Navigating the Currents of Vendor-Supplied Cataloging

Helen Heinrich
Cataloging Coordinator
California State University, Northridge
USA

Abstract

The paper addresses the complexities of integrating vendor-supplied e-journal cataloging into the local database based on California State University, Northridge Library's implementation of MARC service from Serials Solutions in 2006. The paper centers on four stages of the project: *preparation, load, post-load issues*, and *maintenance*. For the *preparation* phase the paper outlines the decisions that libraries need to make prior to the record load. The *load* part deals with the actual ingestion of the vendor MARC records into the catalog, the arising issues of authority control and possible compromises in the quality of cataloging that libraries should anticipate to make. The third, *post-load* stage focuses on the clean-up project and on educating public services on the new type of MARC records in the catalog and their impact on the library's OpenURL link resolver. And finally, the paper explains how much effort it requires to maintain the commercial cataloging service on a monthly basis.

With the advent of electronic journals the online format quickly became not just a convenience but the preferred type of publication for users and libraries alike. The publishers responded to the marketplace by discontinuing print versions, producing thousands of new online publications and bundling them into expensive aggregations. Today libraries spend hundreds of thousands of dollars annually for access to their e-journal aggregations but do they get the return on their investment and can they maximize that return? In order to justify fiscal outlay for electronic journals, statistics have to show high numbers of use. So what can libraries do to facilitate the exposure of their e-journal collections to the user?

Currently there are several methods that libraries use to provide access to their e-journal collections:

A-Z lists

A–Z title lists and subject groupings can be created and maintained by the library or produced by an outside vendor. The advantage of web lists is their ease of use and low expense of production. The disadvantages include access to e-journals that is segregated from the print library collection, lack of relationship with counterparts in other formats, limited searching options and exclusion of e-journals from the OPAC search results.

Electronic Resource Management System (ERM)

ERM provides users' access to electronic resources through the public display record. ERM systems contain information on packages, license agreements, usage, OpenURL linking, subscriptions, etc. Based on the data contained in ERM systems, libraries can provide title listings or subject groupings of electronic materials. However, implementation of ERM is a time consuming process. Some may argue that the effort it takes to implement an ERM combined with the cost of the software may not justify the result.

Online Public Catalog

A library catalog is an effective and proven way to provide access to the collection. Unlike an A-Z list, the catalog offers major advantages in providing access to e-journals. Users rely on the catalog to gain access to the library collection and to find all formats of publications there, whether print, online or any other format. In addition to being a "one-stop-shop," the catalog offers a full range of searching options, displaying family relationships and alleviating a problem of split holdings. However, faced with shrinking resources and increased volume of work, libraries realistically cannot undertake the in-house cataloging of aggregated e-journals. In this case a solution for inclusion of e-journals into the catalog may come from some commercial companies that not only provide cataloging in bulk but also maintain the records with continuously updated title lists in the aggregations and date coverage of access.

In 2002 in his white paper commissioned by NISO, Ed Jones formulated a definition of this type of service and coined a term PAMS (Publication Access

Management Service)[1], meaning an agency offering customers basic and updated data on the publications to which they have access rights, whether these publications are hosted locally or remotely. Examples of PAMS are Serials Solutions and TDNet.

Integration of MARC record sets from PAMS accomplishes several things at once. It provides all the advantages of cataloging without investment of staff time (besides the time of implementation).
Advantages for patrons include:
- "One-stop shop" access to library collection in all formats
- Searchability of electronic journals via full range of indexes
- Findability of related titles in print and online formats.

Advantages for librarians include:
- Instant subject coverage in many disciplines
- Obtaining catalog records for thousands of e-journal titles at once
- Liberation from manual maintenance of title and date coverage information and URLs.

In 2006 the library of California State University, Northridge (CSUN) implemented 360 MARC Updates service from Serials Solutions. CSUN is a medium-size academic library currently holding about 30,000 online journal titles. Its ILS is Innovative Interfaces.

The project of integrating vendor cataloging at CSUN consisted of four phases: *preparation*, *load*, *post-load issues*, and *maintenance.*

Phase 1: Preparation

The best way to prepare for integrating vendor cataloging is to prepare for compromises – the quality of MARC records, accuracy of the data and local policy decisions. Before loading records the library needs to review and/or revise some of its policy decisions:

1. Single- or Multiple-Record Approach

When first e-journals began to appear most libraries opted for a single-record approach, i.e. noting electronic access on the record for the print format. This was

---

[1] Jones, Ed (2002). *The Exchange of Serials Subscription Information: A White Paper Prepared for the National Information Standards Organization, with Support from the Digital Library Federation.* Bethesda, MD: NISO Press.

done primarily for two reasons: patrons favor a single record and it is easier to edit a print record with online information than to create a new one. However, when an institution considers buying record sets, for all practical reasons it should think about choosing a multiple-record approach, i.e. creating separate records for print and online formats. This way the library will be able to avoid the difficulties of the matching process associated with the single record. The easiest match point for records is the OCLC number contained in all Serials Solutions CONSER records. If a match is found, the records need to be merged in order to transfer the URL information and other fields to the existing print record. However, non-CONSER and "brief" records do not have an OCLC number and the potential title/ISSN match points between vendor records and the existing catalog records are unreliable. In addition, the only record available from a vendor may be for an online format not matching the print holdings and therefore reducing e-journals coverage. In such case the records may need to be matched manually. When no record is available from a vendor then the so-called "brief records" are created, containing almost no information besides the tile and the URL. Alternatively, a separate-record approach eliminates the problem of matching records, frees catalogers' time and allows libraries to measure more effectively its usage of electronic journals.

2. Classification

Does your library assign call numbers to electronic journals? For online resources the function of the call number as a shelving device is replaced by virtual shelf listing for the purpose of subject collocation and display of the full range of library's materials regardless of their format. However, only 75 % of Serials Solutions' MARC records have a call number, which means that 25 % do not. If the local practice called for classification of online materials, loading of the vendor records will leave a quarter of the library's electronic journals unclassified. Manually classifying those journals would be time-consuming and useless task since the records may be overlaid by the next update. And thus the library has to be prepared to have a substantial chunk of its e-journal collection unclassified. This may be especially important for academic libraries that use call numbers for the accreditation reports.

3. Direct vs. Hosted Links

Serials Solutions offers two ways to display holdings information in the record: through direct and "hosted" links. Direct linking means that the links to the access providers are aggregated on a single record. The link displays the name of the access provider and coverage dates, public note and has an embedded URL

to the source. The user will be taken directly from the record to the title screen of the journal in the corresponding database:


CSUN. Example of direct links

An advantage of direct links is that no extra clicks are needed when connecting to the journals. A disadvantage of this option is that updates take place only once a month and, strictly speaking, the data is current only for one day a month.

The hosted links option means that links are hosted on the Serials Solutions' server. In this scenario the library's record contains only one link that takes the user to a screen, which lists all access options for the title:
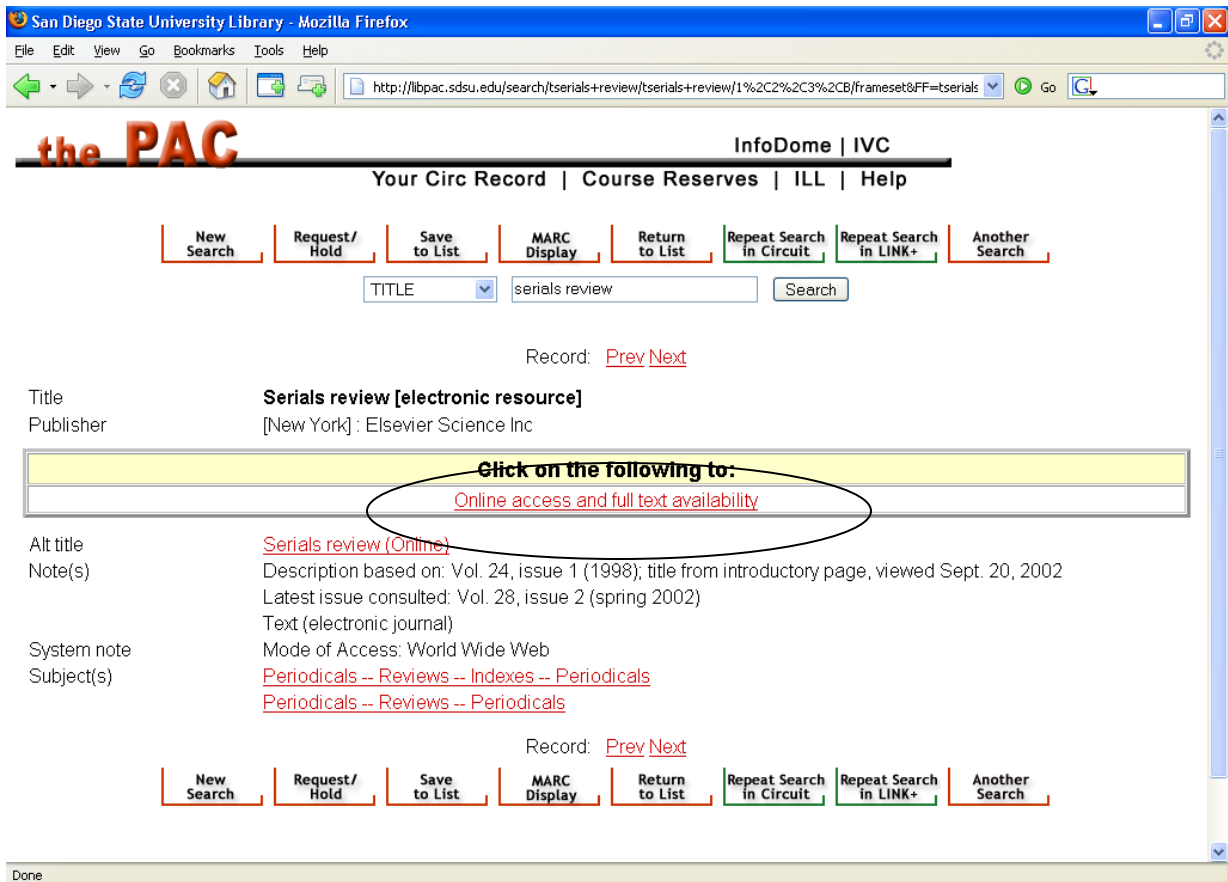
San Diego State University. Example of hosted links. Screen 1.
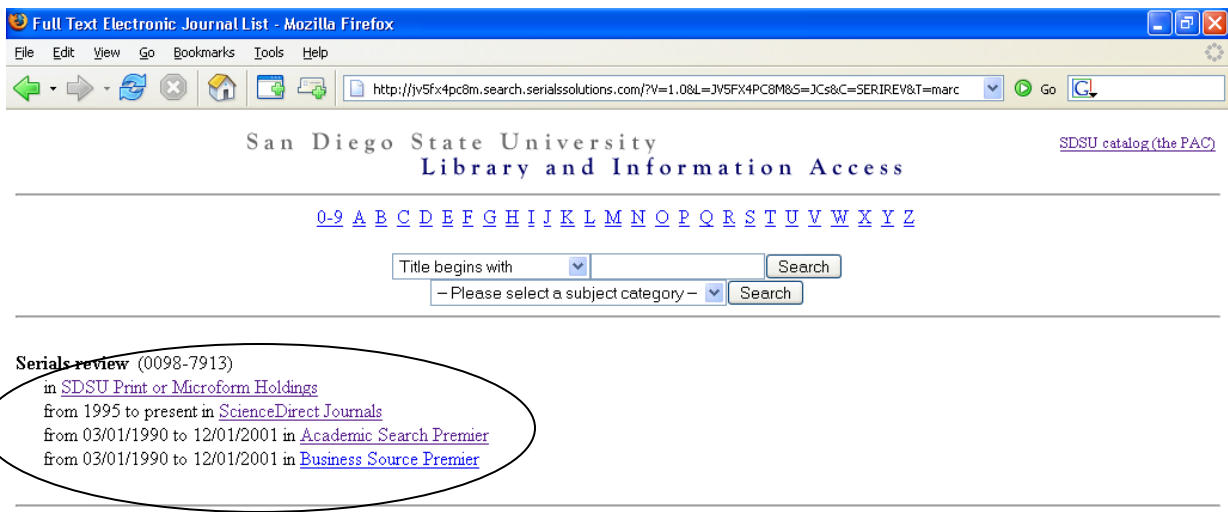
San Diego State University. Example of hosted links. Screen 2.

The disadvantage of this option is an extra screen that users have to pass through on their way to the full text. Many users may find this step confusing and unnecessary since their goal is to get to the full text via the shortest route. However, a great advantage of this option is that holdings data is updated daily since the data resides on the vendor's server. Another benefit is that the number of "change" records in a monthly load will be significantly reduced reflecting changes only to bibliographic records and not to the URLs. In addition, the hosted link option provides title usage statistics since every search is registered in the Serials Solutions client center.

4. Free Journals and Government Documents
Serials Solutions can provide MARC records for a large number of free open-access journals and government documents. The issue here is whether a library should take advantage of free e-resources and add them to its catalog. The SerSol tiered pricing model is based on the number of MARC records supplied to the library. Adding free e-journals would increase the number of records and may increase the cost to the next price bracket.

Once a library decides on the major access and classification issues, it needs to create an e-journal portal in its SerSol client center and fill out the MARC profile which SerSol will use to customize the library's records.

MARC Profile

The next step of the *preparation* phase is to fill out the MARC profile provided by the vendor. This profile can be modified at any time as a result of the test record evaluation process or after the go-live production phase.

Update Method Options
In the MARC profile, it is important to specify the updating method options: whether to update the existing vendor records or delete them, and load the new records every month. If the choice is to modify the records, the library then receives three types of files: "new" records, "change" records and "deletes."

> "New" records are delivered when:
> - A new title has been added to a subscription aggregation.
> - A title that erroneously was sent as a "delete" the previous month and then re-sent the following month as a "new" title. The occurrence of false "deletes" is not uncommon and therefore at CSUN we manually go through the monthly "deletes" file (about 50 titles per update) and verify that we indeed no longer have access to the title.

"Change" records:
- A changed record is defined as one where any or all of the bibliographic contents, coverage dates or URL syntax have changed since the last update received by the library.

"Delete" records:
- A title was dropped from an aggregation (ceased or moved to another aggregation).
- A title was erroneously eliminated from the library's profile (usually due to a bug in the program).

Replacing <u>all</u> records:

Another update option offered by SerSol is to deliver a completely new set of records each time. In this case there is no need to manage "change" and "delete" files because records from the previous load are deleted and replaced by the new ones. However, for some single journals that we had previously cataloged in-house we transferred local notes into the protected fields and the orders to Serials Solutions records. Therefore, deleting all records with monthly updates was not a viable option for us, or else we would lose valuable local information.

Record Selection

Library's preferences for the type of records will define the body of the entire e-journal collection, as every choice will translate into hundreds, if not thousands, of records. It is also one of the places where the imperfections, as well as convenience, of vendor records will show the most. There are not enough perfectly matched records for every online title. A large number of Serials Solutions records were created for the formats other than online and were made "online" by addition of some format-specific elements, such as the uniform title with the "online" qualifier. Some records may have a physical description area (MARC field 300) that details the number of volumes and size of the publication in combination with the title field designated as electronic resource (MARC field 245 *general material designation* in subfield $h). However, all these records carry well-managed URLs and coverage dates, and that is what matters most to the users.

In addition to *CONSER online records*, i.e. authenticated records from the CONSER database created specifically for the online format; *CONSER "neutral"*

records", i.e. the format-neutral records from the CONSER database, created for an unspecified format; and others such as Serials Solutions supplemental "brief" records. The brief records are machine-generated and contain only uniform title, main title entry, customized fields (e.g., genre headings), and URLs:

CSUN. Example of brief record

When first defining the record priorities, CSUN opted out of SerSol supplemental brief records, assuming that the first several types of records would be sufficient to cover our e-journal collection. However, the gap between the 20,000 titles expected to load and the number of records actually loaded was more than 5,000 records. This difference made CSUN reverse the supplemental record decision and by accepting them we achieved 100 % title coverage. In addition to the "poor" cataloging quality, the supplemental records have to be deleted with every update because they do not have permanent matching point. Serials Solutions is working on an enhancement that would rectify this problem.

 Customization Options
Serials Solutions offers an extensive MARC field customization. They are capable of customizing any record field thus enabling libraries to implement local cataloging decisions without much effort or manual work. For example, CSUN requested that non-LC subject headings be stripped and always set the first and second indicators of 776 field (additional physical form entry) to 1 and blank. Both these requests were added as a result of the test records evaluation and were the reflection of the local cataloging practice.

As libraries fill out the MARC profile template provided by Serials Solutions they are assisted by default choices offered by the vendor that are based not only on cataloging rules but also on the common requests from libraries. CSUN has retained certain default choices and changed others to comply with local cataloging practices.

Phase II: Load

Test Load
Once the MARC profile is filled out, SerSol provides a set of sample records, configured according to the library's MARC profile. Usually it is the first 500 records, but it can be a different number if requested.

The process for reviewing records depends on the technological capacity of the library. Some libraries have a test server that allows review of the records, including their public display, without load into the production database. Libraries that do not have an additional server, CSUN among them, have two choices: load the records into the catalog and review them either live or suppressed, or use a MARC reader software, such as MARCEdit, to look at the records without actually loading them.

Evaluation of Test Records: What to Look for and What to Expect:

- Compliance with the library's MARC profile
- General quality of MARC records and extraneous fields

In CSUN's experience all MARC profile choices and customization requests were fulfilled by Serials Solution from the beginning of the process.

With regards to the quality of cataloging, as CSUN found out during the first load, most SerSol online records are based on the record for other formats of the publication, such as print or microform. That created hybrid records that do not comply with "pure" cataloging, however, except for the brief records, all main access points were there, and that was most important.

Source: CSUN Online Catalog

The records also contained some fields that CSUN was not using. For example, there was a plethora of non-LC subject headings, including foreign subject headings, which prompted CSUN to adjust its MARC profile, to strip all non-LC subject headings. For the libraries that do not use LC subject headings the adjustment may be in reverse. Our greatest concern had to do with title changes and whether Serials Solutions would be able to handle them. The worry was based on the fact that SerSol receives its data from the providers, and a title change in library terms may not be considered a change by a vendor and vice-versa. Vendors commonly use latest-entry practice and list all iterations of a journal under the most recent title, or they change the title but continue to use the old ISSN. This results in SerSol using the record for earlier or later titles and adding a link covering both iterations. For example, in the record for *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, the publication dates state 1934–1990. However, the links show range of dates from 1905 to 1990, covering the earlier title *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*

Example of a record where coverage dates for earlier and later titles are placed on the record for the later title.

Another example of the shortcomings of vendor-supplied data is that the records for earlier and later titles may be split between different aggregations. In this case, one of the titles may appear in an aggregation outside the library's subscription scope. An example is *Advertising Age's Creativity*, ISSN 1072-9119, which in 2001 changed its title and ISSN to *Creativity (Crain Communications Inc.)*, ISSN 1541-3403. CSUN subscribes to both titles: the first is available from Lexis/Nexis and Factiva; the latter is available from databases outside CSUN's subscriptions. However, the record for the earlier title provides coverage dates that span both versions of the title, thus giving full coverage for the publication.

Source: CSUN Online Catalog

Here is another example of customization done by Serials Solutions. Most aggregations provide URLs that link from the record directly to the journal. However, LexisNexis and Factiva link to the general search screen. For this very reason, CSUN requested Serial Solutions to order the links in such a way that the LexisNexis and Factiva links will be at the bottom, making them the resources of "last resort".
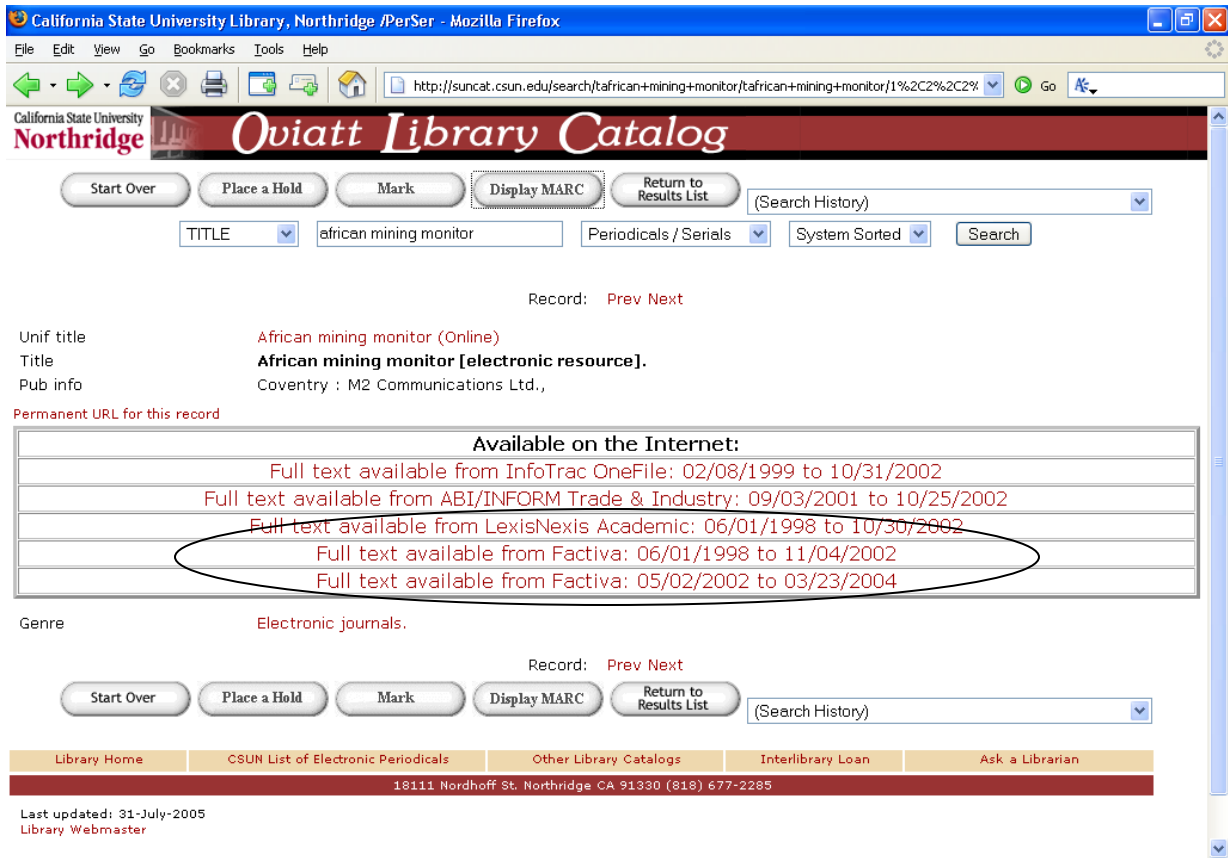
Record with links displayed in customized order

Source: CSUN Online Catalog

Linking

When supplying the test file, Serials Solutions emphasized the need for CSUN to check as many links as possible within and outside the library's IP ranges to ensure that proxy servers worked correctly. Most of the library's URLs were configured correctly and there was no problem linking to the journals' sites. We did detect, though, a number of records that contained duplicate linking: two URLs for the same database and the same coverage dates. These were cleaned up by SerSol for the production load and did not reoccur in future loads.

Production Load

The first production load is the most complex one and requires several decisions.

- How to load: overlay existing catalog records or create duplicate records?

This question arose for us because we had manually cataloged about 5,000 of our electronic journal titles using a separate-record approach. We were considering

an overlay of the existing records with the incoming records from Serials Solutions. For the overlay to occur we needed to find reliable match point(s) for the system to detect duplication. Several parameters were considered, such as ISSN and title combination, GMD (general material designation), etc., but CSUN decided against the overlay. Some of the fields were not indexed by the system, and in the case of ISSN and title the concern was that some e-journals do not have an ISSN and the title may be constructed differently in the local catalog and SerSol records. The issue of the second and all subsequent loads was clear – all vendor records have a unique title identifier enabling reliable overlaying of records. Once it was decided not to use the overlay for the first load, the next question was: What to do with 5,000 titles that would have duplicate records? Our decision was to undertake the clean-up project. The clean-up would entail finding the duplicate records, transferring local fields that the library wanted to protect/retain to the SerSol records and deleting the existing/suppressed records.

- Item records

Most systems, including Voyager and Millennium, have the capability to automatically create an item (or MFHD/item) record for bulk cataloging. CSUN's library Systems Administrator configured the load in such a way that Millennium created and attached item records with predetermined parameters to every SerSol record loaded.

- Authority control

CSUN's authority control procedures include running weekly new headings list reports. Preparing to load about 20,000 new records, we were concerned about the impact this would have on our authorities processes. The number of potential new headings was unknown and there was some trepidation that new headings would overwhelm the capacity of the headings reports. One option was to exclude SerSol records from the authority workflow from the start. Since Serials Solutions records are unstable, they were excluded from authority control.

Once we completed the test file evaluation, it took only a few days for Serials Solutions to generate the production file.

Phase III: Clean-up project

We began the clean-up project by creating a file of e-journal records that existed in the catalog prior to the Serials Solutions load. The file, containing more than 5,000 records, was divided among five staff members. The implementation team developed the procedures, outlining steps in the process:

- Search the title of locally cataloged e-journals and find the duplicate among SerSol records.
- Compare the URLs in old and new records and ascertain that all functional links are listed on the SerSol record.
- Transfer local fields that needed to be protected during the overlay from the local record to the SerSol record.
- Check if order record is present – if yes, transfer it to the SerSol record.

De-duping e-journal titles was a tedious and time-consuming task. Because of other competing work priorities it did not progress as quickly as we had hoped. In the end the de-duping was assigned to a library school intern, who effectively completed the task.

Phase IV: Maintenance

The most time-consuming effort in the process of the MARC records implementation involved the work leading to the first load and then the post-load clean-up. Routine monthly maintenance of the records, in contrast, is an easy task that makes all initial efforts pay off.

SerSol sends a monthly e-mail notifying the library that an update is ready to be loaded. The update consists of three files: "new", "changes" and "deletes." CSUN loads "new" and "changed" records automatically and the "delete" records are handled manually.

Despite the numerous advantages of vendor records, they lack the precision of in-house cataloging, something that reference services and patrons are used to. Once vendor MARC records are loaded in the catalog, they trigger many questions from public services. A typical question is: *"Why do the coverage dates in the record show that our access starts in 1996 but when we connect to the journal our actual full-text access starts in 2000?"* This is a valid question and the answer is that Serials Solutions does not customize coverage dates for the link display notes. The vendor provides information obtained from a publisher or an aggregator that is true to the title in general, and may not apply to a specific library with its own subscription terms.

Another issue that CSUN encountered was the lack of synchronization between its OpenURL link resolvers (SFX) knowledgebase and Serials Solutions knowledgebase. The problem was quickly detected by reference services and

prompted this question: *"When linking from a citation via the SFX button, the SFX menu says that no online full text is available but the next line says that the library has this title – check the catalog. Clicking on the 'check the catalog' the link takes you to the journal that is in fact available online! Why doesn't SFX menu show that we have the full-text online access? We did not use to have this problem – what is happening?"* The answer is that the number of journals in the SFX knowledgebase was at that time substantially lower (by about 10,000 titles) than the number of electronic journals in the SerSol knowledgebase, which caused SFX to miss titles that we had. By default, if SFX does not find a journal in its own knowledgebase, it does a title look-up in the catalog and only then finds the SerSol record. We came across this problem because the SerSol load created about a 10,000-title gap between the two databases. Since the time of our first implementation, the SFX knowledgebase grew considerably and the problem has been alleviated. However, the lack of synchronization between databases of the library's technology tools needs to be taken into consideration. Ideally, the library's link resolver and the e-journal MARC records should stem from the same knowledgebase thus avoiding any gaps in holdings.

eHoldings
After MARC records were loaded into the catalog, an immediate question arose as to whether show our e-journal holdings in OCLC (a bibliographic utility).

In summer 2006 OCLC introduced a new service called "eHoldings" which enables automatic setting of e-journal holdings in the Worldcat database. The service works directly with your MARC record provider and uses the ISSN as a match point. The holdings' setting and updating takes place on the title level and the library has an option of setting "lend" or "not lend" ILL policy. The service is free for PAMS' customers and OCLC subscribers.

The benefits include a higher usability rate of the library's collection through OCLC First Search and Open WorldCat platforms. The library user gets access to online contents by searching an OCLC interface, not just the library's catalog.

It has been two years since CSUN has implemented 360 MARC Updates from Serials Solutions. By now everyone has become used to the convenience of having electronic journals as part of the subject or keyword search results in the catalog. By the same token public services have adjusted to seeing many brief records in the catalog and to verifying coverage dates by checking the e-journal site. The lack of synchronization between SFX and Serials Solutions knowledgebases has been alleviated. In the end we served the best interests of

our users by making the catalog the single point of entry to our library journal collection in both electronic and print format.