

## **Jupiter: a Tool for Cataloging Web Resources**

Judith Emde  
Electronic Resources/Technical Services Librarian  
University of Kansas

Jill Glaser  
Library Web Services Coordinator  
University of Kansas

Holly Mercer  
Coordinator of Digital Content Development  
University of Kansas

### **Introduction**

A favorite choice of students for discovering information is searching freely available sites on the World Wide Web. According to a recent OCLC study, a high percentage (89%) of college students start with search engines in their studies. Students trust the information obtained through search engines “almost equally” as the “information they get from libraries” (OCLC 6-4). To assist students in selecting quality sites, academic libraries have struggled with various ways of providing some kind of access to freely available Web resources through their home pages. A common principle among libraries is that scholarly Web sites complement existing library collections and careful selection of those Web sites has become a collection development activity (e.g., Librarians’ Internet Index, <http://www.lii.org>). In the process of “collecting” Web sites, supplemental metadata adds value to records describing the resources.

One option for organizing descriptive data for scholarly sites is the creation of Web-based subject bibliographies. The bibliographies are inclusive of all formats including print, licensed resources, and quality Web sites. As to the more traditional routes of bibliographic description, the literature includes arguments against cataloging Web sites using MARC format, which is more appropriate for print, while Dublin Core is “grounded in the internet” (Hunt 4). OCLC’s CORC (Cooperative Online Resource Catalog) is a resource-sharing product for electronic resources utilizing MARC or Dublin Core metadata elements.

When a library decides to catalog freely available Internet resources, who is responsible for cataloging and where do the records reside? A survey conducted by Boydston and Leysen included questions that assumed only cataloging staff participated in this task, although practices vary. Responding libraries “offer patrons a Web-clickable list by both title and subject, in addition to creating MARC bibliographic records” for their catalogs (137). Baruth asks if future generations of search engines will have the sophisticated capability of refining search results and adding the value that libraries currently attempt to create. She raises a valid point about inundating library catalogs with records to Web sites (Baruth 2).

The University of Kansas (KU) Libraries decided to take a somewhat different approach to cataloging freely available Internet sites. Subject specialists would select and catalog the resources through a locally developed database, "Jupiter." The resources would be stored and accessed through a separate repository linked through the Libraries' Information Gateway. After two years of cataloging Web sites, the subject specialists were surveyed to indicate how they use Jupiter in their collection development activities.

### **Rationale and Requirements**

In 2003, the Dean of Libraries charged the Managing Electronic Resources (MER) working party with examining some of the issues surrounding the provision of electronic resources for the KU community. Licensed electronic resources were the primary focus for the group, but subject specialists often identified freely available scholarly resources for inclusion on the Libraries Web site and online catalog. The library administration preferred a consistent treatment of resources whenever possible, so MER investigated methods to identify and catalog scholarly open access electronic resources. MER recommended that librarians select and describe open access resources for their assigned subject areas. The Head of Collection Development and the Assistant Dean for Scholarly Communication, in conjunction with MER, created a document titled "Guidelines for Selecting and Processing Open Access Electronic Resources" (University of Kansas Libraries). The document "provides guidelines to selectors on the best mechanism for identifying open-access resources for cataloging." It outlines what types of electronic open access resources are suitable for inclusion in library collections, who is responsible for cataloging them, and where they will be made available on the Libraries Web site. A team from KU Libraries and Information Technology designed and built Jupiter, an easy-to-use tool for subject specialists to catalog freely available Web resources of scholarly value, which are then available to patrons through the Libraries' Information Gateway.

This paper does not address the public interface (<http://infogateway.ku.edu>) for Web resources cataloged by subject specialists in Jupiter. However, the digital library system in use at the time, Endeavor Information System's ENCompass product, influenced many of the initial decisions. KU used the Endeavor products ENCompass for Resource Access and ENCompass for Digital Collections. The ENCompass cataloging interface did not meet the Libraries' requirements for usability out-of-the-box, and would have required labor-intensive customization. Further, ENCompass lacked bulk metadata export functionality. KU wanted an application that would provide both a simple, intuitive cataloging interface and a method for exporting records from what would become the database of record for Web resource metadata records.

KU had other considerations when designing its system to capture metadata about scholarly Web resources. KU used qualified Dublin Core in XML format for descriptive metadata within ENCompass, and therefore wished to use Dublin Core for metadata in the new application. The Libraries wanted a Web-based input form so software installation would be unnecessary. Another requirement was to use some type of authentication to limit access to the system to authorized users only.

Subject specialists required an easy-to-use interface that would not complicate the process of applying metadata. Most subject specialists at KU do not have cataloging experience, and

learning complex cataloging rules and metadata schemes are not part of their assigned responsibilities. At least initially, we did not want a sophisticated system. Jupiter was seen as a temporary solution until the ENCompass cataloging tool improved. As the application evolved, it was evident that Jupiter was better suited to the local needs of the KU Libraries. Subject specialists needed to know what resources had been cataloged previously, whether by themselves or by others, so Jupiter includes some simple reporting features. These include who created or last modified a record, records associated with a specific subject specialist, all resources cataloged under each subject term, and all resources cataloged under each genre term.

Dublin Core is a widely used, interoperable metadata standard. The elements and element refinements of “qualified” Dublin Core provide for simple description of networked resources. The display order of the fields is reproduced in Table 1.

Table 1  
Metadata Fields in Jupiter Resource Record

<i>Dublin Core</i>	<i>Jupiter Label</i>	<i>Required Field</i>	<i>Repeatable Field</i>
title	Title	✓	
identifier	URL	✓	
title.alternative	Other Title		✓
creator	Creator		✓
publisher	Publisher	✓	✓
subject	Subject	✓	✓
description.abstract	Abstract		✓
resource type	Resource Type/Genre		✓
description	Notes		
description	Evaluation		
(not exported from Jupiter)	Last Viewed Date	✓	
(not exported from Jupiter)	Subject Specialist	✓	✓

The Subject fields require subject specialists to select from a controlled list of values. The subject terms used to describe resources cataloged in Jupiter are part of two-level subject taxonomy, and include the same terms assigned to licensed electronic resources. Within the Libraries’ Information Gateway, the terms present a browse-by-subject approach for locating relevant resources. There are 124 terms arranged under nine broad subject headings. MER selected the broader terms during the first stages of the electronic resources public interface redesign in 2003. The narrower terms were selected from the terms used by the University for the Schools, departments, programs, and study areas within the University. Subject specialists were involved in each phase of subject taxonomy development.

Initially, subject specialists could request additions and modifications to the subject taxonomy; MER would review requests twice a year and implement the approved revisions. Currently, the Information Gateway and the subject taxonomy are under review. Jupiter will be updated to reflect resulting decisions.

Little training is required to use Jupiter. Subject specialists can add new resource records, as well as view, modify, or delete existing resources. Each metadata element on the resource input form

includes an information icon; a popup box with a definition or usage examples appears when a user clicks on it. From the Jupiter update form (Figure 1), users can modify any existing record, whether entered by them or someone else. They can also choose to “expire” a resource, which will prevent it from being displayed in Jupiter, and from being exported to the Information Gateway. Jupiter also includes administrative forms for maintaining the subject and genre terms, and a feedback form for subject specialists to ask questions or suggest changes. There is also a Subject Specialist form to add or update contact and descriptive information about subject specialists. Other items linked from Jupiter include the list of subject terms, genre terms with definitions, and the “Guidelines” document.

Figure 1  
Update Resource Form in Jupiter

The screenshot shows the 'Update Resource' form in Jupiter. The form is titled 'Update Resource' and includes a 'Required fields marked with \*' section. The fields and their values are as follows:

- Title:** UniProt : the universal protein resource
- URL:** http://www.pir.uniprot.org/
- Other Title:** Universal protein resource
- Creator:** European Bioinformatics Institute
- Creator 2:** Swiss Institute of Bioinformatics
- Creator 3:** Georgetown University, Protein Information Resource.
- Publisher:** UniProt Consortium
- Subject:** Molecular Biosciences
- Abstract:** UniProt (Universal Protein Resource) is a comprehensive catalog of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR.
- Resource Type/Genre:** Statistics and datasets
- Notes:** Various search tools and analyses are available from the UniProt members and other sites that complement the UniProt databases.
- Evaluation:** Up-to-date resource
- Last Viewed Date:** 2005-01-10
- Date Entered:** 2005-01-10
- Subject Specialist:** Clement, Susanne
- Last Modified By:** Judith Ernde
- Last Modified Date:** 2005-01-20

The form also includes a 'Submit' button and a 'Printer Friendly' link. The page header shows 'The University of KANSAS' and 'Libraries A-Z'. The footer includes contact information for the University of Kansas Libraries and a copyright notice for 2006.

## **Development Process**

The Libraries chose to use ColdFusion and a Microsoft Access database to build Jupiter. Access is easy to use, and is readily available at KU. ColdFusion simplifies the creation of database driven Web pages and forms, and was already available on the main library Web server. The same functionality could be achieved using any database and Web programming language.

Records in the Jupiter database are extracted in XML format, the metadata “massaged” into the necessary format, imported into ENCompass, and made available to our patrons through the Information Gateway. A form in Jupiter allows us to choose new resources added since a specified date, resources modified since a specified date, or all resources for export. The same options are available for exporting subject specialist records, which are also imported into ENCompass. Since Endeavor has discontinued the ENCompass product, the Libraries will build an interface for licensed databases, Web resources, and subject specialist information using a SQL Server database and ColdFusion. This will make it possible to update Jupiter records and have those updates immediately available in the public interface, without any export or import process.

ColdFusion generates a page listing the Web resources from Jupiter. Then Xenu’s Link Sleuth free link checking software performs a URL link check on the page. The page listing all resources is also a form used to generate a spreadsheet listing the URL, Title, and Subject Specialist associated with each record with a broken link. The information is saved for tracking purposes. The same form is used to generate automated emails to the subject specialists to let them know they need to correct the links.

## **Findings**

In May 2006, there were 732 resources described in Jupiter. The average number of subjects associated with each resource was 2.5, and the maximum, thirty-six. On average, each subject had 14.6 resources associated with it; the largest number associated with a subject was 101. Two subject terms had no resources associated with them.

Anecdotal evidence pointed to mixed reactions to the Jupiter tool, the practice of subject specialists cataloging Web resources, and the Libraries’ presentation of freely available Web resources alongside licensed electronic resources in the Information Gateway. Some subject specialists opposed the added workload, or did not wish to perform “cataloging” tasks. However, others said the process was easy, but resource selection was both time consuming and challenging. Subject specialists apply very different criteria for selecting scholarly Web resources and for choosing the subjects with which to associate them. The authors decided to survey the subject specialists to ascertain what they think of Jupiter in terms of its usability and as a collection development tool.

A survey (<http://www.lib.ku.edu/jupiter/survey>) was distributed to the thirty-one subject specialists at KU. To create and deliver a simple and attractive survey, a three-month subscription to Zoomerang, an Internet based survey instrument, was purchased. Subject specialists were contacted via e-mail to participate in the Web survey.

Twenty of thirty-one potential participants (64%) completed the survey with a good representation across the four discipline groups: humanities, international programs and cultures, science/technology, and social sciences. Statistically, there was little difference in the survey responses across the four groups. An overwhelming majority (95%) agreed that the identification and cataloging of scholarly Web sites is an important collection development activity, yet only 40% actively search for resources to add. Methods for identifying Web resources include finding in journals, on discussion lists, and by chance. The top three criteria for selecting sites were scholarly content, quality of work, and authoritativeness. Other criteria included meeting programmatic needs, updating frequency, and lack of advertisements. Most subject specialists have been selective in the number of sites cataloged in the subject areas for which they are responsible. Since the adoption of Jupiter in early 2004, 45% have cataloged five or fewer sites in each subject; 40% have cataloged between six and twenty sites, and only 15% have cataloged twenty-one or more. When selecting which subject terms to assign, some limit selections to their assigned subject areas. Others select the two or three subjects most closely related to the resource. Still others try to assign all the subject terms that may be related to the resource. This has caused some confusion, as subject specialists are neither prevented from nor encouraged to assign terms outside their bibliographic assignments. One survey respondent noted, “[Jupiter] is becoming unwieldy as hundreds of sites are added. Subject liaisons feel they have lost control of Jupiter since many others are adding sites that liaisons have no knowledge of or interest in.” Only 40% review the cataloged Web resources for updates monthly or quarterly with 40% reviewing annually. The remaining 20% never review the sites. As to Jupiter’s ease of use, 20% of the participants found the tool difficult to varying degrees while 55% find it at least somewhat easy to use. Importantly, is the Web Resources “collection” promoted in instructional sessions or in reference interviews? Sixty percent of the subject specialists refer to Web resources in outreach activities, but many qualified their response. Time and relevance for the particular class were cited as factors.

### **Conclusion**

Most of the issues raised by subject specialists had to do with the public interface and the subject taxonomy. Respondents indicate they cannot adequately classify resources with the current subject taxonomy. The 124 subject terms in the subject taxonomy were too limiting, and they would prefer both more terms and more than two levels in the taxonomy. Many subject specialists expressed concern with the lack of control they have over the resources listed in their assigned subject areas. The current policy leaves decisions up to each subject specialist about which subject terms to apply to resources. Resolution will require subject specialists to work toward a common understanding about assigning subject terms to resources.

There was also a desire for closer integration of Web and other resources. Nineteen percent of respondents would prefer to have freely available scholarly Web resources integrated with licensed databases. The current interface presents databases on one page and Web resources on another. Other comments indicated a preference for integrating all electronic and print resources into the online catalog. Still other respondents were in favor of Web-based subject bibliographies in place of the current Information Gateway. Closer integration of all relevant resources might

raise awareness of the Web resources and make it easier for subject specialists to include sites in instruction sessions.

Several respondents indicated they would like to have usage statistics for the Web resources section of the Information Gateway to know whether their efforts to catalog these resources are benefiting those conducting library research.

Subject specialists at KU consider the identification and cataloging of scholarly Web sites to be an important collection development activity, and find Jupiter an acceptable tool for this endeavor. Survey respondents suggested some changes to the functionality and interface, but generally found Jupiter easy to use. The findings of the Jupiter survey will inform future development of Jupiter and the Information Gateway, as the KU Libraries strive to assist college students in their search for information.

### **Works Cited**

Baruth, Barbara. "Is Your Catalog Big enough to Handle the Web?" American Libraries (2000): 56-60.

Boydston, Jeanne M. K., and Joan M. Leysen. "Internet Resources Cataloging in ARL Libraries : Staffing and Access Issues." Serials Librarian 41 (2002): 127-145.

Hunt, Stuart. "The Cataloguing of Internet Resources." Catalogue & Index 141 (2001): 1-5.

OCLC. College Students' Perceptions of Libraries and Information Resources. 2006. OCLC Online Computer Library Center. 8 June 2006  
<<http://www.oclc.org/reports/perceptionscollege.htm>>

University of Kansas Libraries. Guidelines for Selecting and Processing Open Access Electronic Resources. Lawrence, KS, 4 October 2004.