

Adding Free eBooks to the Library Catalog

This project was supported by a Library Services and Technology Act (LSTA) granted to Santa Clara County Library (SCCLD) from the California State Library to fund the exploration of testing the effect of including public domain eBooks in the library's catalog for discovery by users, and to observe whether eBook use would increase for the selected books. The project was conducted between July 2013 and June 2014. The resulting software is available as Open Source at <https://foss4lib.org/package/openlibrary-utilities-scclد>.

Background

There exist today a number of sources of free, open eBooks: Project Gutenberg, Google, Internet Archive, HathiTrust, and others. While some savvy users know to look for these, others are unaware that the book they may be seeking is available at one of these sites. It is also the case that the books most sought by readers are not easy to find among the many millions of digital books in these online databases, not only because of the size of the database but because there is often a mix of public domain books and books that are restricted by copyright.

The circulation of eBooks protected by copyright and vendor licenses is a complex function that cannot be performed by the library catalog or circulation module. Instead, eBooks are delivered through a vendor system that is outside of the library. To support this, eBook vendors created a separate interface for library patrons to use to discover and obtain the eBooks. The confusion generated by these multiple eBook interfaces (each eBook vendor had their own), was somewhat improved when the eBook vendors began providing MARC records so that patrons could discover eBooks inside the library catalog. The links in those MARC records still took the user to the vendor interface for eBook selection and check-out.

The vendor system, which primarily prevents users from making usable copies of eBooks outside of the digital rights management system, is unnecessary for public domain eBooks, which can be freely distributed with no rights management controls, and can be freely copied by users. However, like the early vendor eBook systems, these sources of eBooks are not integrated with the library catalog, and therefore users are generally unaware that the source exists, and are not made aware of the availability of specific titles through the library catalog.

Just as it has become clear that the use of purchased eBooks use is increased when patrons are able to discover them through catalog searches, it is also highly likely that use of open access eBooks could be increased if records for desirable titles were inside the library catalog. Open access books include those books considered

"classics" up to the early 20th century (such as the books of Jane Austen or Mark Twain), as well as still relevant non-fiction in philosophy, history and science.

The Free eBooks Project

Santa Clara County Library (SCCLD) LSTA-funded project was designed to test the effect of including public domain eBooks in the library's catalog. One critical goal of the study was that users would download the open access eBook directly within the library catalog interface, without being sent to a site outside of the library. In particular, the study had the following goals:

1. Discover and Download eBooks Without Leaving the Catalog

While other eBooks services may include some of these same open access titles, most links to eBooks direct patrons to the eBook provider's interface to obtain the eBook. This means that users leave the library catalog and encounter a visibly different site outside of the library. Users may not perceive the library as the source of these eBooks, or may be unclear as to their relationship with the entity providing the eBooks. SCCLD particularly wanted users to obtain eBooks directly from the library catalog environment with which users are familiar.

The LSTA grant project, *Evaluating Effect of Adding Free eBooks to the Library Catalog*, has developed a way to add selected open access eBooks to the library catalog without requiring the patron to leave the library catalog and without requiring the patron to "check-out" the item. Since these titles are in the public domain, each user can download their own copy and keep it or even share it with others. Users also do not need to provide their library card number for the download, which means that this is a service open to all.

2. Selectively Add Titles to the Library's Collection

The eBooks used for this project come from the Internet Archive, whose user interface is the Open Library. The Internet Archive contains over one million free eBooks. The eBooks are available in a variety of eBook formats, and many of the eBooks have an associated MARC record. However, the majority of the eBooks available through the Internet Archive are not considered classic works of literature. It was therefore important to develop a mechanism that allowed libraries to identify the specific titles they'd like to include in their collection. A list of the titles available at SCCLD can be found at <http://www.sccl.org/classics>

Through the usage statistics that were gathered during the project, SCCLD discovered that patrons immediately were finding and downloading the eBook classic records in

the catalog. In March, 2014, the first full month of data, 300 eBooks had been downloaded, with a good blend of ePub and Kindle, and very few other formats. By May of that year, usage had jumped to over 750 downloads per month. This proved the point that patrons will find materials if they are in the catalog, and that they find them easy to use and will download them. Survey results expressed keen interest in the titles, and the hope is that more titles would be added. Statistics on use can give collection development staff information on demand for open eBooks, and they can add other open eBooks to the catalog to meet this demand.

Step-by-Step Project Workflow

The following sections describe the process undertaken by SCCLD in its project, and which may be used by other libraries or modified as needed. All of the programs developed for the project are available at [??Tom?]. This description is intended to provide enough detail to help other libraries determine their own workflow.

1. Create a List of Desired Titles

In order for a library to make use of the scripts that SCCLD has created, they must first generate a list of titles and authors that they'd like to add to their collection. These can be titles the library already holds in hard copy, or it can be additions to the library's collection.

The author/title list of desired titles that the library creates is entered into a spreadsheet with a separate column for authors and titles. SCCLD used Google Docs for this as it allowed staff members to easily share the work on the list. [?? could the program be changed to use a non-online file??]



	A	B
1	Alice's adventures in Wonderland ;	Carroll, Lewis, 1832-1898.
2	My Antonia	Cather, Willa.
3	Don Quixote of La Mancha	Cervantes Saavedra, Miguel de, 1547-1616.
4		

These fields can be copied from entries in the library catalog. It is not necessary to include the full authority-controlled form of the author name to obtain matches. This may be especially useful when including on the list books the library does not already

have in its catalog. The spreadsheet is then input to the program that will search for these books in Open Library.

Before starting, you may want to familiarize yourself with the Open Library collection. When searching for particular titles, you can add the filter to show only eBooks, but this includes both public domain and protected eBooks. You can retrieve the entire set of Open Library public domain books at <https://openlibrary.org/read>. However, note that this is a list of about one million digital books.



2. Select Editions

The spreadsheet is input to a program that searches the Open Library by author and title and returns results with links to the Open Library entries. These results are stored by the program spreadsheet format. (If using Google Docs, all of the input and output takes place in the same spreadsheet.) In many cases there is more than one result returned because popular classics have often been published in numerous editions, and more than one edition has been digitized. From the list of results, entered into the same spreadsheet, the library must choose which edition or editions they wish to add to their catalog. The results include an estimate of the quality of the OCR (optical character recognition) process that was undertaken when the item was added to the Internet Archive. This number is output from the OCR program when it was originally run. It was the judgment of the SCCLD staff that this measurement generally did identify the most readable copy of the book.

Little women	Alcott, Louisa May, 1832-1888.	17	94.02	1915	http://openlibrary.org/books/ia:littlewomenorme
	Alcott, Louisa May, 1832-1888.		93.47	1868	http://openlibrary.org/books/OL1175196M
	Alcott, Louisa May, 1832-1888.		92.18	1919	http://openlibrary.org/books/OL25129264M
	Alcott, Louisa May, 1832-1888.		91.28	1868	http://openlibrary.org/books/ia:littlewomenorme
	Alcott, Louisa May, 1832-1888.		81.90	1915	http://openlibrary.org/books/OL7208979M
	Alcott, Louisa May, 1832-1888.		81.86	1868	http://openlibrary.org/books/OL7060579M
	Alcott, Louisa May, 1832-1888.		81.81	1887	http://openlibrary.org/books/OL7150341M
	Alcott, Louisa May, 1832-1888.		81.76	1896	http://openlibrary.org/books/OL7092213M

Other criteria may be used to select books, for example a library may wish to prefer first editions or illustrated editions. The results list includes the link to the Open Library page for the eBook so that staff can look at metadata for individual editions or download the eBook itself.

Selection is done by copying the URL for the selected edition to another column in the spreadsheet. This column is then input to a program that retrieves the associated MARC record from the Internet Archive.

3. Modify the MARC Record

Once a specific title and edition is selected, a MARC record is generated for the library catalog with software developed for the project. The MARC records generated are stored at the Internet Archive associated with the eBook. These records are often provided by the library whose edition was digitized. Note that not every eBook on the Open Library has a MARC record. If a MARC record does not exist, the SCCLD project did not select the title. (There has been some discussion with the Internet Archive of adding MARC records where they do not exist, but this was outside the scope of our project.)

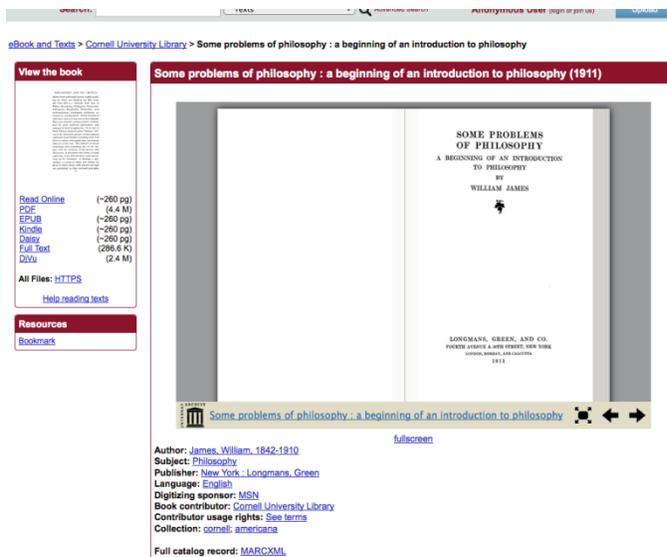
The Open Library has a standard export format for the MARC record for the eBook, which is derived from the record for the hard-copy that was digitized. This exported record is coded for the eBook in the 007 field.

Here's a sample exported MARC record for the eBook before it has been modified with the library's preferences.

LEADER: 01713nam 2200373 4500
 005 20060625140956.0
 007 cr|||||||||
 008 740606s1920 j 0 1 eng
 010 \$a 20015592
 020 \$a0553211285 (pbk.) :\$c\$5.00
 035 \$a003290353\$bGB
 035 \$a740239288
 040 \$dUtOrBLW
 091 \$pJ\$aFIC\$cT
 100 1 \$aTwain, Mark,\$d1835-1910.
 245 14 \$aThe adventures of Tom Sawyer,\$h[electronic resource]\$cby Mark Twain.
 260 \$aNew York,\$cc1920.
 520 \$aThe adventures and pranks of a mischievous boy growing up in a Mississippi
 River town in the early nineteenth century.
 650 0 \$aBoys\$zMissouri\$vFiction.
 651 0 \$aMississippi River\$vFiction.
 856 41 \$u<http://archive.org/details/adventuresoftoms00twai2>\$zFree eBook from the
 Internet Archive
 856 42 \$u<http://www.openlibrary.org/books/OL7183411M>\$zAdditional information
 and access via Open Library

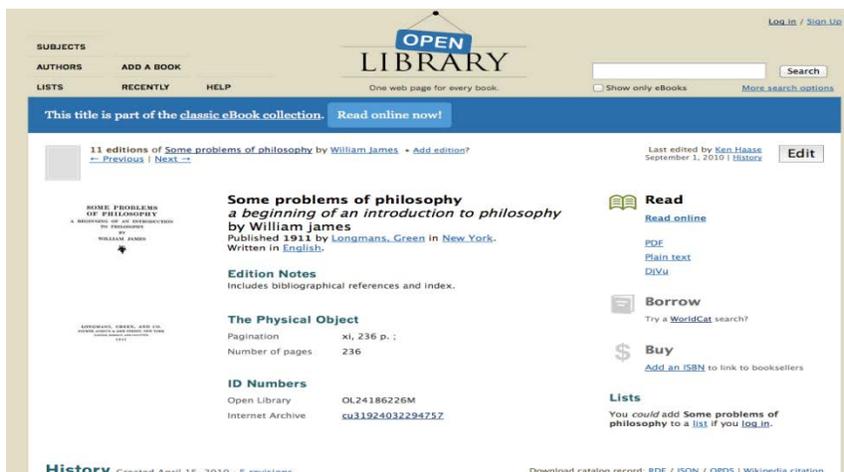
Primarily, the library needs to decide what linking fields (856) are desired, and how these should display in the catalog. The default fields link to the Internet Archive page for the eBook:

There are two links in the default MARC record. The first takes the reader to the Internet Archive page about the eBook with all of the information about the book.



(Click for full page view.)

The second links to the Open Library page for the eBook:



The SCCLD project wanted to lead users directly to an eBook download. SCCLD also had examined the popularity of formats by its users, and determined that by far, ePub and Kindle formats were more used than any other eBook format. Because of this, they decided to include three 856 fields: the first link goes to the the ePub version of the edition on Internet Archive, another to the Kindle link, and a third link takes the reader to a page on the Internet Archive where all the formats are available.

The eBook formats available from the Internet Archive are:

- PDF

- ePub
- Kindle
- Daisy
- Text
- DjVu
- Web-based book reader

What Libraries Need to Do to get Started

1. System Requirements

Before you get started you need to understand the following:

1. What input formats can your library cataloging system take in? Common formats are MARC21 and MARCXML, but your system may also take in MARCEdit edit format, or others.
2. What 856 subfields are used for the user-display?
3. In which displays are 856 fields included? (e.g. results list, short display, long display)
4. If you wish your 856 fields to display in a particular order, determine if your ILS vendor sorts or re-orders fields before they are displayed. If so, you may have options to manage the display order.

When you have the answers to these questions, you probably want to load one or a few test records to make sure that it actually works the way you would like.

2. Cataloging Choices

The original study was done in a catalog that used the "separate record" method. When the library added an eBook version of a print book that they already had, each was represented by its own record in the library database. For libraries that use the "single record" method (adding links to eBooks on the print book record), moving the 856 fields from the incoming MARC record to the library's print book record is not automated in the software developed here.

3. Modifications to the MARC Format

As mentioned above, the MARC record as output from the Internet Archive has two 856 fields that would usually display as:

[Free eBook from the Internet Archive](#)
[Additional Information and access via Open Library](#)

Both of these lead the user to a page outside of the library catalog. If you prefer to give users a direct link to one or more eBook formats, this can be coded in the program that processes the record after it is downloaded from the Internet Archive. The format of the URLs for individual eBook formats is:

```
https://archive.org/download/[OL identifier]/[OL  
identifier]/[OLidentifier].[eBookFormat]
```

An example of three different eBook types for a single edition:

```
https://archive.org/download/adventuresalices00carrich/adventuresalices00car  
rrich.epub  
https://archive.org/download/adventuresalices00carrich/adventuresalices00car  
rrich.pdf  
https://archive.org/download/adventuresalices00carrich/adventuresalices00car  
rrich.mobi
```

The one exception to this is the link to Kindle books, since those must be downloaded through the Kindle service. The link pattern for Kindle books is:

```
https://www.amazon.com/gp/digital/fiona/web-to-kindle?clientid=IA&itemid=[OL  
identifier]&docid=[OL identifier]
```

Using the same book as the above examples, the Kindle edition URL is:

```
https://www.amazon.com/gp/digital/fiona/web-to-  
kindle?clientid=IA&itemid=adventuresalices00carrich&docid=adventuresalices00car  
rrich
```

A library can choose to include links for all eBook formats, or only a select few. The SCCLD project decided to link directly to ePub and Kindle formats, and added a third link to the Open Library page that lists the remaining formats.

In addition to the URLs (that will be entered into the 856 \$u subfield), a display message needs to be included in the 856 in the subfield used by your system for user display. This message can be anything you would like. The SCCLD decided to call these "Always available eBooks", followed by the eBook format:

[Always available eBooks \(EPub\) »](#)

Your choice will require minor modification of the MARC record generation code.

4. Keeping Statistics on Use

If a library wishes to keep statistics on how often the links are followed, it must first determine if the library system can supply those. If not, a method was developed for the SCCLD project using Bitly URLs. If the library system can provide statistics on "clicks," or if there is not a need to quantify the use of the open access eBooks, then one can use the Open Library URLs as described.