

OCLC Response to the WorldCat Local Special Collections and Archives Task Force

Mela Kircher

April 3, 2009, Revised

Summary

The WorldCat Special Collections and Archives Task Force final report includes an executive summary with four major recommendations. Specific issues identified by the task force are prioritized as high, medium, and lower significance. A separate group of problems with data was also identified.

This response to the report begins with background on the task force and then treats each major and specific recommendation. OCLC provides a timeline for development or other activity for all the major recommendations and nine of the fourteen specific recommendations (numbers 1, 2, 3, 4, 5, 7, 8, 10, and 12). The timelines range from April 2009 into the future. Two recommendations (numbers 6 and 9) have been assigned problem reports to be investigated and resolved as part of monthly maintenance. Two recommendations (numbers 11 and 14) have no OCLC action, as the system is working as designed. Recommendation 13 has been fixed.

OCLC is in the process of refining its FY2010 (July 2009-June 2010) development priorities. The Task Force's recommendations are included in a division-wide process and are prioritized along with all the other development for WorldCat.org, WorldCat Local, Group Catalogs and Navigator. As such, not everything that is a high priority for the Special Collections community will constitute a high priority in the inclusive development list. Also, some of the issues will be broadened beyond the Special Collections focus. For example, recommendation one, to display all variable fields, will be broadened to consider which additional variable fields for all types of materials need to be displayed.

OCLC looks forward to receiving further information from the task force in response to our requests for examples. In addition, OCLC will be pleased to respond to any questions that Task Force members may have after reading this report. Please contact Matt Goldner, Executive Director, End User Services at goldnerm@oclc.org.

Background

In August 2008 OCLC convened the WorldCat Special Collections and Archives Task Force to recommend improvements to the discovery of such materials in WorldCat.org generally and in WorldCat Local (WCL) implementations specifically. The task force was made up of eight experts from libraries with WCL and four OCLC staff. Biweekly meetings were held from September through November. The task force final report was submitted to OCLC in December.

The problem offered to the task force was as follows:

Special Collections and Archives metadata records present unique issues within WCL. The types of materials represented by the records have significant differences from other materials. Some of the user needs for discovering, assessing, and accessing these materials are also different than for other materials. As a result, the way WorldCat.org and subsequently WCL manage and display Special Collections and Archives metadata is often not appropriate.

This task force was charged with

- a) identifying the highest priority problems in the user experience with Special Collections and Archives metadata in WCL
- b) proposing solutions to these problems.
- c) classifying solutions in two categories:
 - 1) can be fixed with metadata currently in place in WorldCat.org or services already existing in WorldCat Local,
 - 2) changes the way WorldCat.org or WorldCat Local currently access metadata.

OCLC expresses its gratitude to the Task Force for its thorough response to its charge within the requested timeline. Such generous contributions by experts in our member libraries enable OCLC to be responsive to member needs.

Task Force Major Recommendations with OCLC Responses

The task force recommended that OCLC:

1. Deploy WCL with the underlying philosophy that each institution's local records will be included in full.
2. Index and display all descriptive data and access points in the member institution's local MARC records.
3. Include data not represented in OCLC master records, including that in the "institutional records" of ex-RLIN libraries.
4. Ensure that materials linked to Internet resources can express the format of the original materials, rather than being represented as Internet resources, for purposes of display and limiting of searches.

In response to major recommendations one, two, and three, OCLC expresses its commitment to making significant local data available for searching and display in WCL. Significant local data does not necessarily mean all local data, nor can OCLC commit to including each institution's local record in full. OCLC is committed to displaying more data from the master record and additional data from the local catalog. How that data will be made available to OCLC is the subject of additional

research. OCLC emphasizes that WCL is an end-user interface, not an expert or librarian interface. OCLC understands the need for an expert/librarian interface and targets this project for release in 2010.

An internal group is already researching the key issues raised by the need to store, manage, maintain, and deliver more local data. This group is writing a white paper that outlines what types of data need to be stored, the sources of the data, where the data should be stored, and which OCLC services will use the data. The paper will propose OCLC's solutions to the task force recommendations. The paper will be reviewed by a subgroup of external experts, who have not yet been identified. Their reaction to the paper will indicate to OCLC whether the proposed solution is appropriate. The paper is due in April 2009. The subgroup response will be targeted for completion in May. A development target will be determined based on the confirmed solution(s).

In response to the fourth major recommendation, OCLC concurs with the task force and has undertaken the analysis to correct this problem. At issue is the algorithm used to identify document types at varying levels of granularity. A change to this algorithm needs to be identified, tested, and implemented. This effort will be coordinated with a project to bring all icons used to signify the item types into synch with the item types. This project is targeted for late 2009 or early 2010.

Task Force Specific Issues and OCLC Responses

The task force identified and prioritized fourteen specific issues. Issues one through five are considered by the task force to be of the highest significance. The task force considers issues six through nine to be of medium significance, and issues ten and eleven to be of lower significance. The task force labeled issues twelve through fourteen as "problems with data." Below are the fourteen issues, the task force's recommendations, and OCLC's responses.

1. Record displays are incomplete
 - *Task force recommends:* Display all variable fields found in full WorldCat records.
 - *OCLC response:* Allow dynamic display of and labels for additional fields based on document and material types. This project will be included in the project to provide display of more variable fields in other formats as well. Until analysis is undertaken, specific fields, labels and implementation rules are unknown.
 - *Timeline:* Late 2009 to early 2010.
2. Local/institutional data is not indexed
 - *Task force recommends:* Index all local/institutional record access points in WCL
 - *OCLC response:* See major recommendation number 3 above.
 - *Timeline:* Analysis April 2009 with development to follow.
3. Principal author is not distinguished from added entries
 - *Task force recommends:* Display only 1xx headings at head of the record. Display all 7xx headings in the "About the Authors" section.
 - *OCLC response:* Currently, a solution to this problem is not scheduled because it involves displaying data in a different location (rather than just adding more data to be displayed.)

An effort will be made to include this work in the work needed for issue one, but until analysis is undertaken we don't know if that is possible or not.

- *Timeline:* Late 2009 to early 2010 if it can be included in the display of additional fields. Otherwise, not scheduled.
4. Keyword searching is inadequate for voluminous results sets.
 - *Task force recommends:* Add string searching or browse options
 - *OCLC response:* This recommendation will be carried into the planning for the expert user interface. In the interim, beginning in July 2009, an alternative search strategy may be used. At that time, libraries with a WorldCat subscription will be able to search the WorldCat database (as opposed to the WorldCat.org database) in their WorldCat Local interface. WorldCat will have eighteen indexes, compared to the five indexes in WorldCat.org.
 - *Timeline:* 2010 for an expert interface.
 5. Definition of “type” for archival materials is inaccurate.
 - *Task force recommends:* Adjust system specifications so that all records for archival materials are retrieved when limiting a search by format.
 - *OCLC response:* See Major recommendation number 4 above.
 - *Timeline:* Late 2009 to early 2010.
 6. Archival holdings display as copies and box order is not maintained.
 - *Task force recommends:* May possibly be remedied by basing the "copy" designation on holdings records rather than item records.
 - *OCLC response:* The problem may be the result of the way in which a particular local system returns information to us (if it is Z39.50) or it may be the result of how we're screen-scraping availability from a particular local system. A problem report has been opened for analysis and resolution of the problem. Assignment, prioritization, and scheduling of problem reports occurs monthly.
 - *Timeline:* This problem report is not yet scheduled.
 7. “Buy It” link is inappropriate.
 - *Task force recommends:* Remove Buy It linking functionality for archival materials.
 - *OCLC response:* Buy It links will now only appear when matched on a standard number.
 - *Timeline:* April 2009.
 8. Summary/Abstract field warrants prominence.
 - *Task force recommends:* Move "Summary/Abstract" field up to the basic record display following “Related subjects.”
 - *OCLC response:* Move the summary display to the top of the detailed record display
 - *Timeline:* April 2009.
 9. Multivolume works display zero volumes.
 - *Task force recommends:* Do not display the second indicator of the MARC tag for the 852 field.
 - *OCLC response:* This problem may be the result of the way in which a particular local system returns information to us (if it is Z39.50) or it may be the result of how we're screen-scraping availability from a particular local system. A problem report has been

opened for analysis and resolution of the problem. Assignment, prioritization, and scheduling of problem reports occurs monthly.

- *Timeline:* This problem report is not yet scheduled.

10. Some archival records display zero holdings.

- *Task force recommends:* Staff in several OCLC departments are investigating this problem and hope to resolve it soon.
- *OCLC response:* Staff in RLG Programs are investigating.
- *Timeline:* There is no estimate for a resolution to this problem.

11. Some collection locations are insufficiently specific.

- *Task force recommends:* OCLC perhaps should look at how consortia catalogs such as Melvyl (Univ. of California) and Summit differ from solitary library catalogs and implement a strategy for how locations data should be handled across all WorldCat local implementations.
- *OCLC response:* WCL is designed to surface an institution's collections. When an institution's WCL surfaces the holdings of other related institutions (for example other institutions in a consortia), they first appear as a link rather than detailed holdings. This is because the system is designed to search the remote ILS at the user's request for information (i.e. when the user clicks the link). This is an efficient and cost effective way for WCL to function and in usability testing has proven to be an appropriate user experience.
- *Timeline:* None; system is working as designed.

12. Many NUCMC records have duplicates. In addition, some records list the Library of Congress as an owning institution as a result of the cataloging process.

- *Task force recommends:* OCLC staff should continue to research methods of mitigating this problem.
- *OCLC response:* There are a variety of reasons why searchers may see duplicate records for collections that have been described as part of NUCMC. In some cases, institutions which may have submitted collections to NUCMC in the past have subsequently become OCLC members and have submitted records for their archival materials as part of a retrospective load of all their records. The batchload matching algorithm for archival materials is intended to prevent incorrect holdings from being set. When archival materials are compared, the software requires that the 040 \$c in the incoming record match the WorldCat candidate record. When the 040 \$c of the records do not match, the incoming record is added as a new master record. Duplicate records can be reported to OCLC using the WorldCat Duplicate Record Request form (<https://www3.oclc.org/app/bibdup/>).

The Library of Congress, through NUCMC, enters records in OCLC on behalf of other institutions that are not able to do so themselves. Because NUCMC staff have entered the record, a holding symbol for the Library of Congress's NUCMC is placed on the record. The location of the archival materials is in the MARC 851 field. The 851 field does not currently display in WCL. There are plans to allow dynamic display and labels based on document and material types.

- *Timeline:* MARC 851 field will be included in the project to allow dynamic display of and labels for additional fields based on document and material types. This project will be

included in the project to provide display of more variable fields in other formats as well. Late 2009 to early 2010.

13. Some finding aids have inaccurate multiple links (e.g. link by "web resources" and by "Internet resources found").
 - *Task force recommends:* Remove link for "Web Resources" in main record, as it is duplicated in "Internet Resources Found." Clarify what the difference is between "Web Resources" and "Internet Resources."
 - *OCLC response:* OCLC implemented a fix in December to prevent the display of duplicate links. We believe this problem was fixed. If not, we need examples of the problem for further analysis.

14. Author names are incorrectly matched in WorldCat Identities.
 - *Task force recommendation:* Continue OCLC Research efforts to perfect matching algorithms in Identities. Encourage the cataloging community to report mismatches to assist in this effort.
 - *OCLC response:* Names are passed from WCL to the Identities service. Errors with the identities service can be reported to via the Feedback link on the bottom of WCL pages. The librarian feedback goes to OCLC Customer Support, from which it is directed to the appropriate staff.