

Roles and Reels: RDA and Relationship Designators for Moving Images

**Kelley McGrath
University of Oregon**

**Catalog Management Interest Group
ALA Midwinter 2015**

1

Data about Roles

- **Goal:**
 - **Teach a computer to identify names and roles in free text in 245\$c, 260\$b, 264\$b, 508, 511**
- **Interim Goal:**
 - **Create a file of correct answers for the computer**
 - **Use for training and evaluation**

2

My talk today is based on observations made after looking at a large list of roles from existing MARC records for moving images. This long list of roles is a by-product of a part of a larger project. OLAC has been working on a project to try to extract more structured data from existing bibliographic records for film and video. One of our goals is to analyze free text fields about who did what in MARC records and try to turn that information into something more machine actionable. And we would like to automate that process. To help with this, we wanted to come up with a pool of what we considered correct answers for training and evaluating our program.

Crowdsourcing

written and directed by Charles Chaplin

Language of role or function: We think this film or video is in English.

Name in credit: <input type="text" value="Charles Chaplin"/> <input type="checkbox"/>	Match for name in credit: <input type="text" value="Chaplin, Charlie,"/> <input type="checkbox"/>	Role or function in credit: <input type="text" value="written by"/> <input type="checkbox"/>	English translation of role/function: <input type="text"/> <input type="checkbox"/>
Name in credit: <input type="text" value="Charles Chaplin"/> <input type="checkbox"/>	Match for name in credit: <input type="text" value="Chaplin, Charlie,"/> <input type="checkbox"/>	Role or function in credit: <input type="text" value="directed by"/> <input type="checkbox"/>	English translation of role/function: <input type="text"/> <input type="checkbox"/>

3

We decided to try to crowdsource these correct answers. This is a screenshot of the core of our web form. On the top line you can see the credit or statement about responsibility that we want analyzed. These credits come from real MARC records that we have gathered. The first thing we want the computer to do is identify the names and roles in the credit. So we ask our human helpers to reproduce this process by splitting the information between two boxes--the name in credit and role or function in credit on the bottom line. When possible, we would like to link names up with their authority records. As a first step, we want the computer to see if it can match the transcribed name to a name in a 7xx (or occasionally 1xx) in the same record. To imitate this, we provide a dropdown list with the 1xx and 7xx fields from the record and ask people to pick a matching name if there is one. This is the box labeled "match for name in credit." I'll talk about the rest of the form with the next slide.

Crowdsourcing

Please Annotate This Credit

kyakuhon, Tsuda Fumio

Language of role or function: Japanese ✔ We think this film or video is in Japanese.

Name in credit: Tsuda Fumio ✔ ?

Match for name in credit: Tsuda, Fumio ✔ ?

Role or function in credit: kyakuhon ✔ ?

English translation of role/function: script ✔ ?

4

We also want to know what language the role is in. The dropdown for "language of role or function" has probably been the most misunderstood part of the form. The clue next to it where we say "we think this film is in Japanese" is meant to be a hint for dealing with words that are the same in more than one language. What we want people to choose from the list is the language of the words that they put in the "role or function in credit" box. Finally, in the bottom right we ask for an English translation if the role isn't in English.

Data

CREDIT NAME	AUTH NAME	ROLE LANG	CREDIT ROLE	ROLE TRANS.	ROLE CATEGORY
Joseph Barbera	Barbera, Joseph	english	Produced by		production
Werner Herzog	Herzog, Werner,	german	Produktion	production	production
Elías Querejeta	Querejeta, Elías	spanish	una producción	a production	production
Christophe Rossignon		french	produit par	produced by	production
BBC	British Broadcasting Corporation	no role specified			no role

5

This is an example of the type of data that we are gathering. The far right box in red doesn't come from the web form, but rather from a separate grouping of the terms done after the fact.

80/20 Rule

- On-screen, vocal or musical participation (performers, cast)
- Direction
- Production
- Writing
- Cinematography
- Editing
- Music (sometimes hard to tell if music credit is for composing or performing)

6

Looking at the data, the distribution of roles follows a form of the 80/20 rule. Most of the roles fall into a small number of categories, which coincidentally happen to be the categories that have traditionally been recorded in moving image records.

Variants

- Executive producer
- Co-producer
- Assistant producer
- Associate producer
- Line producer
- Senior producer
- TV producer
- Video producer

7

However, if you look more closely, it's not so simple. There are in fact numerous subcategories within each of these categories. For example, here are some common types of producers.

Non-English Terms

- генеральный продюсер = general producer
- realización = production? direction?
- 总导演 = director-in-chief
- 企画 = planning? development? production?
- mise en scène = staging? direction?

8

There is also the challenge of what to do with non-English terms that possibly don't map cleanly to an English term or to a word used in the Anglo-American film industry. What does a general producer or director-in-chief do? Kikaku in Japanese literally means planning, but that isn't a word that is used in North American film credits to my knowledge.

Not-So-Straightforward

- Segment producer
- Producer in Japan
- Series producer
- Series senior producer
- Producer for BBC
- Producer of U.S. release
- Executive producer of English version

9

There are also roles that have more complicated relationships with the resource. They may pertain to just a part of the resource, usually something time-based like a segment, but it can also be location based, as in the producer for Japan. Alternatively, their responsibility may be at a level higher than the resource, such a series. Producers may also be connected only to a particular version of the resource (which doesn't really line up with RDA's rigid association of producers with works).

The Long Long Tail

- Dream sequences based on designs by Salvador Dali.
- Garden designer
- synchronization director
- tiger trainer
- spider web spinner
- floor drop painted by ...
- Body art & locations
- Movement coordinator

10

After you get past that 80%, there are many usual, but less common (or at least less commonly-recorded) roles. But once you get past those, the long tail is *really* long.

Not to Even Start on Instruments



From accordion to whistles
From clarinet to clavinet

From around the world:

bodhran, conga, kenkeni,
mridangam, tabla,
taiko and so on



11

And it's truly amazing how many musical instruments there are in the world.

Challenges for assigning relationship designators

1. The vague
2. The ambiguous
3. The incoherent

12

Now I'd like to talk about a few challenges other than the long tail that come up when I think about trying to assign relationship designators to some of the transcribed roles in our dataset. I think of these as the vague, the ambiguous and the incoherent.

The Vague

responsible, Lê Mỹ Phương

13

The vague ones seem to be the most common. This is my favorite so far. It just seems like the archetypal statement of responsibility.

The Vague

- action, Shyam Kaushal
- associate, Ved M Rao
- concept, Clive Sugars
- devised & designed by Kamalini Dutt
- idea, Hana Bělohradská
- Series proposed by Benoit Peeters
- supervisor, Dr. Nurdin Perdana
- Team works, Kartawijaya ... [et al.]
- with Wolfgang Brendel
- Dreamed and directed by Randy Redroad

14

However, there are plenty of other examples. What does an associate or supervisor do on a film?

The Ambiguous

- By** Shakespeare [author of play]
By W.A. Mozart [composer of opera]
By John Cleese and Connie Booth
- Bass** [guitar or vocalist?]
Music [performer or composer?]
Songs [writer or singer?]

15

Ambiguity can also be a problem. Just looking at the words in these statements, you wouldn't know what these people did. However, this is probably a bigger problem for a computer than it is in real life as in many cases these kinds of ambiguous statements can be resolved with context.

The Ambiguous

- Musical adaptation and direction, Penella [musical direction or direction?]
- associate director and editor, Alexander Hammid [associate editor or editor?]
- author and singer of songs, Vladimir Vysotski [author or author of songs?]
- graphics and video editor, Michael Seibert [graphics or graphics editor?]

16

In some of these compound statements, it's hard to say if the initial word applies to both roles or not. Is Penella a music director or just a regular director?

The Incoherent

Winstar Cinema release, Shochiku Co., Ltd.
presents a 3H Film Productions of a film by
Hou Hsiao-Hsien

an Old Photo film presents

A Sony Pictures Classics release of Vans "Off
the Wall" Productions presents an ADP
Productions

un film von Xavier Koller

Villealfa Filmproductions esittää ; Aki
Kaurismäen elokuva

17

Then there are incoherent statements that make no sense grammatically. There are also a surprising number of statements that mix languages indiscriminately. Many of the incoherent statements in our dataset are probably transcription errors, but likely not all. The last credit in red proved very puzzling to people helping with translating. It turns out to be inextricably linked grammatically to the previous statement and is nonsensical on its own.

What's Missing in RDA and MARC?

Way to relate names of performers to character names

- Desi Arnaz (*Ricky Ricardo*)
- Sandra Dickinson *as Genie*
- Laura Cervinsky, *Barbarina, Antonio's daughter*
- *Sieglinde*, Jessye Norman

18

The data contains many instances of situations where catalogers tried to convey information that is not possible or is difficult to convey in RDA and MARC in a structured way. For example, catalogers sometimes attempted to link cast members' names to the names of the characters they played. Unfortunately, since there is no standard way to do this, this information is expressed in a wide variety of ways. This makes it harder for a computer to work with. In the final style, where the names are separated by a comma, it is sometimes impossible to distinguish a priori whether one of the names is a character name.

What's Missing in RDA and MARC?

Way to relate roles to parts

- Narration, Kajiwara Shirō ("*Nihon kōgei no hana*"), Kagami Sachiko ("*Hyakka ryōran*")
- *The steadfast tin soldier* (Patricia McBride, Mikhail Beryshnikov)
- [*Lament, Undertow*] camera, James Byrne
- directed by Andy Sommer. *The masterclasses : 6 films* / directed by Allan Miller
- *Little match girl*: Cinematography, Jean Bachelet^o

There are many records that describe more than one work. Again, catalogers use a multitude of strategies to express these connections.

RDA Relationship Designators

Generate endless discussion on RDA-L and elsewhere

- Is there a relationship designator for this?
- What relationship designator do I use for that?

JSC keeps expanding the list of relationship designators

- Never going to keep up with the long tail

- How do we deal with the statements that we don't know how to interpret?

20

Relationship designators are certainly a topic of endless discussion. The length of the long tail means that there will never be enough relationship designators to cover all the roles that catalogers come across. The JSC will keep being asked to add more and more relationship designators. There will always be statements about responsibility that for some reason or other are difficult to interpret or categorize.

IMDb-Style Approach?

Does the Internet Movie Database (IMDb) provide a potential model?

- Limited list of broad role categories
- Very specific, free-text roles

21

Libraries are not the only ones to face these problems. I have spent a lot of time using the Internet Movie Database, or IMDb. IMDb uses a combination of broad categories of roles combined with the ability to expand on or clarify the role using free text.

IMDb-Style Solution Approach?

Romeo and Juliet (1968)

Controlled vocabulary for Broad Terms:

Writing Credits

William Shakespeare ... (play)
Franco Zeffirelli ... (screenplay)

Grab Bag Category:

Other crew [=contributor?]

Gabriella Bernardi ... production secretary
Alberto Testa ... choreographer

22

This is perhaps not exactly the way a library would do this, but it demonstrates the concept. Various roles are bundled together in the category of writing credits. This would be IMDb's equivalent of a relationship designator. This is the level that makes the most sense to use for searching, limiting or faceting. However, for the user looking at the record, IMDb also provides a more specific explanation of what the person did. They also have a category called "Other crew," which contains all the roles that don't fit into another category. RDA could use creator and contributor in a similar way.

IMDb-Style Approach?

Romeo and Juliet (1968)

Names linked to roles:

Leonard Whiting ... Romeo
Olivia Hussey ... Juliet
John McEnery ... Mercutio

23

IMDb also links cast names to roles.

Structured, Connected Data

The Lord of the Rings: The Fellowship of the Ring (2001)

Cast (in credits order) verified as complete:

Alan Howard ... *The Ring* (*voice*)

Makeup Department:

José Luis Pérez ... *key hair stylist*
(as *Jose L. Perez*)

24

Libraries sort of do all these things. However, the key thing IMDb does, which libraries don't do, is connect all these pieces of information in a machine-actionable way. They link up the authorized name, the transcribed name, the role category, the specific role and the character portrayed. The parts shown here in **blue** are from a controlled vocabulary and the parts in *red* are transcribed text.

Help Us Out

Annotate credits at:

<http://olac-annotator.org/>

Discussion list:

[https://lists.uoregon.edu/mailman/
listinfo/olac-credits](https://lists.uoregon.edu/mailman/listinfo/olac-credits)

25

If you're interested in supporting our project, why not go to our website and annotate a few credits? Thank you.

<http://olac-annotator.org/>

[https://lists.uoregon.edu/mailman/
listinfo/olac-credits](https://lists.uoregon.edu/mailman/listinfo/olac-credits)