April 29, 2010

To: NGTS Steering Committee  
Patti Martin

From: Link Type Subgroup  
Linda Barnhart, Chair  
Claudia Horning  
Karl A. Kocher  
Lisa Sibert  
Perry Willett

Subject: Final Report

The charge to this subgroup was to “investigate the current practice of using and managing URLs in MARC records, and make recommendations on whether UC should standardize on a particular type.” Four questions to consider were also presented, and are answered below. To determine current practice, a survey was conducted across the ten UC campuses and SCP; key findings are highlighted here. The survey responses have been collected in a PDF file that is included with this report, and we thank the campuses (and SCP) for their time and effort in providing this data.

Current practices within UC

- There are at least 10 different types of URLs which are used in MARC records by 60% or more of the campuses, plus additional types of URLs used on individual campuses.
- While some campuses have a hierarchy of link type preferences and others don’t, all campuses are striving for stable, persistent URLs in their records.
- While staff rarely has the option to link to archived content, when they do, they use it.
- The format of the material (monograph, serial, integrating resource) largely does not affect the type of link. However, SCP primarily uses SFX Open URLs for serials and PIDs for monographs.
- There is some difference of opinion across UC on where URLs should resolve to (the journal home page vs. the list of all issues).
- There is some inherent conflict between public services (wanting the fewest clicks) and technical services (wanting easy maintenance), but they have the shared goal of wanting stability.
- Should linking be driven through one service/technique as much as possible in order to facilitate accurate usage statistics?
- All campuses, and SCP, seem to use subfields 3, z, and u in the 856 field; some campuses use additional subfields.
- Exploration of new directions has begun on some but not all campuses (e.g., moving links to holdings records; letting the link resolver (only) display coverage information,
local ERMS implementation). The aggregate survey data provides a good start on pros and cons for these ideas.

- Maintenance that takes place outside of the MARC record is preferred, so that records don’t need to be re-distributed by SCP, re-loaded by each campus, or re-sent to Melvyl, resulting in faster access for users.
- For link maintenance, campuses largely seem to rely on the URL checker built into their ILS, and upon reports of broken links.
- A wide range of staff do the actual URL maintenance: the full range from students to librarians. Staff members spend their time on a wide range of problems, some requiring considerable expertise to troubleshoot.
- 80% of the campuses use their local instance of SFX to manage locally-licensed electronic resources. The detailed survey results document a wide range of issues reported by the campuses.
- Campuses provided a broad range of thought in terms of the future of link types in MARC records:
  - Frustration that there are too many types of URLs, resolving in different places
  - Recognition that a persistent URL system is needed
  - Prior to a persistent URL system being developed, standardization within UC is desirable
  - For now, UC should maintain flexibility but also provide best practices in the form of a hierarchy of choices for preferred link type, with explanations of the advantages or disadvantages of using particular types of links
  - More centralized/shared link maintenance (less redundant work) is highly desirable
  - Along with persistent URLs, UC should be archiving and preserving content when appropriate

**Current practices within SCP**

- Overall, SCP use of some broad link types (across all formats) shows that:
  - 48% are vendor URLs
  - 38% are PIDs
  - 7% are “persistent URLs” (DOIs, ARKs, URLs with vendor assurance of permanence (IEEE))
  - 5% are SFX OpenURLs
  - 2% “other”
- There are currently 180,000 PIDs being maintained, and which would need continuing maintenance.
- The main reason the SFX OpenURL number is proportionately so low is they work primarily for serials—and thus are overshadowed by vast numbers of monographs. 89% of the SCP serials use SFX OpenURLs.
- Some SFX OpenURLs are hidden behind PIDs because of the need to get the resources cataloged quickly; initially SCP waited for resources to be added to SFX KB, but it was delaying cataloging by months; now they can catalog the resource quickly using a PID and redirect behind the PID when the KnowledgeBase is updated.
Part of the reason the Vendor URLs percentage is so high is it includes EEBO (Early English Books Online, approx. 100,000 titles). For workload reasons, SCP decided not to make PIDs for these; if the URLs change they will redistribute the whole set. This category may also include other very large monographic sets, like ECCO.

Coincidentally, as the Link Type Subgroup began its work, SCP began an urgent effort to fix over 5,000 broken links generated by an unannounced URL change by CRC Press (who changed both the domain name and the individual identifiers for each monograph). SCP staff were grateful that this sudden and substantial maintenance effort could be carried out through a combination of batch and manual activities in the PID server, because the updated redirect made the content immediately accessible to users on the ten campuses, without the delay and disruption of record re-distribution.

**Will all of the identifier types continue to be necessary for cataloging?**

The short answer is yes. For the most part, UC cannot control the types of links that are used by the large number of outside vendors with whom we interact. For many resources, but not all, we can use the SFX OpenURLs that point through UC eLinks; pre-twentieth century monographs, databases, and open access resources, though, are particularly problematic. While we all might hope for more standardization and fewer link types, we have found no evidence of this as a trend. Therefore, for the foreseeable future, we believe that UC will continue to encounter all of these link types, many of which will be included in catalog records. A set of UC best practices may help to reduce the number of different types of links, but it is highly unlikely that we will be able to restrict usage to a small subset of types.

**Will there be changes for which identifier types are used for particular electronic resources?**

We can hope that the vendor community understands about libraries’ need for permanence and stability and continues to move in this direction, but we have no specific evidence of movement for particular electronic resources.

**What changes might be anticipated in the next five years in the use and management of URLs in MARC records?**

- Campuses may begin to investigate migration of the 856 data to the holdings record. The survey responses provide a good overview of the potential pros and cons of this approach.
- The resolution of the problem of redundant coverage data (between UC eLinks, the new Serials Solutions ERMS at CDL, in the 856 $3 in bibliographic records, and in Local Holdings Records at OCLC) may result in changes to the bibliographic records and their links through analysis of the distinctions to be drawn between linking to available content and entitlement to it. Redundant data leads to redundant effort.
- Link resolver service providers will continue to significantly improve the coverage of eBooks and we expect that the next release of SFX will provide faster turnaround for KnowledgeBase updates.
• URLs have predominantly been located in the 856 field, but there are many other authorized fields where links are permitted. UC may find more URLs in non-856 fields over time, and will need to be aware of the complexity this may add to validation and maintenance.
• We know of no specific impacts on link types by RDA, but UC should watch RDA’s future implementation just in case.
• The Subgroup became aware of questions generated by the SOPAG Shibboleth Task Force regarding WAYFless URLs and their potential use across UC. It is unclear to the Subgroup how easily this approach could be implemented (technically) for UC because it has requirements both for UC and for every content vendor, but this is a development that UC should monitor.

Is additional support/training/software development needed?

Campuses have had ongoing concerns about link validation for BibPURLs because OCLC has abrogated its responsibility for running the BibPURL link checker. (This is a national problem not just a UC problem.) We believe a temporary fix has been implemented through a script authored at UCSD, but may need to keep an eye on this as a potential software development need. Ongoing SFX training may also be useful.

Recommendations

1. For the foreseeable future, UC needs to maintain the ways that we provide persistent URLs in MARC records while being aware of trends and better techniques. This includes maintaining the existing PIDs.
2. Eliminating redundancy is good. Shared maintenance of URLs (fixing once, in one place, for the benefit of all) should be encouraged.
3. A group should be charged to develop best practice guidelines for UC for choosing/constructing the appropriate type of URL.
4. We want to lend support to the idea that the redundant coverage data (between UC eLinks, Serials Solutions, bib records, and LHRs) needs to be simplified and resolved.
5. The detailed results of the survey should be distributed to HOTS and to the staff who responded. There was a clear sense of common problems and shared frustrations, but also the collective wisdom in terms of pros and cons for various ideas could be useful to all.

Note: Separate PDF document included with complete survey results.