

7. Marshall, Catherine and Gretchen B. Rossman (1999) *Designing qualitative research*. 3rd ed. London: Sage Publications.
8. Merrill, P.F. (1996) *Computers in education*, 3rd ed. Boston: Allyn and Bacon
9. Mutshewa, A; K.N. Rao (2000) Enhancing access through electronic resources: the University of Botswana experience. *Library Hi Tech*, 18 (4): 315-320.
10. Mutuwa, S.M. (2001) Financing public universities in eastern and southern Africa: implications for information services. *Managing Library Finances*, 14 (3):116-132.
11. Nwalo, K.I.N. (2000) Managing information for development in the 21st century: prospects for African libraries, challenges to the world. *66th IFLA Council and General Conference*. Jerusalem, Israel, 13-18 August. (<http://www.ifla.org/IV/ifla66/papers/012-114e.htm>) (Accessed 30 April 2007)
12. Rosenberg, D. (1998) IT and university libraries in Africa. *Internet Research*, 8 (1): 5-13
13. Sweeney, L. (1997) The future of academic journals: considering the current situation in academic libraries. *New Library World*, 98 (1132): 4 -15.
14. Wood, P. A and James H. Walther (2000) The future of academic libraries: changing formats and changing delivery: the bottom line. *Managing Library Finances*, 13 (4):173- 182



ENHANCING THE USE OF ELECTRONIC RESOURCES ON THE TEXAS A&M UNIVERSITY LIBRARIES WEBSITE: HERDING ELECTRONIC RESOURCES INTO SUBJECT GROUPINGS

Esther Carrigan, Robin Sewell, Mary Dabney Wilson and Anne L. Highsmith

Medical Sciences Library, Texas A&M University, College Station, TX 77843, USA

E-mail: Ecarriga@medlib.tamu.edu

Rsewell@medlib.tamu.edu

ABSTRACT: *The integration of five separate websites from the campus libraries at Texas A&M University in 2002 resulted in the creation of a unified listing of all electronic resources, with well over 20,000 entries. This was the first time we were able to provide library users with a single, comprehensive list of electronic resources. But it quickly became clear that it was not very usable. In June 2002 the University Libraries Web Implementation Team (WIT) identified a diverse group of library staff to "explore the possible alternatives for providing intellectual access to the subject content of web-presented resources." Expertise and library experience across the group spanned library public services, technical services and systems. This group became known as the WIT Subjects Group.*

The goals for this group were to develop a system that would:

- Support keyword and subject access to web-listed resources and to our website as a whole*
- Support mapping to broader subject groupings, both dynamically created or predetermined*
- Provide a framework and guidelines for the addition of metadata to all locally developed content or external links*
- Make full use of the specificity and richness of existing content data*
- Maximize the use of existing content data with automatic updating capabilities to avoid labour intensive maintenance requirements. It took over a year to develop and release the system for subject presentation of web resources. Efforts are continuing to improve the maintenance process and to refine the actual visual presentation of subject searching results.*

This paper describes the investigation and evaluation of possible alternatives for the subject presentation of electronic resources; the selection and modification of a system for subject presentation; the implementation of the system including maintenance workflows and continuing challenges; communication and training efforts; and plans and hopes for the future.

Context and background

The University Libraries at Texas A&M University include 5 geographically separate campus libraries: the Sterling C. Evans Library, the general library; the West Campus Library, with a collection focus of business; the Cushing Memorial Library, the home of our special collections; the Policy Sciences and Economics Library; and the Medical Sciences Library, supporting the College of Veterinary Medicine, the Health Science Center and the College of Agriculture and Life Sciences. Each of the five libraries had its own website. The redesign of the Texas A&M University Libraries website began in 2001. A primary goal of the redesign was the integration of the five separate campus library websites from a user perspective. This website integration resulted in the creation of the first-ever unified listing of all campus electronic resources. Like many other research libraries, the Texas A&M Libraries licensing activity for electronic resources blossomed beginning about 2000. Also like many other research libraries, we provided access to electronic serials through our web-based catalogs (struggling valiantly, but unable to keep pace with licensing), as well as through separate hand-built listings of e-serials and e-databases on the libraries' websites. Two of the five A&M libraries, the Medical Sciences Library and the Evans Library, had separate listings of electronic resources. The Medical Sciences Library presented a simple alphabetical title listing of about 1900 titles covering medicine, veterinary medicine and biomedical research. The standard MS Word "find" command was

available for words in the title. The Evans Library had a more sophisticated listing which included about 7,000 titles that could be searched by keyword and had been manually enhanced by subject keywords provided by subject specialist librarians. Keyword enhancement depended upon the interest and activity of the individual librarian. It varied greatly across subject disciplines, resulting in highly variable subject access to electronic resources. Both lists provided a simple alphabetic navigation that allowed the separation of sections of the alphabet as an assist to title browsing. In summary, access provided through the web lists was limited and confusing to users.

The Challenge

By summer 2002, being overwhelmed by the rapidly increasing numbers of electronic resources already licensed and encouraged by the commercial content management system that was powering our website redesign, we were determined to move out of the manual, static list arena. We contracted with SerialsSolutions, an electronic serials management vendor, to provide a comprehensive listing of all electronic serials licensed by the Texas A&M libraries, to include complete full-text journals, as well as article level aggregated titles. Preliminary data feeds from the vendor suggested that the final version of our listing could easily include over 35,000 electronic serials. Early data feeds also made clear the limitations of the vendor supplied data. Our plans for a comprehensive web listing of electronic resources included electronic

journals, databases and books. That brought our expected total listing to over 70,000 entries. Our plans also included mechanisms for record enhancement to improve searching capabilities for the web listing. It was obvious that our challenge was to make this enormous list useful and to provide a subject browsing capability. Like many other research libraries faced with a challenge, we created a "committee" to undertake it!

In June 2002 the University Libraries Web Implementation Team (WIT) identified a diverse group of library staff to "explore the possible alternatives for providing intellectual access to the subject content of web-presented resources." Expertise and library experience across the group spanned public services, technical services and systems. This group became known as the WIT Subjects Group. The goals for this group were to develop a system that would:

- Support keyword and subject access to web-listed resources and to our website as a whole
- Support mapping to broader subject groupings, both dynamically created or predetermined
- Provide a framework and guidelines for the addition of metadata to all locally developed content or external links
- Make full use of the specificity and richness of existing content data
- Maximize the use of existing content data with automatic updating capabilities to avoid labor intensive maintenance requirements

The Process

Several members of the WIT Subjects Group were also members of another WIT group, the WIT E-Workflows Group that was tackling the actual creation, enhancement and maintenance of the web listing for electronic resources. This cross-membership helped keep all the balls in the air that it would take to juggle the subject presentation of these resources. For this paper, we will describe the process after both of these groups had agreed that the entries for titles in the web listing would be programmatically enhanced by the addition of metadata from the MARC cataloging records in our Voyager online catalog system.

The WIT Subjects Group agreed early on that we did not want to re-invent the wheel. We would go to the web to look for good examples of subject presentations for electronic resources. Searching for other Association of Research Libraries institutional websites yielded a couple of interesting presentations. One was the University of Washington. Their presentation of electronic serials by subject at that point in time was done manually and was tailored to the departmental organization of the university. We also found the Columbia University's Hierarchical Interface to Library of Congress Classification (HILCC). As indicated by its title, this arrangement is based on the Library of Congress classification, making it capable of representing all fields of knowledge. Because it is comprehensive and tied to a content rich metadata field (i.e., LC classification number) that we already had in

our online catalog records, it appeared to hold the greatest potential for developing an automated method to create and maintain our subject arrangement. The work done at Columbia to designate ranges of LC classification for a hierarchical subject arrangement was quickly adopted as a starting point for our system.

We adapted the Columbia classification ranges to meet the needs and curricular emphases at Texas A&M. Subject specialist librarians were asked to make changes to the LC Classification ranges to adequately categorize the subject areas and departments that were their responsibility. Not surprisingly, some of the same subject selectors who had been most involved in enhancing the earlier version of our title listing for subject access also took the most interest in and made the most changes both in LC range selection and in the names of the hierarchies. Business and life sciences selectors were the most involved in making changes. In Columbia's hierarchy, only two levels of hierarchy and one category were defined for Veterinary Medicine. Because of the importance of veterinary medicine at Texas A&M's, the subject selectors made significant refinements to this area of the hierarchy.

Once the process to enhance the Columbia system to provide the intellectual infrastructure to group the electronic resources by subjects was underway, we turned our attention to the metadata that would be used to populate the resulting subject groupings. The general plan was to extract data from our Voyager online catalog database. We had already planned to extract title and subject metadata. The Columbia system needs only a call number to direct the records into the correct grouping; this is the "herding" process. The "simple" addition of call number to this extract seemed straightforward. The complication was that many records for electronic resources did not contain call numbers. Metadata was extracted from the MARC bibliographic record on a weekly basis to enhance our web listing. It included:

- Title proper and subtitle (MARC field 245 subfields a, n, p & b)
- Alternate and former titles (MARC fields 246 and 247) provide alternate title access.
- Topical subject headings (MARC field 650) to add subject descriptor access and correlate subject terms to call numbers if needed.
- Subject specialist assigned keywords (MARC field 653) to add subject descriptor access and correlate subject terms to class numbers if needed.
- Call numbers (MARC fields 050, 060, 090) to match records against the subject hierarchy. If a bibliographic record contains multiple call numbers, all call numbers are used. This technique allows a single bibliographic record to appear in multiple places in the hierarchy under different subject groupings. The only exception is that NLM call numbers beginning W 1, the standard number for all journals with no subject content representation value, are not used.

The selection criteria that built the extract file were:

- The Voyager holdings record (mfhd record--MARC format for holdings and locations) must have a location code that begins with the characters 'www' or the location code must equal 'virtual'.
- Both the bibliographic record and holdings record must be unsuppressed, i.e. visible to the public user.
- Ultimately, the record must have an LC or NLM call number or must have subject headings that can be correlated with the LC or MeSH call number indexes. If none of these conditions applies, then the record is not included in the extract, because the purpose of the extract is to represent an electronic resource in the subject hierarchy. Records which lack both subject headings and call numbers cannot be so represented.

The real challenge was to develop techniques for handling records that do not contain LC or NLM call numbers; this is the "herding cats" process. Frequently it occurs, especially with government documents, that the bibliographic record does not contain any LC or NLM call number. Let's refer to these records as "orphans". A technique was devised to try to determine an appropriate call number for orphans based on their subject headings so that they could be represented in the subject hierarchy. Here are the key steps in handling orphans:

- Each topical subject heading or selector-assigned subject term in the orphan bibliographic record is used. If a bibliographic record lacking LC and NLM call numbers has three subject headings, for instance, it could potentially appear at three different points in the subject hierarchy.
- The program identifies all records in the catalog that use the subject heading appearing in the orphan record and determine the LC or NLM call numbers associated with those other bibliographic records, if any.
- In the case of Medical Subject Headings (MeSH), the subdivisions "Periodicals" or "Electronic journals" are deleted from the end of the subject term to eliminate matching against W1 call numbers.
- Matching subject heading against call number index is based on the indicator value in the subject heading, so that LC subject headings are matched against the LC call number index and MeSH headings are matched against the NLM index.
- The program counts the number of times a specific call number is used. The most frequently used call number is then assigned to that orphan for representation in the subject hierarchy.

Occasionally this technique leads to inappropriate matches that lead to incorrect representations, as in the "mystery of the brown tree snake." A librarian noticed that an electronic resource about the brown tree snake

was appearing in the subject hierarchy under birds. Here's how it happened.

- There was an electronic resource titled "The brown tree snake on Guam."
- The record had no LC call number because it was a US government document, so the program used the technique described above to determine a predominant call number for the item.
- The record had only one subject heading, i.e. 650_0 |a Brown tree snake |z Guam. Only two other records in the database had this subject heading; one print and one electronic copy of the same book "The scientific bases for preservation of the Mariana Crow."
- Since the "Mariana Crow" was quite properly classed in QL676.57, QL 676.57 was assigned to the brown tree snake for matching to the subject hierarchy. And that's how the snake ended up among the birds.
- The librarian assigned a 090 call number of "QL666.O636" to the item, which put it back where it belonged. Human intervention with class numbers is the only remedy for this problem.

The remaining piece was the presentation layer. We had built an infrastructure hierarchy based on the Columbia model and a list of electronic resources that was enhanced by metadata extracted from our Voyager catalog. Both were fairly complicated; we needed an understandable display. Once again we turned to the web for an answer. The search for an appropriate model for displaying resources led to an agreement that something similar to the Google directory display, <http://www.google.com/dirhp>, held great potential. This shows 15 or 16 main topics and a selection, but not an exhaustive list, of the sub-topics available under each. The transformation of our subject hierarchy into a Google directory display was accomplished by our Systems' content management experts. The key information and expectations we provided them about our hierarchy were:

- Call number ranges within the hierarchy are unique and do not overlap.
- Bibliographic records can contain multiple call numbers for matching against the hierarchy, so an electronic resource can appear at multiple points.
- The data extract provides a classification and class number for each electronic resource. Systems personnel run a program that determines where that combination of classification and class number falls in the hierarchy and assigns it that placement.
- The hierarchy is refreshed weekly with each Voyager extract.

The Results

We accomplished a subject presentation for our electronic resources supported by a process that met

our goals. The hierarchy and the process to create and maintain it:

- Support keyword and subject access to web listed resources.
- Support mapping to broader subject groupings, either dynamically created or predetermined.
- Make full use of the specificity and richness of existing content data from the Voyager online catalog.
- Maximize the use of existing content data with automatic updating capabilities to avoid labor intensive maintenance requirements.
- Provide a framework and guidelines for the addition of metadata to all locally developed content or external links. A workflow was developed to support and encourage the enhancement of bibliographic records in our Voyager catalog, and ultimately, the web interface to resources.

Reactions to the subject hierarchy vary. Those most closely involved in building the system readily acknowledge its limitations and shortcomings. A fundamental problem is that the subject divisions make perfect sense in an abstract universe of knowledge; the reality of electronic publishing makes it clear that we often have made too fine distinctions between subject areas, resulting in subject groupings with very few postings. The logic used to assign a class number to an orphan record can also result in some very strange groupings. We have not done usability studies specific to the subject presentation of electronic resources in our website.

Training and informational sessions for library staff were critical to the understanding, acceptance and enhancement of the subject presentation approach on the website. A single desk icon was created for staff to report access problems or request access enhancements for electronic resources.

The Future

Our work to explore alternatives for providing intellectual access to the subject content of web-presented resources continues to be useful. We have carried its basic approach into the implementation of Metalib, a product that will provide federated searching capabilities and serve as a discovery tool for our users. Our current subject presentation will be replaced by the Metalib interface. Metalib supports only two levels of a subject hierarchy, which has necessitated the reworking of the "Business and Economics" and "Life Sciences" top levels of our Columbia based hierarchy. These are being reviewed for greater refinement since the current arrangement (squeezing four levels into two) puts too many entries into some groupings. Metalib implementation has also corrected some of the failings

of the subject hierarchy project by including usability testing as an integral part of the implementation plan.

This project had several impacts on the cataloging of electronic resources:

- The desire to list a resource in multiple subject hierarchies created increased interest in and requests for bibliographic record enhancements
- The practical need for a class number made moot the discussion of whether classification numbers are needed for cataloging electronic resources
- The disconnect between the cataloging definitions for integrating and continuing resources and what subject specialist librarians (and users) perceived as a database created an unanswerable challenge. There is no possible automatic identification of a "database" category of electronic resource, which led to the creation of a manual database subject list.
- Local policy about the level of granularity at which we will catalog electronic resources led to the cataloging criteria of a durable URL and the ability to search an entity separately if it is to be cataloged.

Other unexpected results of our efforts and commitment to this project were:

- Renewed and heightened appreciation for the metadata richness and retrieval power of online library catalogs
- Increased demands for quality control monitoring for systems programs supporting the website
- Practical experience that provided a reality check before beginning our implementation of SFX (full text article level direct linking) and Metalib (sadder but wiser phenomenon)

The project to develop a subject presentation of electronic resources was an important step in the evolution of our website; it continues to impact our selection and implementation of other tools for managing access to electronic resources.

References

1. Briscoe, G., Nyberg, C.R. (2003) The Catalog vs the home page? Best practices in connecting to online resources. *Law Library Journal* 95(2): 151-174.
2. Cargille, K.ed. (1998) Lost in cyberspace? Issues in subject access to electronic journals. *Serials Review* 24(2): 101-109.
3. Wilson, M.D. (2001) Flying first class or economy; classification of electronic titles in ARL libraries. *Portal* 1(3): 225-240.