



Iowa Research Online  
The University of Iowa's Institutional Repository

---

University of Iowa Libraries Staff Publications

---

6-15-2004

# Loading SFX MARC Records into Aleph: The Iowa Experience

Wendy C. Robertson  
*University of Iowa*

---

© 2004 Wendy C Robertson

Hosted by [Iowa Research Online](#). For more information please contact: [lib-ir@uiowa.edu](mailto:lib-ir@uiowa.edu).

# LOADING SFX MARC RECORDS INTO ALEPH: THE IOWA EXPERIENCE

PRESENTED BY:

WENDY ROBERTSON

ELECTRONIC RESOURCES TECHNICAL SERVICES LIBRARIAN

THE UNIVERSITY OF IOWA

*Note: Due to pressing activities during the last several months, we have not been able to make the progress I had expected when I originally offered to give this poster session. I apologize for only being able to present such preliminary information. I hope that this session can serve as a springboard for discussion.*

## BACKGROUND

The current situation:

- We do not add anything to our local catalog from aggregator databases due to their volatility.
- We catalog all e-journals that we pay for and all selected free e-journals. We use the “single record approach” for improved public display and also due to staff time constraints.
- We do not add any holdings information for e-journals to our catalog.
- We display the URL on the brief record. We prefer having the URL display so that the resource can be found from a printout and so that it is easy to copy and paste for patrons (such as for use in references).
- We are batch loading books and Marcive records routinely. We have not yet batch loaded any serials collections. We have not tried to merge batch loaded records for e-books with existing records for the print. We have not experimented with some of the new matching/merging routines that use ISSN or ISBN.

- We have close to 14,000 titles activated in SFX.
- We have a static list of e-journals generated from SFX updated nightly.
- The University of Iowa Law library has completely separate records for all their non-document titles. We share records for some documents.
- All of our serials already have an SFX (InfoLink) button displaying in the web, whether or not we have any full text content. However, only titles with a holdings for our ELEC sublibrary will be retrieved when limiting results to “online” in our advanced search.
- Example from our catalog:

[ [Display full record](#) ]

**Title** The Journal of interdisciplinary history.  
**Published** [Cambridge, Mass., MIT Press]  
**Format** Journal  
**URL** Current issues: <http://proxy.lib.uiowa.edu/login?url=http://muse.jhu.edu/journals/jih/>  
 Back issues: <http://proxy.lib.uiowa.edu/login?url=http://www.jstor.org/journals/00221953.html>  
**Location** [ELECTRONIC Resource For online access, see URL above](#)  
[MAIN Library No call number](#)  
**Availability** Click location above for holdings and availability  
 InfoLink Check InfoLink for related services

What we would like:

- Every title that is active in SFX to be in our catalog.
- Only one record to display to the public for each serial. When we do not own the print, we would like an aggregator neutral record for the electronic version.
- To minimize the staff time required to maintain records for volatile aggregator databases.

- To keep these records/holdings as current as possible (to match changes in our subscriptions, on the publishers' sites, and in aggregator databases).
- All the URLs to display for every active source.
- To choose the order in which these URLs display.
- SFX availability statements to display in the 856 \$\$3.
- CONSER quality records for the e-journals.

Original discussion:

- Records will either duplicate some existing serials or will have fields added to serials (match based on ISSN or e-ISSN).
- Need to be able to identify the records in some way so they can be removed from the catalog as necessary.
- Adding to existing record will be more difficult to batch load, but better end display.
  - All merged information will need to be removed when we lose all access (due to cancellation or changes in aggregators).
  - As titles are added and dropped from one aggregator, they may remain in another aggregator, so we will need to be able to update merged records.
  - We know we do not have ISSNs in every record that should have one. We have e-ISSNs in relatively few titles, so the matching algorithm will need to use several fields.
- Creating separate records is arguably a worse public display but:
  - We are already doing this with Netlibrary and Books24x7, so we have a precedent.

- It will be important to have these records for the electronic version clearly appear public different than the records for the print.
  - It may be confusing when some print titles have URLs in them and these URLs are repeated in the record for the online version. We may need to remove all the URLs we have put in manually.
  - We may investigate the possibility of using Aleph's union catalog so that there are separate records which display to the public as one record.
- We do not want to use a separate product, such as Serials Solutions, as a source for the records, because we would need to keep our holdings current in multiple places.

### LOADING SFX MARC RECORDS

#### Initial testing:

- In late July 2003, we agreed to work with Ex Libris regarding the technical issues of loading the SFX MARC records into Aleph, using our test database.
- We decided to start with separate records and then try to work on merged records at a later date knowing that we could easily pull all the separate records from the database.
- The initial load found some problems that Ex Libris quickly corrected (such as duplicate leaders).
- In late September, we began discussion with Ex Libris about whether the ISSN or e-ISSN should be preferred when matching on records, and what fields should be used in the algorithm. Because the records from SFX are coming from the LC file and not from WorldCat, the e-ISSN was less useful than we had hoped.
- In November, we loaded 1276 documents to our test file without problem. We made the same basic changes to these records that we made to Book24x7 and Netlibrary records. We did not need to make these changes to load the records

- Records that didn't find a match used the data from the SFX Knowledgebase, which often had the title in all capital letters. The test file had at least 110 such titles (8.6% of the total), a higher percentage than we had hoped for. Most records were for the print version. Some were for the microform version.
- We discovered an 866 in the bibliographic record displays in the web in the same manner as an 866 in the holdings. This has the potential for keeping the loads simpler.

#### Discussion of load:

- Most of the records are for the print version. This will result in a very unsatisfactory display when our record for the print is alongside a 2<sup>nd</sup> record for print that is really for the electronic version. If we do not have the print version, the display will not be ideal but at least would be workable.
- There are not useful URLs in the record. If URLs are present, they are often for sources we cannot access. The URLs come from the LC record and have no connection to our holdings in SFX. We determined that we could work with this if we removed all URLs from the record and then had the text like "click SFX button for web version."
- The holdings display is inadequate. The availability statements appear but the target name does not display. This means that if a title is available from more than one source, there is no way to tell which source goes with which availability statement, making it more difficult to select the preferred site. For this reason we may remove all the 866 fields.
- Because there are not relevant URLs in the record, the availability statement is not connected in any way to the appropriate URL (we had hoped to have the availability statement paired with the appropriate URL for display).
- If we merge records, it could be argued that we would not receive much benefit to the MARC enhancement because, for the most part, we'd actually be tossing aside the bibliographic data since we'd be merging it with our print record. In other words, we could simply load straight from SFX to Aleph because in these cases all we're interested in is the summary e-holdings which is already in the SFX export.

- We could load the SFX MARC records into our catalog only if there is no corresponding record in the catalog.
- We could try to do both of these—load if no record, use SFX data if we have a record. We decided to pursue this idea to see if it would be feasible.

#### Analysis of ISSNs:

- We determined that it would be helpful to know how many titles active in SFX were already in our catalog before we decided how to proceed. This would give us a better indication of the direction(s) we should take.
- We ran a report in late March comparing the ISSNs in the SFX export file to the ISSNs in the 022 \$\$a, 022 \$\$y, and 776 \$\$x (in that priority order). We also wanted to match only on “our” records, not the Law Library’s records or our records for government publications (largely Marcive). Non-matches and duplicate matches were reported.
- Report checked 13,173 ISSNs. Of these:
  - 5388 (40.9%) did not match anything in our catalog.
  - 7044 (53.5%) matched a record in our catalog.
    - 5812 (82.5%) matched the 022 \$\$a
    - 491 (7%) matched the 022 \$\$y
    - 98 (1.4%) matched the 776 \$\$x
    - 643 (9.1%) matched 022 \$\$z, Law or Gov records. These need detailed assessment.
  - 741 (5.6%) matched multiple documents (including Law records).

- 234 (31.6%) matched multiple records, but the duplication seems to be with Law or Gov records.
- 400 (54%) matched with multiple records, not including any Law and Gov records.
- 107 (14.4%) matched multiple records, but only Law or Gov records

Next steps:

- We should pursue loading records for the 40% of our active titles that have no corresponding record in our catalog.
- We should run a report to see how many of the matches on the 022 \$\$a and 776 \$\$x to see how many already have an electronic holdings. We should pursue adding an electronic holdings to those that lack one. We need to be able to remove these holdings as necessary in the future (and will need to be able to treat those added through a load from those added manually).
- We should check the matches on the 022 \$\$y to determine if these are electronic records matching on the print ISSN or if they are matching different titles.
- We need to look more closely at the 643 "other" matches to determine what we should do with these. This will involve talking to Government Publications staff and will probably involve talking with the Law Library.
- We need to check if the 234 matches with multiple titles already have electronic holdings and need to see if we should add electronic holdings to the Law or Gov records.
- We need to determine why 400 ISSNs matched multiple records (not including expected duplicate records).
- We need to talk to the Law Library and Government Publications staff about the 107 that matched more than one of their records.