

Electronic Journal Maintenance Task Force

Preliminary (Phase I) Report to the Technical Services Executive Group 30 October 2002

Executive Summary

In July 2002, the Electronic Journal Maintenance Task Force (David Banush (chair), Bill Kara, Jean Pajerek and Scott Wicks) received a three-part charge from the Technical Services Executive Group (TSEG.) From mid-July through late September, Task Force members met with CUL staff in Technical Services, Collection Development, and Library Administration to gather data on current and historical e-journal cataloging practices, determine the incoming flow of e-journal bibliographic data, and to consider the needs of collection development and administration in e-journal identification and tracking.

Task Force members devised methodologies for identifying electronic journals in the CUL library catalog and for the use of a new 948 statistics code for e-journals. Using Microsoft Access queries, the Task Force has identified (as of Oct. 28, 2002), 19,024 e-journals in the CUL catalog. The logic of the queries is given in the full report. Having identified the e-journals via these queries, the group recommends implementing two new 948 codes (one for retrospective use, the other for prospective application) to identify e-journals in the future. The Task Force is recommending that the new codes be implemented February 1, 2003.

Once the retrospective codes have been added to the existing e-journals and the prospective code is in place, generating a more accurate and complete HTML-based list of CUL e-journals should be feasible. The Task Force recommends that TSEG, in consultation with other functional groups, create a separate group to investigate the issues surrounding use of Voyager data for an HTML e-journal list.

Introduction

In July 2002, the Electronic Journal Maintenance Task Force (David Banush (chair), Bill Kara, Jean Pajerek and Scott Wicks) received a three-part charge from the Technical Services Executive Group (TSEG.) From mid-July through late September, Task Force members have met with CUL staff in Technical Services, Collection Development, and Library Administration to gather data on current and historical e-journal cataloging practices, determine the incoming flow of e-journal bibliographic data, and to consider the needs of collection development and administration in e-journal identification and tracking. Task Force members also devised methodologies for identifying electronic journals in the CUL library catalog and for the use of a new 948 statistics code for e-journals. This report details the work completed by the group during the first phase of the work and (where appropriate) its recommendations.

1. ***Gather information regarding the e-journal cataloging variations in the CUL catalog that have maintenance implications, notably single-record cataloging; machine-generated, abbreviated-level records; and multi-provider coverage.***

Historical overview of e-journal cataloging practices at CUL

In addressing this aspect of the charge, the Task Force reviewed the various ways in which e-journals have been cataloged at CUL over the years, beginning in the mid-1990s. Bibliographic description of e-journals has been evolving for more than seven years and continues to do so. The evolution of e-journal cataloging practice is reflected in the records found in the CUL catalog.

Prior to the implementation of MARC format integration in March 1996, e-journals were cataloged as

"language material" (leader/06=a) and as "serials" (leader/07=s). After format integration, e-journals were described in terms of their physical carrier, and were thus assigned the leader/06 code "m," for "computer file." In most cases, the bibliographic level was still "s" for serial. In OCLC (i.e., the CONSER database of record), e-journal records created in the serials format prior to format integration were converted to the computer file format. The conversion resulted in the early catalog records for e-journals having the leader/06 ("type of record") coded "m," for "computer file."

According to the CONSER cataloging manual, "[i]n June 1997, MARBI redefined leader/06 code 'm' to limit its usage." The new definition of "type of record" code "m" was implemented in February 1998, restricting the use of the computer file 008 to "computer file software (including programs, games, fonts), numeric data, computer-oriented multimedia, and online systems and services." Since then, the narrower use of that code in USMARC and MARC 21 has meant that almost all records for computer file serials (that is, those for electronic serials that are textual in nature) have code "a" for "language material" in leader/06, a serial 008 field, and a computer file 006 field.

Currently, the vast majority of bibliographic records for e-journals in the CUL catalog are coded "as." There are several hundred records coded "ms," with less than 100 records in which leader/06 is coded "m" for computer file and leader/07 is coded "m" for monograph, while the serial characteristics of the item are encoded in a serial 006.

David Banush developed a set of Access queries designed to identify the e-journal titles in the CUL database, taking into account the variations in cataloging practice described above (see section 3a.) There are approximately 20,000 bibliographic records for e-journals in the catalog; about 10% of these are suppressed from public view. Having identified these titles, the Task Force sought to analyze other characteristics of these records. Specifically, the charge asks us to investigate single vs. multiple record cataloging, "sleek" records, and records for those titles supplied by more than one vendor with combined holdings statements.

Single record vs. multiple records

Current CONSER practice offers libraries the option of creating a separate bibliographic record for an e-journal that also exists in print format, or of combining information about print and e-versions in a single bibliographic record. CUL developed guidelines in the mid-1990s that allowed for the creation of combined print/electronic records in cases where CUL's holdings include the print version of a title, although there is a stated preference for separate records "when all else is equal" (cf. CUL's Cataloging Procedures for Networked Electronic Resources). The Task Force met with Cecilia Sercan and Nancy Holcomb, who had conducted analyses of several e-journal aggregators to determine the most efficient and cost-effective way of cataloging them. Decisions on whether to create separate records for e-versions or to use the combined record approach were based on factors such as the size of the collection, the percentage of the collection owned by CUL in print form, the availability and completeness of bibliographic records, and the feasibility of batch-loading them. Of the estimated 20,000 e-journal records in the CUL catalog, approximately 20 percent are combined print/electronic records, with the remainder for the electronic version only.

The existing CONSER guidelines do not permit the creation of a single e-journal record representing electronic versions in multiple aggregations (unless the record also represents the print version), but a survey was recently undertaken to gauge the reaction of the CONSER community to this concept. Survey responses (including CUL's) were overwhelmingly in favor of the single record for e-versions in multiple aggregations. The results of the survey are available online at <http://www.loc.gov/acq/conser/aggrsurvresults.html>.

"Sleek records"

In the spring of 2001, the decision was made to provide title access to large numbers of e-journals in aggregators by creating and adding to the catalog abbreviated, machine-generated records, dubbed "sleek" records. At the time, it was anticipated that full-level cataloging would eventually be supplied to replace the sleek records. Currently, there are approximately 6,300 sleek records in the CUL catalog.

Initially, the sleek records were generated from title lists supplied by vendors, using a locally developed program. In the fall of 2001, CUL contracted with SerialsSolutions to purchase title-level bibliographic data for e-journal aggregators not yet cataloged and for updated data for those e-journal sets already cataloged. SerialsSolutions provides updated data bimonthly

for approximately 14,000 journals in over 80 aggregators.

Adam Chandler and Ed Zieba made a presentation to the Task Force, outlining the SerialsSolutions workflow in detail. (See Appendix 1.) The data supplied by SerialsSolutions (which include journal title, ISSN, database code, URL and other information) are run through a customized PERL script developed by Adam, generating a pseudo-MARC record. The pseudo-MARC records are converted to MARC using a utility called MARCEdit and loaded into Voyager by means of a customized Visual Basic program. Occasionally, because of interface or content changes in a given aggregator package, an entire set of sleek records may be removed. Recently, the Dow-Jones interactive records were removed from the catalog when that database became Factiva. Since early 2002, however, nothing further has been done with the vast majority of sleek records, including the addition of new titles from the bimonthly updates, pending the resolution of maintenance-related issues.

E-journal titles supplied by more than one source

When an e-journal is supplied by more than one source, CUL policy has been to create a single holdings record with the serv,remo location, representing all the e-versions. The holdings statement is compressed to reflect the combined coverage offered by the multiple providers. For example, the American Journal of Philology is part of both JSTOR (for back issues) and Project Muse (for current issues). The serv,remo holdings statement conflates the coverage into a simple statement, v.1 (1880)- . The resulting OPAC display allows users to see at a glance that the library's e-version holdings go all the way back to the first volume of the publication. The lack of granularity in these combined holdings statements presents maintenance problems when, for example, one of the providers discontinues or changes its coverage, and a cataloger has to determine where that provider's coverage ends and the other's begins. This kind of maintenance activity is labor intensive, but without it we risk the possibility that users will be misled as to the true extent of the Library's holdings.

Jean Pajerek developed an Access query designed to ascertain the number of serial records in the CUL catalog that include more than one URL (the implication being that records with more than one URL were likely to represent titles available from more than one supplier). The query includes records with bib format "as" and "ms," but not those with bib format "mm." The query results indicate that there are 520 titles in the catalog with more than one URL that have bib format "as" or "ms." (See Appendix 2.) A complete list, including the "mm" format items, could be more easily generated once the coding plan outlined below is in place.

2. Examine the incoming flow of updated e-journal bibliographic data, primarily from SerialsSolutions, but also from vendors.

At present, CUL receives a bi-monthly file from SerialsSolutions, Inc. that contains analytics for groups of titles we have requested they supply through a web interface at their site. These analytics are supplied in several forms:

- o we receive a spreadsheet of a "snapshot" of all titles we have requested, currently about 14,000. This spreadsheet contains the following data elements: Journal title, ISSN, Start Date, End Date, Provider, URL. If multiple providers serve a single title and Cornell has so requested, we receive information for each provider handling the same title.
- o we receive flat files of HTML for each title to support our e-Journals title list through the Library's web page.
- o we receive a spreadsheet of updates that notifies us of any changes to coverage from the last file supplied two months earlier. This spreadsheet includes the following data elements: Title, ISSN, Journal Action (add/remove), Database Code, Previous Date Start, Current Date Start, Previous Date End, Current Date End, Previous URL, Current URL. To date, no systematic use has been made from this data. The Task Force anticipates such use will be outlined in Phase II (Maintenance) of its charge.

Other sources of data include vendor web sites such as Lexis-Nexis. Data are harvested, parsed into spreadsheets, and then manipulated through software developed by Adam Chandler to construct files of associated bibliographic and holdings records.

In addition, data are pushed from ProQuest, Project MUSE, and JSTOR to alert CUL of changes to holdings and to new, ceased, and changed titles. In the Task Force's next phase, strategies for efficient exploitation of these data will be proposed.

3. Draft a plan for coding CUL e-journal records.

1. Identifying e-journals in the CUL catalog using MS Access queries

As noted in section 1 above, e-journals have been cataloged according to differing practices over the years. Based on a series of Access queries, David Banush determined that our remote electronic serials could be retrieved using the method outlined below.

1. The vast majority of online e-journals, including those cataloged now (both manually and via batch), have a bib format of "as" and a location of either serv,remo or mann,gate.
2. A smaller number of online e-journals have a bib format of "ms." These are chiefly older titles cataloged according to previous CONSER conventions, as noted above.
3. Other e-journals have a bib format of "mm." The format "mm" indicates a monographic computer file. The seriality of these items is expressed in a serial 006.
4. There are a number of other possibilities based on combinations of bib format and 006 (e.g., remote cartographic serials). However, there are no instances of these other combinations on bib records that also have one of the two online holdings locations. This was verified via a series of Access queries testing each combination and checking the bibliographic records in Voyager.

Queries

Please see Appendix 3 for the queries and strategy used to identify the e-journals in the CUL catalog. The queries must be executed in the order specified.

Query results

The results of the queries are as follows: (as of October 28, 2002)

	A	B	C
1		Count (Bib Records)	Format
2		18661	as
3		43	mm
4		320	ms
5	Total	19024	

2. Establish a 948 statistics code for e-journals

In deciding how best to identify e-journal records in the catalog with a code, the Task Force considered the bibliographic and holdings records, as well as a number of MARC fields in those records, including the 856, 899, and 948 fields in the bibliographic record and the 852 field in the holdings record. The group also considered using a new MARC field for recording e-journal information. The Task Force consulted with Adam Chandler to determine which fields have been retrospectively updated in the past. After this consultation and other considerations of how the various fields are currently being used, as well as the relative ease with which we could extract information from them, the Task Force determined that the MARC 948 bibliographic field was best suited for recording e-journal information. The 948 field was selected for the following reasons: it is a field that CUL has updated in the past, and one that can continue to be updated; statistics have already been extracted from the field in the past; and catalogers are already familiar with its use.

The Task Force recommends clarifying the distinction between e-resources and e-journals. Currently, e-resources are identified by the subfield 'f' value 'e' in 948 1_ \$f e. E-journals will be specifically identified by use of the subfield 'f' value 'j' in either the 948 1_ or 948 2_.

3. Propose a largely automated method for retrospectively coding e-journal records

The queries developed by David will identify those bibliographic records that should be coded as e-journal records. Once the query is run and the final table generated, he will give Adam Chandler a list of the associated bib ID numbers. Adam will then run a script to add 948 2_ \$f j to those records, signifying that the records have been edited, not newly added to the catalog. For prospective cataloging, staff would add a new 948 code, 948 1_ \$f j, to any records they handle. Such a code would also be added to any machine-generated records loaded in bulk. Other e-resources will continue to be coded as 948 1_ \$f e. Future queries of the catalog would be run keeping in mind the distinction between 948 2_ \$f j and 948 1_ \$f j; this distinction was made so that current statistic gathering codes would not have to be changed. In the future, if an e-journal or collection of e-journals is re-cataloged or withdrawn, and if any associated bibliographic record is not suppressed and is still being used for other purposes (i.e., non electronic versions of the journal), the 948 1_ \$f j will need to be removed from that record. However, the Task Force is recommending that in future, newly cataloged e-journals use separate records in all cases; thus the removal of the 948 code should not be an issue.

The Task Force is recommending that a date be identified after which all prospective e-journal cataloging would be coded with the 948 1_ \$f j code. Just prior to that date, David's queries would be run to generate the list of Voyager Bib IDs to pass on to Adam. Adam would then add the 948 2_ \$f j to those records. Using queries to identify records with a 948 \$f j (regardless of first indicator value) would then produce at any given moment the complete list of CUL e-journal titles.

The recommended date for implementation of this change is February 1, 2003.

4. *Procedure documentation for prospective use of the e-journal code*

As of an agreed-upon date, Catalogers will no longer use 948 1_ \$f e for e-journals. 948 1_ \$f e will be re-defined for integrating networked electronic resources only. Instead, staff will use 948 1_ \$f j, which will be defined as networked electronic journal.

Macros for adding this new code will be created and distributed before the implementation date. Members of the Task Force will work with Nancy Holcomb, editor of the E-Resource Cataloging Procedures, to ensure that the documentation is ready by the implementation date. Pending TSEG approval of the recommendations, the Task Force will also seek time on the Working Group on Cataloging agenda to publicize the change and address questions. An announcement on TS-Voyage-L or Catalogers-L would also be used to communicate the changes.

The Task Group will work with Adam Chandler and Jim LeBlanc to ensure that the new 948 code will be picked up in the statistics counts, and that distribution of records to the utilities will not be affected.

5. *Collection development, budgetary and statistics reporting needs for e-journal tracking*

The Task Force met with Ross Atkinson and Ed Weismann to discuss CUL reporting requirements and interests. After hearing from Ross and Ed about collection development, budgetary and statistical reporting needs, the Task Force felt that detailed retrospective and future coding that could provide the desired information would require complex mapping schema. Even with such schema in place, the group concluded that such methods would be fairly inaccurate for a large number of titles. In addition, any detailed coding scheme would result in continued maintenance when aggregators were renegotiated or when journal pricing schemes changed (i.e., free with print vs. not free with print). It was also felt that coding would not help sort out e-journal budgetary issues. Given the manner in which most publishers price their journals, it is often difficult (or impossible) to ascertain what payment or portion thereof is for the electronic version of an item. In addition, coding in the Voyager acquisitions fund structure takes this into account. Finally, titles in aggregators are often paid for on a collection level, not on an individual title level. After considering these factors, the Task Force determined that detailed retrospective and future coding of e-journal records for budgetary or collection-development purposes would be cumbersome to implement and would have too limited a benefit to be worthwhile.

It was determined that once the coding plan to identify e-journals was in place, an annual report could be written to examine the set of records for e-journals and determine which ones contain a 506 field (restrictions on access); such a report, while crude, would provide some sense of the number of e-journals that Cornell receives free and the number of e-journals for which CUL pays.

6. *The E-journal coding plan as foundation for HTML list of CUL e-journals.*

The 948 \$f j code could be the basis of an HTML list of CUL e-journals, as noted above. The next section addresses the issue in greater detail.

7. *Strategies for implementing a catalog-based HTML list of e-journals.*

The Task Force met with Adam Chandler to discuss the feasibility of extracting records coded with the 948 statistics code outlined in section 3b above and using them as the basis for an HTML-based list of all CUL e-journals. Adam felt that once the records were coded, extracting them to generate such a list would be possible technically. There are a number of potential complications, however, that would need to be resolved for the list to be of maximum benefit to public services staff and users. Some of these complications are contingent upon other recommendations the Task Force makes as it moves through the next phases of its charge. The issues include providing holdings coverage in the list, how to handle the listing of multiple titles when we have used separate records for each aggregator, how frequently the list should be updated, and other concerns.

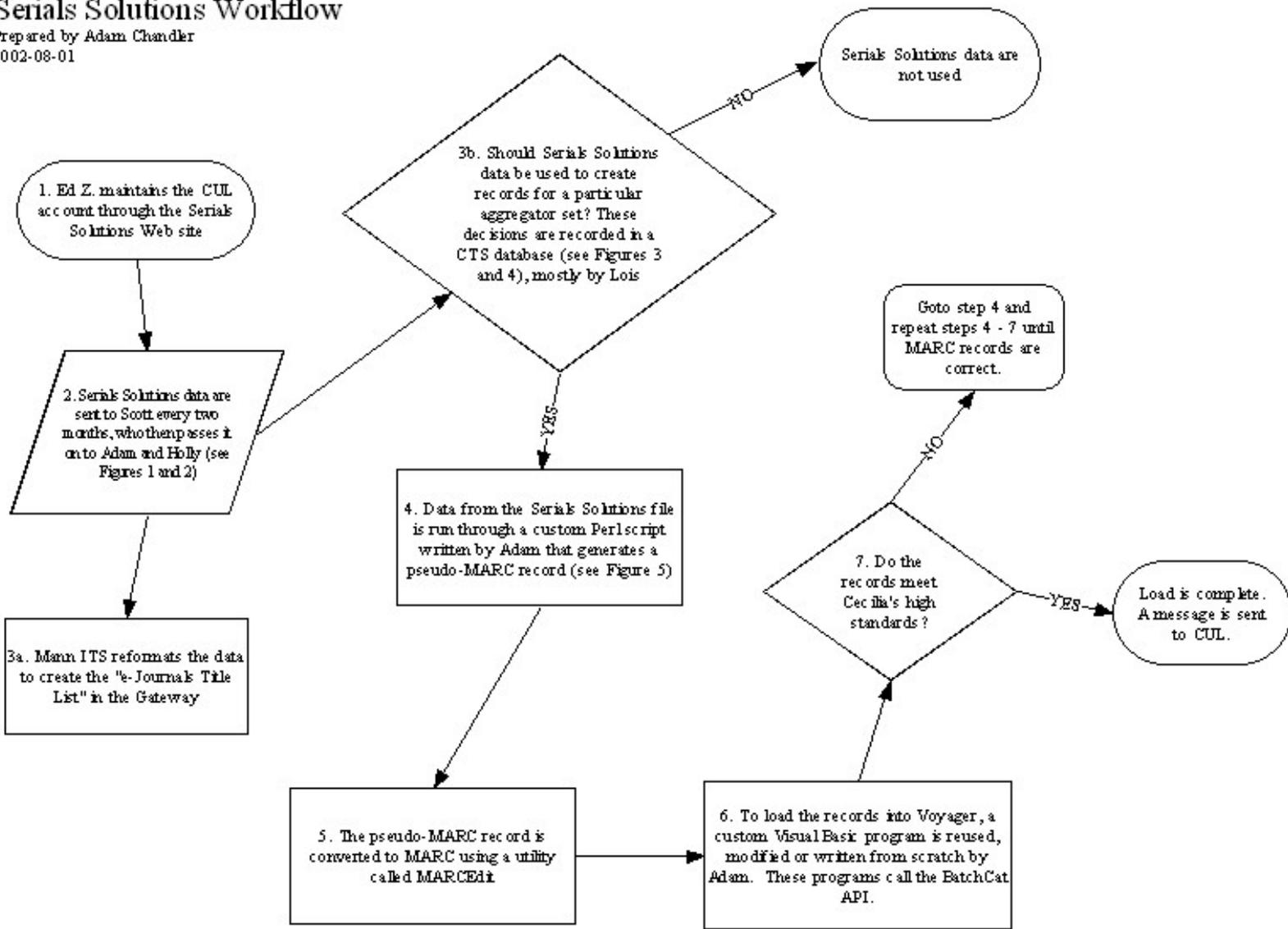
The Task Force recommends that TSEG, working with IRPC Steering and/or other stakeholders, assemble another group to consider the technical and user considerations in implementing a catalog-based HTML E-Journal list. The membership of such a group may include a member of the E-Journal Task Force, but should also include staff with the necessary information technology skills as well as public services staff. The Task Force further recommends that such a group begin its work no earlier than the proposed implementation of the recommendations given in this report (e.g., February 2003.)

Appendix 1: Serials Solutions Workflow

[Printable Version](#)

Serials Solutions Workflow

Prepared by Adam Chandler
2002-08-01



Appendix 2: SQL query to identify serial records with multiple URLs (multiple aggregators on one record)

```

SELECT ELINK_INDEX.RECORD_ID, Count(ELINK_INDEX.RECORD_ID) AS CountOfRECORD_ID,
BIB_TEXT.TITLE_BRIEF, BIB_TEXT.ISSN, BIB_TEXT.BIB_FORMAT, BIB_MASTER.SUPPRESS_IN_OPAC
FROM (ELINK_INDEX INNER JOIN BIB_TEXT ON ELINK_INDEX.RECORD_ID = BIB_TEXT.BIB_ID) INNER
JOIN BIB_MASTER ON BIB_TEXT.BIB_ID = BIB_MASTER.BIB_ID GROUP BY ELINK_INDEX.
RECORD_ID, BIB_TEXT.TITLE_BRIEF, BIB_TEXT.ISSN, BIB_TEXT.BIB_FORMAT,
BIB_MASTER.SUPPRESS_IN_OPAC
HAVING (((Count(ELINK_INDEX.RECORD_ID))>1) AND ((BIB_TEXT.BIB_FORMAT)="as" Or
(BIB_TEXT.BIB_FORMAT)="ms") AND ((BIB_MASTER.SUPPRESS_IN_OPAC)="N"));
  
```

Appendix 3. SQL for queries identifying E-journals in the CUL Catalog

A brief description of each step and query is given below, including the SQL statements. Both the query and table names (given in steps 1-5) must be used for the queries to run properly.

1. The first query is a make-table query called "E-journals and MFHDsMT." This query captures all bibs with a bib format of

"as" and a holdings location of either serv,remo or mann, gate. It excludes all bibs that have been suppressed from public view (i.e., suppressed in OPAC=N.) The results of this query are written to a new table called Ejournals.

2. The second query is called "Ejournals (old) and MFHDS Append." This query captures all bibs with a bib format of "ms" and a holdings location of either serv,remo or mann,gate. It also excludes all bibs that have been suppressed. The results of this second query are then appended to the Ejournals table.
3. The third query is called "006 Blob." This query captures all records with the bib format "mm" and the 006 field and outputs the results to a new table called "006 Blobs."
4. The fourth query is called "006 s Append." This query appends to the Ejournals table all of the records from the 006 Blobs table (see 3 above) that have a serial 006 and also have a holdings location of either serv,remo or mann,gate. It excludes all bibs that have been suppressed. The results are then appended to the Ejournals table.
5. Once all of the records have been appended into the Ejournals table, another query called "Ejournals with unsuppressed MFHDs" is run against that table. This query gathers only those (unsuppressed) bibs with MFHDs that are also unsuppressed. The result is the final count of all publicly accessible e-journals.

1. E-journals and MFHDsMT

```
SELECT BIB_TEXT.BIB_ID, BIB_TEXT.TITLE, BIB_TEXT.BIB_FORMAT, BIB_MASTER.SUPPRESS_IN_OPAC,
MFHD_MASTER.MFHD_ID,
MFHD_MASTER.SUPPRESS_IN_OPAC, MFHD_MASTER.LOCATION_ID INTO Ejournals
FROM ((BIB_TEXT INNER JOIN BIB_MFHD ON BIB_TEXT.BIB_ID = BIB_MFHD.BIB_ID) INNER
JOIN MFHD_MASTER ON BIB_MFHD.MFHD_ID = MFHD_MASTER.MFHD_ID) INNER JOIN BIB_MASTER ON
BIB_TEXT.BIB_ID = BIB_MASTER.BIB_ID
WHERE (((BIB_TEXT.BIB_FORMAT)="as") AND ((BIB_MASTER.SUPPRESS_IN_OPAC)="N")
AND ((MFHD_MASTER.LOCATION_ID)="76" Or (MFHD_MASTER.LOCATION_ID)="128"));
```

2. Ejournals (old) and MFHDS Append

```
INSERT INTO Ejournals ( BIB_ID, TITLE, BIB_FORMAT, BIB_MASTER_SUPPRESS_IN_OPAC,
MFHD_ID, LOCATION_ID, MFHD_MASTER_SUPPRESS_IN_OPAC ) SELECT BIB_TEXT.BIB_ID, BIB_TEXT.TITLE,
BIB_TEXT.BIB_FORMAT, BIB_MASTER.SUPPRESS_IN_OPAC, MFHD_MASTER.MFHD_ID,
MFHD_MASTER.LOCATION_ID, MFHD_MASTER.SUPPRESS_IN_OPAC
FROM ((BIB_TEXT INNER JOIN BIB_MFHD ON BIB_TEXT.BIB_ID = BIB_MFHD.BIB_ID) INNER
JOIN MFHD_MASTER ON BIB_MFHD.MFHD_ID =
MFHD_MASTER.MFHD_ID) INNER JOIN BIB_MASTER ON BIB_TEXT.BIB_ID = BIB_MASTER.BIB_ID
WHERE (((BIB_TEXT.BIB_FORMAT)="ms") AND ((BIB_MASTER.SUPPRESS_IN_OPAC)="N")
AND ((MFHD_MASTER.LOCATION_ID)="76" Or (MFHD_MASTER.LOCATION_ID)="128"));
```

3. 006 Blob

```
SELECT [006 BlobA].BIB_ID, [006 BlobA].TITLE, [006 BlobA].BIB_FORMAT, [006 BlobA].
SUPPRESS_IN_OPAC, GetFieldAll(GetBibBlob([BIB_ID]),"006") AS 006 INTO [006 Blobs] FROM [006 BlobA]
WHERE ((([006 BlobA].SUPPRESS_IN_OPAC)="N"));
```

4. 006 s Append

```
INSERT INTO Ejournals ( BIB_ID, TITLE, BIB_FORMAT, BIB_MASTER_SUPPRESS_IN_OPAC,
MFHD_ID, LOCATION_ID, MFHD_MASTER_SUPPRESS_IN_OPAC ) SELECT [006 Blobs].BIB_ID, [006 Blobs].
TITLE, [006 Blobs].BIB_FORMAT, [006 Blobs].SUPPRESS_IN_OPAC, MFHD_MASTER.MFHD_ID,
MFHD_MASTER.LOCATION_ID, MFHD_MASTER.SUPPRESS_IN_OPAC FROM [006 Blobs] INNER
JOIN (BIB_MFHD INNER JOIN MFHD_MASTER ON BIB_MFHD.MFHD_ID = MFHD_MASTER.MFHD_ID) ON [006
Blobs].BIB_ID = BIB_MFHD.BIB_ID WHERE (((MFHD_MASTER.LOCATION_ID)="76" Or
(MFHD_MASTER.LOCATION_ID)="128") AND ((Mid([006],3,1)="s"));
```

5. E-journals with unsuppressed MFHDs

```
SELECT Ejournal.BIB_ID, Ejournal.TITLE, Ejournal.BIB_FORMAT, Ejournal.  
BIB_MASTER_SUPPRESS_IN_OPAC, Ejournal.MFHD_ID, Ejournal.  
MFHD_MASTER_SUPPRESS_IN_OPAC, Ejournal.LOCATION_ID FROM Ejournal WHERE  
(((Ejournal.MFHD_MASTER_SUPPRESS_IN_OPAC)="N"));
```

[Return to meeting minutes](#)

11/15/02, vwb; rev. 11/20/92, vwb