

# Identifying and Clustering Moving Image Works Found in Manifestation-Based MARC Records

**OLAC Conference 2014**

**Kelley McGrath**

**Metadata Management Librarian**

**University of Oregon**

# Work-Centric View

## 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.

Results focused on movie (work)

### Get from a library:

35 mm film (nitrate) (1931)

Spoken Language: English

Aspect Ratio: Unspecified( Unspecified )

Library: D

Fulfillment options  
below (expression,  
manifestation, item)

DVD (2006)

Spoken Language: English

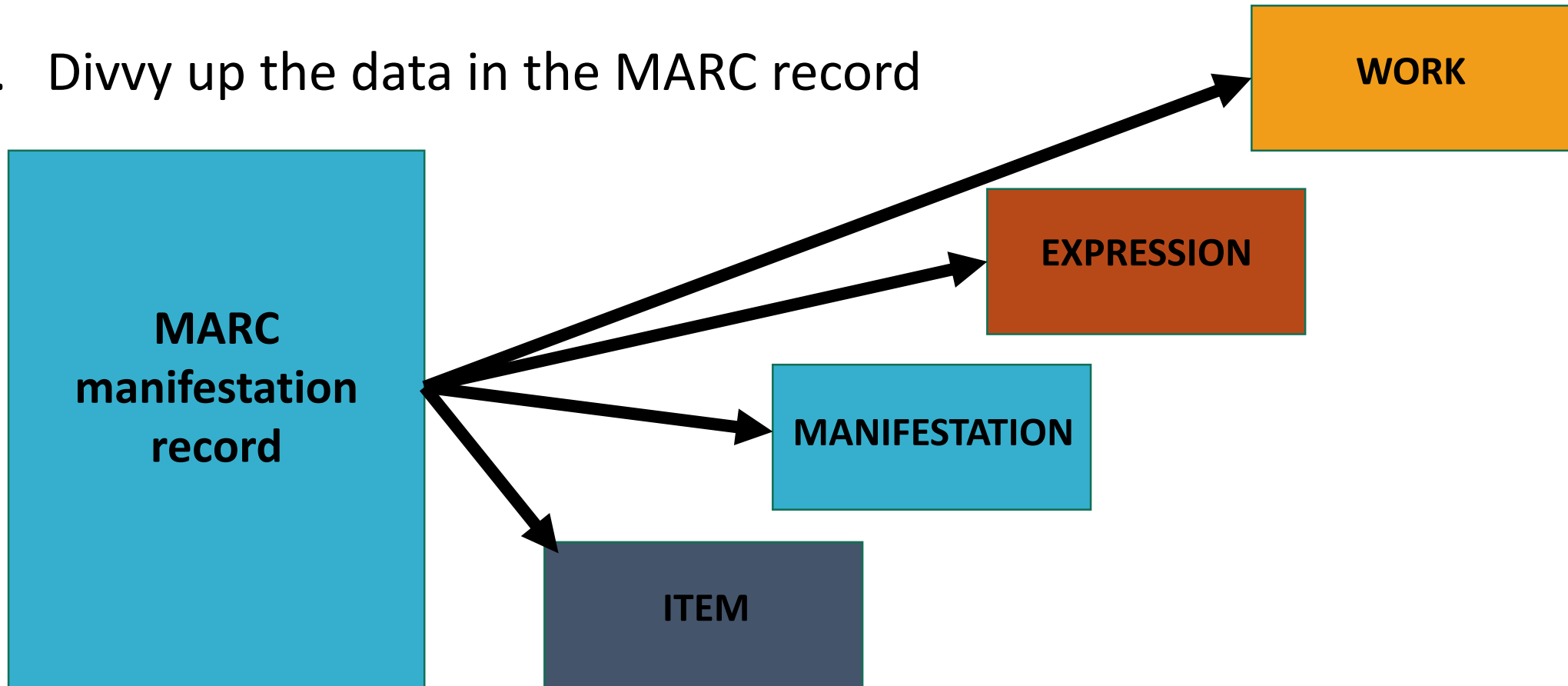
Subtitle Languages: English; French; Spanish;

Aspect Ratio: Full screen ( 1.33:1 )

Libraries: B , D , E ,

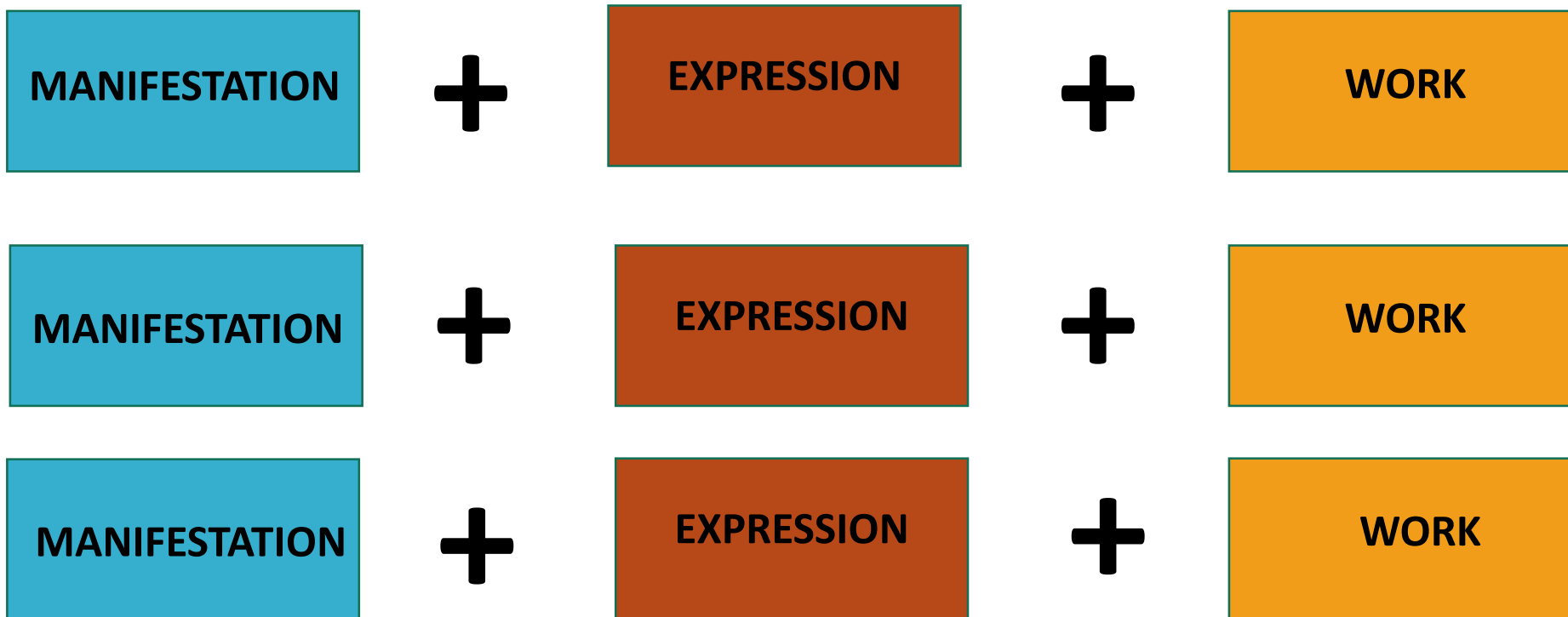
# Manifestation Records → Work-Centric View

1. Divvy up the data in the MARC record



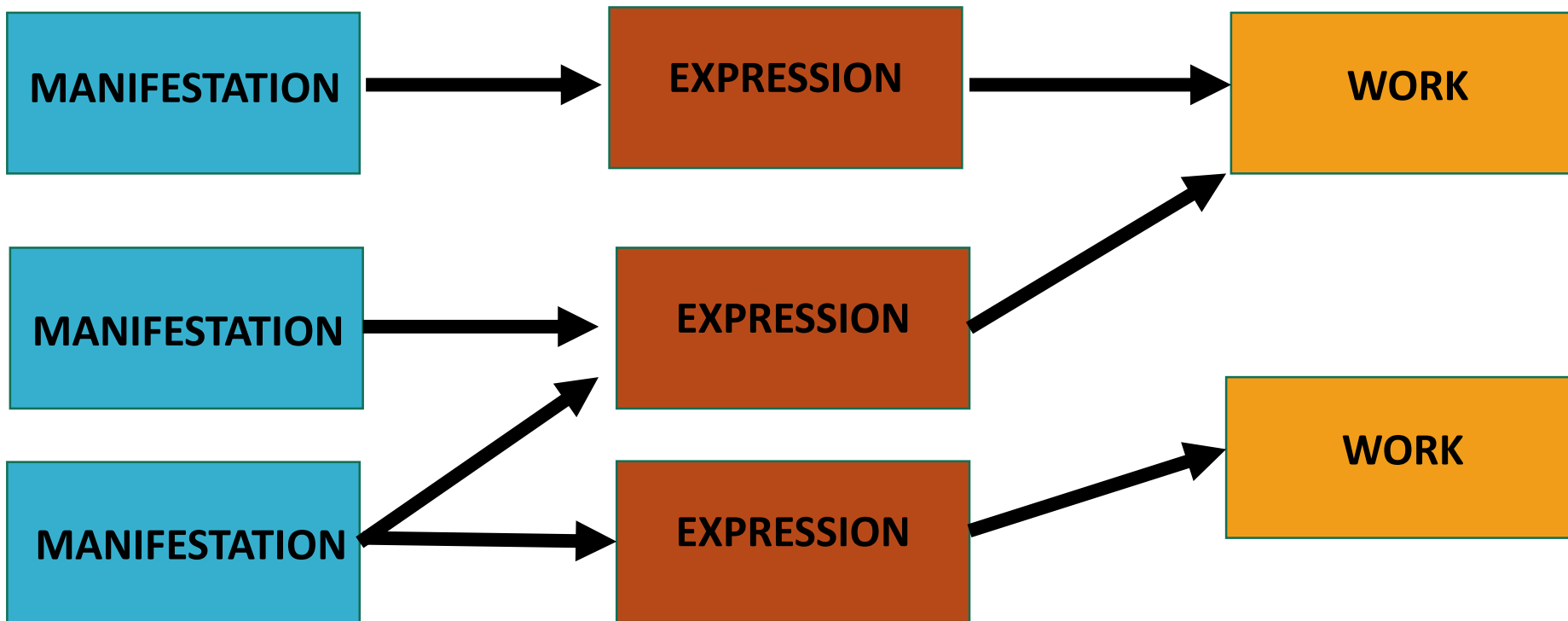
# Manifestation Records → Work-Centric View

1. Divvy up the data in the MARC record



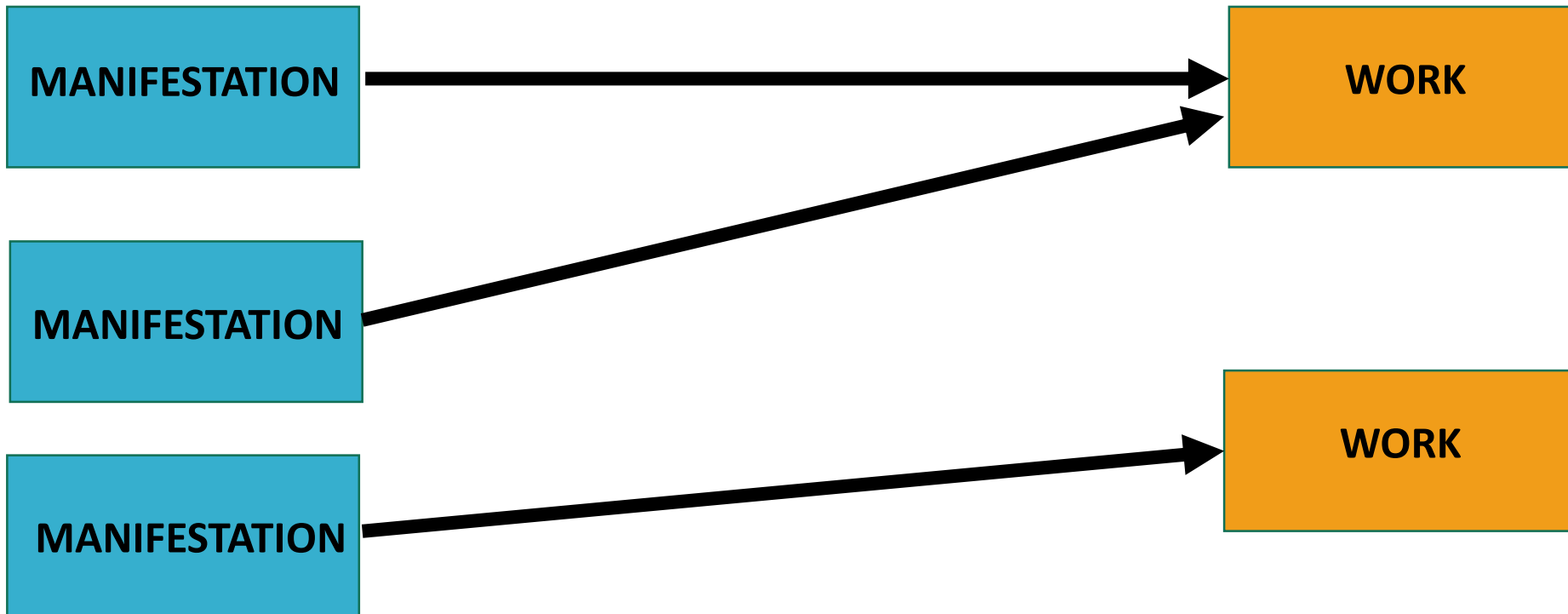
# Manifestation Records → Work-Centric View

2. Group equivalent entities and de-dupe



# Manifestation Records → Work-Centric View

2. Group equivalent entities and de-dupe



# LC FRBR Display Tool

Matches works based on

author + title: 100/110/111 + 240/243/245

title: 130/240/243/245

<http://www.loc.gov/marc/marc-functional-analysis/tool.html>

# LC FRBR Display Tool

- **Author: Not entered under author**
- **Work: Pride and prejudice**
  - *Form: moving image*
    - Edition: 10th anniversary limited collector&apos;s ed.
      - Title: [Pride and prejudice](#)
      - Statement of responsibility: a co-production of BBC Television and BBC Worldwide /
      - Imprint: A&amp;E Television Networks, 2006
      - Physical Description: 3 videodiscs (300 min.) : sd., col. ; 4 3/4 in. + 1 book (117 p.)
      - ISBN: 0767089820 (DVD set)
      - ISBN: 014025157X (book)
      - Publisher Number: AAE-75368 A&amp;E Home Video (set)
    - *Form: moving image - English*
      - Edition:
        - Title: [Pride and prejudice](#)
        - Statement of responsibility: Metro-Goldwyn-Mayer ; directed by Robert Z. Leonard ;
        - Imprint: MGM/UA Home video, 1990
        - Physical Description: 1 videocassette (118 min.) : sd., b&amp;w ; 1/2 in.
        - ISBN: 079280385x



# OCLC FRBR Work-Set Algorithm

| Key Type                      | Occurrences         | Example   |
|-------------------------------|---------------------|---|
| <author>/<title>              | 97,961,220 (70.65%) | bjorling, jussi\1911 1960/opera arias and duets |
| <uniform title>               | 1,569,352 (1.13%)   | 10 commandments                                 |
| /<title>/[one or more <name>] | 26,559,404 (19.16%) | /bergler/bergler, friedrich                     |
| /<title>/<oclc number>        | 12,559,537 (9.06%)  | /britain and antarctica/289903387               |

1. Author + title: 1xx + 240/245
2. Uniform title: 130
3. Title + name(s): 245 + 7xx
4. Title + OCLC number: 245 + 001 (singleton)

<http://oclc.org/research/activities/frbralgorithm.html>

# OCLC FRBR Work-Set Algorithm

| Key Type                      | Occurrences         | Example   |
|-------------------------------|---------------------|---|
| <author>/<title>              | 97,961,220 (70.65%) | bjorling, jussi\1911 1960/opera arias and duets |
| <uniform title>               | 1,569,352 (1.13%)   | 10 commandments                                 |
| /<title>/[one or more <name>] | 26,559,404 (19.16%) | /bergler/bergler, friedrich                     |
| /<title>/<oclc number>        | 12,559,537 (9.06%)  | /britain and antarctica/289903387               |

130 Star wars, episode I, the phantom menace (Motion picture)

Pulp fiction + Quentin Tarantino; Andrzej Sekula; Bob Weinstein ...

# OCLC FRBR Work-Set Algorithm

| Record Number | Names             | Key                     |
|---------------|-------------------|-------------------------|
| Record 1      | “smith” and “doe” | /day in the park/smith  |
| Record 2      | “doe” and “jones” | /day in the park/smith  |
| Record 3      | “jones”           | /day in the park/smith  |
| Record 4      | “harvey”          | /day in the park/harvey |

# marc2frbr Conversion Tool (Norway)

1. Identify the various entities described in the record
2. Select the fields that describe each entity
3. Find the relationships between entities
4. Support normalization by finding and merging equivalent records

# INESC-ID, Portugal

String similarity rather than exact string matching

1. Character-based: good for typographic errors
2. Token-based: good for variation in arrangement of words (“Smith, June” vs. “June Smith”)
3. Phonetic

# Challenges

- Data is incomplete, inconsistent, inaccurate or missing
- Data is in free text (e.g., roles, relationships)

# Things That are Hard to Get Out

- Expressions
  - OCLC found it impossible to accurately identify expressions based just on MARC records; it is necessary to examine some items
- Multiple works within a single manifestation

# Challenges for Moving Images

- Algorithms that identify and cluster works function better when they have more than one match point (1xx + 240/245) or an authoritative match point (130)
- Many moving image records have no 1xx fields
  - Add more moving image preferred/uniform titles!



# More Match Points for Moving Images?

- Title
- Original date
- Director

Could You Help Me with Directors?

<http://tinyurl.com/oc27ng6>