

Report of the PCC Task Group on the Internationalization of the Authority Files

August 15, 2009

Joan Schuitema, Co-Chair
Barbara Tillett, Co-Chair
Eric Childress
Peter Fletcher
Joe Kiegel
Bill Leonard

INTRODUCTION

Charge

- Investigate the feasibility of designing a model for international participation in a global authority file.
- Assess and document the implications for such an endeavor, including costs and PCC governance policies.

Tasks

1. Review the final report of the Joint Task Group on International Participation in the PCC, including the cited IFLA MLAR and FSCH reports that explain a new view of Universal Bibliographic Control.
2. Review the PCC Steering Committee's discussion of the report of the Joint Task Group on the International Participation in the PCC to identify the implications and barriers to the internationalization of the authority file.
3. Review PCC governance documents and recommend changes where appropriate to enable international participation.

Deliverable

A report listing the pros and cons of implementing a model of international participation in a global authority file not limited to the model described above.

- PCC Standing Committees will review the report and make recommendations for appropriate actions.
- The PCC Policy Committee will review the report and recommendations to determine a final action.

Timeline

Start date: August, 2008

Report due date (original): March 31, 2009

Report due date (extension): September 1, 2009

Membership

Co-Chairs: Joan Schuitema, Barbara Tillett

Members: Bill Leonard, Joe Kiegel, Peter Fletcher, Eric Childress

SECTION 1: BACKGROUND

Findings from Past PCC Documentation and Task Group Work

PCC has long held the belief that requiring consistency in rule applications and guidelines yields records that are more easily shared among members of a community that follows those same rules/guidelines. However, it also has been recognized that such requirements naturally exclude other communities that follow different standards.

The Task Group reviewed two PCC documents: the final report of the Joint Task Group on International Participation in the PCC (Sep. 30, 2003) and the PCC Steering Committee response to this report (revised June 10, 2004).

The Joint Task Group report finds PCC standards are a barrier to greater international participation. With regard to name authorities, the use of AACR2 and, in particular, *Library of Congress Rule Interpretations* is problematic in an international context. LC romanization schemes are also a problem. For subject headings, the requirement to use the LC Subject Cataloging Manual is a barrier. These tools are generally a barrier to greater international participation since they are not used outside of the Anglo-American cataloging community.

In its response, the PCC Steering Committee said that it "will seek to develop flexible standards, practices, and procedures (where possible) that are inclusive of non-US AACR environments" (p. 2). In Action Item 5, it states that "the PCC will seek to collaborate with interested parties to monitor and/or begin to develop procedures and parameters for future implementation of technology that will facilitate exchange of data without the need to impose language, cataloging codes or cultural restrictions" (p. 2). Furthermore, PCC will monitor the progress of the Virtual International Authority File (VIAF), which is described in the present document.

Models for International Authority Control

Five models for international authority control were considered by the Task Group:

- Universal Bibliographic Control
- Parallel records, as practiced at Library and Archives Canada
- MACS (Multilingual Access to Subjects)
- Cooperative Identities Hub
- Virtual International Authority File

Universal Bibliographic Control (UBC): rejected as a model

Description: UBC was promoted by IFLA during the 1970s as a practice for national bibliographic agencies. As a concept, UBC is intended as a worldwide system for the control and exchange of bibliographic information. With regard to authority records, it implies a system where a single authorized form of a heading established by a national bibliographic agency is used everywhere.

Pros: Conceptually UBC is attractive because it seems right to have authorized headings in the original language and script, and to share one heading worldwide.

Cons: UBC is not practical for a number of reasons, e.g., users do not know enough languages and scripts to identify names in a range of bibliographic materials, and the lag time to wait for foreign bibliographic agencies to establish headings is frequently too long. It is a fallacy to assume that a Chinese heading established in Beijing would be readable/usable by Western bibliographic agencies or vice versa.

Universal Bibliographic Control is rejected as a viable model for an international authority file used by PCC, which is in accord with recent (since the 1990s) thinking in IFLA. IFLA recognized that not all libraries in the world could read or use authority data in other scripts used worldwide, so agreed to an adjusted view of UBC and supported the idea of a virtual international authority file, where the national libraries of the world would share their authority data and records for the same entity would be linked or merged.

Parallel Records and Translated Records: rejected as a model

Description: In this model, a separate authority record is created for a given heading in each language/script/transliteration/rule base in the cooperative program. The authority records for each heading exist in parallel in the authority file, and may be linked weakly or not at all. An example of the model is the use of parallel English and French headings in Library and Archives Canada.

Pros: Parallel records may work well for two languages.

Cons: Parallel records do not scale worldwide, with a large number of languages and scripts being created by many institutions on separate systems. Although Library and Archives Canada's current system AMICUS uses dynamic links, records in other systems are typically weakly linked, so a user needs to know the form of a parallel heading in order to find the authority record for it. If an authority file is partitioned by language, then access to parallel forms is curtailed. The model requires the programmatic creation of parallel records, whether they are used or not, and this has a high workload for both record creation and maintenance.

A variation of this model is currently in use within PCC. Some libraries, such as the national libraries of South Africa and New Zealand, use 7XX fields to record local forms of names when their decisions differ from PCC practice, so that their automated systems can flip headings. Presumably they keep a parallel record locally and submit the PCC record with the 7XX for their locally authorized heading.

Translated records have also been used for *LCSH* data, for example in Lithuanian, French, Spanish, Japanese, Korean, and other language versions of *LCSH* usually created in brief form by national libraries. These abridged language versions are not currently linked.

Parallel records are rejected as a viable model for PCC because the model does not scale to a large number of languages, scripts, rules, systems, etc.

MACS model (Multilingual Access to Subjects): rejected as a model

Description: According to its Web site, the MACS Project “was launched in 1997 under the auspices of the Conference of European Librarians (CENL) by four national libraries: the Swiss National Library (currently leading the project), the Bibliothèque nationale de France (BnF), the Deutsche Nationalbibliothek (DNB) and the British Library (BL). Most recently the Library of Congress has joined. The objective of this project is to develop a system that allows multilingual subject access to library catalogs using existing indexing languages. The MACS project currently uses the following subject heading systems and languages:

- RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) for French
- LCSH (Library of Congress Subject Headings) for English, and
- SWD (Schlagwortnormdatei) for German.

Pros: Different subject headings schemes are maintained by the originating organizations and linked.

Cons: The linking process is currently a labor intensive manual process for the most part. The method for individual libraries to contribute to a particular vocabulary depends on the vocabulary maintenance agency. For example, LCSH is maintained by the Library of Congress with the assistance of SACO contributors and others worldwide using LCSH.

MACS is rejected as a viable model for PCC. For PCC to expand globally, structures would need to be established to facilitate members contributing to a particular system of their choice.

Cooperative Identities Hub: not recommended at this time

Description: The final report of the RLG Partners Networking Names Advisory group titled *Networking Names* (<http://www.oclc.org/programs/publications/reports/2009-05.pdf>) describes the components of a “Cooperative Identities Hub” that would serve as a place where information about persons and organizations, corporate bodies, and families could be aggregated and made available at the network level. Contribution and maintenance of entities in the hub would be based on social networking models. Users as well as machine applications could extract relevant information for reuse in their own contexts as well as contribute new and updated information.

Pros: This social networking model is conceptually attractive and could serve as a platform whereby PCC expert catalogers would be authorized to not only contribute entities, but also review and update contributions to the hub by others. It could also increase metadata creation efficiency by making better use of existing metadata. It also exposes information about entities established outside of libraries and brings that information into a network level tool.

Cons: It is currently limited to names. Boundaries for PCC participation could be difficult to define, implement, and monitor. Although not yet fully developed as a model, we believe it could have potential, and PCC should follow its development.

The Cooperative Identities Hub is not recommended as a model at this time.

Virtual International Authority File (VIAF): recommended model

Description: Currently the VIAF is a beta service conducted jointly by OCLC Research, Library of Congress, Deutsche Nationalbibliothek, Bibliothèque nationale de France, the National Library of Sweden, the National Library of the Czech Republic, and the National Library of Israel, Bibliotheca Alexandrina (Egypt), Vatican Library, National Library of Portugal, National Library of Spain, National Library of Australia, ICCU (Italy), and the Swiss National Library to match and link authority records for personal names from the partner institutions. [Note: An additional 9 applications are in progress for 2009.]

VIAF provides an environment that encourages loose coordination and reuse of work by various agencies and cooperatives. Record contribution is aggregated through national authority files or similar high-level regional databases from institutions that are partners in VIAF. This is a cost efficient means to sustain the data necessary for this building block for future systems (to be combined with other building blocks as needed).

Pros: VIAF supports headings that make sense to local (national/regional) user groups. It does not require headings to be created unless needed within a cataloging community (i.e., has a low overhead in comparison to other models). The VIAF relies on automation to do much of the work of retrieving and linking headings from different cataloging communities, and is thus cost effective. VIAF presents users with a freely available, single view of authority data from multiple sources.

Cons: While the conceptual model is clear and scales well, there is yet no practical specification/application of it that can be implemented (see suggested functional specifications/applications for VIAF in Appendix). Currently, it is limited to personal names and geographic names, although planned to extend to corporate names and uniform titles.

The Virtual International Authority File is recommended as the preferred model for PCC to encourage regional contributions from international sources using shared cataloging codes, interpretations, and encoding practices.

PCC may wish to work more closely with existing agencies, or in the absence of partnering with an appropriate agency, establish regional/national centers to coordinate the authority file creation and maintenance for the respective region(s) and contribution of authority records to a central system (presumably to OCLC for either their “Networking names” or for VIAF itself).

SECTION 2: RECOMMENDATIONS

RECOMMENDATION 1: PCC should evolve to become a global consortium of agencies, or “branches,” that embrace many cataloging traditions.

Description of a Proposed Structure for PCC

To date, PCC has consisted of a single community of catalogers who share one set of standards (*AACR2*, *LCRIs*, *LCSH*, etc.). As the former Joint Task Group on International Participation found, this presents a number of barriers to the expansion of international participation in PCC. As a solution to the problem, the current Task Group proposes a structure based on the concept of national or regional branches that share common standards within these branches. In brief, PCC should expand to a group of branches, each of which shares a set of standards among its regional members. These branches will share authority data with other branches by using tools that link across language and scripts, cataloging rules, transliteration schemes, etc. This might be through VIAF (for personal names for now) or through similar services now emerging, such as OCLC/RLG’s “Networking Names”.

PCC then becomes an international consortium of branches that preserve local standards that best suit their local users. The current core of the PCC, which follows *AACR2*, *LCRIs*, *LCSH*, *LCC/DDC*, MARC 21, and ALA/LC romanization, would be one branch. Other branches would center around other national or regional standards, with a focus on a particular language/script but following a basic PCC model. The branches would also share training documentation and courses in the language/script of the branch following the PCC models. With the development of RDA, it may be possible that all PCC branches could agree to use this international cataloging standard along with MARC 21.

Change to PCC Governance Document

If the Policy Committee decides to expand PCC into a consortium of branches, then the PCC Governance Document (<http://www.loc.gov/catdir/pcc/governance.html>) will need a thorough review, which is beyond the scope of this Task Group. Nonetheless, we make one recommendation for change in order to expressly enable participation of the PCC in international authority activities.

In Section II.C on NACO, the following sentence should be added as the third one in the paragraph: "PCC may build regional name authority files that would be accessible through the Virtual International Authority File or a system built on a similar model."

The full paragraph thus reads:

"As members of NACO, participants contribute authority records to the national authority file. An individual institution may join this program, or a group of libraries with a common interest may form a funnel project to contribute via a

coordinator who will represent the funnel participants. PCC may build regional name authority files that would be accessible through the Virtual International Authority File or a system built on a similar model. For-profit organizations participate in NACO at the Affiliate level."

Assuming PCC embraces a new role as a consortium of branches, a process providing appropriate representation of libraries in those branches will need to be developed and implemented.

PCC will need to consider and plan for the possible additional costs of supporting a truly global consortium. These costs may include paying for added translations of documents, PCC-supported travel, international conference calls and more.

RECOMMENDATION 2: PCC should support and adopt the VIAF (Virtual International Authority File) as critical infrastructure for the global exchange and enrichment of library authority data.

A New Authority Exchange Vehicle

VIAF as designed and released in beta, and anticipated in enhanced form in the future, will support automatic linking of records in diverse authority files and provide a variety of machine-to-machine services critical to coupling now de-coupled work of a wide variety single-agency and multi-agency cooperative authority programs worldwide, and other parties.

The potential benefits of widely available access to linked authority files are many. Among those potential benefits:

1. Reuse of authority work between communities: For example, an English-language cataloging agency encountering a Spanish language author's name that needs authority work may be able to leverage the intellectual work of Spanish-language agencies by:
 - Reviewing the differentiation of the name in multiple authority records for multiple persons bearing the same name
 - Importing (and possibly modifying) preferred form(s) and cross-references, other researched data such as birth, death dates, etc. from the Spanish-language file(s). [Note: this may require development work to optimize the process] and even link a new English-language authority file record to the correct Spanish-language record.
2. Sophisticated leveraging of equivalence within systems: Searchers potentially will be able to search on a form of name favored in their preferred language community, and systems leveraging VIAF will be able to automatically retrieve all relevant hits including records that may bear alternative forms of the name (including in different scripts). The same system(s) may automatically present all references to the name in the form expected by the user regardless of the form actually present in the underlying records.

For suggested functional specifications and applications of VIAF, see the Appendix.

NACO

For authority headings and records for names of persons, corporate bodies, families, and uniform titles, the Task Group proposes that the Virtual International Authority File be the tool to share data for personal names freely. In each branch, contributions of authority records would be made to a central point, such as a national library, and then shared worldwide through the VIAF. The technical capabilities of the VIAF and applications written for it would make much of the difference in cataloging rules, scripts, etc., invisible to the user of the data. This can be done now with the current VIAF structure and perhaps could be expanded beyond the control of national libraries through the OCLC/RLG projects for “Networking Names” (e.g., a future Cooperative Identities Hub as mentioned above). We note that the PCC structure must not mandate exclusive use of OCLC.

Note that the VIAF is a tool available to the entire world. Should PCC adopt it as a component of a new regionally based structure, the VIAF would still be employed freely by other libraries that choose not to join PCC. In other words, PCC use would be a subset of VIAF use, and not the other way around.

PCC’s relationship to VIAF

For a PCC version of this model, partners in regional branches of PCC sharing the same language/script/cataloging rules/rule interpretations would cooperatively build and maintain their own authority files and contribute records to VIAF. The current PCC partners do this now for current contributions to LCNAF through the Library of Congress.

PCC will need to consider its role with VIAF governance. As the VIAF consortium currently stands, PCC member agencies are represented by the Library of Congress, a principal in VIAF. PCC agencies as members of OCLC are also represented by OCLC, another principal in VIAF. If PCC embraces the VIAF as a critical feature of PCC’s global exchange mechanism(s), should PCC have direct representation in VIAF’s governance?

Investment in infrastructure: Currently the VIAF system is hosted by OCLC Research. Additional investments will be required to ensure VIAF fully realizes the role(s) envisioned in this document. Additionally, fully leveraging VIAF as an aid to NACO authority work would be facilitated by more sophisticated capturing of the available authority data and enabling uploading of new data. What role can/should PCC have in encouraging OCLC and ILS vendors to make investments in VIAF, system authority modules and other systems/services to fully support and leverage VIAF-advantaged features?

RECOMMENDATION 3: PCC should make no changes to SACO at this time.

SACO

For authority headings and records for subjects, the Task Group proposes no change. That is, those institutions using LCSH would continue to follow current SACO practices, and PCC membership should continue not to require SACO participation.

Current explorations to linking LCSH to its counterparts in other languages should be encouraged. Sharing SACO documentation and training materials so they may be translated into other languages also would enable wider use of LCSH.

RECOMMENDATION 4: PCC should form a task group to identify and recommend appropriate changes to NACO practices that unnecessarily limit the utility of LCNAF data for reuse in a linked-data environment.

NACO practices and the future

As internationalization and modernization (in the form of RDA and support for Linked Data) are embraced, questions arise about whether NACO practices as presently written and implemented will need to be adjusted. OCLC Research has identified unfortunate, unexpected consequences to some practices and problems with NACO normalization rules, and other parties have or no doubt will identify other issues.

The Task Group recommends that PCC establish an expert group to identify and recommend appropriate changes to NACO practices that unnecessarily limit the utility of LCNAF as data suitable for reuse in a linked-data environment.

SECTION 3: CONCLUSION

The current Task Group agrees with previous findings that the structure of PCC is an obstacle to internationalization and has recommended a new structure composed of regional branches. The Task Group reviewed several methods of sharing authority information internationally and found that only the Virtual International Authority File model is a viable option for use in a restructured PCC. Recommendations were made for steps needed to move the VIAF from a beta project to a functional tool.

As a next step, this report should be reviewed by the Standing Committee on Standards and other PCC Standing Committees. It should then be forwarded to the Policy Committee and the Steering Committee. The scope of potential changes to the PCC is broad, and the policy committees will need to determine a process for discussion and decision.

APPENDIX

SUGGESTED FUNCTIONAL SPECIFICATIONS/APPLICATIONS FOR VIAF

The Virtual International Authority File (VIAF) is an international project to link name authority records that has great potential for aiding discovery of authority data, and with further add-on applications for also reaching to bibliographic records and digital resources. We attempt to highlight uses for the VIAF model and outline an architecture needed to support them.

Architecture

We envision an architecture with three layers. The bottom level is the VIAF itself (and future extensions of systems in a VIAF model). [Note: VIAF is available as a beta version service (<http://viaf.org/>) with searching and downloading capabilities for MARC formatted records and plans to expand to UNIMARC and possibly RDF]. Additional partners continue to join the project, which is extending coverage of name authority records throughout much of the world.

The second level is an application programming interface (API) (or an internet protocol). The API is a published set of instructions (computer commands) that allows more sophisticated searching and exporting of information from the VIAF and allows machines to utilize the authority data, such as for exporting the authority records to local systems or to provide data for specific Web services and applications in future linked data systems. Essentially, the VIAF layer itself would be invisible to end users of library systems, but visible to those contributing and maintaining the authority data and accessible to machines using the data. It is the API on the second level that allows advanced access to and reuse of authority data. A crucial step in implementing this architecture is to specify a rich API that supports a wide range of uses. Another crucial step is to provide programming support to implement and maintain it.

The third and top level is a variety of Web applications, which will be written by different individuals, organizations, and companies to use the authority data for specific user tasks and displays.

Examples of Uses

Here are some of the possible uses we see:

1. *Library Catalogers as Users:* Catalogers query an application about authority headings or persons, corporate bodies, geographical jurisdictions, and names of works/expressions (uniform titles). If a cataloger keys in a heading or a control number, and a set of parameters (rules, language, script, transliteration), a VIAF application returns an equivalent heading (or a complete authority record to be used locally), when one exists. If no exact match results, the application may return near matches or matches defined by other parameters, as determined by the cataloger. The application also allows general searching by phrase or keyword including qualifiers by parameters, and it returns lists of matches.

2. *End-User Displays on the Semantic Web:* End users in China search in their own script and through connections with VIAF on the Web, locate resources from libraries worldwide that have been cataloged in non-Chinese languages and scripts, but contain Chinese texts or other resources they can read, see, hear, etc. The switching of display characters is enabled by VIAF links of authority records for the same entities worldwide.
3. *Local Library Catalogs:* An ILS uses the application to convert headings in record (and index) displays. At the current time, many catalogs have foreign language interfaces, but only for interface text. A VIAF-based application would allow the user to choose the language in which some cataloging data display. For example, a user could see Arabic script headings or German headings with a different Romanization of Cyrillic.
4. *OCLC:* WorldCat contains records from national libraries with differing rules, etc., for name headings. A VIAF application would allow searching and displays to be unified without changing the underlying data in WorldCat records. A user in Connexion or WorldCat.org could specify parameters for headings as defaults (e.g., English language, AACR2, LC Romanization), and headings would search and display this way whenever an equivalent heading exists in VIAF. Also, the linking feature in Connexion could be expanded to WorldCat records from non-US cataloging agencies.
5. *Digital Repositories:* Repositories such as those maintained by universities could provide alternate displays of headings by using VIAF.
6. *Abstracting and Indexing:* Indexing services could supplement their authority control systems by using VIAF and provide their users with alternate displays of headings.

Specifications

Design a user-facing API or an internet protocol as a Web service to a VIAF-like database. These specifications may be implemented by an API or an internet protocol, depending on the best technical approach (as determined by people with more knowledge than the Task Group). A layer of applications would be built upon the API/protocol, written for specific uses (see examples above).