

Some Thoughts on Authority Control in an International Digital Library

Richard Sapon-White, Oregon State University

Creating a digital library presents challenges on many different levels. Defining the scope of the collection, choosing the technologies to be used in capturing digital objects and harvesting their metadata, and planning an organizational structure for the personnel who will develop and maintain the digital library are amongst these challenges.

In a digital library composed of varied types of institutions from many countries, there are added layers of complexity to these challenges. Aside from differences in language and script, each type of institution has its own standards for describing its collections, such as Anglo-American Cataloging Rules and MARC (Machine Readable Cataloging) in libraries and EAD (Encoded Archival Description) and DACS (Describing Archives: A Content Standard) in archives. In addition, each country may have its own version of MARC or other communication standards. These layers of complexity need not be barriers to sharing metadata amongst institutions, but they do present some thorny problems. I would like to address one rather specific concern for a Maghreb Digital Library for Education, Science, and Culture: authority control.

In the realm of library catalogs, authority control traditionally functions to ensure that access points in a catalog are unique and consistent in content and form. Authority records provide linkages to variant and related headings in a catalog. Authority control contributes significantly to precision and recall. An author search or subject search in the Oregon State University catalog for "Ibn Chaldun" or "Ibn Haldun" is redirected to the standardized form of the access point, "Ibn Khaldun, 1332-1406." The assignment of this standardized access point in the metadata for works by or about this person circumvents problems in retrieval that would be encountered in the absence of authority control or when keywords are the sole method of retrieval.

Effective authority control relies on agreed-upon rules for constructing authorized, variant, and related headings. In the international milieu in which the Maghreb Digital Library would be created, however, there are multiple sets of such rules and authority records already exist in multiple languages and scripts. For example, consider that the Library of Congress' authority record for the 19th century Moroccan sultan 'Abd al-Rahman ibn Hisham has the authorized form of his name (with MARC21 subfield coding) as:

\$a 'Abd al-Ra[m]ān ibn Hishām, \$c Sultan of Morocco, \$d 1790-1859

However, the Bibliotheque nationale de France has the following authorized form of heading (with French subfield coding):

\$a 'Abd al-Rahmān \$e sultan du Maroc \$d 1778?-1859

A Paper submitted for the Digital Library for the Maghreb Workshop

Held in Rabat, Morocco, January 25-27, 2007

The language, form, and content of the heading and differences in MARC subfield coding would need to be reconciled between these two headings in order to use a single authority record in a union database. The process of modifying MARC standards between countries for the purpose of sharing records has been done in the past, but it is something that requires considerable effort and time. The same can be said for changing content rules or merging authority files. The approach of revising rules and records so as to create a single, merged authority file is simply not a feasible solution to this problem.

However, there are other approaches that could prove useful. Recently, the Library of Congress (LC), Die Deutsche Bibliothek (DDB), and OCLC have collaborated to link German and American authority records. The Virtual International Authority File (VIAF) enables German users to see names as established by DDB and American users to see names as established by LC when using the file. With software developed by OCLC, this approach enables each country's authority control system to remain autonomous. Application to a multilingual and multiscript environment is a logical next step.

A different approach was taken by the Linking and Exploring Authority Files (LEAF) Project, where authority records from different institutions were analyzed programmatically and combined into a single, new file. Again, the authority control systems creating the original authority records did not need to modify their practices as the new records were created in an independent file. Yet another approach is taken by ONESAC using RDF and can utilize metadata from MARC21 authority files, Dublin Core records, and the UK Archival Thesaurus.

Whichever method is used, the objective is to utilize already existing authority records to the greatest extent possible. However a second challenge presents itself when authority records do not yet exist. VIAF has attempted to address this by constructing authority records programmatically from bibliographic metadata.

While not insurmountable, the difficulties associated with providing quality authority control in the distributed environment of an international digital library will require careful planning, staff training, and the establishment of best practices.

References

VIAF: The Virtual International Authority File

<http://www.oclc.org/research/projects/viaf/>

Weber, Jutta. 2004. "LEAF: Linking and Exploring Authority Files." *Cataloging & Classification Quarterly*, 38:3/4, p. 227-236.

RDF and ONESAC Shared Authority Control.

<http://www.portia.dk/pubs/ONEX/Misc/ONESAC/RDFandONESAC.pdf>