

An Introduction to the Catholic Research Resources Alliance

This short essay describes the Catholic Research Resources Alliance (CRRA), its purpose, its goals, its functionality, its vision for the future, and some of its existing challenges. [1]

What is the CRRA?

The ultimate goal of the CRRA is to facilitate research in Catholic scholarship. The focus of this goal is directed towards scholars but no one is excluded from using the Alliance's resources. To this end, participants in the Alliance are expected to make accessible rare, unique, or infrequently held materials. Alliance members include but are not limited to academic libraries, seminaries, special collections, and archives. Similarly, content might include but is not limited to books, manuscripts, letters, directories, newspapers, pictures, music, videos, etc. To date, some of the Alliance members are Boston College, Catholic University, Georgetown University, Marquette University, Seton Hall University, University of Notre Dame, and University of San Diego.

Broad three-step process

To achieve this ultimate goal, a very broad, three-step process has been outlined. In the first phase, the Alliance hopes to centrally collect and index metadata describing relevant materials. In this way the CRRA will be functioning like a gateway to content. Hence, a colloquial name for the CRRA is the "Catholic Portal." Once the metadata is indexed, researchers and scholars will have the ability to search and browse the collection, identify items of interest, and contact hosting institutions for access. This goal, and a means for sustaining it, is expected to be accomplished in less than two years.

In the medium-term, the Alliance hopes to make available digitized versions of some, if not most, of its holdings. This will enable scholars to not only search and browse content but also to download items of interest directly to their desktops, and thus, reduce the need to physically visit libraries, archives, special collections departments, etc. The portal may also enhance the "in person"

experience of researchers, who discover materials through the portal that prompt them to visit archives and libraries for further study. The portal will provide greater access to CRRA materials to people all over the world. These digitized materials may be images, items marked-up in various flavors of XML, or combinations of the two.

The long-term objective is to go beyond the role of traditional digital librarianship. Instead of simply providing immediate access to materials in digitized forms, the long-term objective is to provide tools assisting the researcher in doing their work. These tools will enable the scholar to manipulate the materials and do things like comparing & contrasting themes embodied in newspaper articles, tracing the evolution of an idea through or across sets of letters, extracting all the images from a number of manuscripts to create a virtual flip-book for easy scanning, and discussing with other scholars issues surrounding particular materials or concepts.

Possibilities and challenges for Step #1

As outlined above, the immediate objective is to collect and index metadata describing apropos content from Alliance members and constituents. This section outlines some of the possibilities, but it also outlines a number of the challenges.

Much of the content of the CRRA exists in archival collections. Curators of these collections (archivists) increasingly use an XML vocabulary called EAD (Encoded Archival Description) to describe these collections. These EAD files include the metadata the "portal" requires: Names. Addresses. Publication information. Types of holdings. Depth of holdings. Biographical sketches. Controlled subject terms. Authoritative name headings. Etc. Similarly, much of the content of the CRRA also exists in libraries. Librarians usually create MARC records describing their content. While usually not as detailed as EAD files, MARC records also contain names, titles, subjects, notes, etc.

With EAD files and MARC records, one possibility for accomplishing Step #1 is to provide a Web-based interface for acquiring this metadata, ingesting it into a database, and creating reports

against the database supporting browsable and searchable interfaces. This Web-based interface could simply prompt the archivist or librarian for EAD files or MARC data and incorporate it into the system. Additionally, the Web-based interface could provide a set of fill-in-the-blank forms for the Alliance members who do not have EAD or MARC content.

However the content gets into the system, the browsable interface will allow the end user to scan the content of the portal for things like names, subjects, themes, hosting institutions, content formats, etc., and then retrieve items of interest. The searchable interface will allow the end user to enter simple or complex queries and get back a set of matching documents, in turn directing the user to the content.

A prototype implementation of Step #1 has been created, but it is just that. -- a prototype -- because a number of challenges have been encountered along the way. One of the challenges lies in the interpretations of "standards." While EAD and MARC are certainly standards, there are also levels of compliance and interpretation of rules. Must all EAD files and MARC records have unique identifiers? What shape should they exemplify? To what degree must they use the same controlled vocabulary terms, and how do they get created when not all Alliance members have access to the controlled vocabulary systems, or the systems do not contain the necessary terms? What level of detail are files and records expected to exhibit? One set of best practices may work very well for a single institution's needs but not mesh with the needs/desires of the CRRA as whole.

Other challenges come from the realm of computer technology. What is the "best" hardware and software platform to use for implementing this "portal"? Commercial software is expensive and leaves little room for customization beyond look & feel. Open source ("free") software provides almost limitless customization but often requires increased human capital. The portal is essentially an index. Enter a word or phrase. Get back a list of hits. Click on a pointer (URL) associated with a hit to learn more. With the advent of Google, the expected functionality of indexes ("search engines") has significantly changed in the past few years. To what degree

should the portal compete or compliment things like Google? Alternatively, why do data entry at all? Why not create a specialized CRRA Internet "spider" or "crawler" to seek out and harvest content from the Web and index it the way Google, Yahoo, and Microsoft do?

Other challenges are political. Who are Alliance members? What qualifications do individuals or groups need to exhibit in order to participate? Are there levels of membership, and if so, what are those levels and what are the associated responsibilities and privileges? How is the CRRA financially supported? When it comes to content, what constitutes "Catholic"? What limitations, if any, are made on the basis of where the content comes from, the format of the content, or the subjects the content embodies? The charter CRRA members are addressing these and other questions because the stated intent is grow the Alliance as well as content accessible through the portal.

There are not necessarily right or wrong answers to any of these questions, and asking these questions does not impede the process of the CRRA from moving forward. Instead, these are the types of questions asked by any group of people who go about building just about anything. A bridge. A school. A community.

The Catholic Research Resources Alliance (CRRA) -- the "Catholic Portal" -- strives to continue toward its goal of supporting Catholic research and scholarship. Your participation, no matter how small, is encouraged.

Notes

[1] The canonical address of the CRRA is <http://www.catholicresearch.net/>.

Eric Lease Morgan <emorgan@nd.edu>
University of Notre Dame

June 10, 2008