

My Speaking Notes for “What We Talk About When We Talk About FRBR,” Code4Lib 2009

William Denton <wtd@pobox.com>

25 February 2009

(Jodi Schneider and I gave a talk, “What We Talk About When We Talk About FRBR,” at Code4Lib (<http://code4lib.org/conference/2009>) on 25 February 2009. These are my speaking notes and don’t cover Jodi’s sections. The slides are available at http://code4lib.org/files/frbr_code4lib09.pdf and a video is expected to be available by April 2009.)

0. Jodi: Introduction

1. Weak FRBRization

SLIDE: Weak

This is what most people talk about when they talk about FRBR. This is what most vendors talk about when they talk about FRBR. Kind of meager. An entity-relationship diagram with two entities not actually related.

SLIDE: Alexandre Dumas

The entities are Work and Manifestation. THE THREE MUSKETEERS by Alexandre Dumas is a FRBR Work—not any particular translation or printing of it, just the abstract idea. Generally, a novel, a poem, a song, a short story, a program, will be a Work, a “distinct intellectual or artistic creation.”

SLIDE: Alexander Dumas manifestations

Think of a Manifestation as an entire print run of a book. For THE THREE MUSKETEERS there's a Penguin Classics edition, there's an Oxford Classics edition, there was a first printing in France in the 1840s, and those are all manifestations. A manifestation is "the physical embodiment ... of a work."

What a lot of people talk about when they talk about FRBR is a work-set grouping, where when you know one manifestation of a work, you can find other manifestations of it.

SLIDE: xISBN

You've probably heard of OCLC's xISBN service. If you give it an ISBN, it will give you back a list of other ISBNs of the same book. An ISBN represents a manifestation, so when you know one manifestation you can find out the others that OCLC thinks belong to the same work.

xISBN has a nice API, it's easy to format queries, and if you want it'll give you back lots of extra information like language and publication year. It's free for limited use, then they charge.

I gave it the ISBN for the hardcover edition of *Sixty Days and Counting* by Kim Stanley Robinson and it gave me back a set of five ISBNs. It's XML and simple and easy to parse.

SLIDE: OCLC's FRBR Work-Set Algorithm

OCLC doesn't make available the code that does this, but a few years ago they published the algorithm and it's available online. It's under the Apache License, so you can base your own implementation on it and use it in open source or proprietary code.

SLIDE: LibraryThing ... is people

LibraryThing also groups Manifestations together, except that instead of working by algorithm, the users do it by hand. There's an option on every book that allows you to combine it with others into a work grouping, or

SLIDE: LibraryThing separations

separate it if you think it was wrongly grouped. Tim Spalding says that about 6,000 such actions are made each day.

SLIDE: thingISBN

LibraryThing has a web service called thingISBN, which is equivalent to

xISBN.

I put the ISBN for *Sixty Days and Counting* into it and got back these four ISBNs. xISBN had five. The fifth one is an e-book that no-one at LibraryThing has entered. Sometimes one is better than the other but they're both really useful services.

This is where things are mostly at right now when people talk about FRBR: work-set groupings. If you're getting a demo from a vendor or they're talking about how their new system does FRBR, they're probably talking about this. But it's not enough.

SLIDE: jrochkind++

As Jonathan Rochkind put it, "There is much more to the FRBR model than work set grouping ... FRBR is a complete data model that is a new way of looking at our data, not just taking existing records and identifying work relationships."

SLIDE: Weak FRBRization

The problem here is that the manifestations are connected loosely to each other, but there's no way of knowing exactly the work in question. I may have an ISBN, and be able to find out a bunch of related ISBNs, but that doesn't tell me what the Work is, what the book or poem or song that they're all physically embodying, what that really is.

SLIDE: Open Library

Just last week there was an announcement from the Open Library about how they're making progress on this.

This is the page for *Murder on the Orient Express* by Agatha Christie. It shows the different editions that the Open Library knows about, with covers and information and all. At the top it has the title of the work, and the creator, and that it's a novel, and what the original language is. And it says it's work number OL10001W.

This is a Work identifier, and there's been no such thing as a work identifier up to now. Manifestations have ID numbers like ISBNs that aren't hard to find, but there's been no way of uniquely identifying works that we can all agree on. The Open Library's changing that. Now we can start to have work identifiers and we can agree on what work we're talking about in a shared and open way.

SLIDE: Weak

So we're mostly here, at weak FRBRization, with unrelated entities ...

SLIDE: Less Weak

...but the Open Library's getting into something more like this, two entities, Work and Manifestation, with an actual relationship.

This is less weak FRBRization, but what we want is ...

SLIDE: Strong

2. Jodi: Strong FRBRization

3. Complete FRBRization

SLIDE: Almost complete

That's strong FRBRization but we're not at complete FRBRization yet. First we need to handle the two other sets of entities.

First, the Group 2 entities: Person and Corporate Body and Family.

Group 1 and Group 2 fit together with relationships. For example a Work was created by a Person, an Expression was translated by a Person, a Manifestation was published by a Corporate Body, an Item is owned by a Family.

The Group 3 entities are Concept, Event, Object, Place and there are relationships between them and Group 1. They're meant to be the subjects of Works, for example a history of one or more events or a map of a particular place or a poem about the concept love.

So this is what some people talk about when they talk about FRBR. The three groups of entities, the relationships between them all, and also the five user tasks, which we've skipped over: find, identify, select, obtain, and navigate.

This is FRBR. But this is isolated.

SLIDE: Almost complete FRBRization ... with arrows

We don't want FRBR in a box, not connecting with anything. We want it mixed and remixed with everything else out there. Fully FRBRizing the bibliographic universe only using FRBR, only using these entities and rela-

tionships, only talking about a FRBR Manifestation or a FRBR Person, is pointless. We want to grow and link up with everything else out there.

Some of this will happen because of authority records and building FRBR on existing systems.

And a lot of new work is happening now, here and there in various places, but it doesn't all fit together yet. We can't talk about all of this work, so I'll skip over Freebase and some other stuff, but here are a few things.

SLIDE: FRBR in RDF

The first is the representation of FRBR in RDF so we can make it available as linked data.

There are two ways to do that right now. IFLA, where FRBR comes from, is working on getting the official vocabulary in place. It's done, and you can see it in the NSDL Metadata Registry Sandbox, but it hasn't been made official yet because they're arguing over details about how to specify URL paths.

The other way was done by Ian Davis and Richard Newman, and it's become the de facto way of talking about FRBR in RDF. They did it in a couple of weeks in the fall of 2005, which tells you something about the comparative speeds of an international federation of library associations and two guys with a web site.

SLIDE: Libris linked data

You've seen the Swedish Libris catalogue, and this is what one edition of *The Three Musketeers* looks like if you ask for information in RDF. It has bibliographic metadata specified using standard RDF vocabularies, and they added in one of their own, the frbr-related element. It's like xISBN and thingISBN, and lists the ISBNs of other manifestations of the same work.

They implemented this in their own way, which is great and a nice piece of work. To link up the FRBR stuff with other data, they could connect up with what the Open Library is doing.

SLIDE: Open Library linked data

This is the same Open Library *Murder On the Orient Express* information we showed before, but as RDF, reformatted for readability. It tells you the title, that it's a work, gives you a URI representing Agatha Christie, and more. This is data ready to be linked.

Instead of Libris giving a list of all of the other manifestations of the same

work, they could link to the work as represented at the Open Library and say “this is a manifestation of that work.” Then Libris’s data and the Open Library’s data work together, and because Libris or anyone else can improve the Open Library’s data, everyone benefits.

SLIDE: id.loc.gov

Another source will be id.loc.gov. If Library of Congress Subject Headings come back online then the Group 3 entities are taken care of in a lot of catalogues, and if there’s a standard way to refer to Sears Subject Headings, or LibraryThing tags, then that much more falls into place. If authority records for Persons and Corporate Bodies are available, then Group 2 entities fall in place.

One by one, huge chunks of linked bibliographic data and metadata start connecting to each other, all available to everyone in an open way.

SLIDE: Almost complete . . . in a box

And instead of having this, with FRBR in a box trying to reach out, we have

SLIDE: Complete

this, where there is no box around FRBR. This is what we talk about when we talk about FRBR.

If we use an RDF vocabulary to talk about FRBR and the entities and relationships—everything here is prefixed with a frbr namespace to represent that—then instantly every thing on here, the entities and the relationships, is identifiable, linkable, usable, and reusable, and everything can be matched up, made equivalent to, or linked to anything else on the Web, so that the bibliographic universe is now part of the entire universe of knowledge.

4. Jodi: Conclusion