FRBR for Movies
and
Finding FRBR in MARC

OLAC meeting

Kelley McGrath
University of Oregon
January 27, 2013
FRBR for Movies
Users Are Looking for Movies

DRACULA

BY BRAM STOKER
FROM THE PLAY ADAPTED BY
HAMILTON DEANE & JOHN L. BALDERSTON
A TOD BROWNING PRODUCTION
PRODUCED BY CARL LAEMMLE, JR.

COPYRIGHT MCMXXXI BY UNIVERSAL PICTURES CORPORATION.
CARL LAEMMLE, PRESIDENT.
Libraries Describe Publications
Libraries Describe Publications

*Dracula* [videorecording] / Columbia Pictures ; directed by Francis Ford Coppola …

2-disc special ed.

Culver City, Calif. : Columbia TriStar Home Entertainment, c2005.

2 videodiscs (ca. 130 min.) : sd., col. ; 4 3/4 in. + 1 booklet.

Horror film series
## Libraries Describe Publications

<table>
<thead>
<tr>
<th>Dracula</th>
</tr>
</thead>
<tbody>
<tr>
<td>[United States] : Sony Pictures Home Entertainment ; Culver City, Calif. : Columbia TriStar Home Entertainment, c2005</td>
</tr>
<tr>
<td>View full record</td>
</tr>
<tr>
<td>LOCATION</td>
</tr>
<tr>
<td>VIDEO COLL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nosferatu a symphony of horror</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York : Kino on Video, c1991</td>
</tr>
<tr>
<td>View full record</td>
</tr>
<tr>
<td>LOCATION</td>
</tr>
<tr>
<td>VIDEO COLL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nosferatu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indianapolis, IN : KVC Entertainment [distributor], [199-?]</td>
</tr>
</tbody>
</table>
Users Care About Versions

Rg1024: open clip art library
Prototype: Movies & Versions

Funded: OLAC (Online Audiovisual Catalogers)
Developed by Chris Fitzpatrick
Small scale (limited data, few fields and records, simplified data model)

http://blazing-sunset-24.heroku.com
Movie (Mostly Work) Facets

Limit By Movie or Program:

Genre: Horror (12) Fiction (11) Experimental (6) Feature (6) Short (6) Ballet (1) Dance (1) more »

Dates: 1960s (3) 1990s (3) 2000s (2) 1920s (1) 1930s (1) 1950s (1) 1980s (1)

Original Language: English (8) Unknown (2) German (1) None (1)

Country: Unspecified (5) United Kingdom (2) United States (2) Austria (1) Canada (1) Germany (1)

1. Dracula (1931)

Director: Browning, Tod, 1882-1962
Language: English
Country: United States
Genres: Feature; Fiction; Horror;
Description: After a naive real estate agent succumbs to the will of the Count, the two head to London where the vampire hopes to stroll among respectable society by day and search for potential victims by night.

Get from a library:

35 mm film (nitrate) (1931)
Spoken Language: English
Aspect Ratio: Unspecified

Library: D

Fulfillment options below (expression, manifestation, item)

DVD (2006)
Spoken Language: English
Subtitle Languages: English; French; Spanish;
Aspect Ratio: Full screen (1.33:1)

Libraries: B, D, E
Version (Expression/Manifestation/Item) Facets

Limit By Version:

At Library:

C (14) [remove]

Format:

DVD (8)
VHS (6)

Publication Date:

1990s (6)
2000s (6)
1980s (2)

Spoken Language:

English (8)
None (5)
French (1)
Spanish (1)

Subtitle/Caption Language:

English (9)
French (4)
Prototype


Why the FRBR Model?

- to focus displays on original movies while supporting users in selecting and obtaining appropriate versions
- to enable shared maintenance of discrete movie-level records and reduce data redundancy, thereby supporting efficient production of more complete and accurate metadata
Finding FRBR in MARC
Machine-Actionable Data

• Structured data
  OriginalReleaseYear = 2011
  \textit{NOT}
  Originally released as a motion picture in 2011.

• Mapped to FRBR entities and attributes
  OriginalReleaseYear = Date of the Work
Machine-Actionable Data

• Supports *faceted access* and the creation of more *readable, grid-like displays*

• Enables *targeted search* and *flexible display*

**Dracula (1992 : Francis Ford Coppola)**
**Format:** DVD, NTSC
**Languages:** English, German
**Subtitles:** English, French, Spanish
**Accessibility:** Closed-captioned
**Region:** Region 1 (U.S. and Canada only)
ALA Annual 2012 presentation on extracting work data from MARC records:

http://goo.gl/BrpuJ
Names and Functions

Want to link authorized names with controlled vocabulary for functions

Director = Clint Eastwood

directed by Clint Eastwood

700 $a Eastwood, Clint, $d 1930- $4 drt
Identify Individual Statements Based on Punctuation

245 $c$ Metro-Goldwyn-Mayer picture ; screenplay by George S. Kaufman and Morrie Ryskind ; directed by Sam Wood

1. Metro-Goldwyn-Mayer picture
2. screenplay by George S. Kaufman and Morrie Ryskind
3. directed by Sam Wood
Make a List of Terms for Each Function (Synonyms)

aus =

- screenplay
- screen play
- screenwriter
- script

- scriptwriter
- writer
- written by
Make a List of Terms for Each Function (Translations; Unwanted Variations)

\[ \text{drt} = \]

- directed by
- direction
- director
- 監督
- Regie
- режиссер-постановщик
- director of photography
- animation director
Map Transcribed Names

1. screenplay by George S. Kaufman and Morrie Ryskind
2. directed by Sam Wood

1. 700 $a Kaufman, George S. ...
1. 700 $a Ryskind, Morrie, ...
2. 700 $a Wood, Sam, ...
Map Transcribed Roles

1. screenplay by George S. Kaufman and Morrie Ryskind
2. directed by Sam Wood

1. 700 $a Kaufman, George S. … $4 aus
1. 700 $a Ryskind, Morrie, … $4 aus
2. 700 $a Wood, Sam, … $4 drt
Many Possibilities...

700 $a Wood, Sam, 1883-1949 $4 drt

http://id.loc.gov/authorities/names/n85151535.html
= http://www.imdb.com/name/nm0939992/
http://id.loc.gov/vocabulary/relators/drt.html

Director: Wood, Sam, 1883-1949
Regie: Sam Wood
Natural Language Processing (NLP)

“deals with analyzing, understanding and generating the languages that humans use naturally”—Webopedia

– Artificial intelligence
– Automatic summarization
– Machine translation
– Named entity recognition (NER)
Natural language processing toolkits

Named entity recognition (NER)

– “April Stevens”

– “Twentieth Century Fox”

– “an Austrian-French co-production, Wega Film, MK2 Productions and Les Films Alain Sarde, Arte France Cinéma”
Named entity recognition

Approaches to matching

• Start with authorized names and match to statements
• Start with statements and match to authorized names

1. screenplay by George S. Kaufman and Morrie Ryskind
2. directed by Sam Wood

1. 700 $a Kaufman, George S. …
1. 700 $a Ryskind, Morrie, …
2. 700 $a Wood, Sam, …
Hard-Coded Rules vs. Machine Learning

• Rules:
  – Manually-compiled lists and decision trees

• Machine learning:
  – Usually based on statistical models
  – Supervised vs. semi-supervised vs. unsupervised learning
Supervised Learning

• Training data
  – Set of hand-annotated inputs and desired outputs

directed by Sam Wood →
\[ \text{drt} = \textcolor{orange}{a}\textcolor{blue}{ Wood, Sam, d}\textcolor{red}{ 1883-1949} \]

• Computer then generalizes from training data when working on novel data
What You Can Do Soon

• Help us create a hand-annotated set of correct answers for
  – Training data
  – Evaluation

• Online web form coming soon…
What You Can Do Soon

directed by Sam Wood

- English
- Sam Wood
- Wood, Sam, $d 1883-1949
- Person
- Directed by
- Director
What You Can Do Now

Use

• **130** uniform titles
• **257** country of producing entity
• **046 $k** for original date
• **041 $h** for original language
• **1xx/7xx $4/$e** relator codes or terms
What You Can Do Now

130 0- $a Lawrence of Arabia (Motion picture)
257 -- $a Great Britain $a United States $2 naf
046 -- $k 1962
041 0- $a eng $h eng
700 1- $a Lean, David, $d 1908-1991. $4 drt
700 1- $a O'Toole, Peter, $d 1932- $4 act
What You Can Do Now

130 0- $a My neighbor Totoro (Motion picture)
257 -- $a Japan $2 naf
046 -- $k 1988
041 1- $a jpn $a eng $j eng $h jpn
700 1- $a Miyazaki, Hayao, $d 1941- $4 drt
   $4 aus
Overview of Project

1. Develop end-user interface to take advantage of FRBR and facets

2. Extract and transform existing data
   MARC → normalized, FRBR-based data
   Cluster records for FRBR entities
   Create provisional work (movie) records
   Assess and correct errors where possible

3. Create backend interface for ongoing input and management of metadata

4. Develop guidelines and documentation for catalogers
Interested in Participating?

Contact me at

Kelley McGrath
Metadata Management Librarian
University of Oregon Libraries
kelleym@uoregon.edu
(541) 346-8232