
Karen S Calhoun, University of Pittsburgh - Main Campus
John J Riemer
Interim Report
PCC Standing Committee on Automation (SCA)
Task Group on Journals in Aggregator Databases

May 1999

Task Group Members: Jeanne Baker (U. Maryland); Matthew Beacom (Yale); Karen Calhoun (Cornell); Eric Celeste (M.I.T.); Ruth Haas (Harvard) Jean Hirons (LC liaison); Oliver Pesch (EBSCO liaison); John Riemer (U. Georgia) Chair

Introduction
The PCC SCA Task Group on Journals in Aggregator Databases is investigating and making recommendations for a useful, cost-effective and timely means for providing records to identify full-text electronic journals in aggregator databases. The following interim report was presented for comment at the CONSER Operations Committee meeting April 22, 1999. It contains recommendations on the content of a vendor-supplied record, a description of progress with the demonstration project with EBSCO, and information about related projects.

The predecessor of the SCA task group was formed one year ago at the CONSER Operations Committee meeting of 1998. A number of CONSER representatives have mandates from their reference departments to provide better bibliographic control of full-text journals available in services like UMI’s ProQuest Direct, EBSCO’s Academic Search Elite, and other products widely available to library users. The CONSER task force began by surveying CONSER libraries what measures are being taken and what would CONSER libraries like to be to be doing? The results of that survey are available in the July 1998 issue of CONSER Line

Key findings were that CONSER libraries are using numerous methods for providing access to full text titles in aggregations, including the CONSER single record approach for online versions, lists on a Web page, paper guides, and separate catalog records in the library online catalog. Many of those surveyed noted lack of staff time as the biggest obstacle to providing cataloging and maintenance for these titles.

Last fall at the PCC Policy Committee meeting, the CONSER task force was encouraged to survey the broader library community. The survey, which was conducted before ALA Midwinter, indicated that 71% of the 62 responding libraries want records in the OPAC to represent the full-text journals available in aggregators, and 75% are interested in purchasing record sets. Half were willing to pitch in and do some of the work to create the records. Respondents submitted twenty single-spaced pages of comments, indicating a high level of interest in this topic.

Providing access to the full text journals in aggregators made the agenda at the Big Heads (ALCTS Technical Services of Large Research Libraries) discussion group at ALA Midwinter. Aggregators were also the topic of the ALCTS Catalog Management Discussion Group in Philadelphia. Before the speakers at that program could walk out of the room, Oliver Pesch of EBSCO offered his organization’s participation in a demonstration record creation and badging project.
Since the early months of 1999, the majority of our task group's time has been devoted to developing a data element list of what should be included in an aggregator analytic record and a set of working assumptions to guide our efforts to explain our thinking to other outside efforts.

**Working Assumptions**

The work of our task force has been guided by the following working assumptions.

1. Aggregator analytic records need to contain sufficient fields such that they could stand alone in an OPAC as separate records, since for many serial titles the aggregator will be the only local source. (In other words, the analytic record must adequately serve in lieu of the print or other record that will be present in the OPAC of another library.)
2. Records will need to include those fields needed for deduplication against the existing hard-copy version records in an OPAC, for those libraries concerned about avoiding multiple hits for a given title. (In other words, we need to preserve their option to perform an additional processing step upon loading a set of records, toward application of the single-record technique.)
3. Records need to contain data elements that ensure the possibility of their partial or complete removal from the OPAC in the event of a subscription cancellation.
4. Data in records will primarily be a subset of that found in a record for a hard-copy version. Creation of these records by deriving data from other bibliographic records, followed by necessary modifications, may be a strategy that institutions/vendors choose to follow.
5. It will be desirable for interested libraries to obtain a single record in an OPAC, reflecting coverage by multiple aggregators through repeatable fields. It would greatly facilitate local application of the single-record technique if bibliographic utilities would collect all the coverage information onto a single aggregator analytic record. However, some record sets will be available only from a non-bibliographic utility source. Libraries may choose to consolidate 773s onto a single OPAC record, for a fuller list than will exist in a bibliographic utility. A library interested in having separate bibliographic records for every version in an OPAC could theoretically obtain such from a bibliographic utility plus individual vendors; no consolidation steps would be undertaken upon loading record sets.
6. Loading records will involve prior local customization steps such as decisions on classification number fields, selection of a URL for the 856 field(s), deletion/suppression of irrelevant 773 fields, adding other desired fields like 655s, etc.
7. For analytic records that libraries must create, holdings data might be too ambitious at this early stage. If holdings information happens to be supplied by a vendor at this time, an ANSI/NISO 39.71-1999 summary statement could be placed in subfield $3 of the 856. Inclusion of holdings data should be accompanied by a plan to maintain its currency.

**Proposed Data Elements**

**FIXED LENGTH FIELDS**
All Leader and 006/007/008 bytes as appropriate

**CONTROL FIELDS--0XX**
001 Control number
003 Control number identifier
022 International Standard Serial Number
035 System control number(s)

**VARIABLE FIELDS--1XX-9XX**
1XX Main entry
240 Uniform title
245 Title statement (insert $h)
246 Varying form of title
250 Edition statement
260 Publication, etc. (Imprint)
310 Current publication frequency
362 Dates of pub., vol. designation
4XX Series statement
5XX Notes
6XX Subject added entries
700-730 Name/title added entries
773 Host item entry
780/785 Preceding/Succeeding entry
8XX Series added entries
856 Electronic location and access ($3, $u only)
1. Presumably the aggregator analytics will always contain value "b" (serial component part) in the Bibliographic level (Leader/07)

2. The Encoding level (Leader/17) of these experimental records could initially be set at "5," for "partial/preliminary." (Alternative: "2," for "less-than-full level, material not examined")

3. The repeatable USMARC-style 035 field would house control numbers that are critical for deduping against local OPAC records. These numbers include LCCNs and OCLC numbers, respective examples of each, taken from Journal of Academic Librarianship, could look like:

   035 ## (DLC)75-647252
   035 ## (OCoLC)2243594

   ISSN's would go in 022 $a (for an electronic version, if one has been assigned) or 022 $y (for the print version). Example:

   022 $y 0099-1333

   Alternatives (for print version):

   035 ## (DLC-N)0099-1333
   776 $x 0099-1333

4. Field 245 needs to include $h to alert user that item is an electronic journal. Field 130 is too problematic to create across-the-board in records.

5. Field 260 would contain the publisher of the original version. This would identify the serial; it would also be practical in that the analytic record would need to represent potential coverage by multiple aggregators.

6. Field 362 would represent the facts of publication for the original version.

7. Fields 5XX/6XX/7XX would be the same as in the print version record. No special note or added entry would be included for the aggregator(s); the 510 fields would be omitted. Field 655 would not be used.

8. Field 856 would include only subfields $3 and $u, in that order. Passage of a relevant MARBI proposal this summer would move URLs into a subfield of the 773.

**Demonstration Project and Examples**

Oliver Pesch, our task force's liaison from EBSCO, has been working with us since early this year. He derived a set of records for us experimentally based on the task group's instructions. The data in these records is a subset of data from the corresponding records for the print journal found in the CONSER database. The vendor's program also adds some fields to the record.

The examples don't follow exactly the model for data elements we have presented in this interim report instead they reflect the model as it was in mid-conversation about ten days prior to the CONSER Operations Committee meeting. Thus, we view these samples as works in progress, helpful for framing a discussion of the final form of the records. Our goal is to arrive, through discussion and experimentation with records, at a set of requirements for aggregator record sets that we can then present to other vendors of aggregator databases.

Oliver Pesch provided us with a set of a dozen sample records in MARC communications format as an attachment to an e-mail message. For the purpose of displaying them in a familiar interface, we imported them into the Cataloging Microenhancer.

In Figure 1, note that Encoding Level is set to "5" (partial/preliminary). Another possibility is ELvl "2" (less-than-full, material not examined). The task group received some feedback from the CONSER Operations Committee and is still discussing this issue.

Also note that fields 006 and 007 are missing right now but will be added as appropriate (we haven't got around to defining them for EBSCO yet). Placement of the ISSN of the print journal in the machine-generated record is also still under debate. The figure merely illustrates some possibilities.

EBSCO's machine-generation program retains 1xx, 245, 260, and 362 from the print record. Field 245 has $h [computer file] inserted after $a, $n, $p but before all other subfields. The task force has not yet provided instructions about removing the period at the end of 245 $a, but we plan to do so. All 5xx fields are retained except for 510s, and all 6xx fields are retained.
Su, inorder to encode information about the materials specified and the URL.

Figure 1. Machine-generated record, Education Digest

OCLC: 9990416
Repl: 9990416
Type: C
Name: EL
Date: 5
GIns: d
Ctrl: Lang: eng
Rev: 0
Form: Conf: 0
FREQ: m
Repl: n
ISSN: 1
Alph: a
Desc: 0
SrTPI: Cont: DtSt: c
Dates: 1995.9999

003
(DLC-N) 0013-127X
005
(Msw) fsh1567544
090
4b
049
C000
130 0 Education digest (Ann Arbor, Mich.)
245 04 The education digest. #h [computer file]
260 0 Ann Arbor, Mich. : #b Education Digest, Inc. 1995-
310 0 Monthly (Sept.-May)
362 0 Vol. 1, no. 1 (Nov. 1995)-
500 0 Published: Ann Arbor, Mich. : Fraser Publications, <1986->
530 0 Available in microform from University Microfilms.
550 0 Education #x Periodicals.
570 0 #b Academic search elite #x Ipswich, MA : RSC Publishing, 1999- #x 1071-2720
577 0 #x 0013-127X
085 40 #a Full text available: June 1994 - #u http://www.ennet.com/ghost/login.asp

The task force discussed having the program construct a 130 or uniform title field. Public services staff and users have said they like having the title qualified by (Online) because it helps them pick out the electronic version in an index display in the OPAC. In many cases, creating a uniform title automatically would be relatively straightforward the title has no initial articles and the title does not conflict with the title of any other serial.

When the 130 field already has a qualifier, or when a record has no 130 field, writing a program to create or revise a 130 field would not be a trivial exercise. Therefore, in the interest of helping vendors to get record sets ready and available to the library community quickly, our task force has put the requirement for a 130 field aside.

Related ProjectsM
The task force investigated other developments of interest at the University of Tennessee at Knoxville (UTK), the University of Illinois at Chicago, and OCLC.

David Atkins and Bill Britten at UTK were kind enough to provide our task force with detailed information about their projects to bridge the gap between citation databases and journal holdings. One of the aggregations they have treated is Dow Jones (4270 titles); another is Proquest (1500 titles). UTK staff began by harvesting data from the vendor's Web site (lists of titles with ISSNs and coverage dates). They wrote PERL scripts to massage the vendors' lists of full-text journals, then ran the resulting text file through a utility called MarcMaker. This step created MARC records that they imported into their OPAC. Figure 2 provides an example of the public display of one of these records.

Figure 2. Example of UTK Machine-Derived Aggregator Record
Before creating the MARC records with MarcMaker, UTK staff did preliminary work to define tags for storing the data. UTK puts the ISSN, which they need for their hook to holdings, in subfield $9 of the 022 field. They add $h [electronic fulltext] to the title and store it in field 245. Field 506 stores a note about access restrictions; 856 $u stores the URL and $z stores public notes; field 945 contains service/vendor information, dates of coverage, lag time, more access notes, and a control number for the entire set (this serves as the hook to delete the whole set globally). Take a look at more records in UTK's catalog UTK has been able to achieve astonishing turnaround times for providing title-level access to full text and holdings information in their catalog using this technique.

The task force also contacted Karen Zuidera at the University of Illinois at Chicago. Staff there are working on H.W. Wilson Select full text for OCLC's WorldCat Collection Sets project. Zuidera and her staff are creating records in OCLC using a workform and constant data records. OCLC will later send the library a tape of the set for loading. Figure 3 provides an example. The constant data form they are using contains prompts for 022 $y, 043, 110 & 130, 245, 260, 530, 650, 710, 773, 776, 785 (succeeding entry) and 856.

Figure 3. Example from University of Illinois at Chicago Initiative
OCLC's TechPro group has also created records (for Elsevier Science and Academic IDEAL journals) that may become available as WorldCat Collection sets.

**Next Steps**
The task force agreed on the following next steps.

1. Finalize the definition of record content for the EBSCO demonstration project and test downloading the set from EBSCO's Web site.
2. Define record content for a vendor that does not have access to bibliographic records from which to derive their analytic records.
3. Prepare recommendations for maintaining the accuracy of data in aggregator record sets, once a vendor makes a set available.
4. Respond to an expression of interest from UMI to collaborate with our task force on a demonstration project for ProQuest Direct.
5. Prepare specifications for other vendors of aggregator services and make appointments with them at ALA Annual to discuss libraries' needs for record sets.
6. Sponsor a discussion paper on the encoding and placement of ISSNes to propose the definition of a new subfield in field 022 to accommodate aggregators.
7. Continue to raise awareness in the library community of the issues pertaining to journals in aggregator databases. Task force members are scheduled to make presentations at NASIG, the American Association of Law Libraries, and elsewhere.

Prepared by K. Calhoun/J. Riemer 990502

---

Go to:
Program for Cooperative Cataloging Home Page
Library of Congress Home Page

---

Comments: lcweb@loc.gov (09/20/99)