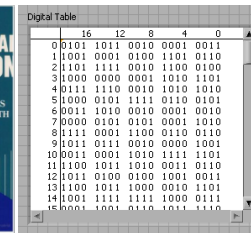
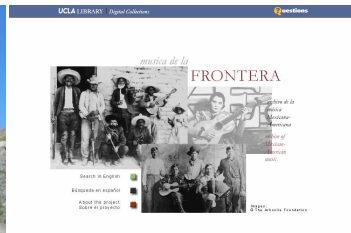
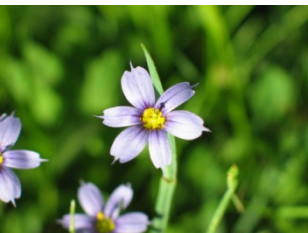


YAMZ: better, faster, cheaper vocabulary standardization

John Kunze
California Digital Library

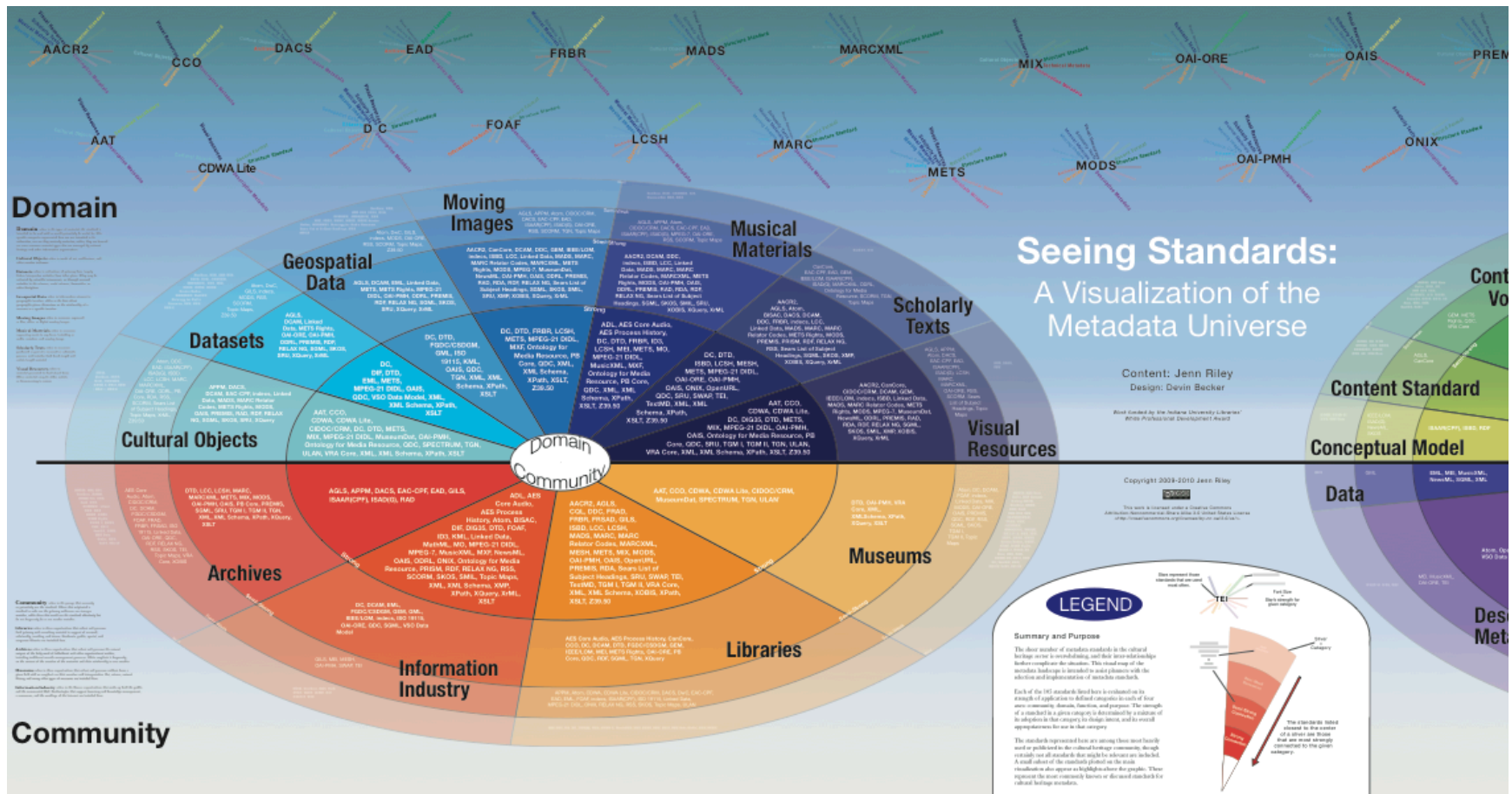


Problem: traditional standardization

ANSI
ISO
IEEE **FGDC** **NISO**
IETF

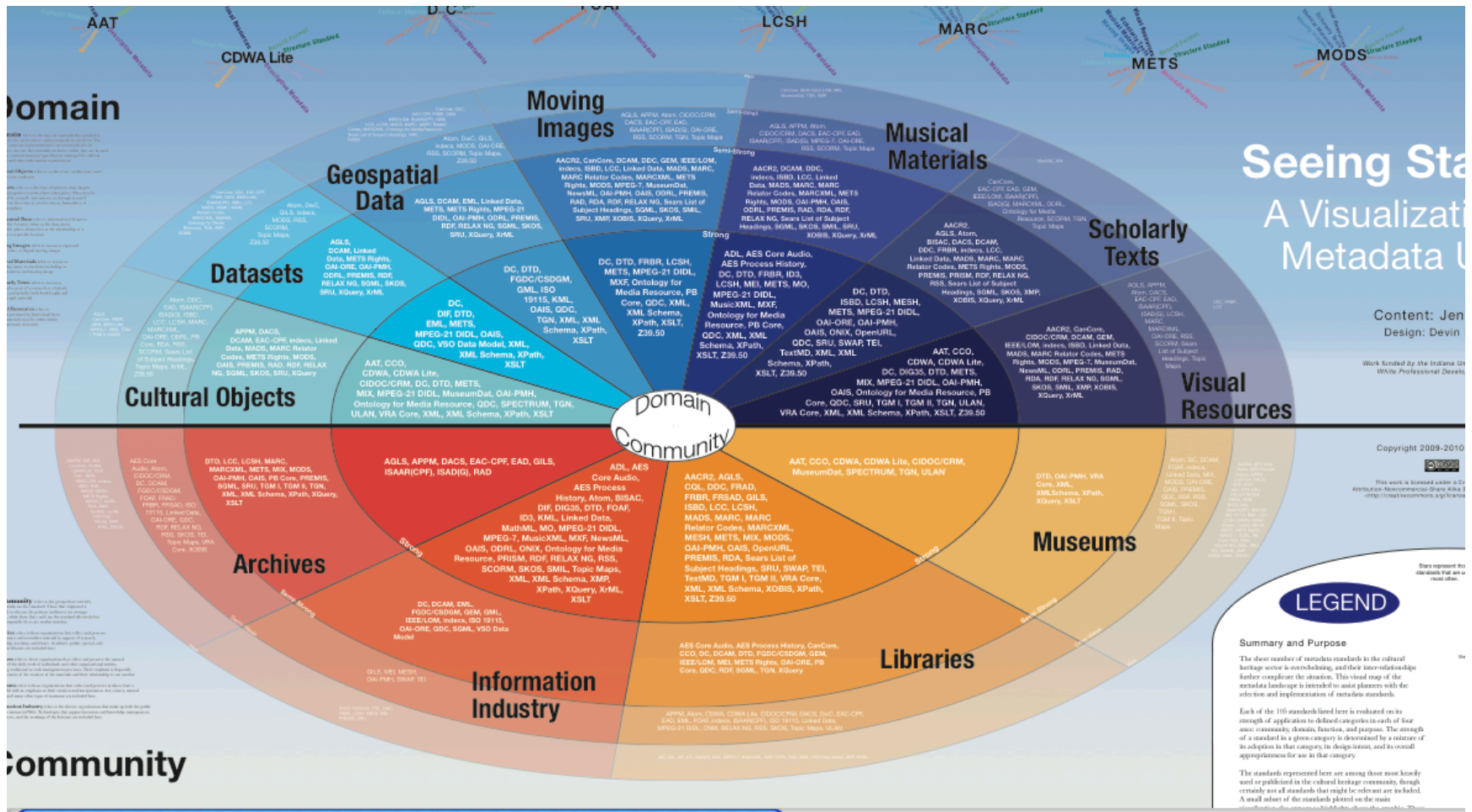
- Metadata? Easy, we use XYZ standard... not really
- Change by committee is ugly, costly, and slow
- Example: Dublin Core, 15 cross-domain terms

The Metadata Universe



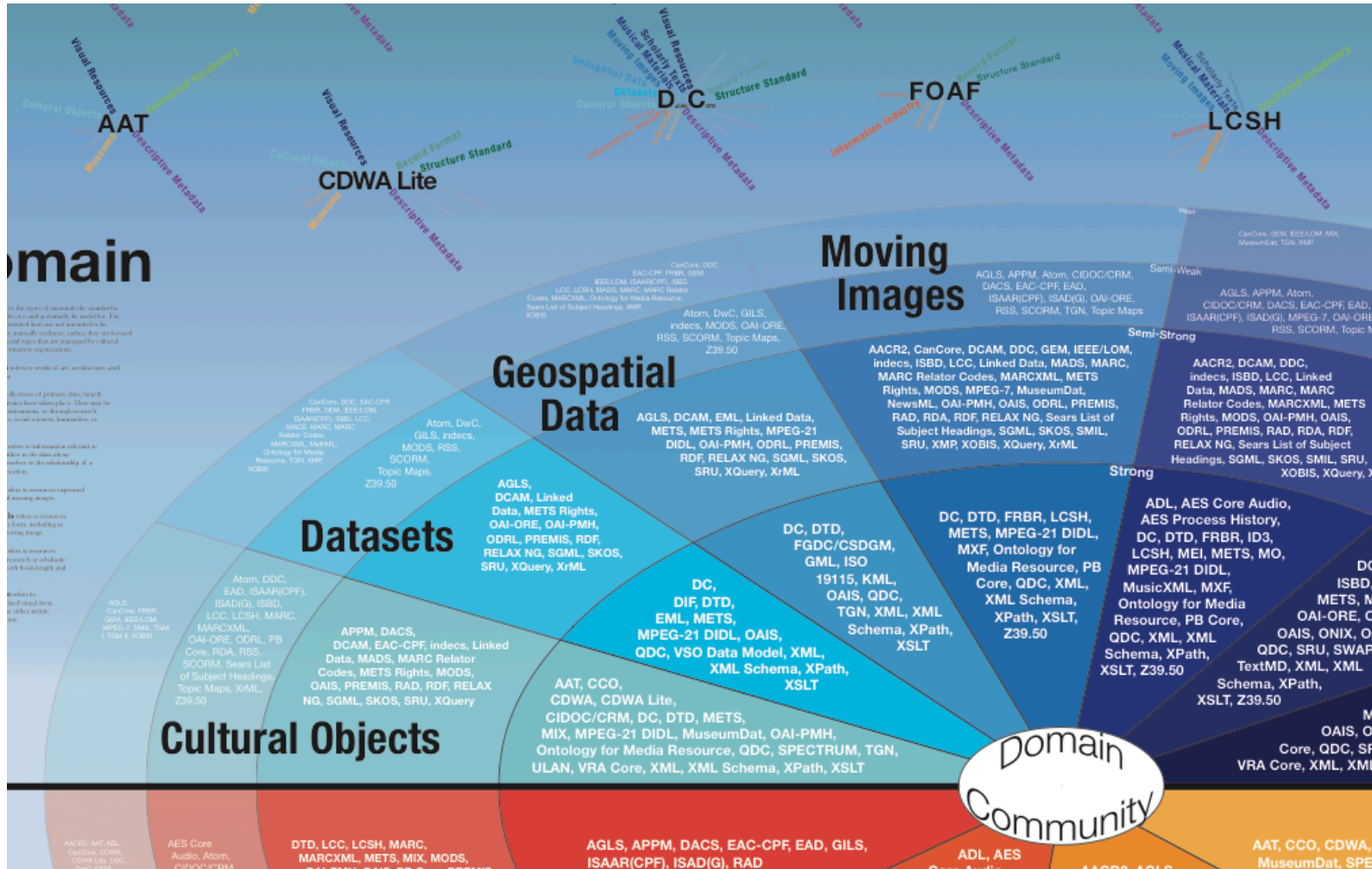
Jenn Riley,
IU

The Metadata Universe



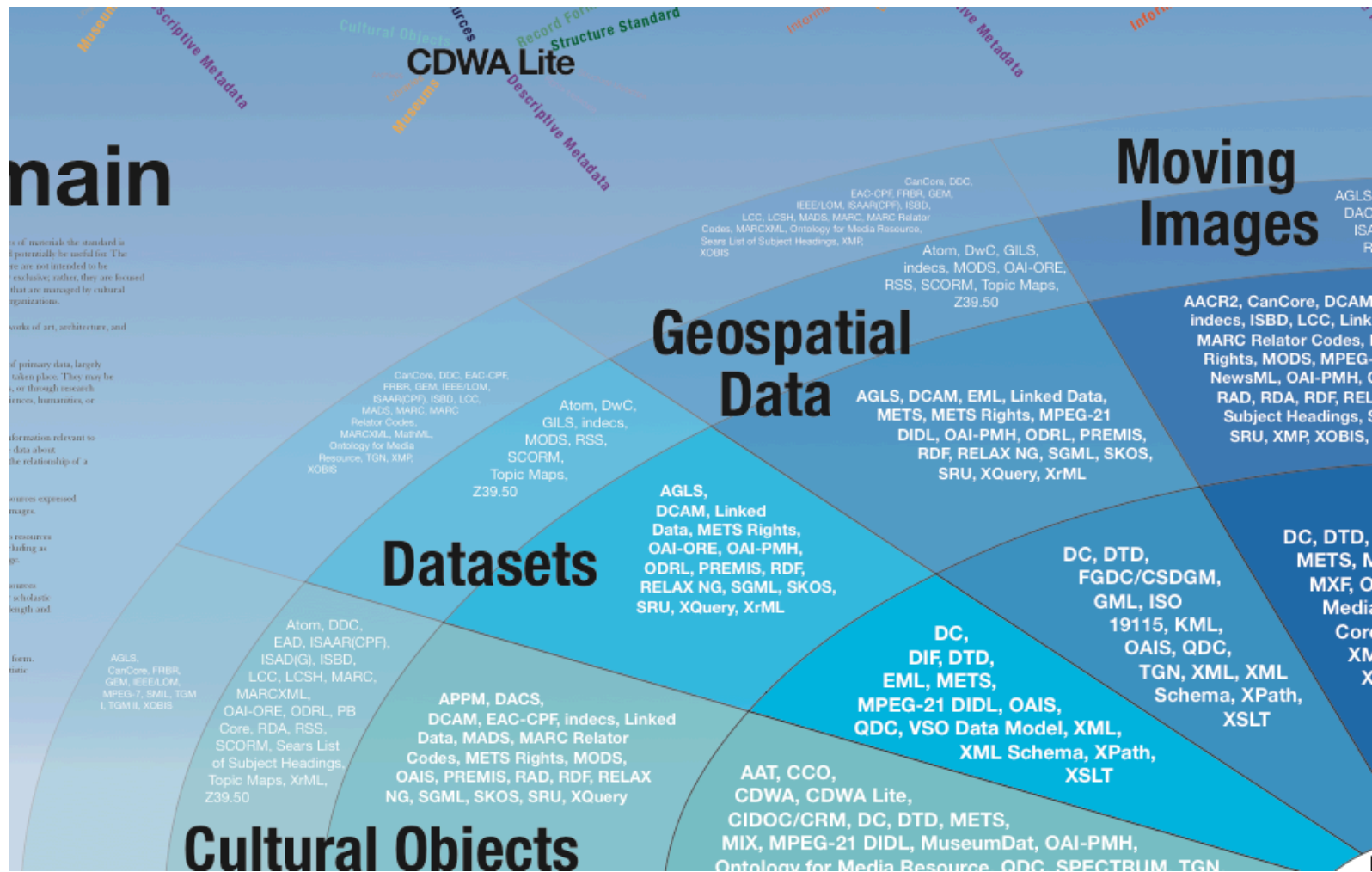
Jenn Riley,
IU

The Metadata Universe

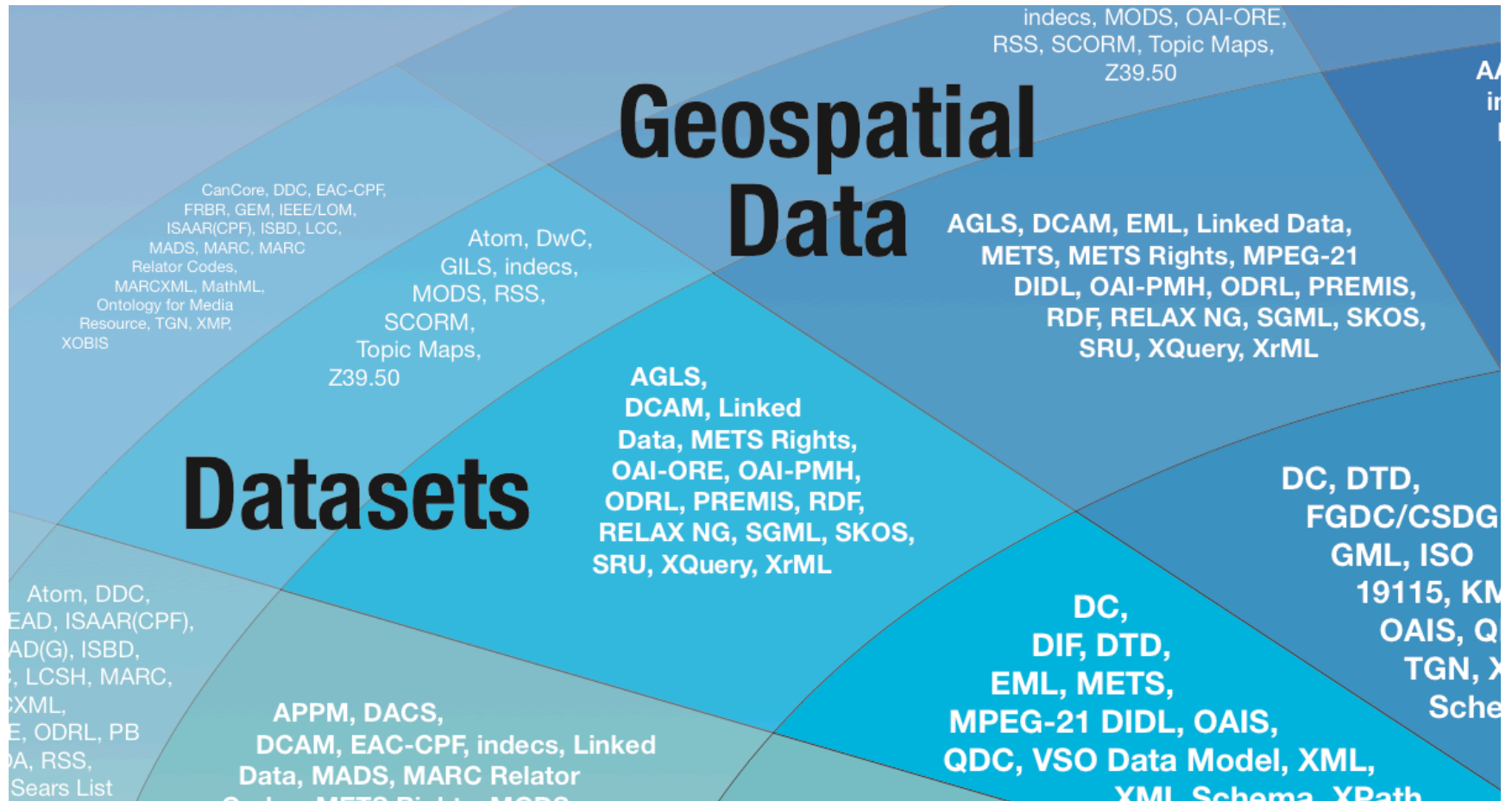


Jenn Riley,
IU

The Metadata Universe



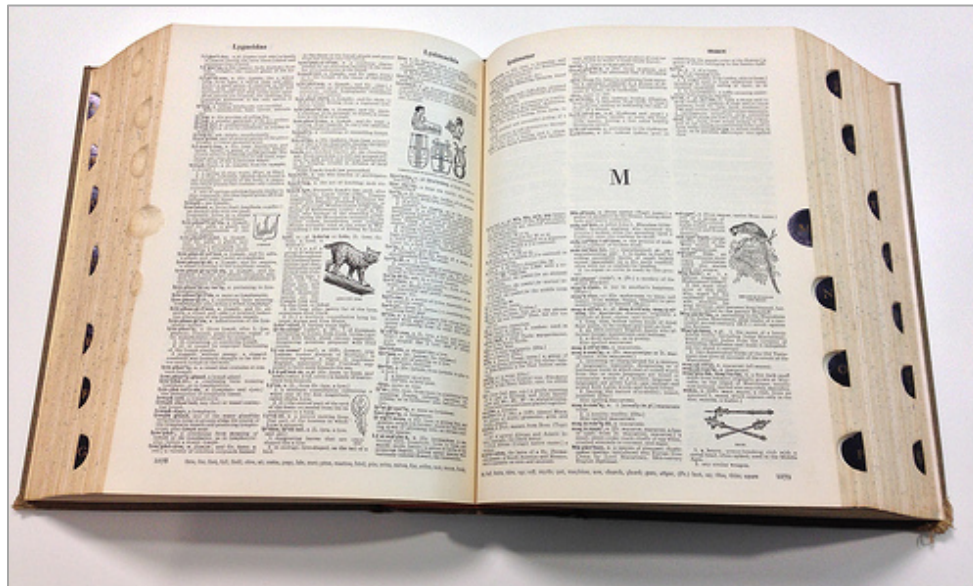
The Metadata Universe



Jenn Riley,
IU

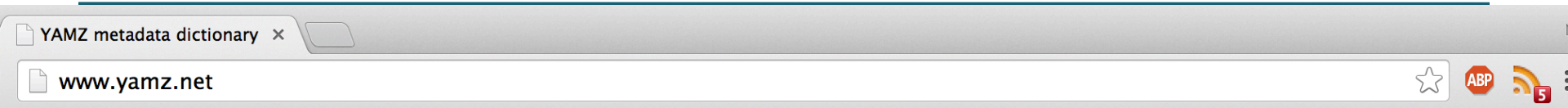
An alternate metadata universe

- Vision: one dictionary, one namespace
- All research domains, any part of “metadata speech”
 - Names, values, units, relationships, ...



SimonRobertson@flickr

YAMZ.net (Yet Another Metadata Zoo)



YAMZ

[Browse](#)

[Add](#)

[About](#)

[Contact](#)

[Account](#)

[Login](#)

Metadictionary

A crowdsourced metadata dictionary. Search for terms, upvote useful ones.

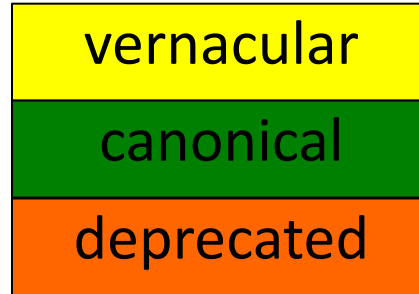
Contributions to the YAMZ metadictionary are dedicated to the public domain under the terms of CC0. By using this site, you agree to Terms of Use and Privacy Policy statements similar to wikimediafoundation.org.



University of California
CDL
California Digital Library

Crowdsourced, but with voting

3 classes
of term



- ← all terms are born here
- ← these don't evolve
- ← so terms never go away

Each term gets a unique persistent id. Example:

identifier: *<http://n2t.net/ark:/99152/h1193>*

term: *oba*

definition: *other (origin: from Tagalog)*

Reputation-based voting resists “gaming”



Karunakar Rayker @flickr

- Meritocracy: strong terms rise, weak terms decline
- Lessons from StackOverflow, Internet standards, and Wikipedia processes

Browse dictionary

[high score](#) | [recent](#) | [volatile](#) | [stable](#) | [alphabetical](#)

Term	Score	Consensus	Class	Contributed by	Last modified
Waveform Frequency	0	low	vernacular	Ashley Adair	17 July 2017
Waveform Duration	0	low	vernacular	Ashley Adair	17 July 2017
hybrid_simulation	0	low	vernacular	Maria Esteva	12 July 2017
global_model	0	low	vernacular	Maria Esteva	12 July 2017
raw_data	0	low	vernacular	Maria Esteva	5 July 2017
test_high_frequency_force_balance	0	low	vernacular	Maria Esteva	19 June 2017
particle_image_velocimetry	0	low	vernacular	Maria Esteva	19 June 2017

YAMZ usage patterns

Search for
terms
(words and
definitions)

find a term you love

great – use it

find a term you kind of love

try it out, comment,
engage with author

no workable term found

instantly enter own term
and watch for comments

find a *word* you love
but a *definition* you hate

“I want that word!”, so
enter a competing term

Add a dictionary term

Here you can propose a new term. You can help us maintain a high-quality metadictionary and minimize redundancies by [searching](#) for your term. Take a look at our community [community guidelines](#) for best practice tips.

Term string:

Definition:

A class of thing, followed by distinguishing characteristics, such as (for 'water'): 'A clear liquid made up of hydrogen and oxygen molecules.'

Example(s):

Examples of usage or related concepts.

Submit

Problem: working group ontology



European Parliament Technology - DG ITEC @ flickr

- Locally handcrafted, artisanal terms
- Hard to find prior art, kindred spirits
- Hard to expose new art for use by others
- What about an existing legacy vocabulary?

Term

Search for a term



Term: **correctable**

Class: **vernacular** (0%)

Created 19 February 2016
Last modified 3 May 2016
Contributed by Scout Calvert

Definition: A value of a Content Variance Profile (CVP) which indicates that previously recorded content may be corrected at any time.

persistence

Examples: CVP: correctable

Permalink:
n2t.net/ark:/99152/h1200

Does "corrected" imply a log of changes made?
what qualifies as "correcting" and how is that different from "overwriting?" - see CVP: dynamic
if this does imply that original content and "corrected" content will both be retained/displayed

"Correctable" implies only that the depositor reserves the right to deposit an object that has had errors corrected. It makes no statement about logs or original content. Implicit in the term is the sense that the object is complete but may be changed only to correct, if errors are found. See also static.

Add comment

Submitted 14 April 2016
by Sandra

Submitted 25 April 2016
by Scout



Tag persistence

Term	Score	Consensus	Class	Contributed by	Last modified
identifier	4	high	vernacular	John Kunze	28 April 2017
RDP_identifier	0	low	vernacular	Scout Calvert	3 May 2016
bestest	0	low	vernacular	Jeremy DeBarry	28 April 2017
PDP	0	low	vernacular	John Kunze	11 May 2016
LCU	0	low	vernacular	Sandra Sweat	11 May 2016
PDP_name	0	low	vernacular	John Kunze	11 May 2016
molting	0	low	vernacular	Jeremy DeBarry	28 April 2017
content	0	low	vernacular	Jeremy DeBarry	28 April 2017
stabler	0	low	vernacular	Jeremy DeBarry	28 April 2017
plunging	0	low	vernacular	Jeremy DeBarry	28 April 2017
RDP_more	0	low	vernacular	Scout Calvert	3 May 2016
LCP	0	low	vernacular	Sandra Sweat	11 May 2016
RDP_mission	0	low	vernacular	Scout Calvert	3 May 2016
NDG	0	low	vernacular	Matthew Hanlon	27 May 2016
subinfinite	0	low	vernacular	Jeremy DeBarry	28 April 2017
extraversioned	0	low	vernacular	Jeremy DeBarry	28 April 2017
business model	0	low	vernacular	Jeremy DeBarry	1 May 2017
waxing	0	low	vernacular	Jeremy DeBarry	28 April 2017
inflection	0	low	vernacular	Jeremy DeBarry	28 April 2017

In conclusion

- Current applications
 - Global Cryosphere Watch (GCW)
 - Citizen Science (Sloan)
 - DesignSafe (UTA)
 - Persistence statements (CDL, UCLA, TACC)
- Why: better, cheaper, faster metadata vocabularies
 - Try out YAMZ at yamz.net
 - Code at github.com/nassar/yamz