Update: A single bullet was added to the budget section to provide cost information for the linking tool that was developed previously for the Harvard University Archives. The new text is in red.

Title: Link-o-matic: Automating Linking of Digital Objects in Harvard's Catalogs

Submitted by: Schlesinger Library (Paula Aloisio, Amy Benson, Joanne Donovan) ; Harvard University Archives (Kate Bowers, Jennifer Pelose) ; Houghton Library (Emilie Hardman, Susan Pyzynski) ; Baker Library (Benjamin Johnson)

Summary
This project seeks to automate the process of linking digital resources held by Harvard libraries and archives and stored in the DRS with associated metadata in discovery systems (HOLLIS and OASIS). Currently, this is a time-consuming process that must be done by hand and is a significant stumbling block for access to digital resources by end-users (faculty, students, and other researchers). The project goal is to create a solution that will make it possible for a curator to create links between digitized objects in the DRS and Harvard's discovery systems in an automated way.

The Problem
As the Harvard Library digitizes more of its collections, it is important to keep pace by developing automated and streamlined means to facilitate prompt access to these materials. Currently, when collection materials have been digitized, the resource owner receives a formatted report that provides the details of the deposit of these digital resources to the DRS. It is then up to the resource owner to create links to these resources in Harvard's catalogs, specifically OASIS and HOLLIS. To do so, each repository must open each report, copy from it one URN at a time and then paste that URN into one or more records in order to provide access to that digital object to the public. Because this is a time-consuming activity (see statistics below) these links are not added to finding aids or HOLLIS as rapidly or efficiently as they might be. This project goes to the heart of one of the Library Lab's stated goals: it benefits both Harvard library users and staff by making content available more quickly and efficiently. And with the anticipated advent of DRS2, now is the right time to develop such a solution.

Statistics
- Harvard University Archives Manuscripts existing backlog: 650 links / ~110 staff hours
- Harvard University Archives anticipated need in FY2013: 422-447 links / ~75 staff hours
- Schlesinger Library Audio anticipated need in FY2013: 800 links / ~240 staff hours
- Schlesinger Library Manuscripts anticipated need in FY2013: 500 links / ~80 staff hours
- Harvard Business School has created approximate 250 links (~40 staff hours) to date
- Harvard Business School anticipated need in FY2013: 500 links (~ 80 staff hours)

The Rationale
The project would focus on the development of a yet-to-be determined means to automate, to the greatest extent possible, the addition of links to digitized objects from Harvard's catalogs for all Harvard's repositories and for all types of digitized material, including documents, books, and audio files. Comments received on the proposal during the public comment phase indicated that in addition to having an automated process to create links at the time of deposit to the DRS, the ideal process would also allow curator-initiated automatic linking in order to account for materials such as audio tapes that may need additional description after digitization before links could be accurately provided and to make it possible to use the tool for materials that are already deposited in the DRS, but that have not yet been linked to descriptive records.

The project team believes that development of a tool or mechanism for repurposing data from the DRS to create links in Harvard's catalogs would fill a demonstrated need and solve a demonstrated problem that affects many of Harvard's repositories. By reaching out to as many of the repositories as possible, and to staff familiar with Harvard's systems to gain input to inform the development process, we hope to develop a solution that would have long-term value and would be scalable across the University. A focus on a solution that does not require special equipment, software, or skills would maximize the applicability and flexibility of the solution. A meeting of staff from several repositories in mid-July has already produced an outline of the problem and desirable features of a widely applicable solution.

In 2009, the Harvard University Archives (HUA) worked with a programmer to develop a tool that performed some of the functions included in this proposal. It was successful in saving the costs of preparing individual links; however, the tool has limited flexibility and supported only encoding practices at HUA for a specific project. This inflexibility prevented broader adoption and use. HUA's experience with this tool demonstrates both its advantages and the need for such tools to be developed cooperatively.

Ideally, the result of this project will be a lightweight solution that has the potential for large-scale benefits across the University. A substantial benefit to Harvard's user community would be faster access to online materials. It would make it possible for Harvard repositories to fulfill their mission of providing maximum access to collection materials faster and more efficiently.

The Plan

We see this as a relatively short-term project. The first step will be to review the entire process and assess available approaches for addressing the problem with an eye to leveraging existing scripts, tools, and systems, and the new functionalities in the upcoming version of WebAdmin2. The next step would be to identify the approach that maximizes effectiveness. It is possible that something as simple as an update, re-write, or expansion of the existing script written for HUA would be a successful solution. It might also be the case that adjustments to existing processes would lead to the most elegant solution. Project partners will consult with programmers and appropriate staff members from LTS to identify a solution that would work within current repository and system capabilities.

Budget

- Each repository will contribute the time of its staff members for analysis, assessment, and testing of solutions.
- The project would require LTS and possibly Imaging Services staff time to consult with about methods of efficiently adding links to Harvard systems.
- Programmer time, the amount of which is to be determined, to assist with the development of a solution. The project team is unsure how to estimate this cost, but using a per-hour cost for programmer time of $100, and a guess that the project would take 60 hours over the course of three months (28 hours per month), the total cost would be $6,000.
    - The existing HUA tool cost $6,000 to develop in 2009.

Assessment

Development of an automated means to repurpose DRS report data will immediately benefit Harvard's users by providing faster access to Harvard's digitized collections. This benefit will only increase as Harvard Library moves to increase its digitization efforts.

Harvard repository staff will benefit by increased efficiency in a necessary and growing area of work, making it possible to focus on more intellectual efforts than the repurposing of data by copying and pasting thousands of URNs one at a time from spreadsheet to finding aid or HOLLIS record. Success will be achieved if Harvard repository staff have an efficient, effective, and easy-to-use tool or mechanism to add links to Harvard catalogs that works for all repositories and all materials. Such a tool or mechanism would provide Harvard's user community with more rapid access to digitized collection materials while effecting efficiencies across the Harvard Library by saving personnel time and increasing accuracy an efficient repurposing of system data.