DEVELOPING AN FRBR-BASED SYSTEM TO SUPPORT USER TASKS

Dr. Yin Zhang
Dr. Athena Salaba
School of Library & Information Science, Kent State University
NOR-ASIST, April 4, 2011

PRESENTATION OUTLINE

1. FRBR Overview
   • FRBR FAMILY MODELS
   • STANDARDS AND PRACTICE
   • APPLICATION
   • IMPLEMENTATION
2. RESEARCH
   • KSU FRBRization Research
   • KSU User Research

Delphi Study Findings
1. FRBR Overview

FRBR Family Models Overview
FRBR FAMILY MODELS OVERVIEW

FRBR AND RELATED STANDARDS

- FRBR Developments:
  - FRBR Review Group (Identifiers, Namespace, Aggregates, Harmonization of Family)
  - FRBRoo – harmonization with ICOM-CIDOC Conceptual Model (ICOM-CIDOC CRM) International Cataloguing Principles

- Other Standards:
  - International Standard Bibliographic Description (ISBD)
  - Resource Description and Access (RDA)
  - Dublin Core Metadata Initiative (DCMI)
  - Machine Readable Cataloging (MARC)
  - RDA/ONIX Framework
  - Resource Description Framework (RDF) and Linked Data

- Delphi Study Findings:
  - The panel raised the largest number of issues in this area
  - FRBR model verification and validation using real data, applied in different communities, stands out as the most critical issue in this area
FRBR APPLICATION

- Settings:
  - Libraries
  - Consortia
  - Digital Libraries, Internet Archives, Institutional Repositories
  - Museums

- Types of resources:
  - Works of art, cultural objects
  - Classical texts
  - Fiction
  - Hand-press materials
  - National literature
  - Live performing arts
  - Moving images
  - Music
  - Continuing resources, aggregates

- FRBR Delphi Study:
  - lack of FRBR application guidelines and examples
  - confusing and challenging points for practical applications:
    - “...understanding separation of work, expression and manifestation;”
    - “...how to know what elements/relationships are significant to different types of materials or different collections”

FRBR IMPLEMENTATION

Overall, current FRBR implementations can be divided into two broad categories:

1. implementing FRBR in online library catalogs (either general library collections or specific types of library collections), and
2. implementing FRBR in nontraditional library settings, such as special collections, museums, digital collections, archives, and Internet resources.
FRBR IMPLEMENTATION: SYSTEMS DEVELOPMENT

- **Full-scale working systems**: Support regular services or functions
  - OCLC Worldcat.org
  - UCLA Library – Film and Television Archive OPAC

- **Prototypes or experimental systems**: developed to simply explore FRBR implementations; they do not support real live library services
  - Libraries Australia’s demonstration system illustrates prototype service with stale data rather than its production service.

- **Supporting software tools, algorithms, and utilities**
  - OCLC FRBR Work Set Algorithm, which converts MARC bibliographic records to conform to FRBR at the work entity level.
  - The Library of Congress has developed the FRBR Display Tool, which allows libraries to display their resources by clustering bibliographic records according to the FRBR model.

- **Delphi Study. Need to**:
  - develop and test tools/software that will facilitate the FRBRization processes;
  - explore, develop, and test various means of FRBR implementation;
  - to address the FRBRization of existing data from a variety of differing standards and practices;
  - to explore, design, and develop effective user interfaces in general, with result displays, in particular, based on the FRBR model.

---

2. FRBR RESEARCH
FRBR Research

FRBR-related efforts in the form of theoretical discussion, exploration, and development as well as evaluation and implementation

Delphi study (Zhang & Salaba, 2009)

Four of the five FRBR research issues are related to user research:

- **Issue 1**: User studies on FRBR-based systems to ensure that implementations benefit end-users;
- **Issue 2**: User research on FRBR-based displays;
- **Issue 3**: Examination of end-user tasks with empirical research;
- **Issue 4**: Automatic processing of databases or full-text electronic resources to facilitate FRBR implementation; and
- **Issue 5**: Development of semiotic frameworks and research to ensure effective communication between users and FRBR systems.

Kent State IMLS FRBR Research Project

FRBRization
**THE COLLECTION**

- For this FRBRization experiment, the collection used was extracted from WorldCat for LC records at the end of December 2007.

- This collection includes:
  - **13,624,251** bibliographic records
  - **7,283,635** authority records

**WORK-LEVEL FRBRIZATION**

- The FRBRization procedures for identifying and grouping works for our project are based on the OCLC FRBR workset algorithm:

- Based on several rounds of evaluations, the KSU FRBR team has added to and refined some of the algorithm’s steps.

- The **13,624,251** LC records were FRBRized into **11,579,660** work-sets:
  - **1,215,654** are multiple-record work-sets (10.50%)
  - **10,364,006** are single-record work-sets (89.50%)
**Expression and Manifestation Level FRBRization – Sample**

- **The sample:**
  - From the multiple-record work-sets, we selected those that contained at least 10 records to explore FRBRization at expression and manifestation level.
  - **Sample size:**
    - 12,579 work-sets (1% of the multiple-record work-sets)
    - 273,866 records (2% of the entire LC collection bib records)

**Expression and Manifestation Level FRBRization - Algorithm and Procedure**

**Previous work and basis:**
- Mapping and importance ratings of attributes and MARC fields on user tasks:
  - Functional Analysis of the MARC 21 Bibliographic and Holdings Formats by Network Development and MARC Standards Office, Library of Congress
  - IFLA FRBR Report, Tables 6.1-6.4
  - KSU FRBR project evaluation based on cataloging standards & practice
- **MARC records analysis (data availability and data format)**
- **FRBR work-set analysis**
### Expressions in a Workset Distribution

<table>
<thead>
<tr>
<th>Expressions-per-workset</th>
<th>Number of worksets</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,794</td>
<td>30.2%</td>
<td>30.2%</td>
</tr>
<tr>
<td>2</td>
<td>2,919</td>
<td>23.2%</td>
<td>53.4%</td>
</tr>
<tr>
<td>3</td>
<td>1,480</td>
<td>11.8%</td>
<td>65.1%</td>
</tr>
<tr>
<td>4</td>
<td>1,227</td>
<td>9.8%</td>
<td>74.9%</td>
</tr>
<tr>
<td>5</td>
<td>927</td>
<td>7.4%</td>
<td>82.3%</td>
</tr>
<tr>
<td>6</td>
<td>615</td>
<td>4.8%</td>
<td>87.1%</td>
</tr>
<tr>
<td>7</td>
<td>423</td>
<td>3.4%</td>
<td>90.5%</td>
</tr>
<tr>
<td>8</td>
<td>302</td>
<td>2.4%</td>
<td>92.8%</td>
</tr>
<tr>
<td>9</td>
<td>226</td>
<td>1.8%</td>
<td>94.6%</td>
</tr>
<tr>
<td>10</td>
<td>154</td>
<td>1.2%</td>
<td>95.8%</td>
</tr>
<tr>
<td>11</td>
<td>103</td>
<td>0.8%</td>
<td>96.6%</td>
</tr>
<tr>
<td>12</td>
<td>74</td>
<td>0.6%</td>
<td>97.3%</td>
</tr>
<tr>
<td>13</td>
<td>56</td>
<td>0.4%</td>
<td>97.7%</td>
</tr>
<tr>
<td>14</td>
<td>48</td>
<td>0.4%</td>
<td>98.1%</td>
</tr>
<tr>
<td>15</td>
<td>35</td>
<td>0.3%</td>
<td>98.4%</td>
</tr>
<tr>
<td>16</td>
<td>37</td>
<td>0.3%</td>
<td>98.7%</td>
</tr>
<tr>
<td>17</td>
<td>31</td>
<td>0.2%</td>
<td>98.9%</td>
</tr>
<tr>
<td>18</td>
<td>21</td>
<td>0.2%</td>
<td>99.1%</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>0.2%</td>
<td>99.3%</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
<td>0.1%</td>
<td>99.4%</td>
</tr>
<tr>
<td>&gt;=21</td>
<td>77</td>
<td>0.6%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>12,579</td>
<td></td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Overall distribution in the sample

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>276</td>
<td>3.6</td>
</tr>
</tbody>
</table>

### Manifestations per Expression Distribution

<table>
<thead>
<tr>
<th>Number of Ms in an E</th>
<th>Frequency</th>
<th>%</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20,495</td>
<td>44.8%</td>
<td>44.8%</td>
</tr>
<tr>
<td>2</td>
<td>5,915</td>
<td>12.9%</td>
<td>57.7%</td>
</tr>
<tr>
<td>3</td>
<td>2,092</td>
<td>5.3%</td>
<td>63.0%</td>
</tr>
<tr>
<td>4</td>
<td>1,655</td>
<td>4.1%</td>
<td>67.1%</td>
</tr>
<tr>
<td>5</td>
<td>1,327</td>
<td>2.8%</td>
<td>70.0%</td>
</tr>
<tr>
<td>6</td>
<td>1,111</td>
<td>2.4%</td>
<td>72.4%</td>
</tr>
<tr>
<td>7</td>
<td>962</td>
<td>2.1%</td>
<td>74.5%</td>
</tr>
<tr>
<td>8</td>
<td>953</td>
<td>2.2%</td>
<td>76.7%</td>
</tr>
<tr>
<td>9</td>
<td>1,090</td>
<td>2.4%</td>
<td>80.1%</td>
</tr>
<tr>
<td>10</td>
<td>1,872</td>
<td>4.1%</td>
<td>84.2%</td>
</tr>
<tr>
<td>11</td>
<td>1,598</td>
<td>3.9%</td>
<td>88.1%</td>
</tr>
<tr>
<td>12</td>
<td>1,007</td>
<td>2.2%</td>
<td>90.3%</td>
</tr>
<tr>
<td>13</td>
<td>717</td>
<td>1.6%</td>
<td>91.9%</td>
</tr>
<tr>
<td>14</td>
<td>566</td>
<td>1.2%</td>
<td>93.1%</td>
</tr>
<tr>
<td>15</td>
<td>408</td>
<td>1.0%</td>
<td>94.1%</td>
</tr>
<tr>
<td>16</td>
<td>356</td>
<td>0.8%</td>
<td>94.9%</td>
</tr>
<tr>
<td>17</td>
<td>307</td>
<td>0.7%</td>
<td>95.6%</td>
</tr>
<tr>
<td>18</td>
<td>246</td>
<td>0.5%</td>
<td>96.1%</td>
</tr>
<tr>
<td>19</td>
<td>226</td>
<td>0.5%</td>
<td>96.6%</td>
</tr>
<tr>
<td>&gt;=19</td>
<td>2,261</td>
<td>5.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Total</td>
<td>45,775</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Overall distribution in the sample

<table>
<thead>
<tr>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,395</td>
<td>6.0</td>
</tr>
</tbody>
</table>
FRBRIZATION ISSUES AND CHALLENGES

○ Algorithm related
  • Processing titles
    ○ Trailing “English”, “Tragedy of” and “Comedy of” in titles
  • Other titles not used in mappings
  • Partial matching
  • Source of data
    ○ Coded data vs. free text

○ Related to legacy data, current cataloging practices and standards
  • Typos and coding problems
  • Missing important elements
    ○ 240s in AT pattern
    ○ 130s in bibliographic records
    ○ 041s for translated works
  • Current cataloging practice
    ○ Musical works – Genre headings
    ○ Some TV and radio programs have been controlled and collocated with the use of uniform titles while others have not
  • The algorithm cannot recognize duplicate records, and therefore they are considered separate manifestations.

KENT STATE IMLS FRBR RESEARCH PROJECT

FRBR USER RESEARCH
GAPS IN FRBR USER RESEARCH

- FRBR user research has been the least addressed facet of FRBR research and development.
  - There has been a lack of user validations of the model
  - Very few FRBR implementation projects conducted or reported user studies
  - Very few evaluative comparisons of existing FRBR prototype systems.

- To a great extent, the current FRBR application and implementation efforts have reflected the views of the designers and researchers with user considerations rather than the user views.

Recent user research:

- Pisanski & Žumer (2010a, 2010b): The research focused on non-librarians’ mental models of the bibliographic universe and compared them with the FRBR model.
- FRBR user research for the Kent State IMLS-funded FRBR project →

USER STUDY OF EXISTING FRBR PROTOTYPES

**Systems included:**
- WorldCat.org ([http://www.worldcat.org/](http://www.worldcat.org/))

**Methods used:**
- Screen captures
- Eye-tracking
- Think-aloud protocols
- Survey interviews
- Focus groups
  - academic library users (hereinafter AL)
  - public library users (hereinafter PL)
USER TASKS

Search tasks are designed around the following general user task categories as defined by FRBR:
   a) to find materials corresponding to the user’s stated search criteria;
   b) to identify an entity;
   c) to select an entity appropriate to the user’s needs; and
   d) to acquire or obtain access to the described entity (IFLA, 1998).

In addition, participants were asked to search their regular online catalogs and the FRBR systems for their own tasks.

FINDINGS: FEATURES USERS LIKE

System features users like based on think-aloud protocols and post session interviews:

<table>
<thead>
<tr>
<th>Feature</th>
<th># of users</th>
<th>Fiction Finder</th>
<th>WorldCat</th>
<th>Libraries Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface (user friendly, easy to use, appealing)</td>
<td>37</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Refining options</td>
<td>35</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Additional information (ratings, reviews, price,</td>
<td>21</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>audience, summaries, format icons, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sorting options (by date, format, language, etc.)</td>
<td>21</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability/holdings information</td>
<td>20</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Single search box</td>
<td>19</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Terminology and labeling (clear, easy to understand)</td>
<td>14</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Organization and presentation of information (results page, description)</td>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>FRBR display (by format, language, editions)</td>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Searching in specified fields</td>
<td>8</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced search option</td>
<td>6</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling or search suggestion “Did you mean...”</td>
<td>5</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Subject clouds option</td>
<td>5</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>External links (to Amazon, etc.)</td>
<td>2</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Keyword search option</td>
<td>2</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Browse option</td>
<td>1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Return query in search box</td>
<td>1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Subject search option</td>
<td>1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
## Findings: System Improvements

Suggestions for system improvements based on think-aloud and post session interviews:

<table>
<thead>
<tr>
<th>Suggested Features</th>
<th># of Users</th>
<th>Fiction Finder</th>
<th>World Cat</th>
<th>Libraries Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add more links to related works (collocation via “more like this,” hyperlinked subjects, hyperlinked authors, series links, linked keywords, linked cited works, linking other formats from a desired work)</td>
<td>39</td>
<td>12</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Grouping and displaying information in results page (group into categories; arrange by language, date, author in alphabetical order, title in alphabetical order if an author search, multiple criteria)</td>
<td>34</td>
<td>7</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>Add sorting options (by multiple criteria, age groups, date, language, relevance, etc.)</td>
<td>34</td>
<td>11</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Interface display (less clutter, larger font, color to differentiate sorting and refining options, aesthetically pleasing, use column display, use bold letters for library information, highlight searched terms in results page, highlight resource summaries)</td>
<td>25</td>
<td>7</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Labeling and terminology used (e.g., “find edition” should change to “libraries,” clarify “language” choices for site interface so they are not confused as to material/resource language, change terms in advanced search options, etc.)</td>
<td>24</td>
<td>13</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

## Designing a FRBR Prototype Interface

- Participatory design study
- FRBR displays - WEM
  - Work-level display of results
    - Author search
    - Title search
    - Subject search
  - Expression level display of results
    - Form as primary, language as secondary
    - Language as primary, form as secondary
  - Manifestation display
- Structured interview survey
- Participants: 25 academic library users
MAJOR FINDINGS AND REVISED PROTOTYPE

- 72% (18 out of 25) of the participants had an overall positive impression of the prototype catalog in displaying results.
- 88% (22 out of 25) of the participants felt that a catalog designed this way would be more helpful to users.
- 84% (21 out of 25) of the participants found the prototype catalog easy to navigate through.
- Based on user input on various WEM displays and WEM navigation, a revised FRBR prototype has been developed for user testing (in process):
  - Comparative approach: FRBR prototype vs. Koha
  - Interactive WEM tasks

ACKNOWLEDGEMENTS

For their help during the project, we would like to thank the following people and groups:

- Thom Hickey and Jenny Toves, for their clarification of the FRBRization process and algorithm;
- OCLC, for providing the LC records that were used;
- The project advisory team,
  - Dr. Barbara Tillett, Library of Congress
  - Dr. Edward T. O’Neill, OCLC
  - Dr. Diane Vizine-Goetz, OCLC
  - Dr. Thomas B. Hickey, OCLC
- Research team:
  - Lei Xie (Programmer)
  - Jake Schaub (Graduate Assistant)
  - Vicki Ceci (Graduate Assistant)
  - Theda Schwing (Graduate Assistant)
  - Matt Shreffler (Graduate Assistant)
  - Melissa Higey (Graduate Assistant)
  - Alison McCarty (Graduate Assistant)
  - Joanna Samuelson (Graduate Assistant)
  - Grace Brinker (Graduate Assistant)
MORE INFORMATION

Project website:
http://frbr.slis.kent.edu

Contact information:
Yin Zhang (yzhang4@kent.edu)
Athena Salaba (asalaba@kent.edu)

Thanks!!

Questions and Suggestions?