FRBR and Facets Provide Flexible, Work-Centric Access to Items in Library Collections

Kelley McGrath  
University of Oregon Libraries  
1299 University of Oregon  
Eugene, OR 97403-1299  
+1-541-346-8232  
kelleym@uoregon.edu

Bill Kules  
School of Library and Information Science  
The Catholic University of America  
Washington, DC 20064  
+1-202-319-6278  
kules@cua.edu

Chris Fitzpatrick  
Stanford University  
Digital Library Systems and Services  
Stanford, CA 94305  
+1-650-724-5085  
cfitz@stanford.edu

ABSTRACT
This paper explores a technique to improve searcher access to library collections by providing a faceted search interface built on a data model based on the Functional Requirements for Bibliographic Records (FRBR). The prototype provides a Work-centric view of a moving image collection that is integrated with bibliographic and holdings data. Two sets of facets address important user needs: “what do you want?” and “how/where do you want it?” enabling patrons to narrow, broaden and pivot across facet values instead of limiting them to the tree-structured hierarchy common with existing FRBR applications. The data model illustrates how FRBR is being adapted and applied beyond the traditional library catalog.

Categories and Subject Descriptors
H.3.7 [Information Search and Retrieval]: Digital Libraries

General Terms
Design, Standardization.

Keywords
FRBR, faceted search, user interface design, moving images, movies, film, video

1. INTRODUCTION
Patrons have trouble browsing for videos in libraries for many reasons. Academic libraries often have video collections that are kept behind service desks so patrons must use the catalog. The catalog usually provides little guidance as to the type and extent of a library’s collection or what search terms would be most effective. Even when libraries have browsable collections, these physical collections are generally only arranged by a single characteristic, usually genre. A given DVD can be put in only one category so a foreign comedy must be placed either with the foreign films or with the comedies.

Library users often have information needs that combine attributes of the various levels of the Functional Requirements for Bibliographic Records (FRBR) model. For example, a patron might want to see all the Blu-ray discs (Manifestation) at her local library (Item location) and then decide to limit her search to comedies (Work). A language teacher might want movies produced in France with the original French soundtrack while a social studies teacher might wish to limit the same set of French movies to those with English subtitles. A 1950s science fiction film fan might want to browse those movies and then limit to DVDs with 2011 publication dates to see what is newly available.

Typical library catalogs mix these different FRBR levels in a single, publication-focused record, hindering users. FRBR is designed to ease this problem, but existing FRBR implementations employ a hierarchical tree structure in which users must first select a specific work, then a version, a publication and only finally a particular item in a particular location.

The Online Audiovisual Cataloger (OLAC) prototype discovery interface (see Figure 1; available at http://blazing-sunset-24.herokuapp.com/) combines a Work-centric, FRBR-inspired data model and faceted navigation to support flexible browsing and navigation of the FRBR levels. Our modified implementation of the FRBR model collapses FRBR to two levels for presentation and provides facets that answer two sets of questions (1) “What do you want to watch?” (A Russian movie? An action flick?); and (2) “How do you want to watch it?” (on Blu-ray? With English subtitles?) plus “Where is it located?” Our approach allows us to highlight and present separately information about the original movies and programs that users are most interested in. Because these two levels interact dynamically, users have a wide variety of potential paths to narrow their search to desirable and suitable materials. The use of facets exposes the collection so that it is easy for users to understand the scope of the collection and to explore it without any prior knowledge of the collection or the indexing vocabulary.

2. FACETED SEARCH INTERFACE
Facets provide patrons with the power to explore collections and refine result sets through clickable categories that allow users to refine and narrow their search without re-entering a query. Facets are commonly seen on the sidebar of ecommerce websites. Facets for attributes like language, format, and LC Subject Headings are increasingly appearing in library catalogs. Facets can be used to improve the browsing experience by giving users an idea of the size and scope of a library’s collection and supporting them in exploration when they don’t have a specific title in mind.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL’11, June 13–17, 2011, Ottawa, Ontario, Canada.
Copyright 2011 ACM 978-1-4503-0744-4/11/06...$10.00.
Facets also support multiple values for a given attribute and the ability to search across multiple characteristics simultaneously. Facets do require controlled values for characteristics of interest that are often not present in current library data. However, the library world is moving in the direction of recording more machine-actionable data and we have demonstrated that some existing free-text data can be mapped to controlled values [3].

In our prototype, we display two types of facets to the end user. The top level is labeled Movies or Programs. The second level, labeled Versions, provides a set of fulfillment options, such as library, format and soundtrack language. To help make the two sets of facets visually distinct, we placed the facets that answer the question “What movie or program do you want?” at the top and “How do you want to get that movie or program?” (e.g., on DVD, with English subtitles, from the most convenient library) on the right side (Figure 1). Since some of the general types of information are repeated in both sets of facets, we distinguish them with descriptive labels, e.g., original language (Movies) vs. spoken and subtitle/captions labels (Versions). Preliminary feedback has suggested additional ways to distinguish labels in the Movies section, such as changing “Date” to “Original Date.”

Figure 2 shows a portion of the search results, with the Movie (Work) record at the top and Version options, such as format, languages and location, below.

These two levels of interconnected facets are technically challenging to implement with the open source tools we used, but provide much more flexible browsing options than more typical hierarchical FRBR implementations in which users must first pick a movie, then a version, and only then see locations where that version of the movie on Blu-ray is available. In our prototype users can easily search just within a preferred library or a preferred format or, alternatively, cast a broad net. Our interface also offers more powerful searching than typical library catalogs where records are focused on publications and information about the original movie is not consistently recorded in a retrievable form.

![Figure 1. Movie (Work) facets are at the top with documentary selected. Movie facets answer the question “What do you want?” Visually-distinct Version limiter facets are on the right. Version facets answer the question “How do you want it?”](image1)

Figure 2. Hit list focuses on the movie (Work) with fulfillment options for publications at various library locations below. Version facets on the right (not shown) offer options to narrow the range of versions, publications and locations presented.

3. THE FRBR DATA MODEL

The Functional Requirements for Bibliographic Records (FRBR) model [7] is a conceptual, entity-relationship model developed in the late 1990s for library catalog data. The model is intended to more effectively meet library users' needs and better represent relationships between different types of bibliographic information, but there is a recognized need for theoretical and practical development of the model [2, 8]. The FRBR model includes four levels of bibliographic entities:
**Works** are the most abstract level and contain information about movies or programs as intellectual or artistic creations. This is similar to what is described in an Internet Movie Database (IMDb) record. Spalding proposes an informal “cocktail party” test where two books belong to the same work if two people casually talking about them think they are discussing the same thing even if they read the book in different languages or editions [6].

*Expressions* are intellectual *versions* that vary in content. The theatrical release of a movie vs. a director’s cut or unrated version is an example of an expression-level variation.

*Manifestations* are publications that vary at the level of the packaging of the content. Changes in format (DVD vs. Blu-ray) or publisher are examples of manifestation-level differences. Library cataloging records are focused on the manifestation level.

An *Item* is a particular physical manifestation. It is a unique physical thing with a certain barcode and coffee stain on the cover.

### 4. MODIFYING THE FRBR DATA MODEL

We have used this division of the bibliographic universe to present information about movies and programs to end users in a more effective form and as a pivot point to connect information about the FRBR entities to holding libraries. Current library records focus on publications (Manifestations). Although these publication-focused records include information about the original Work (or movie), that information is often scattered in a given record and hard for users to search for or identify. Our Work-centric prototype privileges Work information and makes it easy for users to find both specific movies and classes of movies (e.g., recent releases or films from Germany).

In order to optimize the storage and display of moving image data, we modified the FRBR model (Figure 3a) by combining the top three bibliographic entities into two levels (Figure 3b).

1) We combined information about the abstract Work with the Primary Expression or Version (usually the original public release version)
2) We combined information about the particular Expression (version) with the Manifestation (publication).
3) We added the Library entity to record library holdings by linking Libraries with individual Items.

Expressions are thought to be the most difficult of the FRBR bibliographic entities to identify and model [4]. We chose to combine information about the Primary Expression with the Work because the Primary Expression describes important and unchanging facts about the history of the Work that we would want to note in conjunction with any future Expression or version. This eliminates repetition of information about the Primary Expression in multiple Expression (version) records [5]. The abstract nature of the Work in the FRBR model means that some Primary Expression information, such as original language, is not considered an attribute of the Work entity. In addition, *RDA: Resource, Description, and Access* [1], the new library cataloging rules based on FRBR, considers some roles, such as actors and costume designers, to belong to the Expression level where a Work is realized. We can minimize redundancy by recording these functions at the top Work/Primary Expression level.

![Figure 3. Entity-relationship diagrams for (a) the FRBR model and (b) the adapted model used for the prototype. In the adapted model, the Movie table contains information about the basic Work and its primary Expression (Version). The Version/Publication (Expression/Manifestation) table records information about publications and the content versions they contain. The combination of Version/Publication and Item tables are labeled Versions for the end user.](image)

There are some limitations to modeling Expressions (versions) and Manifestations (publications) at the same level and we will be reevaluating this approach. Because some Manifestations (publications) can contain more than one Work and those Works can have different Expression (version) values, there is not a one-to-one relationship in all cases. An example in our sample data is a DVD that contains both the English and Spanish language Dracula movies from 1931. This requires two Expression/Manifestation records with redundant Manifestation data, as shown in Table 2.

<table>
<thead>
<tr>
<th>Work record 1</th>
<th>Expression/Manifestation record 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dracula (1931)</td>
<td>1 VHS videocassette (1985) [Manifestation]</td>
</tr>
<tr>
<td>Tod Browning</td>
<td>OCLC#: 13754402 [Manifestation]</td>
</tr>
<tr>
<td>English</td>
<td>Audio: English [Expression/Version]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work record 1</th>
<th>Expression 1</th>
<th>Manifestation 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dracula (1931)</td>
<td>Audio: English</td>
<td>1 DVD (1999)</td>
</tr>
<tr>
<td>Tod Browning</td>
<td>Subtitles: English or French</td>
<td>ISBN: 0783227450</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work record 2</th>
<th>Expression 2</th>
<th>Manifestation 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dracula (1931)</td>
<td>Audio: Spanish</td>
<td>1 DVD (1999)</td>
</tr>
<tr>
<td>George Medford</td>
<td>Subtitles: English or French</td>
<td>ISBN: 0783227450</td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work record 1</th>
<th>Expression 1</th>
<th>Manifestation 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dracula (1931)</td>
<td>Audio: English</td>
<td>1 DVD (1999)</td>
</tr>
<tr>
<td>Tod Browning</td>
<td>Subtitles: English or French</td>
<td>ISBN: 0783227450</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. SYSTEM IMPLEMENTATION

The OLAC prototype was developed using the free open source tools Ruby on Rails, Solr, and the Blacklight and Hydra plugins and a small sample of records featuring a few selected representative fields. Code for the prototype is available at https://github.com/cfitz/olac. Because we were able to build on functionality already in Solr and in Blacklight, a plugin that supports library requirements for discovery interfaces, much of the development of the prototype was straightforward. The most challenging issue, and one not faced by most faceted interfaces, is derived from the way that the separate Movies and Versions facets need to work interdependently on the results for Movies and Versions. We had to build facets for both the Movies table and the consolidated Versions table, which have a many-to-many relationship. Movie and Version results needed to be aggregated and reflect the correct facet count for their respective sets of records.

In order to make this work, we added a text field to each of the Solr documents that contains all the possible permutations of facets for that Movie or Version. When the results are being returned, an additional search query on this text field is used to filter the documents to display only the relevant Movies and Versions. We then group the results by Movies and display each Movie with its related Versions. Since this approach requires multiple Solr requests for each request the user makes, it slightly reduces query response time.

For example, the search “([format_facet][]=DVD& [genrename_facet][]=Fiction& [library_facet][]=D)” is faceting on (Movie) Genre = Fiction, (Version) Library = D, and (Version) Format = DVD, but it’s also doing a filter on a text field looking for the values genre_Fiction, library_D, and format_DVD. Without this filter, unexpected results would be returned. For example, a Movie would be returned if it had a DVD in library E and a VHS in library D. The filter field forces the facets to apply to the same Version so only DVDs in library D are returned.

6. DISCUSSION AND CONCLUSION

Tightly coupling a faceted display with a FRBR-based data model benefits both patrons and libraries. For patrons, it provides a Work-centric view of collections that is integrated with bibliographic and holdings data. Patrons can start their search at any point in the FRBR hierarchy, from Item (location) to Work (genre, date), and easily transition between search and browse strategies, using facets to broaden or narrow their results and pivoting on facet values. The facets address two important user needs: “what do you want?” and “how/where do you want it?” Libraries and catalogers also benefit. Libraries currently create records based on publications. Many moving image materials are released in numerous published versions. For major motion pictures, extensive information about the original film is repeated in each publication-based record. This leads to redundancy of effort and inconsistency of data. Creating a single Work record for the original helps eliminate data redundancy and improve accuracy and completeness.

This project, although grounded in the domain of moving images, contributes more broadly to ongoing FRBR development. The modified FRBR data model and the working prototype contribute to a recognized need to “verify and validate the FRBR model against real data and in different communities to make sure the model is valid and applicable” [8]. The faceted interface demonstrates a FRBR application not limited to the tree hierarchy used by most current FRBR applications.

There are a number of open challenges. Additional bibliographic data must be presented in a manner that is comprehensible, meaningful and effective for the user. The facet behavior also needs to be refined, e.g. to permit searchers to specify multiple acceptable locations or formats. Additional research is needed to refine the data model to effectively accommodate complex, real-world relationships, like a publication that contains multiple Works (movies) and Expressions (content versions). The prototype’s solution for the multiple levels of facets needs to be revised to efficiently scale up to larger data sets. Commercial tools such as Endeca’s have begun to address this, but this is still a technical challenge in the open source tools we used.

7. ACKNOWLEDGMENTS

We thank OLAC for funding the development of the prototype end-user interface. We would also like to acknowledge the numerous contributors to the broader project, many of whom are listed at http://blazing-sunset-24.herokuapp.com/page/credits. Finally, we are grateful to Andruide Kern for his suggestions, which have greatly improved the paper.

8. REFERENCES


