries, its inclusion in the main directory of the MASC website, deemed a quality site by Google, has resulted in Google bots indexing the page.

According to website statistics, one of the results of these dissemination efforts is that the majority of researchers are locating NWDA finding aids directly through Google (and other search engines), not by using the NWDA search interface. For example in September 2007, researchers retrieved 2,622 EAD finding aids through the NWDA site, but retrieved 7,728 EAD finding aids through search engine or web pages using ARK identifiers. Of the hits of NWDA EAD finding aids from search engines, more than 70% come from Google. [8] The consequence of this pattern along with three rounds of usability testing led us to redesign both the NWDA website and the finding aid style sheet. [9]



Figure 3. The earlier version of the NWDA front page on top. To search the database, visitors had to select the small search tab to go to a search page.

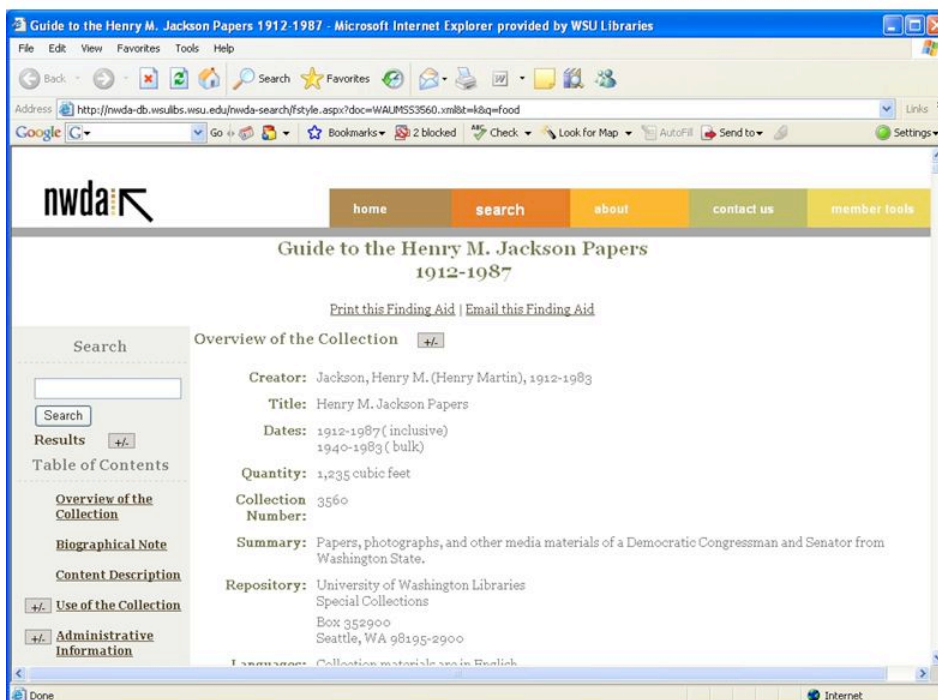
With the redesigned page to the right, we have attempted to make the page more user-centered with a large search box and bigger tabs.



The initial design of the NWDA was geared toward the project itself. The main page described the NWDA database, but required researchers to select a small tab to search the database. One of the key recommendations from the NWDA usability group was to add a large search box on the front of the site and increase the size of the navigation tables. With most researchers proceeding directly from Google search results to individual finding aids thereby bypassing the NWDA results page, we added new features to the finding aid stylesheet including two prominent options at the top of the finding aid to e-mail this document (utilizing the ARK identifiers discussed above) and a printer friendly link to print the finding without extraneous graphics. Since researchers could not access their search terms as “hits in context,” a new search within document option was created. Usability testing also indicated that researchers wanted to scroll quickly through the finding aids so institutional logos were removed and we added expand and contacting (displayed as +/- with the default the + or open option; selecting – collapses the sections) the front matter. Usability testing also revealed that users were annoyed by the presence of subject and controlled access terms appearing before container lists so the revised NWDA style sheet displays subject headings at the bottom of the finding aid. Researchers expressed a desire for the consistent display of data so the images included in some finding aids (but not all) were hidden by the stylesheet. [10]



Figure 4. The revised NWDA stylesheet provides links to a printer friendly version of the finding aid, a new search within the document option, and the choice to expand or collapse sections. Subject headings are displayed at the end of the finding aid.



## Adding digital content

Perhaps the most significant technical need for the existing NWDA database and application is improving the ability to link NWDA finding aids with digital objects, such as digitized photographs and documents. In approaching this development work, project staff members are reviewing systems that support this integration, such as the University of Illinois' Archon application and the California Digital Library's XTF software. During the project's period of NEH funding, project staff developed a web application that enables the creation of linkages from finding aid component list items to externally-available digital content. This works well for larger institutions in the NWDA that have the resources and expertise to support a digital media server, but not for smaller institutions in the consortium. The current need for the NWDA is to support linking to digital content from finding aids, including both externally- (in existing digital content management systems) and internally-stored digital objects. Providing easy integration of digital content with finding aids, along with the search engine integration described above, promises to increase dramatically both the number of NWDA users and use of the consortium's resources.

## Conclusion/Sustainability

As part of the second phase of the NEH grant, we needed to develop a long-term sustainability plan. After deciding against a model of charging users for website access,

we looked for potential partners (the NWDA was too large and complex for one academic institution to take on). One of the challenges of sustaining the NWDA is that we envisioned not only the ongoing maintenance of the existing database including license fees and server replacement, but also funding a dedicated program coordinator and future programming support. In other words, we are striving to sustain the spirit and energy of the NWDA, not merely maintaining the database on life-support.

After exploring several possibilities, in July 2007, the NWDA joined the Orbis Cascade Alliance: a consortium of 34 libraries in Oregon and Washington. [11] Joining the Orbis Cascade Alliance, a very successful regional consortium that provides among other services a union library catalog with self-initiated borrowing, was a natural fit for archival repositories in Oregon and Washington already members of Orbis Cascade Alliance, but for NWDA institutions from Alaska, Idaho, and Montana the move was less certain. However, the benefits for NWDA members in shared administrative costs and support, the challenge of finding other willing partners, and the possibility of new NWDA members from Orbis Cascade Alliance libraries, outweighed any concerns.

In 2007, the NWDA developed an annual budget for the fiscal year 2008 of roughly \$100,000 to fund a half-time administrator and technical support. The budget was divided between NWDA member institutions using a model of a 50% flat fee and a 50% based on the number of finding aids contributed to the NWDA. Certainly this was not an ideal model for determining costs; it penalized the largest contributors to the NWDA. To develop a new, long-term funding model, the NWDA steering team and Jodi Allison-Bunnell, NWDA Program Manager, developed a series of possible options. After numerous meetings, surveys, and a vote, the NWDA membership agreed to membership fees based on institution type. The estimated membership dues for fiscal year 2010 will range from \$1,800 for public two-year institutions to \$6,600 for large four year academic institutions with more than 15,000 full time enrolled students. [12] To encourage the participation of institutions with nascent archival programs or limited collections, a special category was adopted based on the same institutional categories, but with a 25% discount for those repositories with 25 finding aids or less. For current and prospective members faced with membership fees, Jodi Allison-Bunnell developed a membership page outlining what institutions get for joining NWDA. [13]

From the end of the second NEH grant to the writing of this article membership in the NWDA has increased from 17 members involved in the second NEH grant to 31. We will continue to enhance the NWDA database with new finding aids and digital content. Future plans also include adding new members, continuing to expose NWDA metadata to search engines, and working with other regional EAD providers on developing a national archival network.

#### Selected list of standards employed by the Northwest Digital Archives

*The ARK Persistent Identifier Scheme, July 24, 2007*  
[<http://www.cdlib.org/inside/diglib/ark/arkspec.txt>]



*Describing Archives a Content Standard (DACS)*

[<http://www.archivists.org/governance/standards/dacs.asp>]

*Encoded Archival Description, Version 2002* [<http://www.loc.gov/ead/> ]

*Extensive Markup Language (XML) 1.0* [<http://www.w3.org/TR/2006/REC-xml-20060816/> ]

*Google Sitemap Protocol, Version 0.9*

[<https://www.google.com/webmasters/tools/docs/en/protocol.html> ]

*The Open Archives Initiative-Protocol for Metadata Harvesting, Version 2.0*

[<http://www.openarchives.org/OAI/openarchivesprotocol.html> ]

*XML Path Language (XPath), Version 1.0* [<http://www.w3.org/TR/xpath> ]

## References

California Digital Library (2006) *Preservation Glossary*

[<http://www.cdlib.org/inside/diglib/glossary/?field=glossary&action=search&query=preservation>]

Cornish, A. (2004) "Using a native XML Database for Encoded Archival Description search and retrieval," *Information Technology and Libraries* 181-184

[<http://www.lita.org/ala/lita/litapublications/ital/volume23a/number4a/communications.pdf>]

Hauck, J. (2007) "The ABC's of MPLP: How to apply the 'more product less process' method in your archives. *Easy Access*, Northwest Archivists Vol. 33 issue 3.

[<http://www.lib.washington.edu/nwa/documents/EasyAccess--09-07.pdf>]

Kiesling, K. (2001) "Metadata, metadata everywhere - but where is the hook?" *OCLC Systems and Services* Vol. 17, No. 2, 84-88

Pitti, D. (1999) "Encoded Archival Description: An introduction and overview." *D-Lib Magazine* Vol. 5, No. 11 [<http://www.dlib.org/dlib/november99/11pitti.html>]

Roth, J. (2001) "Serving up EAD: An Exploratory Study on the Deployment and Utilization of Encoded Archival Description Finding Aids." Master's paper for the MS in LS degree [<http://ils.unc.edu/MSpapers/2715.pdf>]

Tibbo, H. and Meho, L. (2001) "Finding Finding Aids on the World Wide Web: Evaluation of Six Popular Search Engines." *The American Archivist* Vol 64, No. 1

Yakel, E. and Kim, J. (2005) "Adoption and Diffusion of Encoded Archival Description" *Journal of the American Society for Information Science and Technology* Vol. 56, No. 13 1427-1437

## Notes

[1] The authors wish to thank Jodi Allison-Bunnell, Program Manager of the Northwest Digital Archives, and the anonymous reviewers for *JoDI* for their suggestions.

[2] Dennis Meissner and Mark Greene developed a table for the NWAPI consortium so that members could apply MPLP processing methods in their institutions. The table lists traditional processing methods along side adequate (MPLP) methods so that individual archivists can make processing decisions based on local circumstances. This table may be found on page 5 of the September 2007 *issue of Easy Access*, Northwest Archivists Vol. 33 issue 3. A listing of the participating members of the NWAPI consortium and reports may be viewed at: <http://www.orbiscascade.org/index/northwest-digital-archives-reports>.

[3] The NWAPI regional finding aid standard may be accessed at: <http://www.orbiscascade.org/nwda/tools.html>.

[4] Addition background information on the selection of TEXTML may be found in, Cornish, Alan, "Using a Native XML Database for Encoded Archival Description Search and Retrieval," *Information Technology and Libraries* (December 2004).

[5] For more details on the ARK standard see: <http://www.cdlib.org/inside/diglib/ark/>.

[6] The NWDA's OAI instance can be accessed from URL <http://nwda-db.wsulibs.wsu.edu/oaiprovider/oai.aspx> (base URL), <http://nwda-db.wsulibs.wsu.edu/oaiprovider/oai.aspx?verb=Identify> (example request using the Identify verb).

[7] For one view of the complexity of mapping Encoded Archival Description documents to the OAI-PMH, see Christopher J. Prom's and Thomas G. Habing's paper, "Using the Open Archives Initiatives Protocols with EAD," published in the *Proceedings of the 2<sup>nd</sup> ACM/IEEE-CS Joint Conference on Digital Libraries*.

[8] Technical reports on the NWDA site including sources for document retrievals, the distribution of search engine driving results, and Wikipedia Articles that link to NWDA finding aids are available at: <http://nwda-db.wsulibs.wsu.edu/reports/>.

[9] The first round of usability testing occurred between July and October 2004 with a sample size of 30 respondents who completed an online form. Five respondents (not directly involved with the NWDA) participated in a second round of testing at the annual meeting of the Northwest Archivists, May 18-20, 2006. Eleven respondents participated in the third round of usability testing which took place between February and April 2007

on the Oregon State University Campus. The third round focused especially on assessing users' perceptions and reactions to finding aids. All three usability testing reports are available on the NWDA/Orbis Cascade Alliance web site:

<http://www.orbiscascade.org/nwda/reports.html>.

[10] While this article was under review, the usability testing group completed a fourth round of testing to evaluate the changes made from the previous studies. These new features include a search box added to each finding aid, an expand/collapse function, the re-sequencing of finding aid elements, a direct link from the finding aid to the relevant homepage of the each repository, and search terms highlighted in yellow. One strong trend noted among the participants was their preference for more images and fewer blocks of text. User recommendations included a site-wide reference inquiry form, links explaining archival terminology, and the ability to customize their view of finding aids. For an executive summary of the report, please see:

<http://www.orbiscascade.org/index/northwest-digital-archives-reports>.

[11] For additional information on the Orbis Cascade Alliance see:

<http://www.orbiscascade.org/>.

[12] This budget is available on the Orbis Cascade Alliance web site at the bottom of the Membership page: <http://www.orbiscascade.org/nwda/membership.html>.

[13] This fee structure will be implemented over two years starting in fiscal year 2009. In 2009 the fees will range from \$1800 for public 2-year academic to \$6600 for public 4-year academic. For extensive notes on the guiding principles, methodologies, and the adopted fee model, see: [http://orbiscascade.org/index/cms-filesystem-action?file=nwda/files/nwda\\_cost\\_dist\\_notes\\_20080312.pdf](http://orbiscascade.org/index/cms-filesystem-action?file=nwda/files/nwda_cost_dist_notes_20080312.pdf).

[14] NWDA membership benefits: <http://www.orbiscascade.org/nwda/membership.html>.