

Complexities of Authors and Authorities in KOSs

Heather Lea Moulaison
The iSchool at the University of Missouri
ASIST Annual Meeting
November 3, 2013

The Role Authors Play

- Authors aren't and never were the focus of knowledge organization systems (KOSs) in libraries: library materials were (and are)
- Authors, creators, etc. function as attributes of materials, much as a title or a subject.
 - Authors are not a focal point in the system, just an access point to collocate **materials**

People Are Complex

- Authors are much more complex than
 - People
 - The character strings that represent them
- In Functional Requirements for Authority Data (FRAD) and the FR-family of models
 - People have attributes
 - People/attributes have relationships other people/attributes
 - These relationships are at the heart of the interest with linked data in libraries
- How much of the complexity of authors is captured in KOSs?
 - In access points
 - In authority records

Richness and Complexity of Access Points in Surrogates

- Persons who contribute to the work or its expression are represented by access points in MARC records in fields 100 and 700
 - Access points are comprised of one or more subfields
- OCLC work on fields used in surrogates
 - “MARC Usage in WorldCat” (Roy Tennant)
 - Ability to analyze the use of subfields in WorldCat records for personal names associated with the work or the expression

Example Access Point

DATABASE: Library of Congress Online Catalog

YOU SEARCHED: Keyword (match all words) = lawrence, d.

SEARCH RESULTS: Displaying 1 of 10000.

◀ Previous Next ▶

Brief Record

Subjects/Content

Full Record

MARC Tags

Sea and Sardinia [by] D. H. Lawrence; with an introduction by Richard...

Relevance: ●●●●●

LC control no.: 68134828

LCCN permalink: <http://lccn.loc.gov/68134828>

Type of material: Book (Print, Microform, Electronic, etc.)

Personal name: [Lawrence, D. H. \(David Herbert\), 1885-1930.](#)

Main title: *Sea and Sardinia [by] D. H. Lawrence; with an introduction by Richard Aldington.*

Published/Created: London, Heinemann, 1968.

Description: xi, 208 p. 20 cm.

CALL NUMBER: [DG975.S3 L3 1968b](#)

Copy 1

-- **Request in:** Jefferson or Adams Building Reading Rooms

-- **Status:** Not Charged

Personal Name Access Points (100 and 700) Subfields

((R)=Repeatable (NR)=Nonrepeatable)

Subfield code	Description
‡a	Personal name (NR)*
‡c	Titles and other words associated with a name (R)**
‡d	Dates associated with a name (NR)**
‡e	Relator term (R)***
‡q	Fuller form of name (NR)**
‡0	Authority record control number (R)***
‡4	Relator code (R)***

Key: Mandatory*, Required if applicable**, Optional***

Complexity of Personal Name Headings (i.e. 100 and 700) in OCLC's WorldCat

‡	Description	Total occurrences	%*
a	Personal name (NR)	304,147,357	100.00
d	Dates associated with a name (NR)	89,405,928	29.40
4	Relator code (R)	46,950,525	15.44
0	Authority record control number (R)	28,089,203	9.24
e	Relator term (R)	22,216,662	7.30
q	Fuller form of name (NR)	10,675,174	3.51
c	Titles and other words associated with a name (R)	9,305,730	3.06

*inferred N=304,147,375

Observations

- Content is available in the personal name access point beyond the simple character string that represents the surname/forename of the individual.
 - Dates, qualifiers, titles, etc. are an easy way for users to differentiate between different authors with the same forename/surname in a string of hits
 - This disambiguation works best IF dates, qualifiers, titles, etc. are familiar
 - i.e. users must understand about the author for this mechanism to be truly meaningful in the search for materials.

Caveats

- Interoperability can be hindered; automated character-string matches between systems may not interpret other information in access points correctly, leading to mismatches (i.e. Open Library).
- Much of the data input is not necessarily visible to patrons (relator info, authority record control numbers...)

RDA and Authors

- Through *Functional Requirements for Authority Records* (FRAD), RDA is able to record elements (i.e. attributes) relating to authors in *authority records*
 - FRAD specifies 14
 - RDA/MARC currently proposed 10 fields of attributes

Richness and Complexity of Author Data in Authority Records

- Data in authority records, until present,
 - Helps catalogers choose the correct access point
 - Can be used in library systems as see/see also variant access points
- Other content embedded in authority records, beyond access points, has the potential to be searchable through library systems
- Imagine providing the ability to search the system for
 - materials written by Nobel laureates who have been affiliated with both Harvard and Stanford
 - translated works of German-speaking authors born in Vienna
 - materials written by authors who have had two genders associated with their person

Author Attributes in Authority Record for D. H. Lawrence

.....x."x.jx x"x•xœx

Birth date: 18850911

Death date: 19300302

Place of birth: Eastwood, Nottinghamshire, England

Place of death: Vence, France

Profession or occupation: Novelist Poet Playwright

Special note: Machine-derived non-Latin script reference project.

Non-Latin script references not evaluated.

Found in: His The white peacock, 1911.

Kim, C.M. Oĭ†dum ul†i pulkkot, 1988: t.p. (Lorensul†)

Moksa ul†i ttal tul†l, 1991: t.p. (D.H. [rom.] Lorensul†)

Angliil†skaii, ai,j literatura 1900-1914 godov, 1993: t.p. (D.G. Lourens)

Leĭ,dijas CĭEaterlejas miĭ,ĭĭ|aĭ,kais, 1991: t.p. (D.H. Lorens) colophon (Deil†vid Gerbert Lorens)

Caccipoĭ,yina manisĭfi, 1946: t.p. (Dĭfi. Ec. Laĭ,rensu)

Lii, ui,jbovnik ledi Chatterleil†, 2000: t.p. (Deĭ†vid Gerbert Lourens)

Wikipedia, Nov. 19, 2012 (D.H. Lawrence; David Herbert Lawrence; born 11 September 1885 Eastwood; died 2 March 1930 in Vence, France; English novelist, poet, playwright, essayist, literary critic and painter)

Associated language: eng

Attributes Fields in Personal Name Authority Records

MARC Code	Description
046	Special Coded Dates (R)
370	Associated Place (R)
372	Field of Activity (R)
373	Associated Group (R)
374	Occupation (R)
375	Gender (R)
376	Family Information (R)
377	Associated Language (R)
378	Fuller Form of Personal Name (R)
678	Biographical or Historical Data

Richness of Attributes Fields in Personal Name Authority Records in a Cluster of Academic Libraries

*A case study of attributes (RDA elements) in MERLIN
personal name authority records*

- MERLIN is the cluster of academic MO libraries,
mostly near the center of the state

An enormous debt of gratitude is owed to Felicity
Dykas, Head of Digital Services at MU's Ellis Library
for pulling this data

Merlin Cluster in MOBIUS



- Missouri University of Science and Technology
- University of Missouri (MU)
- MU Law
- University of Missouri - Kansas City
- UMKC Law
- University of Missouri - St. Louis
- University of Missouri Archives and WHMC

Method

- Names and attributes from all personal name authority records used as authors (not as subjects) in the cluster's surrogates were pulled the last day in September, 2013
 - After 6 months of RDA cataloging
 - 1,156,316 records' data pulled
 - Data pulled as 4 text files (file 4's contents, 256317 records, unusable for calculations in the aggregate)
 - 21058 records (8.2%) were drawn at random from the corrupted file and were examined to make sure data analyzed was representative
- Results presented are from usable data (899,999 records)

Results of Field Use in MERLIN Authority Records for Persons

MARC tag	Description of element	# records with content in the field	% of total (N=899,999)
046	Special Coded Dates (R)	63938	7.10
678	Biographical or Historical Data	25576	2.84
375	Gender (R)	16743	1.86
374	Occupation (R)	16030	1.78
370	Associated Place (R)	13432	1.49
377	Associated Language (R)	11935	1.33
378	Fuller Form of Personal Name (R)	9656	1.07
372	Field of Activity (R)	8114	0.90
373	Associated Group (R)	8049	0.89

No. of elements (as MARC fields) in records	No. of authority records with elements (N=899,999)	% of total
0	810051	90.01
1	63440	7.05
2	7984	0.89
3	2532	0.28
4	4432	0.49
5	4466	0.50
6	4038	0.45
7	2318	0.26
8	685	0.08
9	53	0.01

Observations

- The richness of data available in authority records is minimal, overall
 - The vast majority (90%) have no additional elements
 - Only 7% have additional content in one MARC field
 - Only 3% have additional content in more than one MARC field
- If we consider that consistency is an indicator of metadata quality, the data that we are maintaining on authors is not of high quality because it is not consistently supplied.

Discussion

- KOSs used in libraries are working hard to provide additional information about authors and other contributors
 - In the access point (going beyond the forename, surname character string)
 - In the authority record
- Yet providing additional info in the access point may not make a lot of sense if our data is to be interoperable
- And too little data appears in the body of the personal name authority records
 - With no mechanism for ensuring that the attributes are recorded.

Conclusion and Future Study

- The complexities of authors are coming to the fore and are being recognized
- KOSs are beginning to record data in a way that allows that complexity to be better managed, possibly improving the
 - Potential for search
 - Potential for interoperability
 - Potential for re-use in trans-disciplinary domains
- Systems will continue to provide access to materials
- Future study: Is it reasonable to think in the era of participatory content creation, that the FRBR WEMI model adequately takes collaboration into consideration?