SILS Cataloging & Metadata Harmonization Pilot Report (October 2019)

History

The Cataloging & Metadata Harmonization Pilot was established in June 2019 and met monthly through October 2019. The membership was made up of library staff members at each of the 10 UC Libraries and Shared Cataloging Program (SCP), who were familiar with the cataloging practices around local fields. The pilot was intended to both test the process of attempting harmonization across the campuses, as well as to try to determine possible areas for harmonization. Thus, the original scope of the pilot was set as investigating alignment of local fields, specifically the 9XX range used in bibliographic records. This focus was chosen because it was initially anticipated to be relatively uncontroversial and manageable in a short period of time.

The pilot group was facilitated by the chair of the SILS Harmonization (SILS H) group, reporting to the SILS Working Group. This cat/met group shares a relationship with the ILS Data Cleanup (ILS) group, whose work is to identify tasks that can be done prior to migration to a SILS. Several members overlap between SILS H, the pilot group, and ILSDC. This pilot group's work concludes with this report and any work that was not completed by this group will be turned over to the ILSDC for management.

Observations

We recognized early on that harmonization is a large undertaking with the potential to become overwhelming with regards to details and scope. Because of this, the group felt that one of the main drivers for considering harmonization, whether of individual fields or as a larger concept for SILS as a whole, was the impact that it would have on various stakeholders. There was a strong consensus that harmonization should be undertaken when it can be done in a meaningful way, especially with regard to users. Local fields that directly affect user interaction with catalog data provide some of the most potentially impactful areas for harmonization and would be well worth the effort. By "users" we are including those who work with the catalog (internal library stakeholders) as well as those who interact with the catalog via OPACs (students, staff, faculty, researchers, general public). This understanding of impact manifested in the creation of the first principle of the Harmonization Principles.

Along those same lines, while final decisions about how to format information in a given field are generally made by cataloging and metadata staff, those decisions are made in consultation with various stakeholders in other areas. For example, harmonization of fields that may impact work done by Acquisitions or Public Services staff will require significant consultation with representatives from those areas. In a SILS environment, this may mean input at both local and consortial levels.

Campuses heavily use 9XX fields for workflow and project management. The group made a distinction between the responsibility for use of a given 9XX field in a given workflow and the wholesale re-evaluation of that workflow. For example, whether cataloger statistics are recorded in a 910 or 911 is distinct from whether those statistics need to be recorded at all. The recommendations surrounding harmonization don't presume to prescribe what data campuses should record in their own bibliographic records, but instead provide guidance about where to record it in a shared environment.

In addition, the technical limitations or restrictions imposed by the new SILS will have a great deal of impact on future harmonization efforts and guidelines. We consistently felt that many decisions around full-scale harmonization depend on which system we get. The various current practices surrounding 9XX and other local

fields are often the direct results of the various systems the campuses use or used in the past, and so the solution may be just as system-dependent.

At the beginning of the pilot, limiting the scope to the the 9XX fields seemed like a reasonable way to approach harmonization, however it soon became clear that in order to do the work, we would need to understand the full ramifications of all local fields. While doing the inventory of local 9XX practices, it became clear that a subset of those fields included data that may be represented instead by other bibliographic MARC fields. Unfortunately, our pilot's focus on 9xx fields precluded a thorough review of the same category of information if it was recorded in other marc fields by other campuses. For example, the same information that is represented at one campus in a 9XX might be represented in another as a 69X or 79X. Also, the focus on 9xx unintentionally limited Merced's participation in the project because Merced does not use 9xx fields—although it may use some of the same categories of information other campuses record in 9xx fields.

We therefore decided to focus on individual/groups of field functions versus strict codes. For example: collection and package names vs. 973 fields. We developed a set of categories (see Table 2 below) to identify the function that was being performed by the data in the 9XX field. Aggregating fields by function allowed us a new opportunity for analysis, and improved our ability to make recommendations about changes. One of the goals that became clear during the development of the Harmonization Principles was: if we have similar functions we should be open to doing the same thing in the same way. By creating functional categories we hope it will be easier to identify areas for future alignment.

Finally, we recognize that it is possible to use the SILS migration for an opportunity for cleanup and to harmonize that way. We have an opportunity now to identify and tackle "low-hanging fruit," we are going to have to touch a lot of workflows in the midst of migration, and it may be cheap enough to make changes now. That said, it is also getting easier to do batch work on this data as technology improves - it may be possible to do cleanup in the new system much easier than in 10 different systems. The work of a future harmonization team will therefore be to determine how to prioritize these tasks and when to actually undertake them.

Recommendations:
1. **We should avoid using the same 9XX across campuses** if it will create collision, entangle data, or have a negative impact on configuration of discovery in our consortial environment. For example, if two campuses both use the 910 field, but for different purposes. We recognize that the extent to which 9XX fields are harmonized is system-dependent. If we have a system that allows for multiple campuses to define, migrate and manage their 9XX fields (on the backend and for discovery and data reporting) independent of other campuses, then there will be less need to harmonize all of our 9XX fields.
2. **A subset of 9XX fields should be reserved for systemwide use**. A subset of the 9XX fields (listed below in Table 1) have been identified by this group as not currently being in use at any of the campuses. Campuses should be instructed: (1) to verify that they are not already using any of these 9XX fields in their current systems (for current or legacy data) and (2) not to begin using these fields unless otherwise instructed as part of the SILS implementation. For now, the communication and tracking of this work will happen as part of the ILS Data Cleanup group. The appropriate decision-making bodies should determine and document how these fields will be used in the new SILS going forward.
3. We recommend that **local fields used as access points be converted to the appropriate MARC fields**. We recognize that these are not limited to 9XX fields and therefore were partially out of scope for this pilot project but there is a lot of overlap here. Harmonizing access points that are to be used for discovery

purposes will create efficiencies in configuration for search, display and harvesting of data within our discovery layer. For now, the communication and tracking of this work will happen as part of the ILS Data Cleanup group. For reference, we have attached documentation from UC Berkeley on local access points as a suggested starting place for making these changes: https://asktico.lib.berkeley.edu/copy-specific-information-in-millennium/.

4. **9XX fields that are being used for creating "workarounds" in the current systems (i.e. marking records for deletion) should be reviewed closely** with the expectation that they will not be used in the new SILS unless they are meeting a need for which there is no other solution. Ideally, we want to make use of our new system's capabilities in efficient and sustainable ways. The review of these will need to take place when we know what system we are going to in order to determine how that system is designed to handle each of those needs. This work will therefore be done by future harmonization teams.

5. We learned that in order for harmonization to be successful any limits on the **scope of the topic need to be flexible**, and that **harmonization members have to be empowered to consider all necessary facets of the problem space**. By restricting our analysis to only 9XX fields, we limited our ability to fully understand the needs of local fields. We recommend that future harmonization teams be authorized to expand or contract the scope as needed.

6. We recommend that **the use of 9XX fields in the SILS be documented**, including identifying and explaining the function of the data in the field, the form of the data in the field, and the primary stakeholders for the function of this field. This is in line with MARC standards for other non-local fields, and would make future analysis and changes easier to make.

Membership Roster

Marcia Barrett, UCSC
Paromita Biswas, UCLA
Randal Brandt, UCB
Catherine Busselen, UCSB
Jared Campbell, UCD
Rebecca (Becky) Culbertson, SCP
Josh Hutchinson, UCI

Wanda Jazayeri, UCI
Yoko Kudo, UCR
Cathleen Lu, UCSF
Elizabeth Miraglia, UCSD
Latasha Means, UCM
Lisa Rowlison de Ortiz, UCB

Christine Barone, SILS H
Lynne Grigsby, SILS H
Cathy Martyniak, SILS H
Caitlin Nelson, SILS H

Appendices
Appendix A: Cataloging & Metadata Harmonization Pilot - Statement of Work (below)
Appendix B: Inventory of UC 9XX Local Practices (XLSX)

Table 1: 9XX fields with no known UC Libraries data

906	915	921-933	962-969	989
908	917	937	976-978	991
909	918	941-944	983	992
914	919	951	986	996

Table 2. Cataloging and Metadata Harmonization Pilot: 9xx field categories

Acquisitions (see also: Vendor fields)	Information used in the acquisition process (often from vendor, may include gift information). Information may be encoded. **Example:** $b 54.99 ‡g 1 ‡z USD
Statistics	Contents may include cataloger's initials, load source, etc. Purposes may include workload measurement/production evaluation, PCC statistics, etc. **Example:** cmc mono $b pvf $c 20160421 $d 2 $h e $k lccoop/cmc crs ‡b pxb ‡c 20190626 ‡d 3 ‡e in ‡g 856 ‡h ca ‡i 1n ‡j ‡k rcv:20160626
SCP	Code or string pertaining to UC's SCP and related loading processes and record elements. **Example:** 20160419|bSCP Monographs
System numbers	Identifying number unique to the local system. These are usually indexed. These numbers may be historical and include legacy data from previous systems. **Example**: UCRb5000220x $b n $c -
Access points	Searchable string in the OPAC/discovery layer. Subject, name, collection, etc. These are generally indexed. Other campuses may be placing these in non-9xx MARC fields $5 (such as the 690, 710 or 790 fields). Due to the scope of this project, we don't know how many UC campuses provide this functionality using non-9xx fields—further investigation required in a future project. **Example**: $a Western Art and Artists
Integration point	Needed for functional integration with an non-ILS system such as Aeon or the campus billing system. **Example**: Special Collections (Rare)|aF864|b.J216 1953|so *(this allows the call number to be reported through the Aeon title request system)*
Workflow marker	Codes or notes that indicate status of record in a workflow. Permits collocation of batches of records; may have some overlap with 'Project note' category. These may be indexed. **Example**: |a<yyyymmdd>|bUnlinked analytic ; |a<yyyymmdd>|bGovernment publication

Project note	Contains a note identifying a project or a workflow element within that project. The project may be short-term or long-term. May have some overlap with 'Workflow' category. These may be indexed. **Example**: |a govdoc backlog cataloging 2019 |b 31175037249318
Vendor info (see also: Acquisitions)	Local information that is sent in the bib record from a vendor. Usually used to populate other fields in other record types (eg, order records, item records) within the local system. Sometimes this information is not used at all. **Example:** $b 54.99 ‡g 1 ‡z USD

Appendix A: Cataloging & Metadata Harmonization Pilot - Statement of Work

Prepared June 4, 2019 by SILS H

Introduction

Harmonization is a process where the campuses, RLFs, and CDL come together to investigate and decide the practices, procedures, policies, and workflows that need to be aligned within the consortium in order for the SILS to be successful.

Why do a pilot? Moving forward, harmonization will involve complex topics with complex histories. Before tackling these tough topics, SILS H is proposing to run pilots on two "easier" topics to develop the process. We will tackle real topics with real stakes as well as using the opportunity to improve the harmonization process as well. In other words, pilot success is just as much about creating a successful process as it is about the outcome of the harmonization recommendation.

Reporting

The Harmonization Team will report to SILS H. SILS H reports to the SILS Working Group (WG). SILS H Chair will update the Working Group regularly on pilot progress.

Pilot Work and Scope

This team will:

1) Identify specific local bibliographic record fields that can be investigated for harmonization.
   a) Currently proposed: 9XX fields or subset thereof.
2) Compare current local practices for those fields.
3) Identify stakeholders for those local practices (who has input, authority, etc.)
4) Identify which fields would be appropriate for alignment, if any (Recommendation #1).
5) Of those candidate fields, recommend how alignment among the UC Libraries consortium would best proceed (how & when) (Recommendation #2).

Pilot Deliverables and Timeline

1. Pilot topic chosen and approved by Working Group (early June 2019)
2. Harmonization team members identified (early June 2019)
3. Deliverable #1: Inventory of local practices and stakeholders (due early July 2019).
   Of the set of fields:
   a. What fields do we use
   b. How do we use those fields? What data goes in those fields?
   c. Why? For what purpose?
   d. What happens to that data? What processes are run, reports, etc. Is that info made available to the public in discovery?

    e.   Who has a stake in this information? Who needs this info? Who decides what changes can be made to these fields?

        i.   "Deciders"

        ii.   End users - are we going to try to do UX testing?

4.   Compare and investigate local practices with an eye toward Recommendation #1 (due mid-August).

    a.   Recommendation #1 answers the question: "Is it worth aligning these fields across the UC Libraries? Why or why not? If not, might it be worth revisiting in the future?"

5.   For those fields that can be aligned, discuss with stakeholders, and prepare Recommendation #2, (due end of October).

    a.   Recommendation #2 answers the question: "For those fields that should be aligned, what is the best way to proceed to do so? Why? When?"

    b.   Who needs to be consulted to make this recommendation, and why?

        i.   Stakeholders with input on how that data gets used

        ii.   Decision-makers with input on changing the fields at all

        iii.   Stakeholders who would need to be the ones actually changing their workflow

        iv.   etc.

Commitment / Frequency

Members should expect to meet monthly, unless otherwise needed and decided by the group. Some writing and editing will be required to produce the recommendations. Work will be needed to compile current practices and to engage with local stakeholders. Estimated effort is 1-2 hours per week.

Membership of the Pilot Harmonization Team

Caitlin Nelson, SILS H Chair, facilitator
Other members, experts representing the 10 campuses, the RLFs, CDL Acquisitions, and the Shared Cataloging program as appropriate.

Communications and Workspace
- Listserv
- Slack channel
- Google Team Drive