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, even 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 hosts the data of the full member institutions:

- Duke University [www.duke.edu](http://www.duke.edu)
- Library of Congress [www.loc.gov](http://www.loc.gov)
- New York University [www.nyu.edu](http://www.nyu.edu)
- Smithsonian Institution [www.si.edu](http://www.si.edu)
- Stanford University [www.stanford.edu](http://www.stanford.edu)
- The British Library [www.bl.uk](http://www.bl.uk)
- University of Alberta / NEOS Library Consortium [www.ualberta.ca](http://www.ualberta.ca)
- University of Chicago [www.uchicago.edu](http://www.uchicago.edu)
- University of Michigan Ann Arbor [www.umich.edu](http://www.umich.edu)
- University of Pennsylvania [www.upenn.edu](http://www.upenn.edu)
- Vanderbilt University [www.vanderbilt.edu](http://www.vanderbilt.edu)
- Yale University - level 1, former Cohort member, [www.yale.edu](http://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:

- Cornell University [www.cornell.edu](http://www.cornell.edu)
- Frick Art Reference Library [www.frick.org](http://www.frick.org)
- Harry Ransom Center [www.hrc.utexas.edu](http://www.hrc.utexas.edu)
- Harvard University [www.harvard.edu](http://www.harvard.edu)
- Northwestern University [www.northwestern.edu](http://www.northwestern.edu)
- Princeton University [www.princeton.edu](http://www.princeton.edu)
- Texas A&M University [www.tamu.edu](http://www.tamu.edu)
- University of California, Davis [www.ucdavis.edu](http://www.ucdavis.edu)
- University of California, San Diego [ucsd.edu](http://ucsd.edu)
- University of Colorado Boulder [www.colorado.edu](http://www.colorado.edu)
- University of Minnesota [https://twin-cities.umn.edu](https://twin-cities.umn.edu)
- University of Washington [https://www.washington.edu](https://www.washington.edu)
Share-VDE Conversion pipeline: from records to entity

Entity Resolution / Clusterization tool

External sources

Auth/Bib records

The entity as a cluster

Share-VDE members

Share Entity Model

Auth/Bib records

Retro-conversion from entity to Marc (enriched)

Cluster Knowledge Base (CKB)
Library cooperation: the cluster as the set of many faces

Sources

P1

P2

P3

...

Pn

Data

Cluster Knowledge Base

Virtual Discovery Environment
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.
The new revolution: from record to entity

Shakespeare, William, 1564-1616
Шекспир, У. 1564-1616 Уильям
Saixpēr, Gouilliam, 1564-1616

As you like it
Come ti piace
Comme il vous plaira

Fathers and daughters
Padri e figlie
Pères et filles

Cambridge University Press
Cambridge Press
Cambridge Univ. Press
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


---

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. Guidance > Entity boundaries 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 Work, Entity boundary and Diachronic works, Transformation boundaries.

https://www.loc.gov/aba/rda/sgd/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.
Profile: the information unit that expresses the identity of a particular and unique entity

Pier Paolo Pasolini

Pasolini director

Pasolini writer
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

- **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, Ali dagli occhi azzurri, Teorema La Divina Mimesis, Amado mio...

- **URI LOC**: [http://id.loc.gov/authorities/subjects/sh85114250](http://id.loc.gov/authorities/subjects/sh85114250)
- **Exact Matching Concepts from Other Schemes**: [https://lod.nal.usda.gov/nalt/21953](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](https://www.wikidata.org/wiki/Q4022)
  - many others
- **Equivalent**: [https://thes.bncf.firenze.sbn.it/termine_eng.php?id=7750](https://thes.bncf.firenze.sbn.it/termine_eng.php?id=7750)
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,
- *å* => the rest (a smaller stream would be called *bäck*).

Furthermore, in Finnish culture, 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

Source: https://www.ontotext.com/knowledgehub/webinars/build-narratives-connect-artifacts-cultural-heritage/
The *owl:sameAs* problem

“The built-in OWL property *owl:sameAs* 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".*

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

*https://www.w3.org/TR/owl-ref/#sameAs-def*
The *owl:sameAs* problem

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*

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

*Conteggio alla data del 02/02/2023*
## The Share-VDE Cluster Knowledge Base - Numbers

<table>
<thead>
<tr>
<th>Total number: 12,789,833</th>
<th>Distinct clusters that have at least one <code>owl:sameAs</code> property, of which:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>● meeting 403923</td>
</tr>
<tr>
<td></td>
<td>● organization 1953026</td>
</tr>
<tr>
<td></td>
<td>● person 9005860</td>
</tr>
<tr>
<td></td>
<td>● family 32699</td>
</tr>
<tr>
<td></td>
<td>● opus 1222548</td>
</tr>
<tr>
<td></td>
<td>● place 153954</td>
</tr>
<tr>
<td></td>
<td>● occupation 14626</td>
</tr>
<tr>
<td></td>
<td>● language 516</td>
</tr>
<tr>
<td></td>
<td>● role 267</td>
</tr>
<tr>
<td></td>
<td>● agentType 4</td>
</tr>
<tr>
<td></td>
<td>● form 33</td>
</tr>
<tr>
<td></td>
<td>● genre 2263</td>
</tr>
<tr>
<td></td>
<td>● format 57</td>
</tr>
<tr>
<td></td>
<td>● availability 34</td>
</tr>
<tr>
<td></td>
<td>● SubjectType 15</td>
</tr>
<tr>
<td></td>
<td>● OpusType 7</td>
</tr>
<tr>
<td></td>
<td>● PublicationType 1</td>
</tr>
</tbody>
</table>
Data enrichment in an integrated environment
**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

Book / Print

---

**New York University**

<table>
<thead>
<tr>
<th>Title</th>
<th>Abroad with Mark Twain and Eugene Field; tales they told to a fellow correspondent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>Twain, Mark, 1835-1910 -- Travel -- Europe  &gt; Field, Eugene, 1850-1895 -- Travel -- Europe  &gt; Authors, American -- 19th century -- Biography  &gt;</td>
</tr>
<tr>
<td>Publisher</td>
<td>New York : Nicholas L. Brown</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Publication Date</td>
<td>1922</td>
</tr>
</tbody>
</table>

**University of Alberta**

| Language | German |
| Authors | Fischer, Henry W. (Henry William), 1856-1932 |
| 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 |
| Other Authors | Johnson, Merle, 1874-1935 |
Data enrichment in an integrated environment
Abroad with Mark Twain and Eugene Field: a correspondent

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

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

Identifier: p26066566924945487
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
System control number: 00271020, ocn73954015, (OCoLC)271020

<table>
<thead>
<tr>
<th>No.</th>
<th>Subject</th>
<th>Concepts</th>
<th>Subject provenance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PRO Rischbieth, Robert Nevin, family of (donor) (Culture Class Collection copy) (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>2</td>
<td>CHR 1922 (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>3</td>
<td>Authors, American - Biography - 19th century (Topical subject)</td>
<td>Authors, American (Topic) Biography (Genre) 19th century (Date)</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>4</td>
<td>PRO Rascoe, Burton, 1892-1957 (inscription) (Culture Class Collection copy (AC9 F5237 922a)) (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>5</td>
<td>PRO Fischer, Henry W. (Henry William), 1856-1932 (autograph) (Culture Class Collection copy (AC9 F5237 922a)) (Topical subject)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>6</td>
<td>Authors, American - Biography - 19th century (Topical subject)</td>
<td>Authors, American (Topic) Biography (Genre) 19th century (Date)</td>
<td>University of Alberta</td>
</tr>
<tr>
<td>7</td>
<td>Mark Twain (Person)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>8</td>
<td>Twain, Mark, 1835-1910 - Travel - Europe (Name subject)</td>
<td>Mark Twain (Person) Travel (Topic) Europe (Place)</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>9</td>
<td>Field, Eugene, 1850-1895 - Travel - Europe (Name subject)</td>
<td>Eugene Field (Person) Travel (Topic) Europe (Place)</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>10</td>
<td>Eugene Field (Person)</td>
<td></td>
<td>University of Pennsylvania</td>
</tr>
</tbody>
</table>
The subject management in Share-VDE
The subject management in Share-VDE - Examples

W1
- The adventures of Alice in Wonderland -- 1865

W2
- The Secret of Hegel -- 1865

W2
- Carroll, Lewis -- Daresbury -- Alice in Wonderland -- 1865

https://svde.org/dates/3923
https://svde.org/opuses/99392
https://svde.org/agents/202
https://svde.org/places/829
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](https://svde.org/agents/202)

Daresbury
[https://svde.org/places/829](https://svde.org/places/829)

Alice in Wonderland
[https://svde.org/opuses/99392](https://svde.org/opuses/99392)

1865
[https://svde.org/dates/3923](https://svde.org/dates/3923)

Riconciliabili con **FAST, Wikidata** (e simili)
Subjects in Wikidata - The concept as a bridge to other sources

I1

The adventures of Alice in Wonderland

https://www.wikidata.org/wiki/Q92640

1865

https://www.wikidata.org/wiki/Q7708

I2

The Secret of Hegel

https://www.wikidata.org/wiki/Q7763139

1865

https://www.wikidata.org/wiki/Q7708
The Share-VDE Cluster Knowledge Base - Numbers*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17,997,199</td>
<td>Total number of <strong>subject string</strong> clusters (distinct)</td>
</tr>
<tr>
<td>6,428,599</td>
<td>Total number of <strong>concepts</strong> extracted from subject strings (distinct)</td>
</tr>
</tbody>
</table>

*Conteggio alla data del 02/02/2023
| ✔️ | Data enrichment with subjects from other sources (libraries that did not originally have the subjects benefit from this enrichment) |
| ✔️ | attribution of URI to each subject string |
| ✔️ | extraction from each string of the "concepts" that constitute it, and attribution of URI to each concept |
| ✔️ | attribution of multiple Provenances to the same concept |
| ✔️ | 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) |
| ❌ | machine learning processes |
| 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 |
| more effective data presentation, including presentation of all Provenances that have contributed to a concept |
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." — Wikipedia

Options ▾

Original works about this place
Events that took place here
Related people and organizations
Publications published in

Filter... Year Creator Genre Sort by: Title A-Z

47 results

<table>
<thead>
<tr>
<th>No.</th>
<th>Image</th>
<th>Concept</th>
<th>Type</th>
<th>Description</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Selma to Montgomery marches</td>
<td>Demonstration</td>
<td>1965 march for African-American citizens to exercise their constitutional right to vote</td>
<td>Selma, Alabama (USA)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>Battle of Selma</td>
<td>Battle</td>
<td>Battle of the American Civil War</td>
<td>Selma, Alabama (USA)</td>
</tr>
</tbody>
</table>
Martin Luther King Jr.

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 activist, 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 central figure in the modern civil rights movement in the United States. — [Wikipedia]"
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

Works about this

Topics and events related to this

Related people and organizations

Filter:

Sort by: Title A-Z

341 results

1. "Sibley to Montgomery marchers" Video recordings
   - Description: 1965 march for African-American citizens to exercise their constitutional right to vote
   - Location: Selma, Alabama (USA)

2. "I Have a Dream" Speech
   - Description: 1963 speech delivered by Martin Luther King Jr.
   - Location: Lincoln Memorial, Washington, D.C. (USA)

3. "Old Man's Race" Poem
   - Description: Dispute over integration at University of Mississippi
   - Location: Oxford, Mississippi (USA)

4. "Civil rights movement in South Carolina" Social movement
   - Description: 1956-1970 social movement in South Carolina, United States
   - Location: South Carolina (USA)

5. "Civil rights movement in Oregon, Nebraska" Demonstration
   - Description: 1965 march for African-American citizens to exercise their constitutional right to vote

6. "Council of Federated Organizations" Coalition
   - Description: Coalition of advocacy groups in the Civil Rights Movement

7. Freedom Summer
   - Description: 1964 volunteer campaign in the Mississippi

8. Chicago Freedom Movement
   - Description: Movement led by Reverend Martin Luther King Jr., James Bevel, and Al Raby

End of list
Grazie

Tiziana Possemato
tiziana.possemato@atcult.it
tiziana.possemato@casalini.it

Share Family community