Look beyond

Indicizzazione per soggetto

delle risorse non librarie

6 febbraio 2023

Organizzato da

AIB CILW; ISKO Italia

In collaborazione con

Sistema Bibliotecario Sapienza

Le stringhe e la loro semantica Il trattamento dei soggetti nei processi di modellamento delle entità

Tiziana Possemato

Abstract

The transition from the concept of *cataloging* to that of *entity modeling* in the cataloging field requires some tailoring of the mental approach to the wording and management of information, before that of practical tools. Entity modeling is based on the idea that every object in the real world, whether concrete or purely conceptual, can be well identified and described, with the main purpose of being reused, <u>even</u> in very different contexts. The most up-to-date cataloging rules, for example the Official RDA Toolkit, introduces some terms that clearly indicate this shift of perspective towards entity modeling: the term *entity boundary*, for example, clearly suggests the concept of identifying the boundaries of an entity, to steer and direct cataloging decisions.

But the processes of identification, which are effective on some types of things, instead show all their weaknesses when the boundaries of the object become uncertain or even fluid: semantic indexing becomes prone to manipulation or distortion of meaning in this regard, being largely influenced by cultural contexts, linguistic variants, different political and social approaches of the responsible institutions. The problem of using the *owl:sameAs* property to connect individuals with the same identity but in different datasets, which isn't new in the field of Artificial Intelligence and the Semantic Web, becomes huge when the entities to be connected are difficult to place in a precise semantic frame. One of the possible approaches to capture and identify the meaning that a string conveys is the creation of a sort of *semantic chains*, a possible, albeit onerous, method in entity modeling processes that use different, heterogeneous but somehow connectable sources.

Share-VDE Institutions

Share-VDE full members

Share-VDE platform A hosts the data of the full member institutions:

- Duke University www.duke.edu ₽
- Library of Congress www.loc.gov ₽
- National Library of Finland www.kansalliskirjasto.fi/en

- The British Library www.bl.uk
- University of Alberta / NEOS Library Consortium www.ualberta.ca
- University of Chicago www.uchicago.edu №
- University of Michigan Ann Arbor www.umich.edu ₽
- University of Pennsylvania www.upenn.edu ₽
- Yale University level 1, former Cohort member, www.yale.edu

Cohort libraries

The institutions taking advantage of the Share-VDE Linked Data Lifecycle Support and contributing to the Share-VDE Cluster Knowledge Base are:

- Frick Art Reference Library www.frick.org
- Harry Ransom Center www.hrc.utexas.edu ₪
- Harvard University www.harvard.edu №
- National Library of Medicine www.nlm.nih.gov ₽
- Northwestern University www.northwestern.edu ₽
- Princeton University www.princeton.edu
- University of California, Davis www.ucdavis.edu ₽
- University of California, San Diego ucsd.edu ₽
- University of Colorado Boulder www.colorado.edu ₽
- University of Minnesota https://twin-cities.umn.edu/ ₪

Share-VDE Conversion pipeline: from records to entity



Library cooperation: the cluster as the set of many faces



A cluster is an entity, is a prism



From record to entity

Each record has metadata useful to derive properties in order to build entities. But they are hidden.

		@ Where to Beques
000	023	@ <u>where to reques</u>
000	157	700402
005	201	100204124659.0
008	000	162452009 enks b 000 0 eng
010	0.50	la 2000025024
015	_	la GB4959347 12 hnh
016	7	a 015291646 1211k
020	-	la 9780521519748 (hbk.)
020	_	la 0521519748 (hbk.)
020	_	la 9780521732505 (obk.)
020	_	la 0521732506 (pbk.)
035		la (OCol C)ocn416716652
040	_	a DLC d VDXCP d BWKUK d BWK d FDA d UKM d BTCTA d CEF d CDX d DLC
050	00	a PR2803.A2 b H35 2009
082	00	a 822.3/3 2 22
100	1_	a Shakespeare, William, d 1564-1616.
245	10	a As you like it / c edited by Michael Hattaway.
250	_	a Updated ed.
260	_	a Cambridge, UK; a New York: b Cambridge University Press, c 2009.
300		a xv, 240 p. : b ill. ; c 24 cm.
490	1_	a New Cambridge Shakespeare
504	_	a Includes bibliographical references (p. 240).
505	0_	[a Introduction. Journeys ; Plays within the play ; Theatrical genres ; Pastoral ; Counter-pastoral ; The condition of the country ; Politics ; 'Between you and the women the play must please'; Gender ; Nupfials ; Sources ; Date and occassion ; Stage history ; Recent critical and stage interpretations – Note on the text – List of characters – The play – Textual analysis – Appendixes : 1. An early court performance? ; 2. Extracts from Shakespeare's principal source, Lodge's Rosalind ; 3. The songs.
600	10	a Shakespeare, William, d 1564-1616. t As you like it.
650	_0	a Fathers and daughters av Drama.
650	_0	la Exiles Iv Drama.
655	_7	a Pastoral drama. 2 gsafd
655	_7	a Comedies. 2 gsafd
700	1_	la Hattaway, Michael.

Full Record

MARC Tags

The new revolution: from **record** to **entity**



From cataloging to entity modeling

Entity boundary in the Official RDA Toolkit

Glossary: entity boundary: A set of criteria that is applied by an agent who creates metadata to determine if a new entity is being described. *Use for:* boundary

https://access.rdatoolkit.org/en-US rdaregistry.info-termList-RDATerms-1194



When to Create a New Work

(Determining Entity and Transformation Boundaries)

To determine whether or not a new RDA Work is needed, one must first determine whether there are Entity boundaries. <u>Guidance > Entity boundaries</u> in the RDA Toolkit provides an overview of how Entity boundaries are used for various RDA entities. This section of the MGD highlights Entity and Transformation boundaries for Diachronic works. This requires understanding the CONDITIONS and OPTIONS located at <u>Work. Entity boundary</u> and <u>Diachronic works</u>. Transformation boundaries.

https://www.loc.gov/aba/rda/mgd/mg-basicCatalogingDecisions.pdf

Entity Resolution (ER) and Entity modeling



Lewis Carroll => Alice's Adventures in Wonderland



Charles Lutwidge Dodgson => *The game of logic*

L'Entity Resolution (ER)

Entity Resolution: the processes that make it possible to identify the same and unique entity among a thousand different expressions, representations, profiles, identities.

Task of Entity Resolution: identify all mentions representing the same entity within the same knowledge base or in multiple knowledge bases.

Entity Resolution

Profile: the information unit that expresses the identity of a particular and unique entity



L'Entity modeling

In order for a machine to **identify an object**, it must first know its **profile**: it must know what to look for and how it expects to find that object. It must know **how that thing is modeled**.

"look for laptop backpack, size 15", size 45 x 30 cm, in waterproof fabric, blue color"



From concrete to abstract entities

The more concrete the entity, the more the boundaries are definable



.

Pier Paolo Pasolini

- Place of birth: Bologna
- Date of birth: 5 marzo 1922
- Place of death: Ostia
- Date of death: 2 novembre 1975
- **Qualifications:** Italian intellectual, poet, writer, director, screenwriter, actor, playwright, painter, novelist, linguist, translator and essayist
- Works: Ragazzi di vita, Una vita violenta, Donne di Roma, L'odore dell'India, Il sogno di una cosa, Alì dagli occhi azzurri, Teorema La Divina Mimesis, Amado mio...
- URI LOC: <u>http://id.loc.gov/authorities/subjects/sh85114250</u>
- Exact Matching Concepts from Other Schemes: https://lod.nal.usda.gov/nalt/21953
 - Closely Matching Concepts from Other Schemes:
 - Bach (https://d-nb.info/gnd/4140549-3)
 - Fließgewässer (https://d-nb.info/gnd/4017551-0)
 - Fluss (https://d-nb.info/gnd/4131972-2)
 - river (<u>https://www.wikidata.org/wiki/Q4022</u>)
 - many others
- Equivalent: <u>https://thes.bncf.firenze.sbn.it/termine_eng.php?id=7750</u>

How difficult is it to define a "thing"

French:

- *fleuve* => a river that runs into the sea
- *rivière* => a river that runs into a lake or another river.

Swedish - there's three words but the differences are sort of more vague:

- *Älv* => a large river in Sweden, Norway or Finland,
- **flod** => a large river elsewhere,
- $\mathbf{a} =>$ the rest (a smaller stream would be called *bäck*).

Furthermore, when naming rivers, the convention is that **rivers north of Götaälv** (a specific river in Sweden) are named '**älv**' and **rivers to the south** are '**å**' and finally **rivers outside of Scandinavia** are '**flod**'.

*Examples taken from a conversation with Matias Frosterus, Information Systems Manager at the National Library of Finland, around analysis and subject studies in a multilingual and multicultural context, to build Finto (centralized service for interoperable thesauri, ontologies and classification schemes for different subject areas - https://finto.fi/en/) and the general Finnish ontology.

The Web of Data as a decentralised knowledge representation



The owl:sameAs problem

"The built-in OWL property **<u>owl:sameAs</u>** links an individual to an individual. Such an owl:sameAs statement indicates that two URI references actually refer to the same thing: the individuals have the same "identity"*



This interlinking is typically established by asserting that two names denote the same real-world entity



Whenever multiple names are used to denote the same thing, *owl:sameAs* statements are needed in order to link the data and foster reuse

The owl:sameAs problem

<rdf:Description rdf:about="#Alessandro_Pertini"> <owl:sameAs rdf:resource="#Sandro Pertini"/> </rdf:Description>

```
<owl:Class rdf:ID="FootballTeam">
```

```
<owl:sameAs rdf:resource="http://sports.org/US#SoccerTeam"/>
```

</owl:Class>

This statement indicates that **every property asserted to one name will also be inferred to the other**. Hence, allowing both names to be used interchangeably in all contexts.

Entities that are *identical* are considered to share *the same properties* _____ (Indiscernibility of Identicals, Leibniz)

The Share-VDE Cluster Knowledge Base - Numbers*

165,417,388	Total clusters/entities generated from approximately 65,000,000 bibliographic records and approximately 14,000,000 authority records.
	Types of clusters: Person: 10,428,626 Family: 46,369 Organization: 5,443,637 Conference: 675,985 Original work: 30,699,652 Work: 35,702,678 Instance: 46,600,234 Item: 3,491,938 Subject: 17,190,875 Other entities: 15.137.393
32,854,501	owl:sameAs properties present in the Share-VDE CKB

The Share-VDE Cluster Knowledge Base - Numbers

 meeting 403923 	owl:sameAs property, of which:
 organization 1953026 person 9005860 family 32699 opus 1222548 place 153954 occupation 14626 language 516 role 267 agentType 4 form 33 genre 2263 format57 availability 34 SubjectType 15 OpusType 7 	 meeting 403923 organization 1953026 person 9005860 family 32699 opus 1222548 place 153954 occupation 14626 language 516 role 267 agentType 4 form 33 genre 2263 format57 availability 34 SubjectType 15 OpusType 7

Data enrichment in an integrated environment



Library of Congress

- 100 1_ |a Fischer, Henry W. |q (Henry William), |d 1856-1932.
- 245 10 |a Abroad with Mark Twain and Eugene Field : |b tales they told to a fellow correspondent / |c by Henry W. Fisher.
- 260 ____ |a New York : |b Nicholas L. Brown, |c 1922.
- 300 ____ |a 246 p. ; |c 21 cm.
- 500 ____ |a Edited by Merle Johnson.
- 500 ____ |a In addition to the trade ed., there is a limited ed. of 325 numbered copies on Blandford Antique Laid paper.
- 500 ____ |a LC has copy of trade ed. |5 DLC
- 600 10 |a Twain, Mark, |d 1835-1910 |x Travel |z Europe.
- 600 10 |a Field, Eugene, |d 1850-1895 |x Travel |z Europe.
- 650 _0 |a Authors, American |y 19th century |v Biography.
- 700 1_ |a Johnson, Merle De Vore, |d 1874-1935.
- 710 2_ |a Roy J. Friedman Mark Twain Collection (Library of Congress) |5 DLC



Ŕ

NICHOLAS L. BROWN

MCMYXII

NEWYORK

University of Alberta

Language:	German
Authors:	Fischer, Henry W. (Henry William), 1856-1932
Publication Information:	New York, N. L. Brown, 1922.
Publication Date:	1922
Physical Description:	p. cm.
Publication Type:	Book
Document Type:	Non-fiction; Individual biography
Subject Terms:	Field, Eugene, 1850-1895 Twain, Mark, 1835-1910
Notes:	committed to retain 20170930 20421231 HathiTrust I
Other Authors:	Johnson, Merle, 1874-1935

Duke University

Abroad with Mark Twain and Eugene Field; tales they told to a fellow correspondent

by Henry W. Fisher.

Fischer, Henry W. (Henry William), 1856-1932 New York, N. L. Brown, 1922. Book / Print

New York University

Abroad with Mark Twain and Eugene Field : tales they told to a fellow correspondent
Henry W. Fischer (Henry William), 1856-1932. >
Merle De Vore Johnson 1874-1935. >
Twain, Mark, 1835-1910 Travel Europe >
Field, Eugene, 1850-1895 Travel Europe >
Authors, American 19th century Biography >
New York : Nicholas L. Brown
English
1922
246 p. ; 21 cm.

Data enrichment in an integrated environment



ShareVDE BETA 💄 Options 🗸 Search all content... 😫 Henry W. Fischer 🔹 "Abroad with Mark Twain and Eugene Fi... 🕥 💿 "Abroad with Mark Twain and Eugene Field : t...

Publication ①

Abroad with Mark Twain and Eugene correspondent

Volume. Published in English in 1922 in New York and New York b

Contributors

Culture Class Collection (University of Pennsylvania), American Culture Class Collection and Roy J. Friedman Mark Twain Collection (Library of Congress)

Identifier: p26061656924945467

Notes: Edited by Merle Johnson.; Edited by Merle Johnson.; "Editor's note" signed: Me Bibliographic level: Monograph, Monograph Cataloging source: DLC, DLC Descriptive cataloging form: AACR, AACR, Local Dimension: 21 cm. Encoding level: Full, Full, Full not examined Extent: 246 p. ; Record status: Corrected or revised, Corrected or revised Responsibility statement: by Henry W. Fisher., by Henry W. Fisher. System control number: 00271020, ocn739054015, (OCoLC)271020

Available at 4 libraries 🗸

More options ∨





The subject management in Share-VDE - Examples

https://svde.org/dates/3923



The subject in Share: the string and the concept (something more than a term)

Sistemi pre-coordinati

Carroll, Lewis -- Daresbury -- Alice in Wonderland -- 1865

https://svde.org/subjects/897429874

Riconciliabile con LCSH (e simili)

Sistemi post-coordinati

Carroll, Lewis *https://svde.org/agents/202*

Daresbury https://svde.org/places/829

Alice in Wonderland https://svde.org/opuses/99392

1865 https://svde.org/dates/3923

Riconciliabili con FAST, Wikidata (e simili)

Subjects in Wikidata - The concept as a bridge to other sources



The Share-VDE Cluster Knowledge Base - Numbers*

17,997,199	Total number of subject string clusters (distinct)
6,428,599	Total number of concepts extracted from subject strings (distinct)

Subjects Road Map - What has been done

\checkmark	Data enrichment with subjects from other sources (libraries that did not originally have the subjects benefit from this enrichment)
\checkmark	attribution of URI to each subject string
\checkmark	extraction from each string of the "concepts" that constitute it, and attribution of URI to each concept
\checkmark	attribution of multiple Provenances to the same concept
\checkmark	Interlinking with external sources (LCSH, LCGFT e FAST)

Subjects Road Map - What hasn't been done (yet).

×	concept clustering: we do not enrich the concept with elements from different sources, so we do not have the concept of the cluster understood as an entity with (for example) equivalent labels in different languages
×	interlinking between concepts (linking the concept to external sources such as Wikidata or FAST)
X	machine learning processes

Subjects Road Map - Work in progress

	expansion of resource search possibilities by extracting the relationships present in the subject strings or in the CKB
	concepts clustering using external sources, such as Wikidata
Ó	more effective data presentation, including presentation of all Provenances that have contributed to a concept

• Place Selma, Alabama (USA)

City in and county seat of Dallas County, Alabama, United States

"Selma is a city in and the county seat of Dallas County, in the Black Belt region of south central Alabama and extending to the west. Located on the banks of the Alabama River, the city has a population of 17,971 as of the 2020 census. About 80% of the population is African-American." — <u>Wikipedia</u>



Wikimedia

Options 🗸

Original works about this place		Events that took place here		8 Related people and organization	ions () Publications	O Publications published in the second se	
Filter. 47 re :	 sults	Q Year	Creator Genre	Sort by: Title A	Z ♥		
No.	Image	Concept		Туре	Description	Location	0
1	教社	Selma to Montgomery	marches	Demonstration	1965 march for African-American citizens to exercise their constitutional right to vote	💽 Selma, Alabama (USA)	
2		Sattle of Selma		Battle	Battle of the American Civil War	🔕 Selma, Alabama (USA)	

Search all content...

Related topics and events

• Person Martin Luther King Jr.

O Publications by

American civil-rights activist and leader (1929–1968). Born in Atlanta (Ga.). Died in Memphis (Tenn.).

"Martin Luther King Jr. was an American Baptist minister and activits, one of the most prominent leaders in the civil rights movement from 1955 until his assassination in 1968. An African American church leader and the son of early civil rights activist and minister Martin Luther King Gr., King advanced civil rights for people of color in the United States..." — <u>Wikipedia</u>

Options ~

Original works by



5 3

Q

Q Sort by: Title A-Z 🗸 Filter... Туре Location 47 results Ô No. Image Concept Туре Description Location O COINTELPRO Government Series of covert, and often illegal, projects conducted by the program U.S. FBI and the second Assassination of Martin Luther King Jr. Political murder 1968 murder of Martin Luther Memphis, Texas (USA) King Jr. 1 Martin Luther King Jr. Day Topic United States federal holiday O USA Suneral of Martin Luther King, Jr. Topic Funeral following the assassination of Martin Luther King Jr. 5 O I've Been to the Mountaintop Speech Popular name of the last speech delivered by Martin Luther King, Jr. O I Have a Dream 6 Speech 1963 speech delivered by Martin Luther King Jr. O Speech on the Assassination of Martin Speech "Ladies and Gentlemen, I'm only Luther King going to talk to you just for a minute or so this evening ... "

Original works about

6 O Civil rights Topic 7 O African Americans. Topic

Show more

Civil rights movement

1954-1968 U.S. social movement against institutional racism

"The civil rights movement was a political movement and campaign from 1954 to 1968 in the United States to abolish

institutional racial segregation, discrimination, and disenfranchisement throughout the United States. The movement had its

origins in the Reconstruction era during the late 19th century, although it made its largest legislative gains..." - Wikipedia

O Topic

Options ~

Search all content...

B Related people and organizations

O Topic

Civil rights movement

1954–1968 U.S. social movement against institutional racism

"The civil rights movement was a political movement and campaign from 1954 to 1968 in the United States to abolish institutional racial segregation, discrimination, and disenfranchisement throughout the United States. The movement had its origins in the Resconstruction en a during the late 19th centry, although it made its largest legislative gains." <u>— Witkingfit</u>

O Topics and events related to this

Options 🗸

Filter...

47 results

Works about this

Wikimedia



Wikimedia



Q Type Location Sort by: Title A-Z ~

	Juiceo				
No.	Image	Concept	Туре	Description	Location O
1		Selma to Montgomery marches	Demonstration	1965 march for African-American citizens to exercise their constitutional right to vote	 Selma, Alabama (USA) Montgomery, Alabama (USA)
2	2	O I Have a Dream	Speech	1963 speech delivered by Martin Luther King Jr.	Lincoln Memorial, Washington, D.C. (USA)
3		Ole Miss riot of 1962	Riot	Dispute over integration at University of Mississippi	Oxford, Mississippi (USA)
4		O Civil rights movement in South Carolina	Social movement	1950–1970 social movement in South Carolina, United States	South Carolina (USA)
5		Civil rights movement in Omaha, Nebraska	Demonstration	1965 march for African-American citizens to exercise their constitutional right to vote	
6		Council of Federated Organizations	Coalition	Coalition of advocacy groups in the Civil Rights Movement	
7		• Freedom Summer	Political campaign	1964 volunteer campaign in the Mississippi	
8		Chicago Freedom Movement	Торіс	Movement led by Reverend Martin Luther King, Jr., James Bevel and Al Raby	Chicago, Illinois (USA)

🕷 End of list

Grazie

Tiziana Possemato <u>tiziana.possemato@atcult.it</u> <u>tiziana.possemato@casalini.it</u>

Share Family community