CuLLR the Catalog

Holly Mistlebauer, MLIS
Cornell University
Albert R. Mann Library
Ithaca, New York 14850
USA
E-mail: hlm7@cornell.edu

Abstract:

CuLLR (Curated List of Library Resources) was developed and implemented by the Cornell University Library to primarily address the discovery needs of patrons without access to a physical library. CuLLR is a process whereby print and electronic resources in the library catalog (in the form of metadata), and from other sources to be determined, are extracted in accordance with the subject areas for a specific unit library. These resources may then be annotated to identify the subject areas for which the resource is useful and other attributes as determined by the library. CuLLR has been implemented for Cornell's Engineering Library, Physical Sciences Library, and Entomology Library.

Introduction

Cornell University, like many universities, has unit libraries for various subject areas (e.g., Engineering Library, Fine Arts Library, Mathematics Library). Over the last two years Cornell has closed the physical facilities of three of these unit/subject libraries: the Engineering Library, the Entomology Library, and the Physical Sciences Library. Not having a physical library means that a virtual library must be established to assist the patrons in discovering the appropriate materials without librarian assistance. The approach Cornell chose is to create a sub-catalog that contains only the materials for the subject area of the unit library and to annotate this sub-catalog's resources in such a way as to make their usefulness obvious to patrons. By adding

relevance searching and the weighting of resources, specific key resources appear higher in the search results so that patrons become aware of their existence. Without these features, all catalog resources appear to be equal to patrons. These changes could not be made within the catalog, so they had to be done externally. In addition to assisting patrons in finding pertinent materials, the sub-catalog allows each of the closing libraries to maintain an identity and allows the subject-area librarians to have more local control over what is presented to the patrons. An additional bonus is that when an entomologist searches for "apis" (the genus of honeybees) they won't have to sift through all the search results related to computer software APIs. All of the "apis" search results will be relevant to the entomologist.

What is CuLLR?

CuLLR (Curated List of Library Resources) is the name of the tool Cornell University Library has developed to facilitate subject access. CuLLR collects subject metadata for a particular unit/subject library into one database, representing what the physical library used to contain plus all related electronic resources.

CuLLR consists of a database and interface by which to update the database. It is not a website, although its data is used on websites. The initial data for inclusion in the CuLLR database is identified by a unit library's subject librarians using call number ranges, catalog record identifiers, and/or ebook package identifiers for resources within the subject area. After the selected catalog records have been imported into the given library's CuLLR database, additional non-catalog data about the resources is added, some in an automated manner and other data manually.

How was CuLLR developed?

The CuLLR project team was unable to locate relevant research about related projects at other libraries; consequently, we worked entirely on our own. The team was led by the author, and included a librarian representing each of the three closing libraries as well as a librarian representing the libraries that are not slated for closure. The team was rounded out by a metadata librarian and two programmers. By including the four librarians, we hoped to develop a product that would work for all the libraries here at Cornell. (While the project was initiated due to the needs of the closing libraries, the product itself was felt to be useful for any library.) The CuLLR project team started working on the first phase of CuLLR in late February of 2011 and finished in mid-August of the same year. The end result of the first phase of the project was the CuLLR database ready for implementation on library websites.

CuLLR implementation for the Engineering Library

The first implementation of the CuLLR database was within the Engineering Library's website at http://engineering.library.cornell.edu/ (**Figure 1**). This project was initiated in the mid-summer of 2011 and publically launched in January of 2012. The Engineering Library uses data from their CuLLR database for many of its features and functions. The "Browse by Subject" option is created using the "Discipline" (i.e. subject) field in CuLLR. "Search for Engineering Resources"

searches the CuLLR database and provides the search results in order by relevance. The pages linked to by the options "Journals & E-Journals", "Books & E-Books", "Handbooks", "Standards", "Course Guides", and "Research Guides" are all search results from the CuLLR database that have been produced by predefined searches.
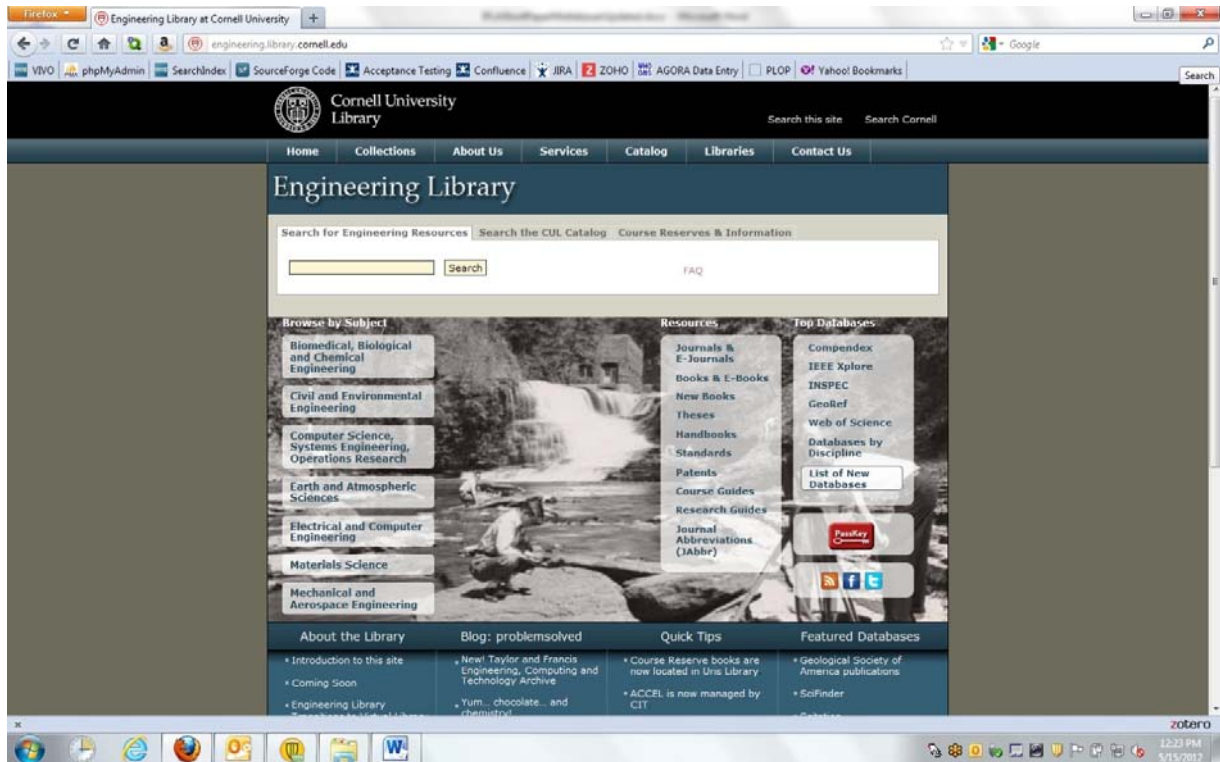

Figure 1: Screenshot of the home page of the Engineering Library website (http://engineering.library.cornell.edu/)

After selecting a "Browse by Subject" option, searching the Engineering Library, or selecting one of the predefine searches (i.e. "Journals & E-Journals", "Books & E-Books", "Handbooks", "Standards", "Course Guides", and "Research Guides") the "Format", "Content", and "Discipline" facets will display **(Figure 2)**. These facets are used to narrow down the results and all come from CuLLR. Sorting options of "Year" and "Title" are also available.

As an example, let's say you want to find journals that cover materials science. You are working from home, so you want to find electronic journals only. You would simply select the "Journals & E-Journals" option from the home page, click on the "ejournal" option under "Format" on the page that appears, and then click on the "materials science" option under "Discipline." This quickly reduces the approximately 23,000 engineering journals to a more manageable 2,200. At any point during the faceting, a search may be performed to further reduce the results to a relevant few journals.
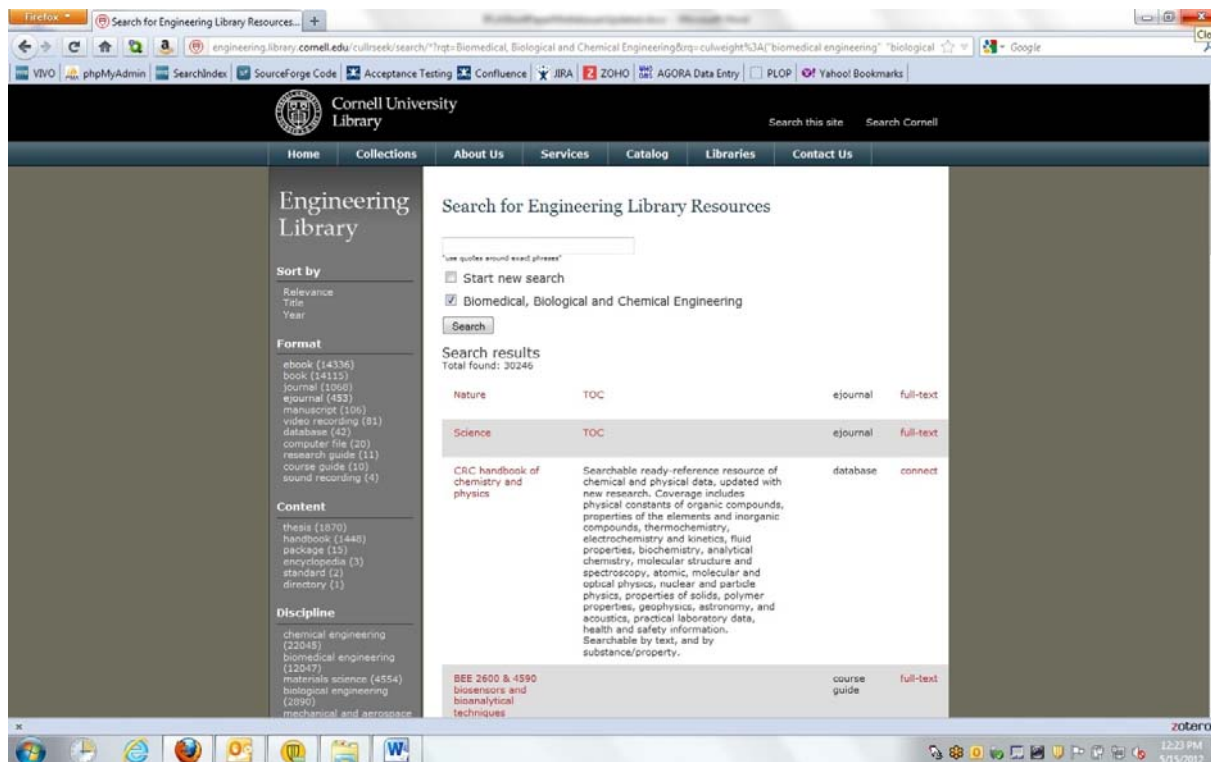
3

Figure 2: Screenshot of the "Biomedical, Biological and Chemical Engineering" subject browse within the Engineering Library website
(http://engineering.library.cornell.edu/cullrseek/search/*?rqt=Biomedical,%20Biological%20and%20Chemical%20Engineering&rq=culweight%3A%28%22biomedical%20engineering%22%20%20%22biological%20engineering%22%20%20%22chemical%20engineering%22%29)

CuLLR implementation for the Physical Sciences Library

CuLLR was next implemented within the Physical Sciences Library website at http://physicalsciences.library.cornell.edu (**Figure 3**). We have completed only the first phase, which was started in January and implemented in March of 2012. (There are three more phases being planned.) You will notice that the Physical Sciences Library implementation of CuLLR is very different from that of the Engineering Library. The intention of CuLLR is to provide data for use by each unit/subject library as they see fit. The patrons of the Physical Sciences Library expressed very different needs from those of the Engineering Library.

In phase one, the Physical Sciences Library developed an "eJournal Finder" and an "eBook Collections" option using data from their CuLLR database. The "eJournal Finder" presents a list of all Physical Sciences electronic journals with a link to the table of contents of the latest issue, if available. The "eBook Collections" option provides a list of all ebook collections covering Physical Sciences subjects. As part of the next phase, both options will have faceting by subjects

within Physical Sciences. We will also be adding a "Specialized Database Finder" and a "Properties Finder".
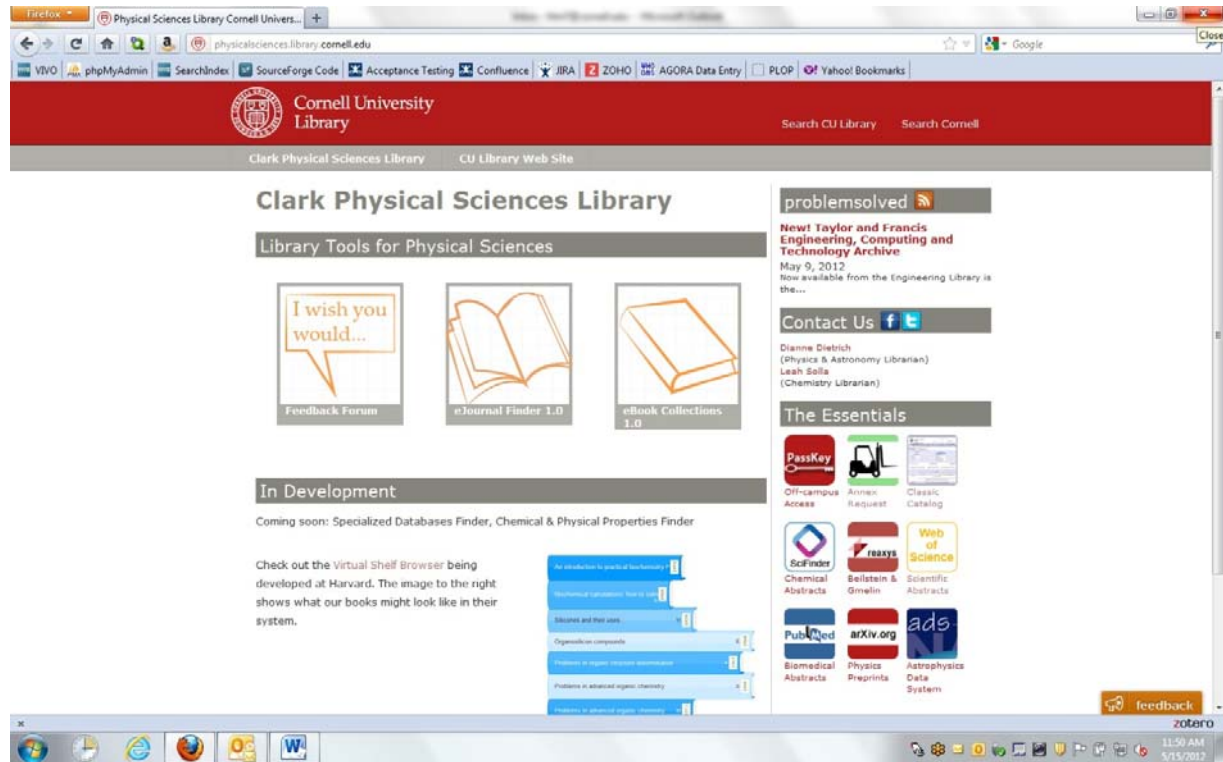
CuLLR implementation for the Entomology Library

The final planned implementation of CuLLR is the Entomology Collection at http://entomology.library.cornell.edu/ (**Figure 4**). The Entomology Library implementation of CuLLR is defined as a collection rather than a website. This is because the Entomology Library has been merged with the Agriculture and Life Sciences Library. As such, Entomology did not need an entire website of its own.

The Entomology Collection is made up entirely of data from CuLLR. The "Search the collection" option searches the Entomology CuLLR database and the "Journals", "Reference Collection", "Theses and Dissertations", and "New Books" options are all based on predefined searches within the same CuLLR database.

Once again you will note that the Entomology Library implementation of CuLLR is different from the CuLLR implementations for the Engineering Library and the Physical Science Library. The subject experts for Entomology determined that, in addition to the usual "Journals" list, an

electronic "Reference Collection" and a list of Cornell "Theses and Dissertations" were
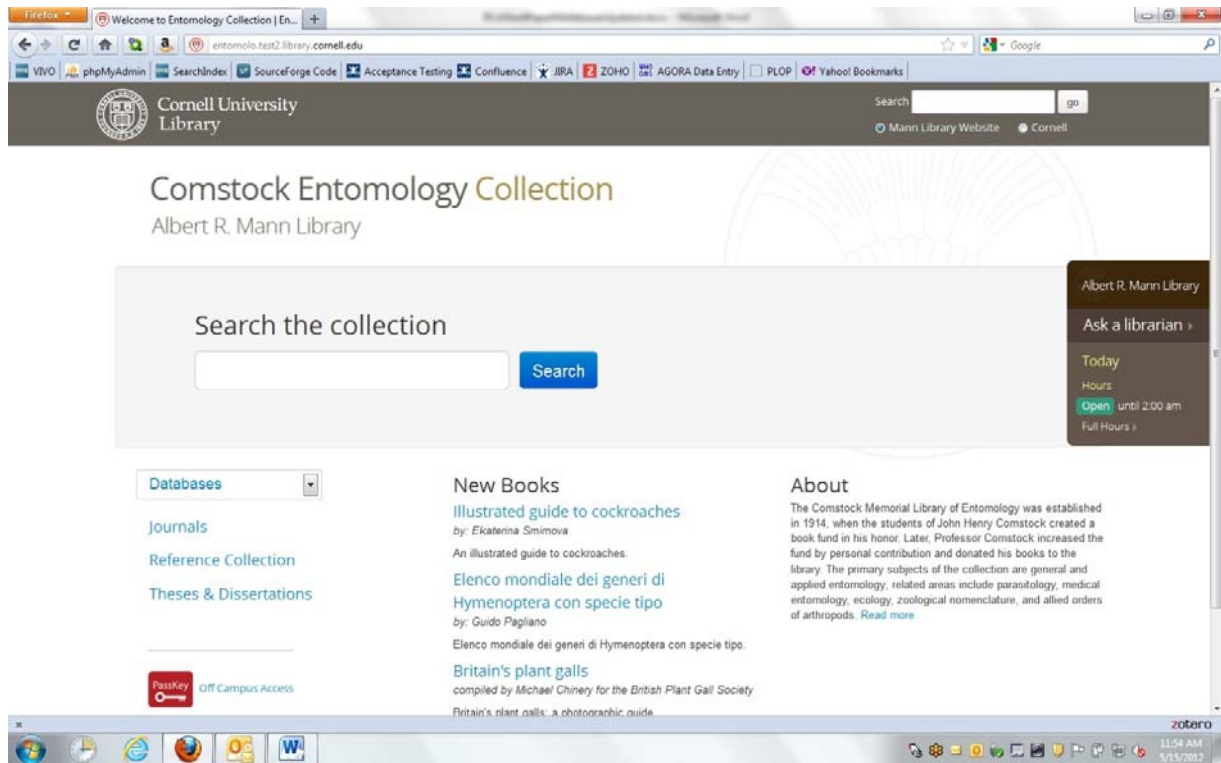important to their patrons.

The CuLLR Process

The CuLLR process is not complex, but it is made up of multiple pieces that may give the appearance of complexity (**Figure 5**).
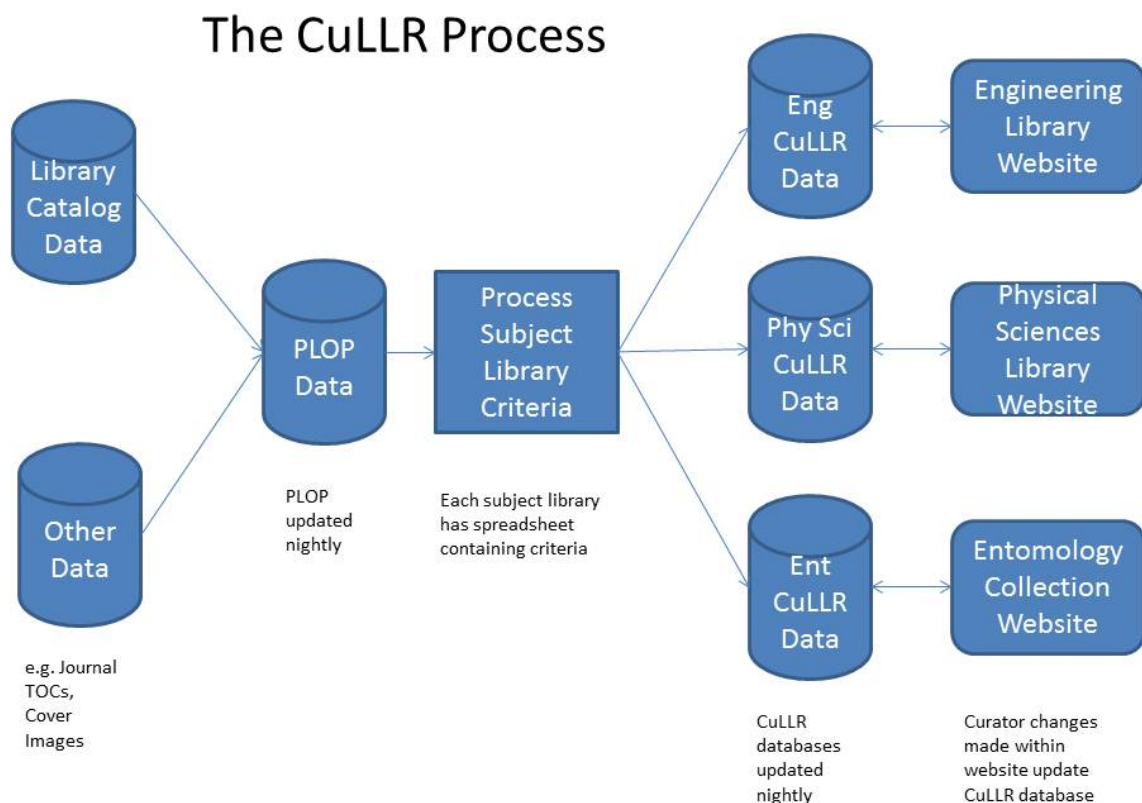

The CuLLR Process

The steps are as follows:

1) The entire library catalog is pulled into a database called PLOP (Persistent Library Object Pages) for use by the CuLLR process. This is done so that additional derived data can be added, and also for future growth as we add non-catalog items into CuLLR. Data elements needed by CuLLR are automatically created at this time, such as "Format" (e.g., book, ebook, journal, ejournal, database). Additional derived data about the resource that does not come from the catalog may be added to PLOP in an automated fashion. For example, we have identified the need to add the table of contents for journals. We obtain the table of contents for a journal by using the journal's ISSN from PLOP to get the table of contents link from JournalTOCs[1] and add it to PLOP. Similar processing will be done

---
[1] CuLLER uses the JournalTOCs API at http://www.journaltocs.ac.uk/api/.

for cover images, chemical/physical properties, and material/compound types. The data in PLOP is the common data available to all unit/subject libraries.

2) The CuLLR extract for the unit/subject library is created using a spreadsheet that contains an identifier (e.g., catalog record identifier, call number range, ebook package identifier) plus the "Discipline" for that item or group of items. The "Discipline" has a controlled vocabulary established by the unit/subject library, and is used for faceted browsing and searching. The controlled vocabulary is a list of terms selected by the unit/subject library to be used as "Disciplines"[2]. Each resource added to CuLLR must have one or more "Disciplines" associated with it. For example, the Engineering Library has set the call number range of GC 1 to GC 1581 to a "Discipline" titled "environmental engineering". HC 79 has two "Discipline" fields identified, "environmental engineering" and "civil engineering". For individual items not included in a call number range or ebook package, the catalog record identifier is used instead. Cornell receives catalog records for ebooks from a variety of sources, including Serials Solutions, for which no call number is provided. To work around this issue, package identifiers are used for ebook packages in order to include ebooks that may not have a call number assigned. For example, the IEEE/IET Electronic Library (IEL) package contains 3,009 ebooks. Of the 3,009 ebooks, Cornell's catalog only has call numbers for only 390. If we did not use package identifiers to pull records, we would miss out on including 2,619 ebooks from this package.

3) Librarians are able to use the curator interface to CuLLR to manually add annotations, free text keywords, internal notes, and search weighting. Annotations are useful in directing patrons to the materials best suited for their research or classes. Search weighting allows specified materials to "float to the top" of the search. For example, if the unit/subject librarian for Engineering would like for the *Journal of Seismology* to appear on the top of the search results when a patron searches for "seismology", he/she would add a high weight to the resource[3]. How this additional data is used is up to the website designers and implementers.

4) The public uses the CuLLR data via a web interface. Cornell University Library uses Drupal for the development and maintenance of our websites. As mentioned above,

---

[2] The Engineering Library has over 15 "Disciplines" defined for use, including: Aerospace Engineering, Applied and Engineering Physics, Biological Engineering, Biomedical Engineering, Chemical Engineering, Civil Engineering, Computer Engineering, Computer Science, Earth and Atmospheric Sciences, Electrical Engineering, Environmental Engineering, Materials Science, Mechanical Engineering, Operations Research, and Systems Engineering.

[3] There is relevance searching without weighting. This is determined by the unit/subject librarian in conjunction with the development team. Normally if the search term is found in the title or author field of a resource it will appear at the top of the search results. Next in the search results would be resources with the search term in the description field. Every field in the database is able to have its own relevance setting. Initially we had settings for every field in the database, plus individual weighting for resources. This proved to be too confusing so we now use resource weighting in a very limited manner.

CuLLR has been implemented for Cornell's Engineering Library at http://engineering.library.cornell.edu/, Physical Sciences Library at http://physicalsciences.library.cornell.edu/, and Entomology Library at http://entomology.library.cornell.edu/.

Challenges

Feedback from the users of CuLLR has been good, as were the results of our usability tests. Only minor changes have been required thus far.

Despite the successful implementation of CuLLR, this project has not been without its difficulties. As mentioned above, the lack of call numbers for ebooks had a major impact. How do you determine the subject area covered by a book without a call number or subject headings? Also, over time cataloging processes have changed resulting in the need for thorough research and understanding of multiple processes. We also discovered that some catalog records were in error. For example, resources that were databases were being identified by CuLLR as serials. The project team worked closely with the cataloging department to fix the catalog records so that the CuLLR process would run correctly.

Future Plans

The first phase of CuLLR as a product is complete and has been implemented within three library websites. CuLLR's future plans include allowing records from PLOP to be added to the unit/subject library's CuLLR database via the curator interface, rather than an Excel spreadsheet. We also want to be able to add resources from other sources besides the library catalog. Cornell University Library is currently in the process of improving the overall discovery of materials and where CuLLR fits into this scenario is not yet known.

The CuLLR project has spawned two related projects. We have initiated a project to investigate how the ebook catalog records could be improved to include call numbers and subject headings. We are hoping to find an on-line source for the missing call numbers and subject headings. We are also working with the Harvard Library Innovation Laboratory on a virtual shelf browser to display CuLLR data graphically and to provide information about how the Cornell community engages with the collection. For example: How much is the item used? What type of user is using the item (e.g., undergraduate, graduate, faculty)? For which classes is the item being put on reserve? What are the patron's comments about the item? Both of these projects will enhance CuLLR significantly.

Conclusion

Despite being initially developed for communities that no longer have a physical library, there is great interest in using CuLLR for Cornell unit libraries that do still exist. This is due to the subject guidance CuLLR provides and the local control allowed by the separation of the subject-related materials from the official library catalog. Cornell supports many remote users, so developing subject access that is available 24/7 is important to all Cornell libraries.

CuLLR is an example of how one library has dealt with providing subject access to its collection. CuLLR is **surprising** us daily in its flexibility and potential for growth. It is **inspiring** in that we continue to identify new ways to use the CuLLR data, such as with the virtual shelf browser described above. CuLLR is **empowering** in that it has given control of the catalog records (in the form of a sub-catalog) to the subject experts which enables our patrons to meet their research needs.

Bibliography

Clark Physical Sciences Library. (2012). Retrieved May 15, 2012, from http://physicalsciences.library.cornell.edu/

Comstock Entomology Collection. (2012). Retrieved May 15, 2012, from http://entomolo.test2.library.cornell.edu/

Engineering Library. (2012). Retrieved May 15, 2012, from http://engineering.library.cornell.edu/

JournalTOCs. (2009). Retrieved May 15, 2012, from http://www.journaltocs.ac.uk/

Searching for Engineering Library Resources. (2012). Retrieved May 15, 2012, from http://engineering.library.cornell.edu/cullrseek/search/*?rqt=Biomedical,%20Biological%20and%20Chemical%20Engineering&rq=culweight%3A%28%22biomedical%20engineering%22%20%20%22biological%20engineering%22%20%20%22chemical%20engineering%22%29