auroraDL and responding to end-user digital library needs

G Benoit
Computer Science & GSLS
Simmons College
300 The Fenway, Boston, MA USA
01 617-521-2879
benoit@simmons.edu, benoit@fas.harvard.edu

ABSTRACT
This reports on functions and extensions to a digital library record creation, search, and collection-building tools, auroraDL, as a response to academic and professional focus groups’ interest in digital library/content exploration tools.

Categories and Subject Descriptors
H.4 [Information Systems]: Applications; J.5 [Computer Applications] Arts and Humanities, Arts, fine and performing; K.3.1 [Computing Milieux]: Computer Uses in Education

General Terms
Management, Design, Human Factors, Standardization.

Keywords
auroraDL, digital libraries, internationalization, Java, CLIR.

1. INTRODUCTION
auroraDL is a full-featured, modularized, and customizable digital library record creation, digital object management tool, end-user data exploration platform, and test bed for cross-language retrieval techniques, user-created interactive interfaces, and modular programming for dynamic addressing of local resource needs and work-flow issues, relying on Java 1.6, XML (for DC, VRA, EAD, local schema), MySQL, and Apache Tomcat. The app was developed originally in response to end-user experienced with other DL and CMS systems, to be a test bed for many user, information retrieval, and DL research activities. The project homepage is http://gslis.simmons.edu:8080/auroraDL/index.html

1.1 Motivation
Several large research libraries, museums, archives and academic programs in K-20 education have complained about the difficulties they’ve experienced with specific functions of commercial and OpenSource digital library applications. As the result of focus groups conducted between 2008-2010, several themes emerged, addressed in auroraDL. This DL application is being tested in the United States, Iraq [1] and being considered in Greece [Pontika, personal communication].

2. FOCUS GROUPS
2.1 Brief description of main concerns
The focus groups members felt other products lacked in six areas: integrating DLs more successfully in education, ease of collection building and sharing, record product and standards, flexible retrieval results and interactive display, and contextualizing collection elements. The literature about implementing certain popular DL projects [2, 3, 4, 5, 6] focused on technical issues such as platform requirements and compatibility issues of compiler versions and platforms [7], issues of evaluation [8, 9, 10, 11, 12, 13], or issues of non-Latin scripts [14, 15]; few considered the “on-the-ground” implementations and use issues by either non-trained library, museum, and archival staff or users outside such environments [16] and streaming retrieval results to existing interfaces, Flash, XML, and using XML-FO to .pdf files.

2.1.1 Internationalization
Internationalization was one of the primarily reasons auroraDL was created, using Java, relying on UTF-8 and UTF-16 encoding and Unicode-compliant fonts. To encourage participation, auroraDL employs user-translated resource text files (UTF-8, stored in ISO693-3 titled folders (e.g. “arb” for Arabic) for UI labels and option boxes. As a test, untrained users created Greek and Chinese translations of the default English texts, and so were able to tailor auroraDL to their language and script immediately. This suggests the model of end-user translations of such support files, though computationally less effective, is considerably more effective.

2.1.2 Tailoring to local and professional needs
[2] identifies that integrating both OpenSource and commercial DL applications suffer from variety of installation issues (e.g., compiler incompatibility, features not ported). auroraDL relies on a default set of JPanels for data input and to enforce standards {agent, bibliographic, subject, physical description (or image metadata), notes} and user-defined, purpose-specific panels (e.g., for video records), that can be loaded at run time, using JRE Classloader, and user-created and pre-defined schema (EAD, DC, VRA). On-demand loading is used to ensure modularity of system operations and because they are associated with input screens, end-user-created classes provide integration to local work flows and data needs.

2.1.3 Notes – Contextualizing digital objects
Most of the participants who would create, or have students create, records wanted also to annotate individual items with “public” data, that is, brief articles written by experts, and “private” files to store the history of the use of the object and other administrative notes. This option was used by the academic participants to create encyclopedia entries to situate the object’s value to the subject and class. Curiously, the Iraq test case asked to use this option to segregate Sunni from Shia supplied comments [Rashid, personal communication].
2.1.4 Work Flow
There were significant differences in concepts of workflows between the “information professionals” and the academics. The professionals’ perspective emphasized “working linearly” or “horizontally” through the default JPanel menu options, creating complete records, one item at a time. Academics thought in terms of single work functions, or “working vertically”, remaining on a single function, usually subject cataloguing and the using auroraDL’s automatic thumbnail creator and metadata extraction to process entire folders of several hundred images in minutes. They saw this as an “easy way to build collections”. Similarly, combining LCSH, non-English subject tracings and locally-assigned descriptors enabled searching cross-collections, items, and languages. Every item in the selected folder is processed by creating a thumbnail, unique item-level identifiers, and whatever creator, subject, and title data are active in the record, an XML record in the appropriate standard is also created, such as Visual Resources Association’s Core Descriptors version 4 (VRA4).

2.1.5 Integration with existing systems
Another persistent concern was that records created not become lost effort should the institution adopt other information systems. Of the fifteen participants, only two - the museum specialist and professional photographer - knew that records could be exported and ingested into other relational databases and that XML records could be processed into other formats, including those that can be integrated into other systems. auroraDL includes a parser for integrating MARC, archives, and tailorable locally-assigned fields into the collection.

2.1.6 Interfaces and interaction
Finally, all participants noted the difficulty of locating, storing, and then manipulating their own selections from any retrieval system. Consequently, the architecture of auroraDL supports user-defined retrieval results display interfaces and options to stream results as single html, xml, and text data to be ingested into other user-created interfaces, particularly as parameters for applets, Flash, XML-FO, or other RDBMS.

3. FUTURE RESEARCH ACTIVITIES
End-users felt the data to be “closer” to them, altering expectations of systems. Three emphasis areas emerged as continuing projects.

3.1.1 Interfaces and Collection Building
The focus groups were very interested in designing how they want to work, e.g., “light tables”, drag-and-drop interfaces, pre-sort materials, and to create shared collections. These include labeling collections by use history: (term, professor, etc.). Some wanted also to pursue how learning outcomes could be integrated into the items’ records to map to state standards. We will, then, explore the influence of user-participation and greater integration of user values in interface design of DL. Based on group input, the project will release an image-driven retrieval system, to encourage user exploration of data and context.

3.1.2 Work Flow and Integration
Horizontal vs. vertical work behaviors will be examined at three volunteer sites (two public research libraries and one college).

3.1.3 Cross-Language Searching
Using server data, the project will capture queries to map between controlled vocabularies of one language to the CVs and locally-assigned descriptors to consider the retrieval results of mixed-language, mixed-thesauri collections. Future work in this area would aid cross-language retrieval algorithm design to ensure all relevant materials can be identified regardless of source language. The multi-lingual support of the application is being applied to explore the integration of standardized and non-standardized subject tracings on digital object repurposing and in comparing mixed language retrieval algorithm performance [Tsai Bartley, personal communication].

4. REFERENCES