Collaborative Batch Creation for Open Access E-Books: A Case Study

Philip Young, Rebecca Culbertson, Kelley McGrath

Abstract

When the National Academies Press announced that more than 4,000 electronic books would be made freely available for download, many academic libraries expressed interest in obtaining MARC records for them. Using cataloging listservs, volunteers were recruited for a project to identify and upgrade bibliographic records for aggregation into a batch that could be easily loaded into catalogs. Project organization, documentation, quality control measures, and problems are described, as well as processes for adding new titles. The project's implications for future efforts are assessed, as are the numerous challenges for network-level cataloging.

Introduction

In June 2011, the National Academies Press (NAP) issued a press release announcing that portable document format (PDF) versions of their books could be freely downloaded from their website.[1] While about 65 percent had previously been available for download, and almost all were available to read online through a web reader, over 4,000 books were now downloadable, as would be most books issued in the future.

NAP books are primarily reports from scientific panels on a variety of topics, and often ordered in print by academic libraries. A link to the online version is frequently added to the print record in either the library catalog or in OCLC. A library's catalog web page for a book can also include a link to the e-book via a link resolver or JavaScript that uses elements of the MARC record to create a search. However, these methods depend on the presence of the print version in the catalog, and most libraries do not have all of the available titles. The announcement by NAP presented an opportunity to fill those gaps and add nearly all of the content in electronic form to a library's catalog.

In several academic libraries, collection development librarians expressed interest in providing access to the e-books in their catalogs. A week after the press release, a catalog librarian made an inquiry on the Batch cataloging listserv[2] to discover whether records were available for the newly accessible e-books. An OCLC collection set could be purchased for 2,580 books published through 2008, but the records were for the print version with a link added. In 2008 the library contributing these print records suspended cataloging records for this set because they were changing in their local catalog to the separate record technique and could no longer support adding links to print records. At this time, most libraries use separate records for print and electronic versions of monographs. One respondent to the inquiry contacted NAP and was told that MARC bibliographic records were not available from them, though records could be ordered through NetLibrary or Ebrary. Neither vendor answered inquiries about the availability

and cost of a NAP record set. A third vendor was discovered but offered only a subset of the available titles.

Soon, another listserv respondent offered to organize a project to create a list of OCLC record numbers for the NAP e-books that any library could batch search and download. The record batch would therefore be free (that is, no cost above the OCLC subscription), weeded of the duplicates that plague e-book cataloging, and there would be an opportunity to upgrade the records. Doing so at the network level would save individual libraries from the work of batch searching and weeding, as well as the quality control often necessary after loading records into the local catalog. While lacking an explicit cost, the record batch would depend on time volunteered by catalog librarians to create it.

Libraries in general add fewer online open access resources to their catalogs than might be expected. The most commonly added materials in this category are government documents. Most libraries receive government document records in batches through a service such as Marcive or OCLC. Many libraries load records for e-journals in the Directory of Open Access Journals, especially since the records are often available from vendors who provide journal records for electronic resource management systems (ERMs). Some libraries also add selected websites to their catalogs. Although there are many freely available online e-books, these are not often added to library catalogs in large numbers. One reason for this is the lack of organized record batches that could be quickly loaded into the catalog. This is in contrast to vendors who often provide MARC bibliographic records as an inducement for customers to buy a particular collection of e-books. However, there are potentially many advantages to including open access e-books in library catalogs. In addition to expanding a library's collection, open access e-books can also be used as a collection weeding tool[3] and libraries will likely want to provide access to the increasing number of open access textbooks.[4]

Unlike some very large open access e-book collections, the size of the NAP collection is small enough that collaborative batch creation seemed a reasonable, attainable goal. The creation of a curated record batch would ensure record quality and reduce the burden on any individual library wishing to provide local access to this collection. Additionally, the collection consists of recent and ongoing scholarship, whereas larger freely available collections tend to be dominated by older public domain content. Since all libraries have access to the content, the NAP e-books and similar collections seem ideally suited for large-scale cooperative cataloging of record batches.

Literature Review

Academic consortia most frequently report collaborative work on record sets, usually involving quality control of vendor records. Cary and Ogburn describe the origins of what may be the earliest effort, involving a group of Virginia academic libraries with access to the same content.[5] In an attempt to avoid duplication of work, they contacted other consortia, but no other shared cataloging agreements were discovered. Their first project involved a set of vendor records that were improved and shared via file transfer

protocol (FTP). Catalog librarians in the consortium differed in skill and experience, and the project revealed "a significant need for training and help in interpreting and applying cataloging rules and standards." Shieh, Summers, and Day subsequently provided a more detailed account of the same project, including difficulties in loading, record quality, and authority work.[6] They note, "further research is needed on administrative implications of cooperative cataloging in consortia, addressing equitable allocation of personnel, scheduling in conjunction with local projects, and cost/benefit for participating institutions."[7]

Martin and Mundle relate a consortial effort to improve vendor records through communication rather than through record editing.[8] Record problems were reported on a discussion list by libraries in the consortium, then aggregated and forwarded to the vendor. They found that communicating with each other and the vendor to improve records before receipt (by reviewing sample record sets) was the best way to ensure quality metadata, assisted by the added influence of the consortia as opposed to a single library. Contrary to the pre-distribution quality control employed by the NAP project, the authors suggest that libraries may best serve their users by "working to improve accuracy, completeness, and discoverability after access has been established."[9]

Chew and Braxton describe an Illinois consortium using a shared system and its effort to establish consortial standards for cataloging electronic resources.[10] Among the problems mentioned are vendor restrictions on record sharing and the importance of record identifiers, particularly for vendor records.

Preston focuses on how cooperative e-book cataloging work was "organized, negotiated, and divided among project participants" in an Ohio consortium.[11] Work on specific record sets was negotiated by members at bimonthly meetings, and was largely dependent on the skills needed. Issues of fairness can arise when only a small minority contributes but all benefit.

Cataloging work can be distributed in a variety of ways. A post on the blog All Things Cataloged described a method used by the Bavarian library network, in which

> for one year one library 'adopts' one e-book package, taking responsibility for improving that package's metadata (which includes adding subject headings, doing authority work and, where possible, linking print version and e-book). These automatic and manual improvements are then shared cooperatively.[12]

Such consortial efforts toward batch record improvement, however, are rarely shared on a wider scale.

Little research has occurred about record batches for open access e-books. Beall discusses loading a record set for a very large collection of open access e-books (Mbooks, now HathiTrust) into the catalog.[13] Records were stripped of metadata due to the requirements of OCLC's member agreement , made available via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard, then crosswalked back

into MARC using the editing software MarcEdit. Records were then improved using global update in the catalog. Despite low metadata quality, the author felt that providing access to the content through the catalog was far more important than record accuracy and completeness.

The Cooperative Online Serials group (CONSER) completed the first year of an Open Access Journal Project begun in April 2010 to catalog any e-journals in the Directory of Open Access Journals (DOAJ) lacking a CONSER record.[14] The project recognizes the increasing importance of open access resources as libraries undergo journal cancellation projects and provide support for open access publishing initiatives. Journals in DOAJ must meet certain criteria, such as scholarly content, peer review, and assignment of an ISSN. The DOAJ project is similar to the NAP project in size, collection growth, and multidisciplinary content. Records already exist for most of the content, and the project "decreases duplicative cataloging efforts." Titles were assigned by cataloging expertise (e.g., language or subject knowledge), and a frequently asked questions (FAQ) page on the project website is provided to assist participants. [15] This project was so successful that CONSER libraries signed up for a second round of cataloging new DOAJ titles.

Hellman points to several large open access e-book collections and notes that libraries have not done a good job of including them in their catalogs.[16] One effort singled out is the University of Pennsylvania's Online Books Page, edited by John Mark Ockerbloom.[17] Over a million open access e-books are indexed, including large collections such as HathiTrust and Project Gutenberg. The metadata needs for these collections are sometimes great, and "libraries can make significant contributions, especially when they work cooperatively."[18]

Project Organization

Once catalogers were in agreement about the significance of the project, attention quickly turned to implementation. Considerable work was required to provide participants with a spreadsheet of the NAP titles. An initial title list was established based on a coverage load for SFX (3,860 titles through some time in 2008). SFX, an ExLibris product, is best known as an OpenURL link resolver, but also contains a knowledgebase. The SFX spreadsheet included only minimal metadata: title proper, provider name and URL. The NAP identifier was extracted from the URL and student labor was used to fill out the spreadsheet with ISBNs for searching and title availability status. NAP identifiers do not appear to be issued sequentially and there are unused numbers in the sequence. Newer titles were identified by student workers who checked a range of NAP identifiers on the NAP website. They began checking with a number that was deemed to be low enough to have sufficient overlap with the identifiers provided by SFX not to miss anything and stopped when they encountered a long series of unused numbers. All titles were sorted into the following categories.

1. Available in PDF and assigned an ISBN
2. Available in PDF but not assigned an ISBN

3. Available only in HTML/openbook format
4. Forthcoming and prepublication titles
5. Various categories for which a free ebook is not available

This information was imported into Microsoft Access, which was used to track the status of the project, sort titles into categories, and generate Excel spreadsheets for participants.

Early in the project, there were about eight participants, but when progress was slower than hoped for, a second call for volunteers was made on the Batch and Autocat listservs. The number of participants then swelled to about twenty, though the amount and quality of work varied by participant.

Project documentation was prepared by the organizer and evolved through four versions as participants added suggestions (see appendix). Other guidance provided by the organizer included a procedure for batch searching on OCLC Connexion, and directions for using a macro that converted e-book records to the provider-neutral standard.[19] The project had no explicit decision-making process, and the project direction depended on e-mail feedback to the organizer's questions.

One discussion centered on whether the URL should be standardized, and if so, which form of it should be used. A wide variety of NAP links have been attached to OCLC records (see table 1). A standard URL would be easier to manipulate in MarcEdit or other batch editing programs. Some forms of the URL lead to NAP's web reader, where it is less clear that a downloadable PDF is also available. These forms were rejected. A search on OCLC determined the form that was used with the greatest frequency, and this became the standard.

Table 1. Examples of NAP link variation in OCLC

nap.edu/catalog.php?record_id=9999	**General e-book page (selected)**
books.nap.edu/catalog.php?record_id=9999	General e-book page
nap.edu/catalog/9999.html	General e-book page
nap.edu/books/NI000136/html/	HTML/openbook version
nap.edu/books/030907603X/html/	HTML/openbook version
nap.edu/openbook.php?record_id=9999	HTML/openbook version
nap.edu/catalog.php?record_id=9999#toc	HTML version table of contents

Project Workflow

Upon receiving a spreadsheet of 50 titles, participants were to carry out several tasks. One of the most important was identifying the best record for a title and recording its OCLC record number. At the project's end, this would enable any library to import a text file of all the record numbers into Connexion's batch searching module to retrieve a record batch that could then be exported to the library's catalog. The high number of

duplicate records made selection difficult. A few participants reported duplicates to OCLC, but this was a very time-consuming process.

Once a record was selected, participants verified that the URL for the NAP version was present in the agreed-upon form, and working properly. Other edits ensured that any e-ISBNs, where present, were recorded in the MARC field 020 $a, and that any print ISBNs were in a 020 $z. Participants also checked that headings were authorized, in order to save work at the end of the project. Some edits were optional, such as making a record provider-neutral or adding Medical Subject Headings (MeSH) and National Library of Medicine (NLM) classification. Macros were available for automating provider-neutral record conversion, and for deriving an original e-book record from a print version. While PDF download does require registration on the NAP site (at a minimum, users must provide an e-mail address), no note about this was added to records. Individual libraries may choose to add a note, if desired, after records are retrieved.

Participants were also asked to note special situations or problems. Some OCLC records needing editing were Program for Cooperative Cataloging (PCC) records, and non-PCC libraries were restricted from editing them. The PCC status of records was noted so that they could be gathered and sent to a PCC participant for editing. A recent change in OCLC policy allows those with a Name Authority Cooperative (NACO) authorization to edit PCC records. This might have improved project workflow had the change taken place before the project began. In a very few cases, an e-book record for a title did not exist, and an original record was created. Name headings (usually for committees or conferences) without authority records (those that could not be "controlled" in OCLC) were also noted for NACO participants to address later.

When all of the spreadsheets had been returned to the organizer, it became apparent that the varying skill levels of the participants resulted in quality control problems. A follow-up project was begun using selected volunteers. Record numbers were searched in batch mode in OCLC Connexion and sent to the local save file where a macro was used to identify record errors and anomalies for cleanup. The macro focused on identifying information that would affect the usability of the records, such as lack of a NAP URL in the agreed-upon form and the presence of uncontrolled headings that might not be supported by authority records. Many of the name headings had authority records, but a few participants had not been familiar with the "control" function in OCLC which links a heading to its authority record. Some participants did not understand that the instructions from the first part of the project asked them to report headings that could not be controlled in the notes column of the spreadsheet so that they could be followed up on. In addition, there were many headings of the form

710 2b $a National Research Council (U.S.). $b. Committee on Fire Research.

where the base part of the heading in $a could be controlled, but the whole heading was not controllable. Some participants did not understand that these should be reported as problems. Because the Connexion macro command *CS.IsHeadingControlled* identifies

these partially controlled headings as controlled, they could not be identified during the second stage. Although a goal of this project was to support all headings with authority records and control the headings in Connexion, this goal was not met. This is primarily because not all of the uncontrolled headings were flagged at the point where a person was looking at the record, and it is not currently possible to retroactively identify them by automated means.

A second obstacle to our goal of comprehensive authority control is that records in OCLC Connexion are not static. Even if the record was complete and accurate at the time one of the participants last edited it, record quality may be enhanced or degraded before a library retrieves it. Due to record merging, unauthorized headings were introduced into some of the project's previously cleaned-up records.

Completed reviews were reported to the organizer, along with any problems encountered. A second follow-up project addressed titles that were part of multi-volume sets or had been cataloged as serials in print. While it is common for multi-volume sets to be cataloged as individual volumes as e-books, a decision was made to use set records where the Library of Congress had done so for the print version. However, for cases where the print version was on a serial record, such as Biographical Memoirs, cataloging as individual volumes was thought to be more practical. Authority record creation was also completed during this stage.

OCLC record numbers were then compiled into a text file and uploaded to the web. A separate text file was made available for multi-volume set records. Availability of the files was first announced to project participants for testing before wider distribution. The records downloaded from OCLC based on this list will need some editing by each library for loading into the local catalog. At a minimum, non-NAP URLs present on these provider-neutral records will need to be removed and information needed for the loading process will have to be added. This takes little time with tools such as MarcEdit, but is an extra step that will have to be performed. The record batch will also be available as a WorldCat collection set available to both OCLC and non-OCLC libraries, though at a cost. Collection set records are pre-processed by OCLC so if they are properly set up, the records can be loaded immediately. Usage data from text file downloads and WorldCat Collection Set purchases could give an indication of the project's usefulness.

Future Plans

Plans for updating the record batch are ongoing. According to NAP, about 200 new titles are issued per year. Options for keeping up with new titles include the NAP weekly e-mail newsletter,[20] the new books web page[21] (although it only lists new books for the last 30 days), and possibly a vendor knowledgebase. Volunteers could be assigned to titles on a monthly basis. However, different skills are required for creating new records (i.e., "original" cataloging). Deriving new records from the print version would not be possible in most cases, since the online version usually precedes the print version. Therefore the pool of qualified volunteers will likely be smaller. Rather than create new

records, another option would be to periodically search OCLC for new records entered by others, and add those OCLC record numbers to an updated text file.

An additional problem is that many of the new titles are issued in a prepublication version before being replaced with the published version. This process can take a few months. Should titles be cataloged as prepublication versions, or should catalogers wait for the final publication version? Advantages of the former are that catalogers would be providing timelier metadata for available content that will likely change little. The vast majority of the description, including the URL, would remain the same once a publication version was issued, although NAP does appear to use different ISBNs for the prepublication and published versions. However, records would need to be marked in some way (either through a MARC field or a list kept by the project) so they could be finalized against the publication version and the "description based on" language and physical description changed. While print prepublication versions continue to exist, online prepublications often do not. While a prepublication PDF could be downloaded by a library, this seems unlikely. If records are not updated, then they will describe a manifestation that no longer exists in its online form, and another record for the final version could be created. NAP notifications for replacement of the prepublication version with the published version could be used to update the record. Closer collaboration with NAP could also help in distinguishing titles with and without PDF versions, and in identifying any removed items.

Discussion

The NAP batch creation project was far more time-consuming than expected, and placed inordinate demands on the organizer. The project began in June 2011 and the OCLC numbers for PDF e-books with ISBNs were distributed in February 2012. Work is ongoing on the NAP e-books for PDFs without ISBNs. It has not yet been decided if the project will incorporate e-books available only in HTML format. Except for very recent releases, there were e-book records available in WorldCat for all the NAP PDF e-books with ISBNs. The larger problem was sorting through multiple records to identify the best record. The project did not offer explicit guidelines for selection of the best record, although these were implied in the instructions (see appendix) and participants were expected to have sufficient expertise to evaluate records. In e-mails among the volunteers, selecting the record with the most holdings was frequently suggested. As long as a record met criteria in the instructions, or was upgraded to that standard, record choice was not crucial.

In contrast, a significant proportion of the NAP e-books without ISBNS lack e-book or even print records in WorldCat. Even where records exist, they are often of lesser quality. Completing the process of identifying and upgrading or creating these records requires a different and more advanced skill set than the initial part of the project and it is not clear that the current pool of volunteers has the necessary resources or is willing to make the required time commitment.

The project proceeded more slowly than anticipated in part because the organizer lacked time to devote to the project at key points and became a bottleneck in the workflow. Likewise, some participants did not complete as many batches nor work as quickly as would have been necessary for a more timely finish. This reflects the reality that many catalog librarians have extensive demands on their time and that this is a volunteer project added to participants' regular duties. Other factors delaying completion were the relatively low number of volunteers and the need for error checking. The second review of the records identified errors made by a few volunteers without the appropriate skills, as well as record problems that are inevitably and inadvertently missed in a project on this scale.

It is difficult, if not impossible, to anticipate the problems encountered in such a project, and some policy changes were made after work had begun. Therefore the editing done by participants was not always consistent from beginning to end. Consistency was also affected by making some edits optional, such as adding MeSH and NLM classification desired by some participants.

Most batch creation projects will likely require too much work for a single organizer. Duties should be well distributed in order to prevent overload and project bottlenecks. A model such as the CONSER project for the DOAJ could be implemented by the PCC for open access e-book record sets such as those included in the new Directory of Open Access Books,[22] though many skilled catalogers not at PCC institutions would be excluded. Documentation proved hard to write, and it was even more difficult to get a disparate group of participants to follow it. The project revealed wide variation in the skill and knowledge of its volunteers. No assumptions should be made in this respect, particularly given the lack of an accountability mechanism. This has been a long-standing problem that was noted in the earliest consortial effort in collaborative batch cataloging.[23]

Though crowdsourcing batch creation through a global cataloging network has tremendous potential, it is difficult to ensure that quality work will result, even when specific guidance is provided. One potential way to improve the existing project directions would be to include more "before and after" examples of editing records, including screenshots. It might also be possible to create a macro for participants to run after editing a record that would alert them immediately to areas of the record needing attention. Of course, this is limited to the sorts of errors that are amenable to automated identification. It would also be wise, although more time-intensive, to initially distribute a few records to participants as a test to see if any cataloging misunderstandings exist, which would enable faster feedback. Spreadsheets would only be distributed after participants had demonstrated the ability to meet the project's standards for record quality. Alternatively, if batch e-book projects were taken on by the PCC, organizational responsibilities could be distributed, and well-trained participants would be ensured.

The ever-changing nature of OCLC's bibliographic database presents the practical problem of maintaining record consistency. In many of the consortial projects described in the literature review, edits were made to records received from vendors and then

distributed directly to consortial members. In this scenario, the quality of the distributed records can be closely controlled. For our project, since the current OCLC record use policy[24] does not allow the redistribution of records, we are limited to distributing OCLC record numbers that OCLC members can re-search to obtain the records. The updating and building upon existing metadata in OCLC records is usually a positive development, for example when different subject vocabularies or genre headings are added to records. However, the extensive record merging taking place in OCLC frequently changes record content, and sometimes for the worse. In addition to the introduction of unauthorized headings mentioned earlier, the wide variety of URLs used in OCLC records to access NAP e-books means that the links probably will not remain standardized.

The pre-distribution quality control employed by this project differs from common practice, where records are improved locally, whether before or after record load.[25] Local quality control is a vastly inefficient process due to the duplicative work done by each library receiving the records. Pre-distribution quality control, even using mostly manual record editing, saves a tremendous amount of duplicative effort by ensuring that records are largely error-free. One challenge for the local editing model has been lack of a method to batch upload record improvements to the network level once they have been distributed to individual libraries. Locally, efficient batch editing can take place via MarcEdit or global update in an individual catalog, but transferring these improvements to OCLC would require editing individual records one at a time. There is a conflict between the need to quickly load records for immediate access to electronic resources, and the desire to reduce duplication of effort in the editing and authority control of the records.

While network-level cataloging work remains the exception rather than the rule, encouraging recent steps have been taken by OCLC toward that ideal. These include the expansion of the pilot Expert Community Experiment into a permanent program, the extension of PCC record editing privileges to NACO members, an algorithm to programmatically perform heading control on new and existing records, and WorldCat Local. At the same time, network-level cataloging is hindered by OCLC's record use policy and the proliferation of other record sources. It is also difficult or impossible in OCLC's Connexion interface to do the kinds of efficient batch editing of records supported by MarcEdit and many local ILSs. Newer bibliographic utilities such as SkyRiver and Biblios.net freely share records, but OCLC, MARC record services, and some vendors place restrictions on record sharing. Cataloging may become more network-like yet remain in restricted silos. Truly network-level cataloging will require freely sharable records.

Future Research

Possibilities for new metadata elements emerged during this project. A code for the identification of open access resources would serve two purposes for libraries. First, it would enable catalogers to find and gather resources for adding to the catalog. Second, if the code was part of the catalog display, education and advocacy about open access in

academic libraries might be furthered. Also, a metadata element to indicate publication status could help solve the problem with prepublication versions in this project. This indicator could alert a cataloger that the record description needed updating when the final publication version was available. The result would be faster delivery of electronic content to catalog users. Finally, a metadata element to describe e-book formats is badly needed. The term "e-book" has been used for a wide variety of online textual material, but users need to know whether it is downloadable (and if so in what format) or only readable in a web browser (thus requiring an internet connection).

Further work is needed to ensure metadata consistency between formats. In the course of upgrading NAP e-book records, participants often consulted the print record. Print records sometimes contained desirable metadata not present on the e-book record, such as MeSH and NLM classification. Conversely, e-book records often contained contents notes and summaries not present on the print record. It was also more common for e-book records to link to the print version (using the MARC 776 tag) than vice versa. The two formats occasionally presented conflicting metadata in the form of a differing main entry or call number. This metadata divergence for identical content does not help the catalog user. Implementation of FRBR-aware catalogs, in which metadata at the work level can be applied to all formats, may solve this problem. FRBR may also help in situations where multi-volume sets were cataloged on a single record in print, but each volume on a separate record as an e-book.

The case for open metadata should be made more clearly and forcefully. Freely sharable metadata will mitigate the effects of the competing MARC record silos that are developing. Open metadata may be a requirement for the linked data environment the library world is currently exploring,[26] but there is no need to wait until then.

Conclusion

The NAP e-book project represents a unique and successful collaboration resulting in a batch of over 3,500 records available for loading into local catalogs. No other cataloging project, to our knowledge, has been accomplished with such a wide variety of volunteers. While this aspect of the project hindered consistency, future projects can implement suggested controls. The CONSER project for DOAJ could serve as an organizational model. As open access resources increase in numbers and prominence, libraries will need to devote greater attention to metadata for them.

Elimination of duplicative record editing is badly needed. Lacking a mechanism to upload a batch of corrected records, this project employed pre-distribution quality control at the individual record level. This quality control was affected by the skills of the volunteers as well as the dynamic nature of a large bibliographic database. Open metadata is needed if network level cataloging is to be realized. Despite the problems encountered in this project, organized batch creation projects are an effective way to provide access to important collections.

Notes

1. "The National Academies Press Makes All PDF Books Free To Download; More Than 4,000 Titles Now Available Free To All Readers," accessed February 28, 2012, http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=06022011

2. Batch listserv, http://listserv.vt.edu/cgi-bin/wa?A0=BATCH.

3. Kirstin Steele, "Free electronic books and weeding," *Bottom Line: Managing Library Finances* 24(3) (2011): 160-161, http://dx.doi.org/10.1108/08880451111185982.

4. Steven Ovadia, "Open-Access Electronic Textbooks: An Overview," *Behavioral & Social Sciences Librarian* 30(1) (2011): 52-56, http://dx.doi.org/10.1080/01639269.2011.546767.

5. Karen Cary and Joyce L. Ogburn, "Developing a Consortial Approach to Cataloging and Intellectual Access," *Library Collections, Acquisitions, & Technical Services* 24 (2000): 45-51, http://dx.doi.org/10.1016/S1464-9055(99)00095-0.

6. Jackie Shieh, Ed Summers, and Elaine Day, "A Consortial Approach to Cooperative Cataloging and Authority Control: The Virtual Library of Virginia (VIVA) Experience," *Resource Sharing & Information Networks* 16:1 (2002), 33-52, http://dx.doi.org/10.1300/J121v16n01_04.

7. Shieh, Summers, and Day, "A Consortial Approach," p. 48.

8. Kristin E. Martin and Kavita Mundle, "Cataloging E-books and Vendor Records: A Case Study at the University of Illinois at Chicago," *Library Resources & Technical Services 54*(4) (2010): 227-237, http://alcts.metapress.com/content/h1455767637633x8/.

9. Martin and Mundle, "Cataloging E-books," p. 235.

10. Chiat Naun Chew and Susan M. Braxton, "Developing Recommendations for Consortial Cataloging of Electronic Resources: Lessons Learned," *Library Collections, Acquisitions, & Technical Services* 29 (2005): 307-325, http://dx.doi.org/10.1016/j.lcats.2005.08.005.

11. Carrie A. Preston, "Cooperative E-Book Cataloging in the OhioLINK Library Consortium," *Cataloging & Classification Quarterly* 49 (2011): 257-276, http://dx.doi.org/10.1080/01639374.2011.571147.

12. All Things Cataloged, "Publisher e-book metadata," (July 7, 2011), accessed February 28, 2012, https://allthingscataloged.wordpress.com/2011/07/07/publisher-e-book-metadata/.

13. Jeffrey Beall, "Free Books: Loading Brief MARC Records for Open-Access Books in an Academic Library Catalog," *Cataloging & Classification Quarterly* 47 (2009): 452-463, http://dx.doi.org/10.1080/01639370902870215.

14. CONSER, "Cooperative Open Access Journal Project Planning Group Report, April 30, 2010." http://www.loc.gov/acq/conser/Open-Access-Report.pdf .

15. CONSER, "Open Access Journal Project FAQ." Updated August 26, 2010. http://www.loc.gov/acq/conser/Open-Access-FAQ.html.

16. E.S. Hellman, "Open Access E-books," *The No Shelf Required Guide to E-book Purchasing, Library Technology Reports* 47:8 (2011): 18-27, http://alatechsource.metapress.com/content/r7u235k327mm3q3h/.

17. "The Online Books Page," edited by John Mark Ockerbloom, accessed February 28, 2012, http://onlinebooks.library.upenn.edu.

18. Hellman, "Open Access E-Books," p. 24.

19. Program for Cooperative Cataloging, *Provider-Neutral E-Monograph MARC Record Guide.* (Prepared by Becky Culbertson, Yael Mandelstam, George Prager, includes revisions to September 2011). Retrieved February 7, 2012 from http://www.loc.gov/catdir/pcc/bibco/PN_Guide_20110915.pdf.

20. "Subscribe to the NAP Newsletter," The National Academies Press, accessed March 6, 2012, http://www.nap.edu/updates/index.html.

21. "New Releases," The National Academies Press, accessed March 6, 2012, http://www.nap.edu/new.html.

22. "A New Service for Open Access Monographs: the Directory of Open Access Books," Open Access Publishing in European Networks, accessed March 5, 2012, http://project.oapen.org/index.php/news/46-doab-press-release.

23. Cary and Ogburn, "Developing a Consortial Approach," p. 50.

24. "WorldCat Record Use Policy," OCLC, accessed February 28, 2012, http://www.oclc.org/worldcat/recorduse/default.htm.

25. Elaine Sanchez, Leslie Fatout, Aleene Howser, and Charles Vance, "Cleanup of NetLibrary Cataloging Records: A Methodical Front-End Process," *Technical Services Quarterly 23*(4) (2006), 51-71, http://dx.doi.org/10.1300/J124v23n04_04.

26. Raymond Bérard, "Free Library Data?," *Liber Quarterly* 20:3/4 (2011), 321-331, http://liber.library.uu.nl/publish/articles/000512/article.pdf.

Appendix

National Academies Press Free Ebook Project

Goals

- Update OCLC ebook master record for free NAP ebooks to include NAP URL in the form http://www.nap.edu/catalog.php?record_id=?????
- Perform basic quality control on the master record
- Compile a list of OCLC numbers that can be used by participants and others to batch search, attach holdings and load the complete set of records into local catalogs

Initial set up

A list of NAP title IDs, titles, and (where available) ISBNs has been prepared covering the free NAP ebooks. Spreadsheets have been prepared that include batch search strings and 856 fields that can be cut and pasted into Connexion records.

Perform batch search on assigned record range

Each participant in this project will receive one or more spreadsheets with a list of titles to search and upgrade. The spreadsheet has columns with various search strategies that can be used for batch or individual searches and a column with URLs for pasting into Connexion.

The columns that can be used for batch searches are: ISBN, ISBN limited to records held by Ebrary, title keyword combined with "National Academ*" as publisher (to pick up National Academy or National Academies) both as a plain search and as a search limited to records held by Ebrary. Each search is limited to mt:cai (cataloged as internet resource) and ll:eng (for English language records).

To use for batch searching, select and copy the cells from the column you intend to use and paste the results in a Notepad text file. Separate instructions are provided for using the text file with the selected searches to generate a batch search in Connexion if you are not familiar with this process.

Searches from any of these columns can also be copied and pasted into the command line search box and run individually.

Select record to use and make the following edits

If more than one record is retrieved, select the best record. Once you have selected a record to use, make the following edits.

Try to select a record that has an LC call number and LC subject headings that appears to be based on the print record. If the record does not have an LC call number and LC subject headings, add them if possible (note: I am finding some records with 588 Description based on print record with no 776 and no print record that I can find in OCLC; these probably should not be cataloged as "based on print record.")

Choose a record for an online item (no print records with URLs)
Encoding level I or L if possible (upgrade if you feel comfortable)

Check and fix if needed:
008/23/Form = o
006 = m\\\\\\\\d
007 = cr [leave any additional codes if already on the record, but it is not necessary to add
        them unless it is your local practice
020: add $z in front of any print ISBNs
050 _4 This is preferred to the 090 for LC call numbers not assigned by LC; ditto for the
        060 _4
245|h = [electronic resource] [If an AACR2 record; there is at least one RDA record in
        this set]
Control headings if possible
Provider-Neutral tidbits: If you are cataloging as P-N, then you will either be using 500
        Title from PDF t.p. (National Academies Press, viewed July 1, 2011) OR 588
        Description based on print version record. In either case, with P-N, delete the 538
        Mode of access note. *[NOTE: The 500 field in the sentence above has now been
        changed to: "...then you either be using 588 Description based on online
        resource; title from PDF t.p. (National Academies Press, viewed..."]*

Make sure there are LCSH subject headings in the record; would be nice to add
        NLM/MeSH headings if they are available and you have time
Delete any institution-specific or proxied 856 fields
856 40 Remember that the second indicator is zero
Add National Academies Press catalog record URL with $3 for National Academies
        Press (can copy from spreadsheet; might be a good idea to make first URL). Do
        not include $z
        URL should be in the form: http://www.nap.edu/catalog.php?record_id=12815
        A good practice would be to make this the first URL (since it will be accessible to
            all users) and to delete other URLs going to the NAP site (so that we have
            consistent results for manipulation with MarcEdit)
If the record is not provider-neutral, you may choose to make it into a provider neutral
        record, but this is not required.

IMPORTANT Click on the URL in the 856 and make sure it works.

After making changes, replace master record.

Update spreadsheet and return to coordinator

Update your copy of the spreadsheet with the relevant OCLC numbers. Just insert the plain OCLC numbers; it is not necessary to add any prefixes such as * or #. Add any questions or concerns or describe any unusual situations in the notes column. For example, it would be useful to note RDA records in the notes column. Return your completed spreadsheet to the coordinator by email. This will be used to compile a list of OCLC numbers that will be distributed to participants and posted publicly somewhere.