Cataloging & Linked Data

From Strings to Things

Michael Panzer

Assistant Editor, DDC
OCLC

@MichaelPanzer

Topics for today

- Cataloging from a linked data perspective
  - Describing vs. identifying
  - Mapping vs. linking
- *Interlude: Technical basis for linked data*
- Library Linked Data: 2 Examples
  - Dewey.info
  - WorldCat.org + schema.org
- Why should librarians care?

Define “metadata”

Implications for descriptive cataloging

- produces surrogates
- transforms textual information

“Metadata is data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics.”

Dempsey/Heery (1997)

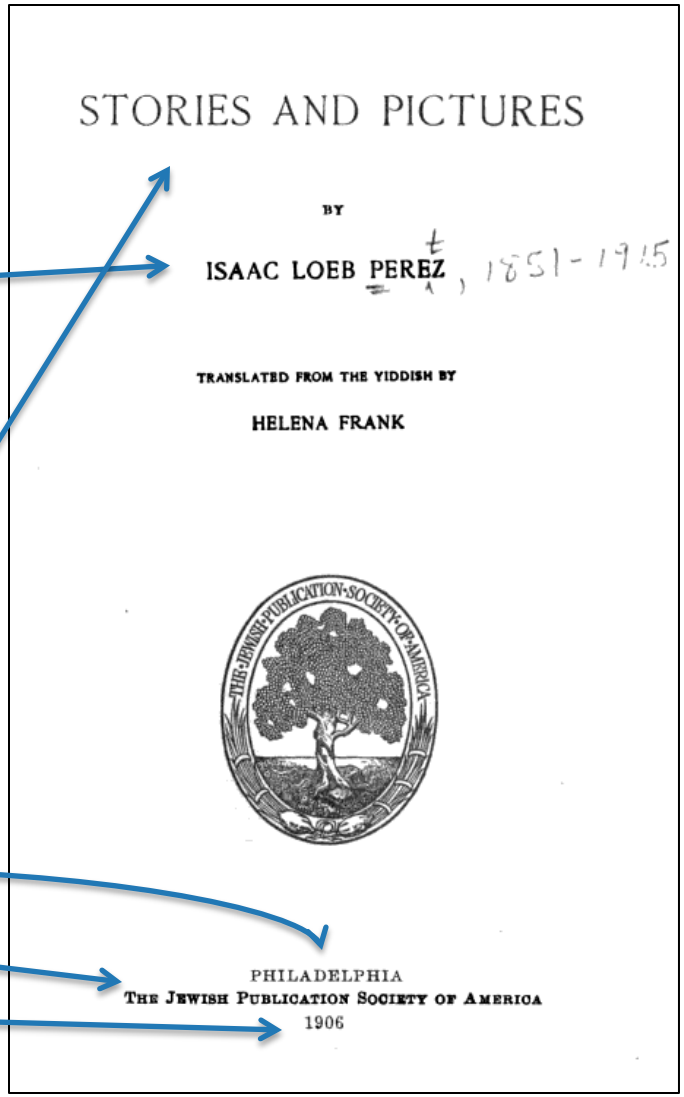
Cataloging as string transformation

100 1# $a Peretz, Isaac Leib,
$d 1851 or 2-1915.

240 10 $a Short stories. $l English.
$k Selections

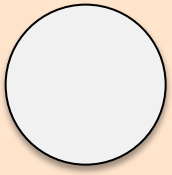
245 10 $a Stories and pictures / $c by Isaac Loeb Perez ; translated from the Yiddish by Helena Frank.

260 ## $a Philadelphia : $b Jewish Publications Society of America, 1906.



Strings bundled as records

100 1# $a Peretz, Isaac Leib, $d 1851 or 2-1915.
240 10 $a Short stories. $l English. $k Selections
245 10 $a Stories and pictures / $c by Isaac Loeb Perez ; translated from the Yiddish by Helena Frank.
260 ## $a Philadelphia : $b Jewish Publications Society of America, 1906.



Cataloging as string transformation still operates (cum grano salis) within the confines of a card catalog

- Focus on record as “end product”
- Record as mark-up of text
- But: transformations can be extremely sophisticated!

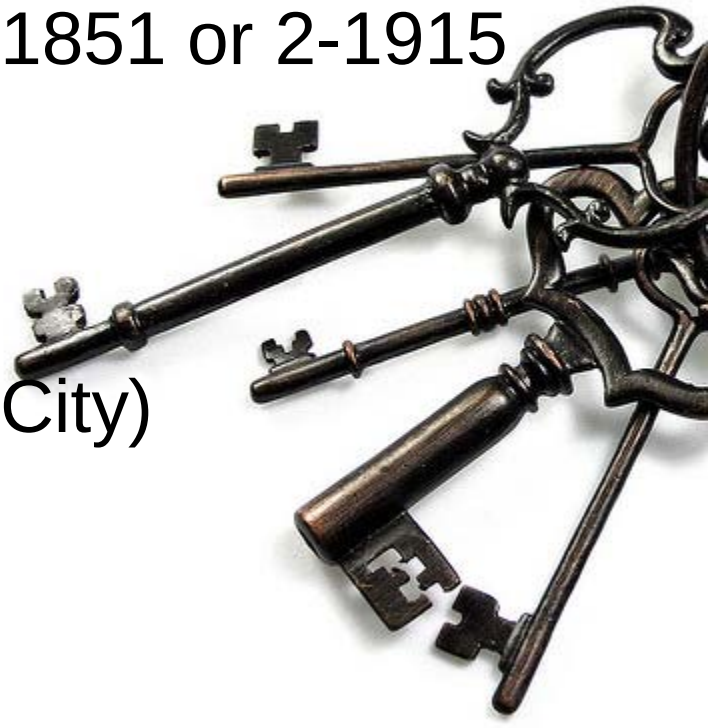
Strings as keys

Personal name:

100 1# $a Peretz, Isaac Leib, $d 1851 or 2-1915

Geographic heading:

151 ## $a Grand Canyon (Ariz. : City)

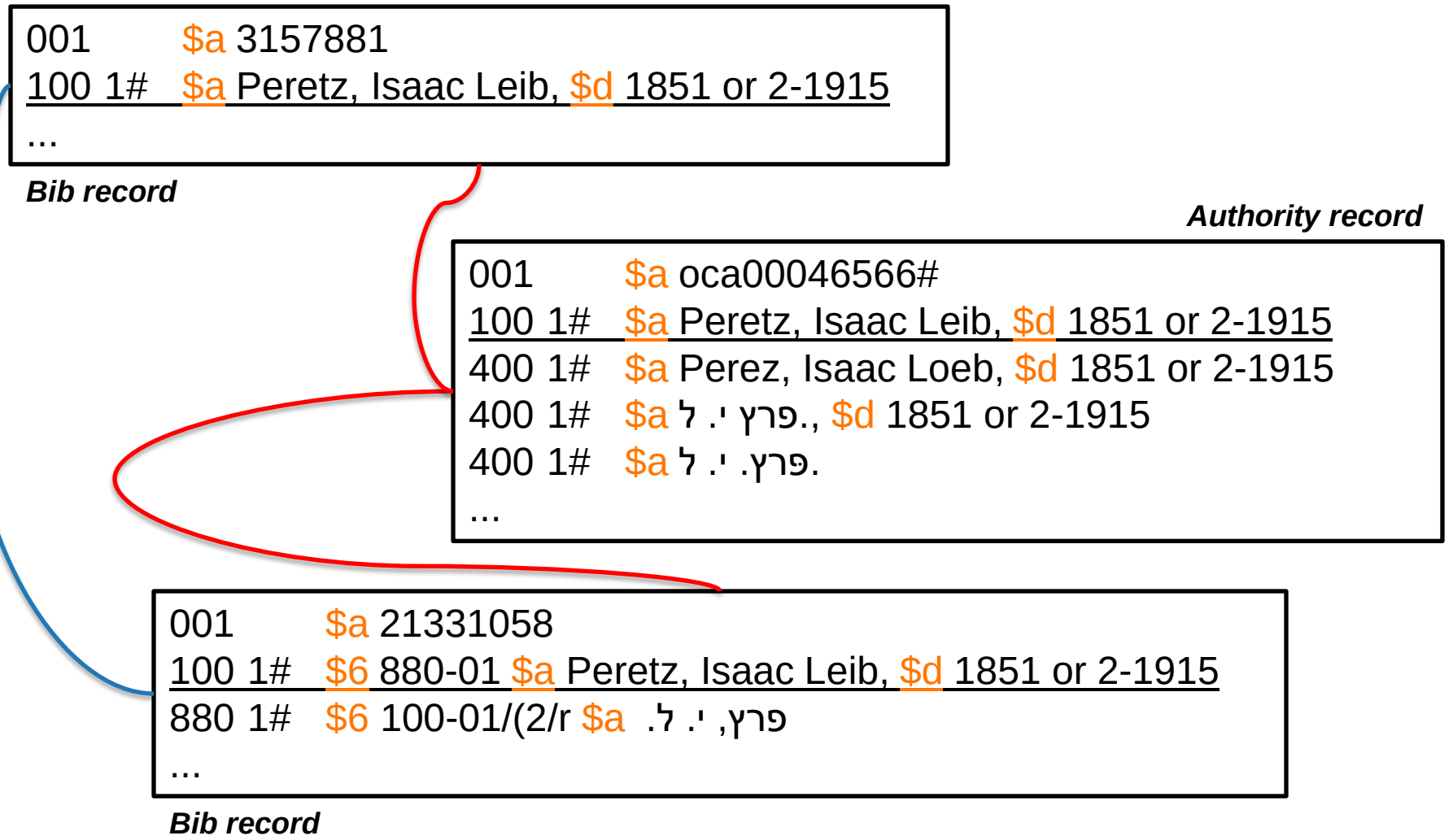


Identification by description

- Heading construction based on elimination of ambiguities
  - Parts are added to increase specificity of description to avoid conflicts with other headings
- Use of abbreviations, punctuation, special delimiters
- Form of heading carries meaning (dates, etc.)
- Headings relate to other headings
- Keys act as access points for *records*, not *described entities*

≈ cutting a key based on a vague description of a lock that may not even exist

Clustering of records by co-occurrence of complex text strings



Links as keys

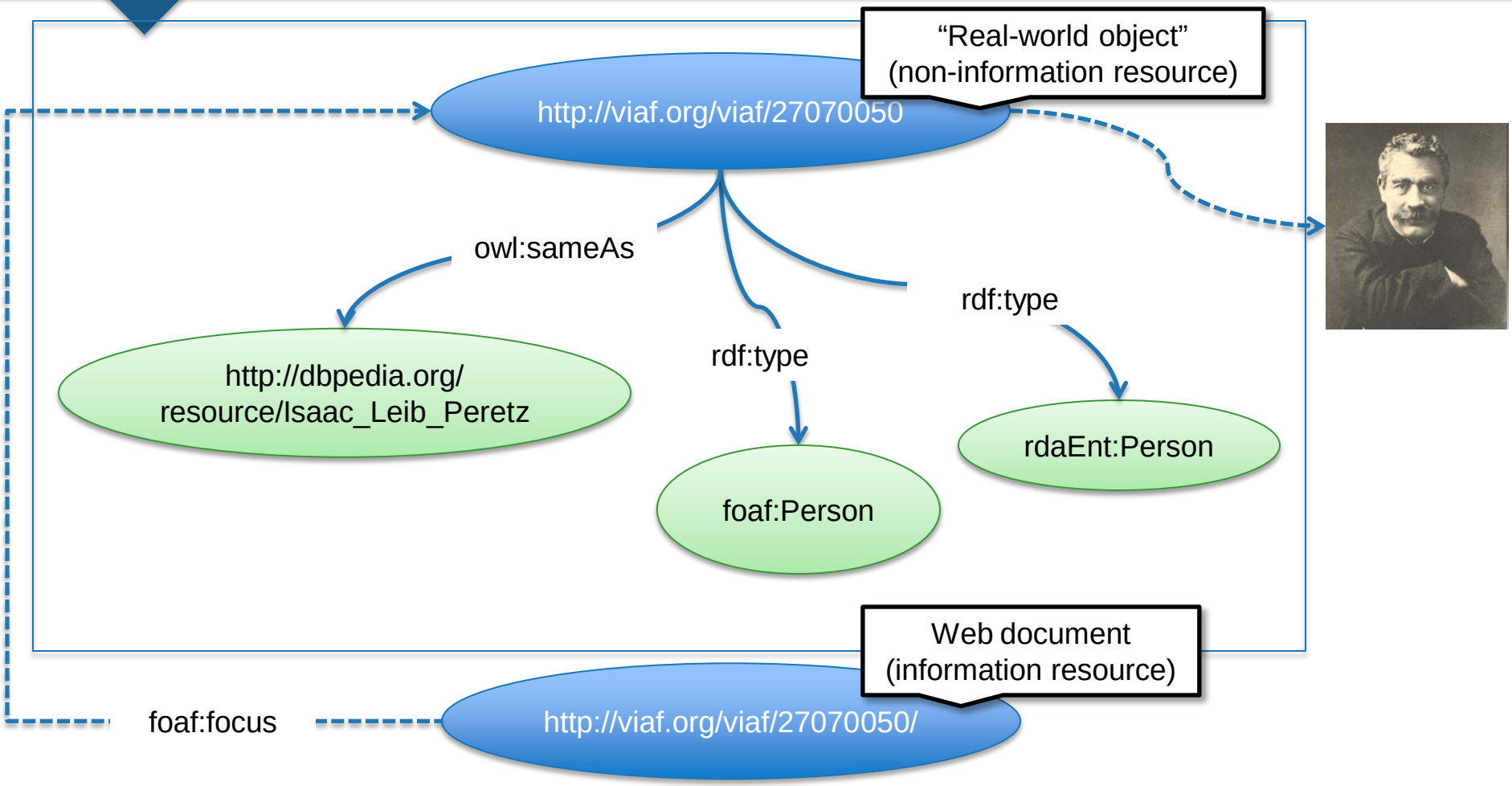
Personal name:

http://viaf.org/viaf/27070050

Geographic heading:

http://sws.geonames.org/5296391/

Identification by denotation

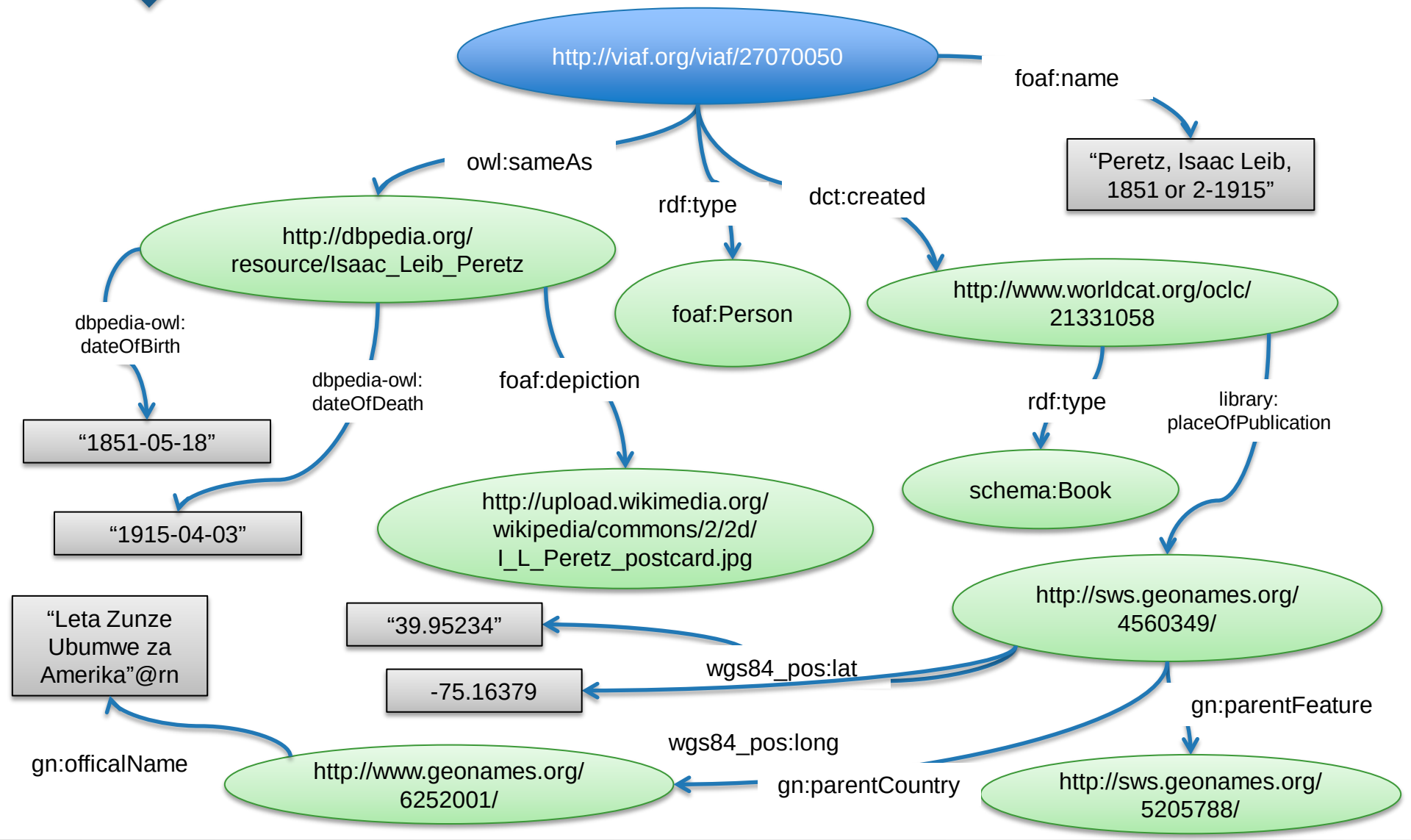


Identification by denotation

- Identifier acquired from defining data source
- Actionable/verifiable by dereferencing through web protocols
- Opaque
- Link directly to things
- Allow targeted knowledge acquisition across domain boundaries

≈ Key is inseparably connected to an entity (person, place, object, concept, …) and unlocks further knowledge

Exploring the knowledge graph



Metadata as textual data (aka “the well-wrought urn”)

- “Record” as paradigm
- Emphasis on creation of “literal” descriptions
- Structured, but flat
- Schema-dependent
- Assembled from strings
- Collection of complex values with meaning also carried by syntax and relative position

Metadata as linked data (aka “the deconstructed urn”)

- Graph as paradigm
- Emphasis on creation of properties between entities
- Structured, but open-ended
- Schema-free
- Assembled from building blocks
- Collection of knowledge statements

I‘m confused; *this* is linked data?

Technical Interlude

Linked data principles: Rules, stars, and bags of chips

Four rules (Tim Berners-Lee):

1. Use URIs for things (identification)
2. Use http-URIs so people can look up those names (dereferenceability)
3. When someone looks up a URI, provide useful information, using the standards
4. Include links to other URIs, so that they can discover more things (relationships)

Let‘s explore the first rule for a second

- As seen, there is an important distinction between identifiers for things and for descriptions of things
- URLs for things that have no web representation
- Reusing web infrastructure to resolve URL for non-information resources
- Web of data overlays web of things
- Big idea of early semantic web; linked data approach more prudent

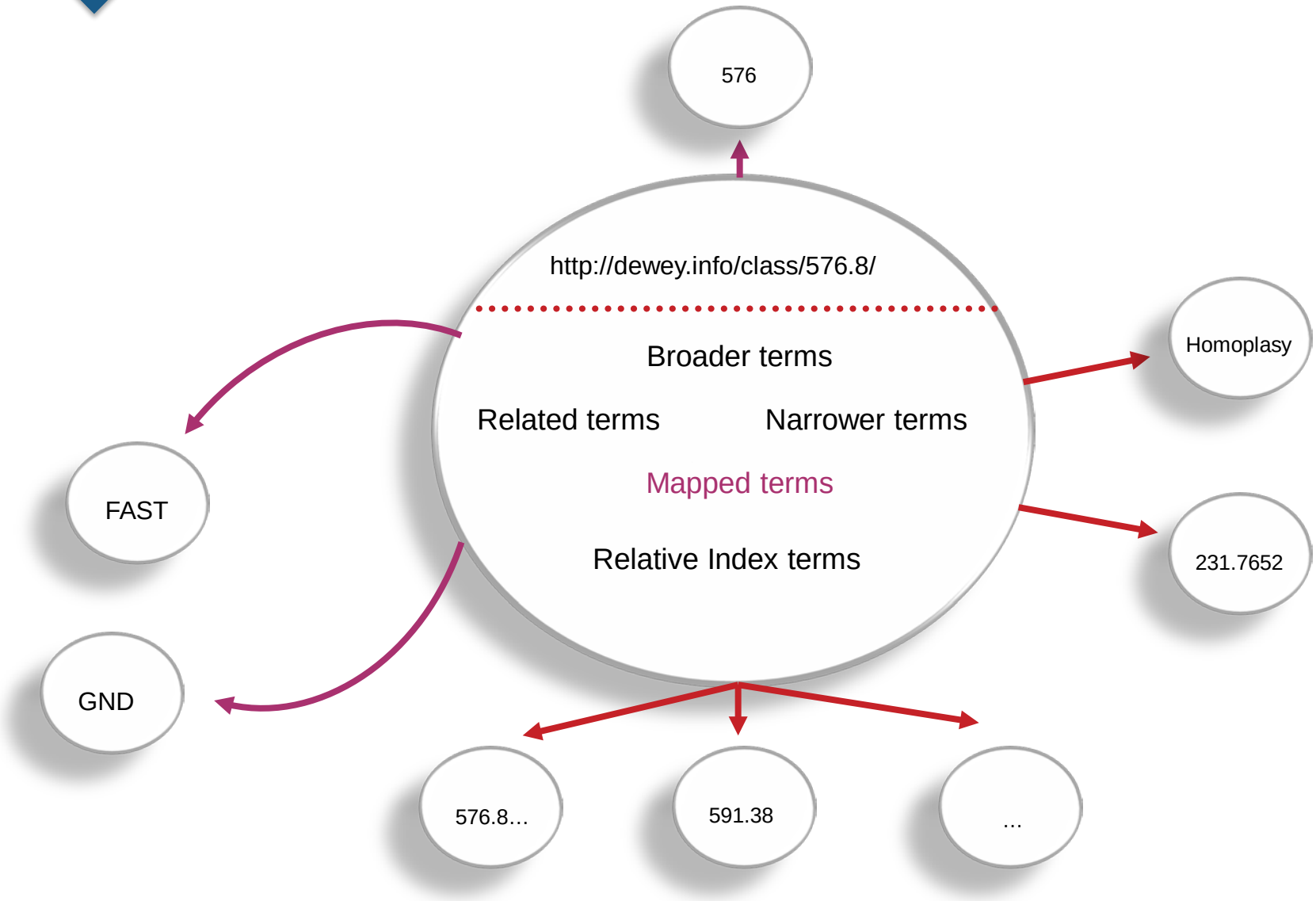
Examples for publishing library linked data

DDC 23 as linked data

Dewey URIs: Sample URIs

- http://dewey.info/class/641/
- http://dewey.info/class/641/about.rdf
- http://dewey.info/class/641/about.en
- http://dewey.info/class/641/about.it
- http://dewey.info/class/641.5/a14/
- http://dewey.info/class/641/e23/about.en.rdf
- http://dewey.info/class/641.23/e23/2012-06-14/about.en

Literals to Links



Dewey.info service (2009) (html view)

af

940 Geskiedenis van Europa

945 | Italiaanse Skiereiland & aangrensende eilande

http://dewey.info/scheme/e22/

ar

940 تاريخ أوروبا.

945 | شبه الجزيرة الإيطالية والجزر المجاورة.

http://dewey.info/scheme/e22/

de

940 Geschichte Europas

945 | Geschichte der italienische Halbinsel und benachbarter Inseln

http://dewey.info/scheme/e22/

en

940 History of Europe

945 | Italian Peninsula & adjacent islands

http://dewey.info/scheme/e22/

Dewey.info service (2009) (RDF/SKOS view)

```
- <rdf:RDF>
  - <rdf:Description rdf:about="http://dewey.info/class/945/2009/08/about.en">
      <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
      <xhv:license rdf:resource="http://creativecommons.org/licenses/by-nc-nd/3.0/"/>
      <cc:attributionName>OCLC Online Computer Library Center, Inc.</cc:attributionName>
      <cc:attributionURL rdf:resource="http://www.oclc.org/dewey/"/>
      <dct:isVersionOf rdf:resource="http://dewey.info/class/945/"/>
      <dct:language rdf:datatype="http://purl.org/dc/terms/RFC4646">en</dct:language>
      <skos:notation rdf:datatype="http://dewey.info/schema-terms/Notation">945</skos:notation>
      <skos:inScheme rdf:resource="http://dewey.info/scheme/2009/08/about.en"/>
      <skos:prefLabel xml:lang="en">Italian Peninsula & adjacent islands</skos:prefLabel>
      <skos:broader rdf:resource="http://dewey.info/class/94/2009/08/about.en"/>
      <cc:morePermissions rdf:resource="http://www.oclc.org/dewey/about/licensing/"/>
    </rdf:Description>
  </rdf:RDF>
```

Dewey.info service (2010)

http://dewey.info/scheme/a14/

945 Italian Peninsula and adjacent islands Italy

945.09	Italy since 1900
945.4	San Marino
945.6	Vatican City
945.8	Malta, . . .

http://dewey.info/scheme/a14/

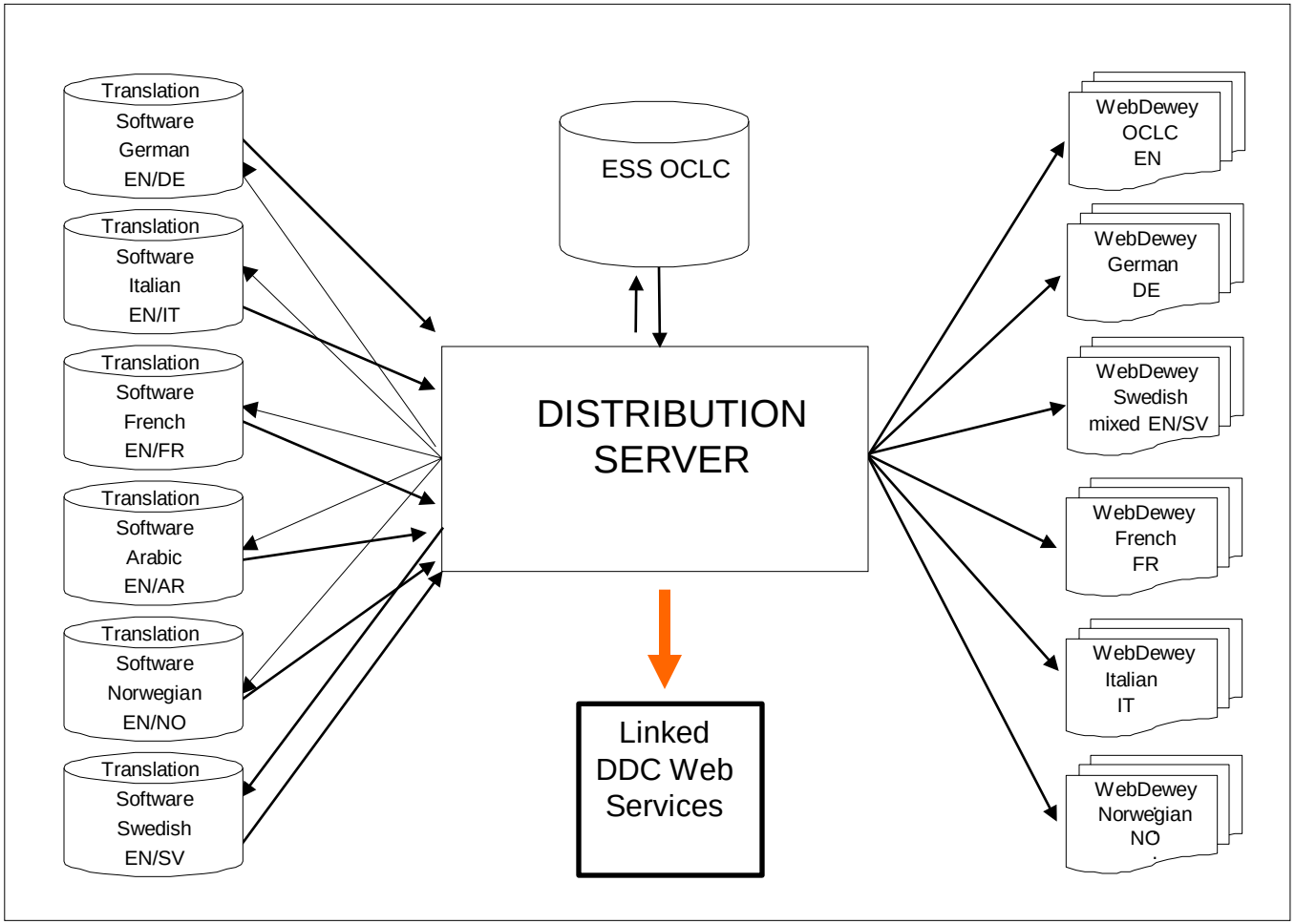
945 | Penisola italiana e isole adiacenti Italia

945.01	Primo periodo, fino al 774
945.02	Periodo della dominazione carolingia (franca), 774–962
945.03	Periodo degl'imperatori germanici, 962–1122
945.04	Periodo comunale, 1122–1348
945.05	Periodo dalla peste nera alla discesa di Carlo VIII, 1348–1494
945.06	Periodo delle lotte per l'egemonia tra Francia e Spagna, 1494–1527
945.07	Periodo della dominazione spagnola e austriaca, 1527–1796
945.08	1796–1900
945.09	1900–
945.1	Italia nordoccidentale Regione del Piemonte
945.2	Regione della Lombardia
945.3	Italia nordorientale Regione del Veneto
945.4	Regione dell'Emilia-Romagna e San Marino
945.5	Regione della Toscana
945.6	Italia centrale e Città del Vaticano
945.7	Italia meridionale
945.8	Regione della Sicilia e isole adiacenti
945.9	Regione della Sardegna

Dewey Linked Data Timeline

2009	2010	2011	2012
dewey.info launched with DDC Summaries in 11 languages	Abridged Edition 14 in 3 languages	12th language added to DDC Summaries in dewey.info	**DDC 23 in dewey.info**
	New data format and distribution environment introduced for Dewey data		**New data format and distribution environment extended to dewey.info**
	British Library DNB	FAO DDC in LOD Cloud	

Dewey Distribution Environment



Dewey.info service (2012)

it

http://dewey.info/scheme/a14/

641 Cibi e bevande

641.5 | Cucina

641.502 Miscellanea
641.508 Cucina in riferimento a categorie di persone
641.509 Storia, geografia, persone
641.59 Cucina tipica di specifici ambienti geografici, cucina etnica

en

http://dewey.info/scheme/edition/e23/

641 Food and drink

641.5 | Cooking

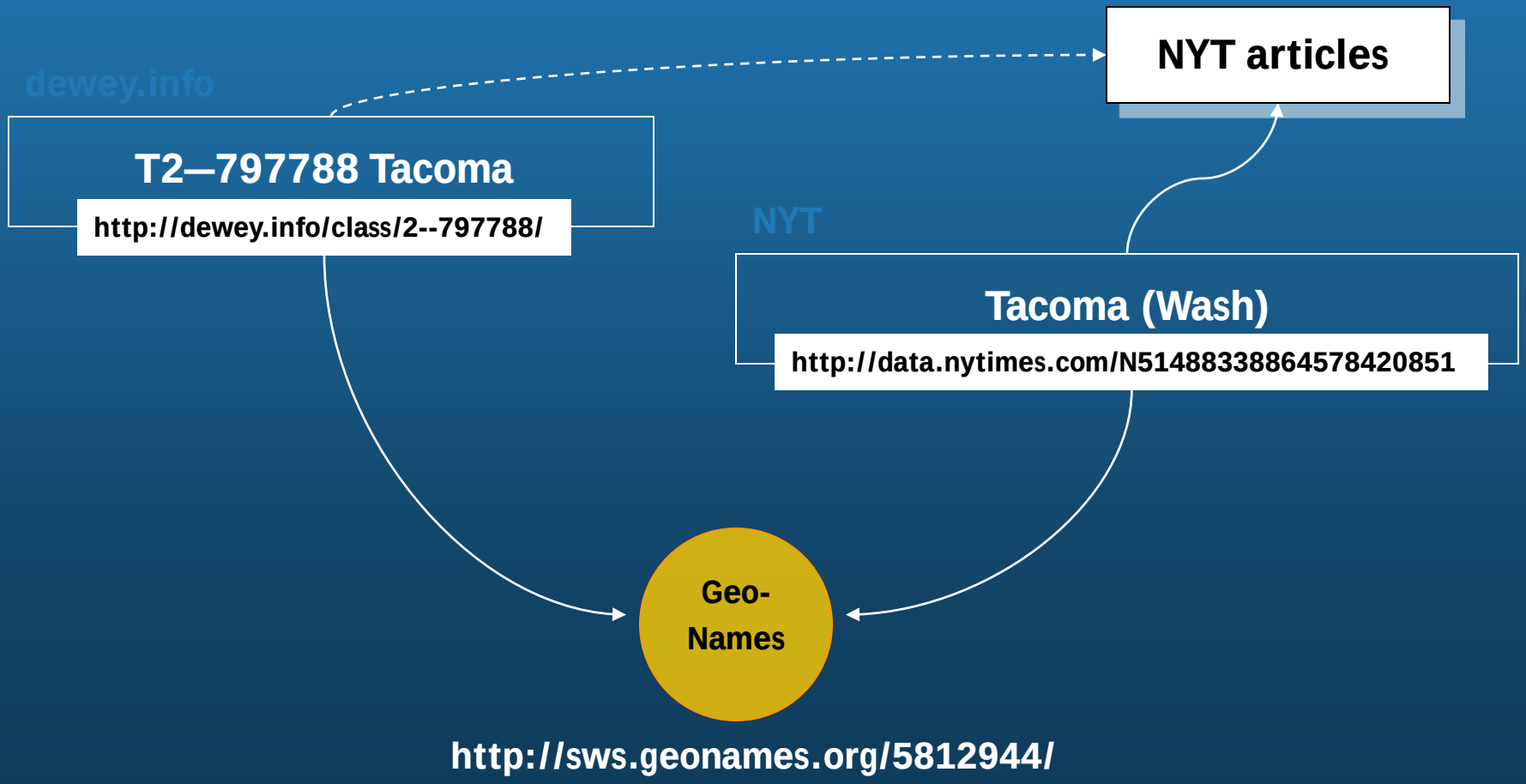
641.502 Miscellany
641.508 Cooking with respect to groups of people
641.509 History, geographic treatment, biography
641.51 Beginner and gourmet cooking
641.55 Money-saving and timesaving cooking
641.56 Cooking for special situations, reasons, ages
641.57 Quantity, institutional, travel, outdoor cooking
641.58 Cooking with specific fuels, appliances, utensils
641.59 Cooking characteristic of specific geographic environments, ethnic cooking

Dewey Linked Data Timeline: Next steps

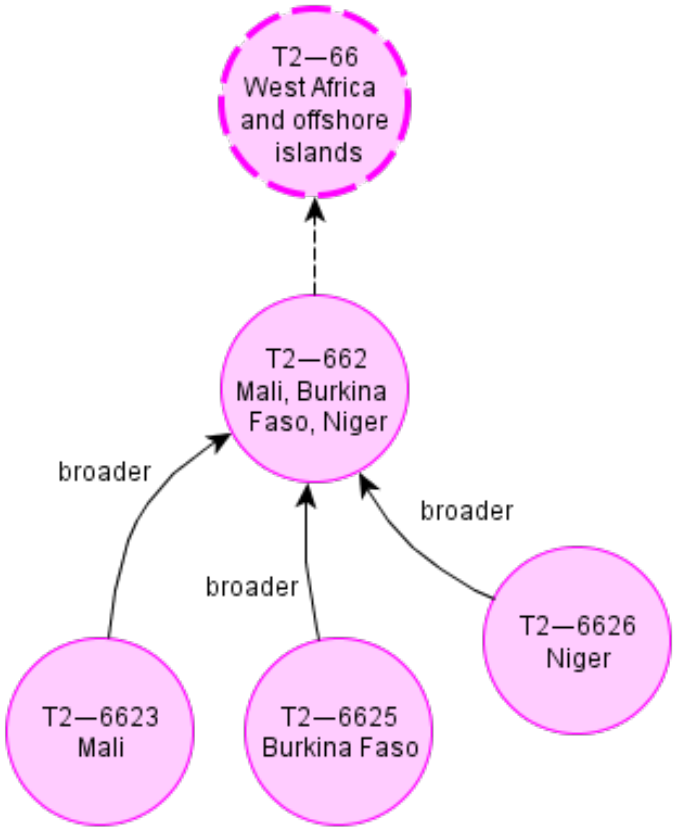
2009	2010	2011	2012	2013
dewey.info launched with DDC Summaries in 11 languages	Abridged Edition 14 in 3 languages	12th language added to DDC Summaries in dewey.info	DDC 23 in dewey.info **Table 2 + GeoNames**	**DDC 23 facets** **FAST** **MSC**
	New data format and distribution environment introduced for Dewey data		New data format and distribution environment extended to dewey.info	
	British Library DNB	FAO DDC in LOD Cloud		

Use Case:

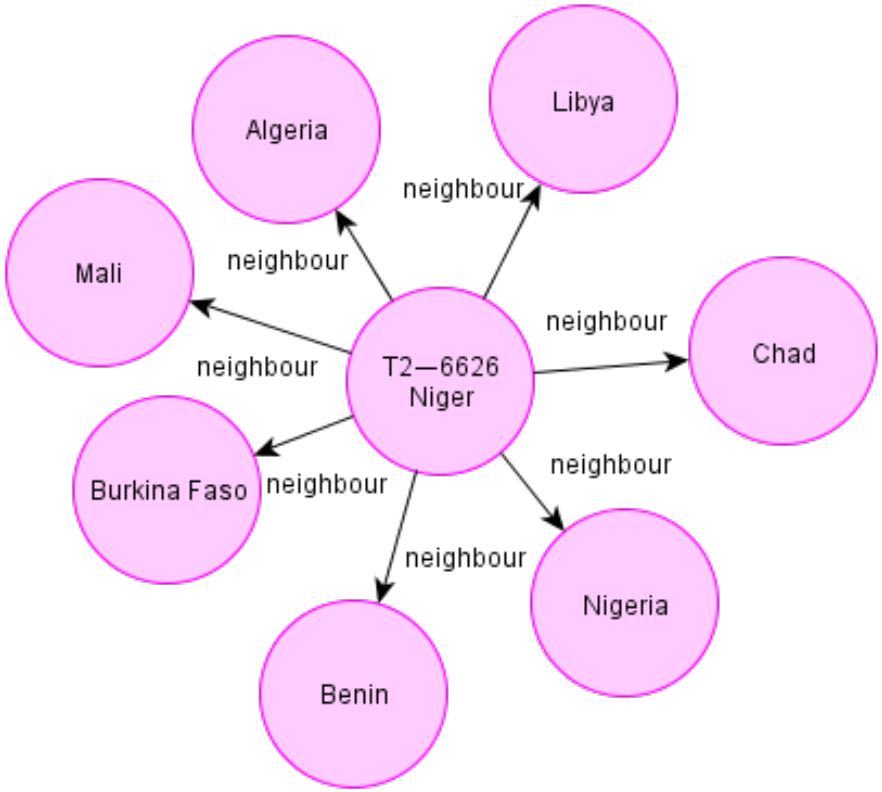
Links to GeoNames



Two Views of T2—6626 Niger

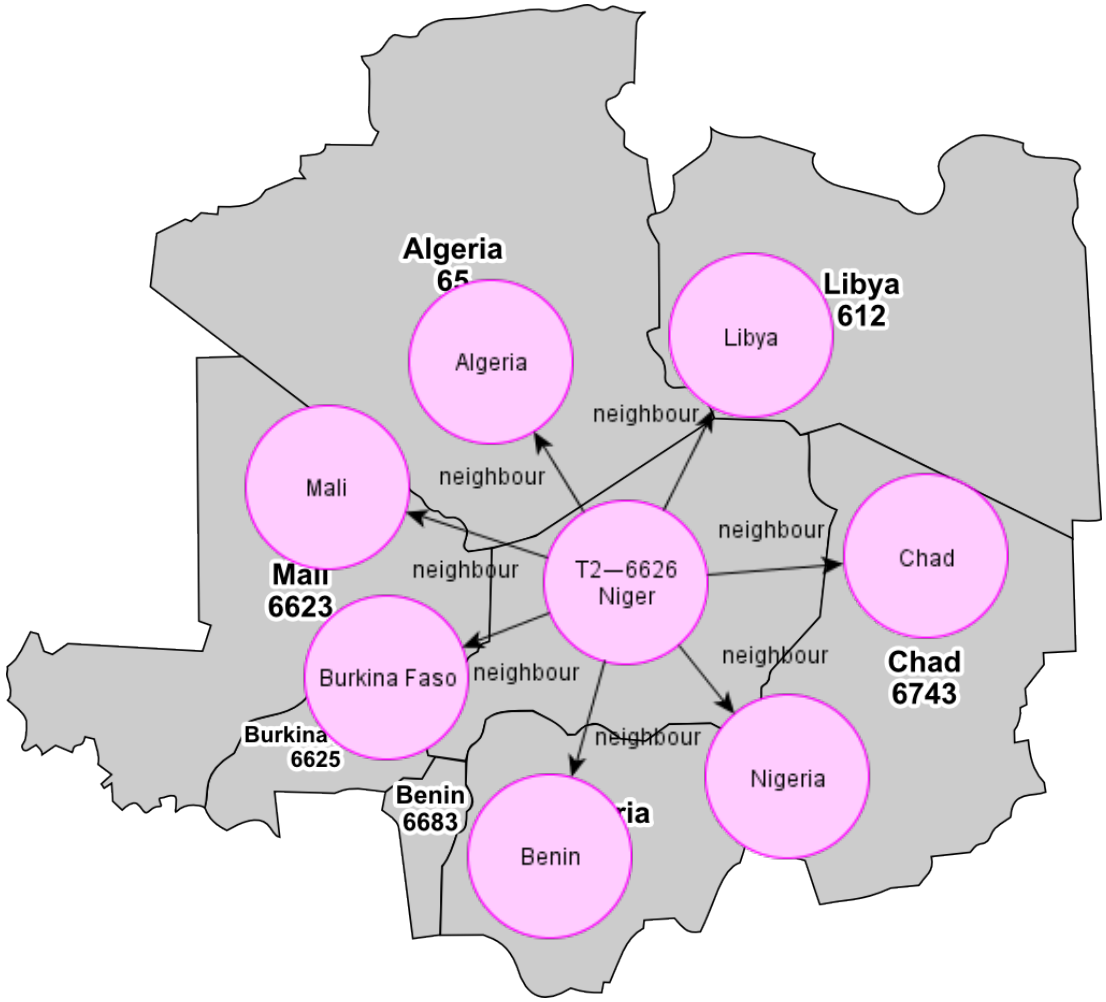


Dewey "neighborhood"



Geographic "neighborhood"

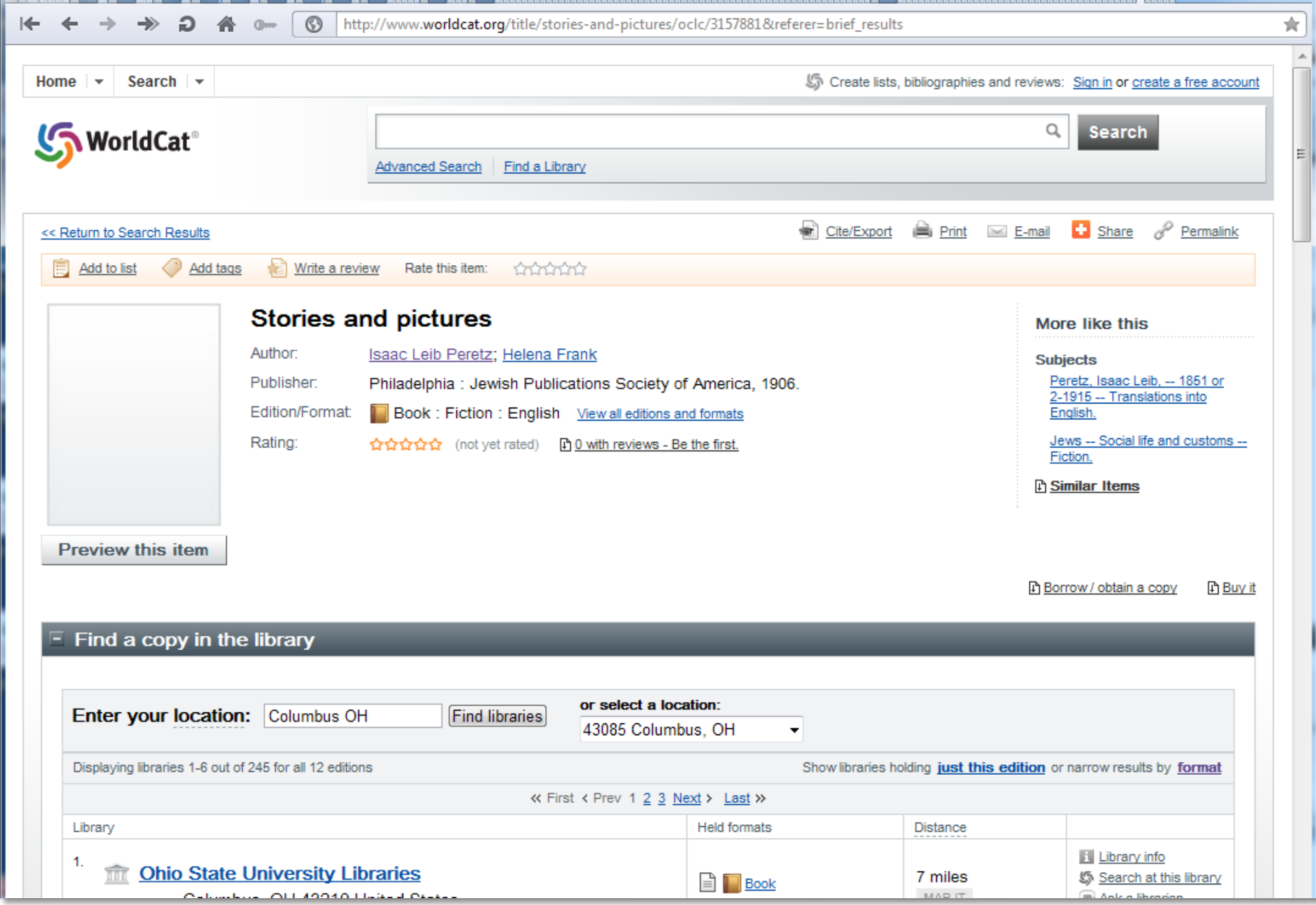
DDC + Geographic Viewpoints



Geographic "neighborhood"

Examples for publishing library linked data

Schema.org microdata in WorldCat.org


http://www.worldcat.org/title/stories-and-pictures/oclc/3157881&referer=brief_results
Home
Search
Create lists, bibliographies and reviews: Sign in or create a free account
WorldCat
Search
Advanced Search
Find a Library
<< Return to Search Results
Cite/Export
Print
E-mail
Share
Permalink
Add to list
Add tags
Write a review
Rate this item:
Stories and pictures
Author: Isaac Leib Peretz; Helena Frank
Publisher: Philadelphia : Jewish Publications Society of America, 1906.
Edition/Format: Book : Fiction : English
View all editions and formats
Rating: (not yet rated)
0 with reviews - Be the first.
More like this
Subjects
Peretz, Isaac Leib, -- 1851 or 2-1915 -- Translations into English.
Jews -- Social life and customs -- Fiction.
Similar Items
Preview this item
Borrow / obtain a copy
Buy it
Find a copy in the library
Enter your location:
Columbus OH
Find libraries
or select a location:
43085 Columbus, OH
Displaying libraries 1-6 out of 245 for all 12 editions
Show libraries holding just this edition or narrow results by format
« First ‹ Prev 1 2 3 Next › Last »
Library
Held formats
Distance
1.
Ohio State University Libraries
Book
7 miles
Library info
Search at this library

Linked Data

More info about Linked Data

<http://www.worldcat.org/oclc/3157881>

library:holdingsCount	"142"
library:oclcnum	"3157881"
library:placeOfPublication	rdf:type: schema:Place schema:name: "Philadelphia :"
rdf:type	schema:Book
schema:about	madsrdf:isIdentifiedByAuthority: <http://id.loc.gov/authorities/subjects/sh85070361> rdf:type: skos:Concept schema:name: "Jews--Social life and customs"
schema:about	<http://viaf.org/viaf/27070050> madsrdf:isIdentifiedByAuthority: <http://id.loc.gov/authorities/names/n50011063> rdf:type: schema:Person schema:name: "Peretz, Isaac Leib, 1851 or 2-1915" schema:name: "Peretz, Isaac Leib, 1851 or 2-1915."
schema:about	<http://id.worldcat.org/fast/983364> rdf:type: skos:Concept schema:name: "Jews--Social life and customs"
schema:about	rdf:type: schema:Person schema:name: "Peretz, Isaac Leib, 1851 or 2-1915"

Closing the circle, traversing the graph

Why should librarians care?

Better living through LOD: role of libraries and librarians

- Librarians are uniquely qualified, they just have to get more involved
  - From producing metadata records (back?) to extracting and engineering knowledge
- Do we really want to leave the field to the engineers?
  - Borges vs. schema.org

Borges' taxonomy vs. schema.org: but which one is which?

Animals

a) belonging to the Emperor
b) embalmed
c) trained
d) piglets
e) sirens
f) fabulous
g) stray dogs
h) included in this classification
i) trembling like crazy
j) innumerables
k) drawn with a very fine camelhair brush
l) et cetera
m) just broke the vase
n) from a distance look like flies

- Intangible
  - Enumeration
  - JobPosting
  - Language
  - Offer
  - Quantity
  - Rating
  - Structured Value
- Organization
  - EducationalOrganization
    - CollegeOrUniversity
    - ElementarySchool
    - HighSchool
    - MiddleSchool
    - Preschool
    - School

Better living through LOD: role of libraries and librarians

© John Kovalic

- Librarians are uniquely qualified, they just have to get more involved
  - From producing metadata records (back?) to extracting and engineering knowledge
- Do we really want to leave the field to the engineers?
  - Borges vs. schema.org
- Librarians can continued to do what they have always done best ... We have just been handed sharper tools!

Michael Panzer
panzerm@oclc.org

Questions?