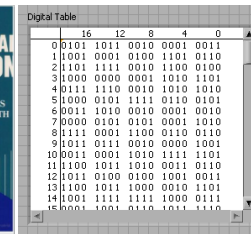
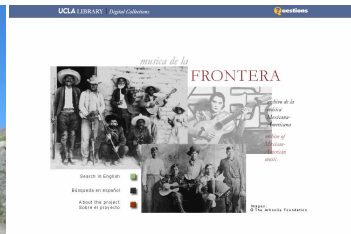
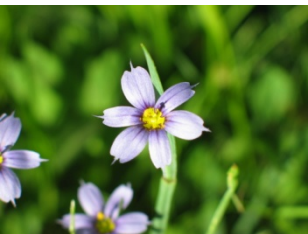


YAMZ: better, faster, cheaper vocabulary standardization

John Kunze
California Digital Library



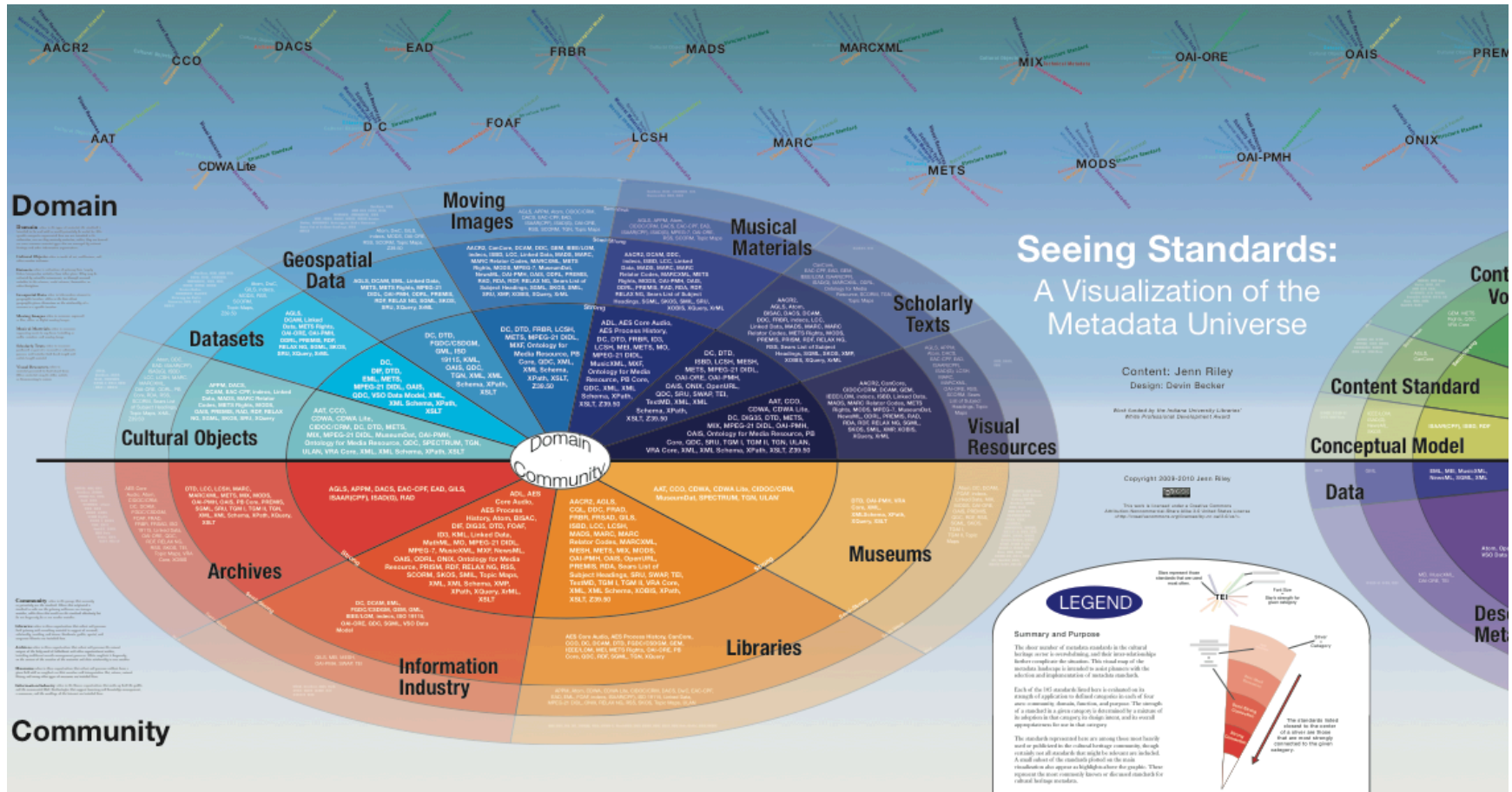
Problem: traditional standardization



European Parliament Technology - DG ITEC @ flickr

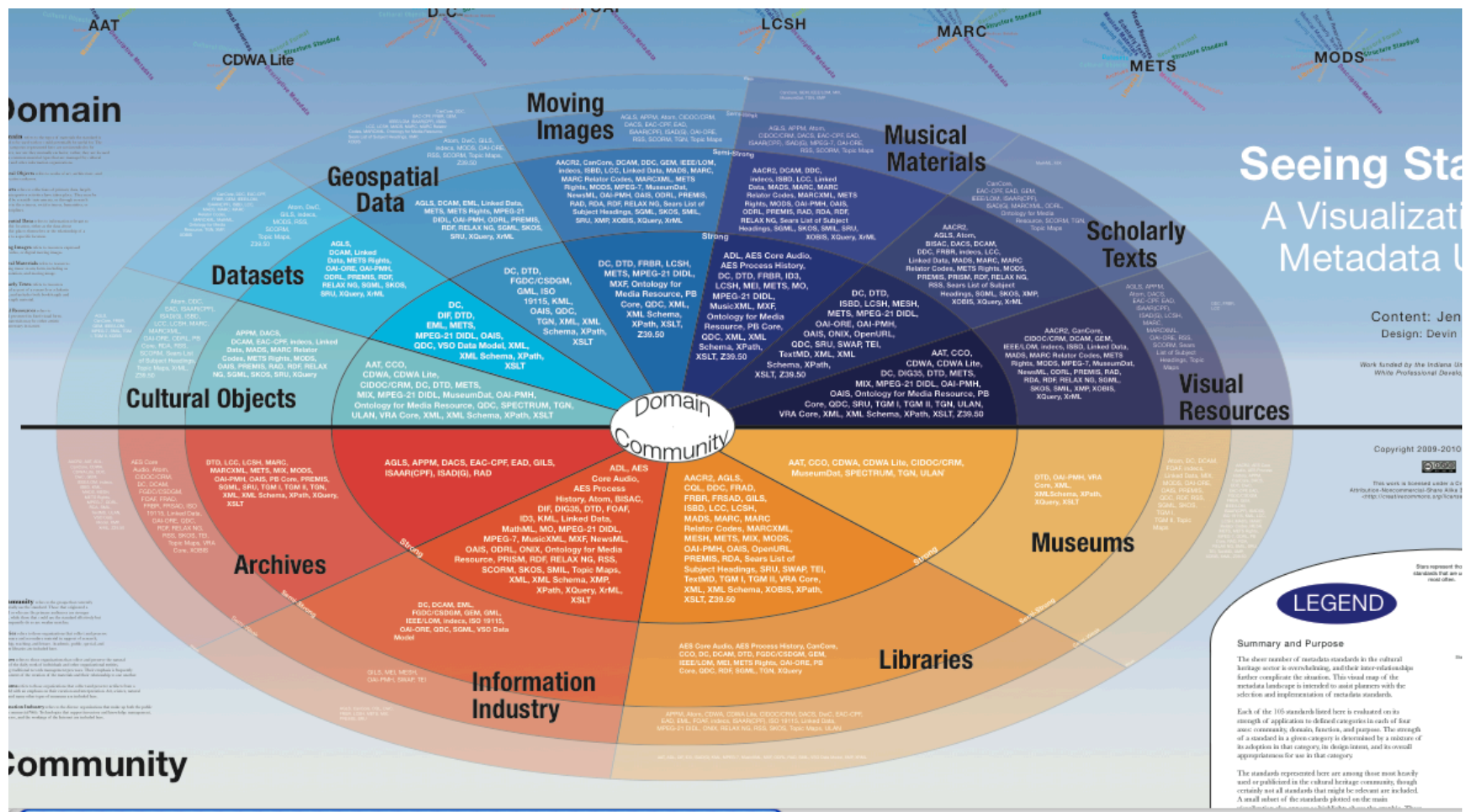
- Change by committee is ugly, costly, and slow
- Example: Dublin Core, 15 cross-domain terms

The Metadata Universe



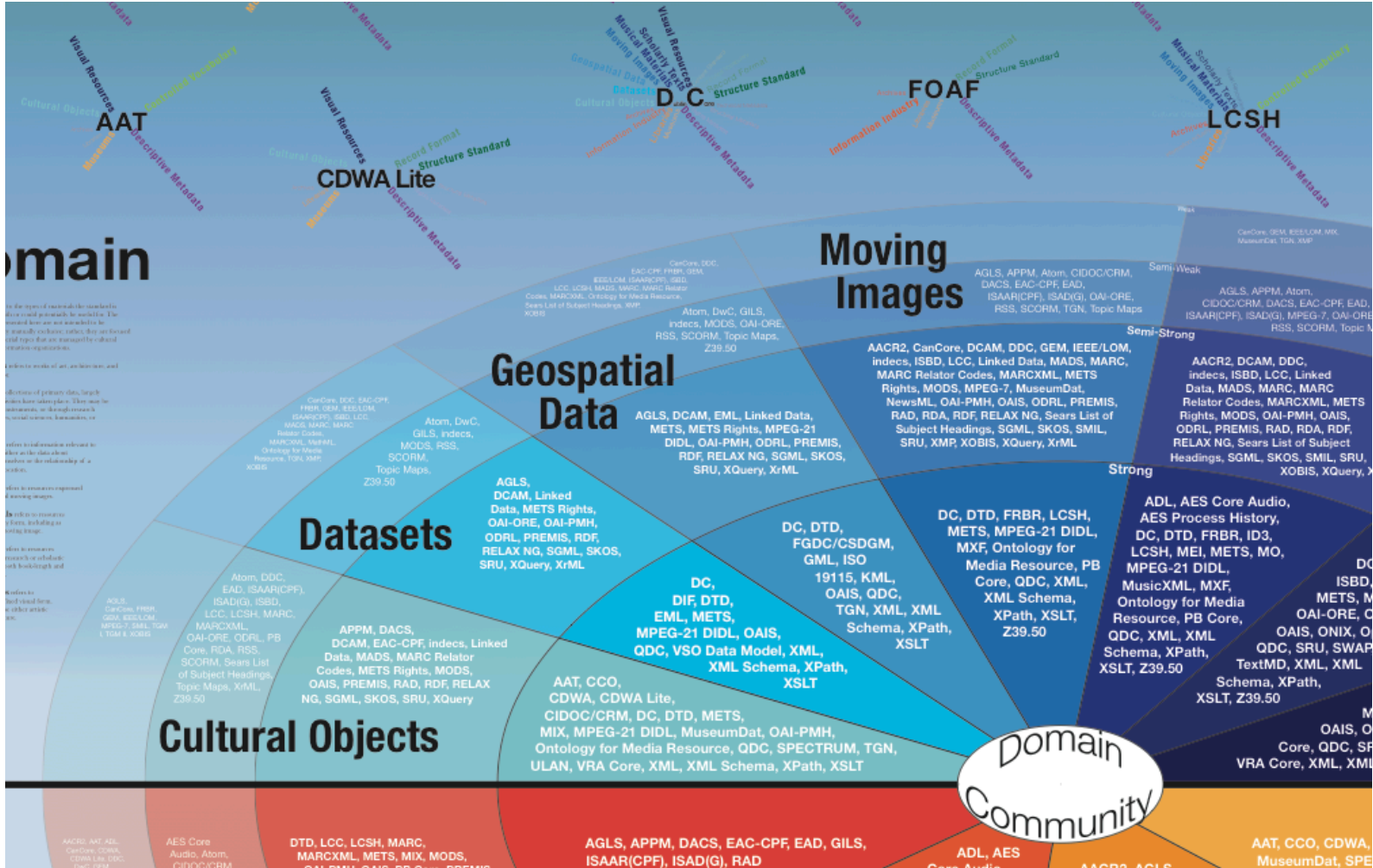
Jenn Riley, IU

The Metadata Universe

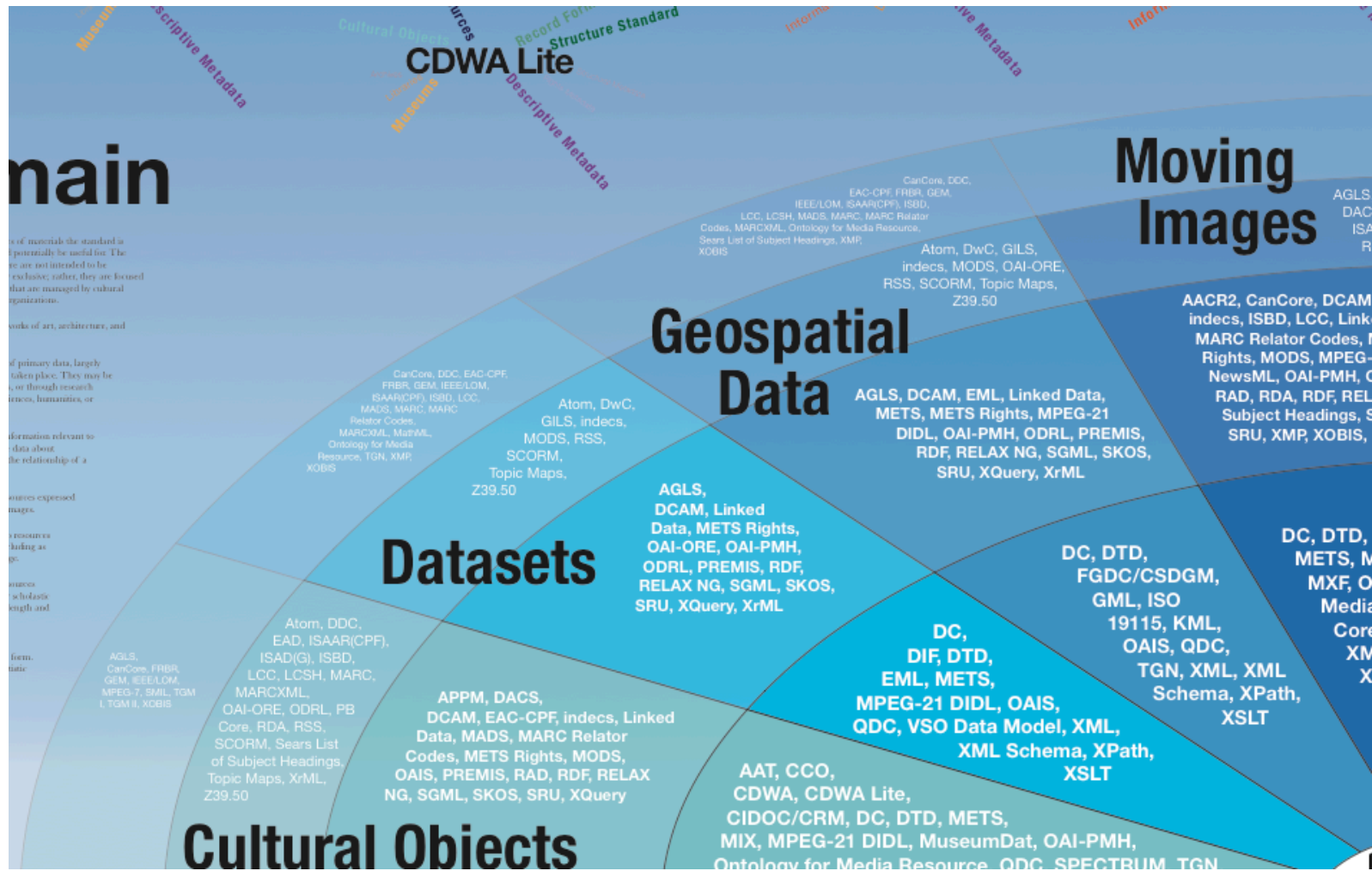


Jenn Riley, IU

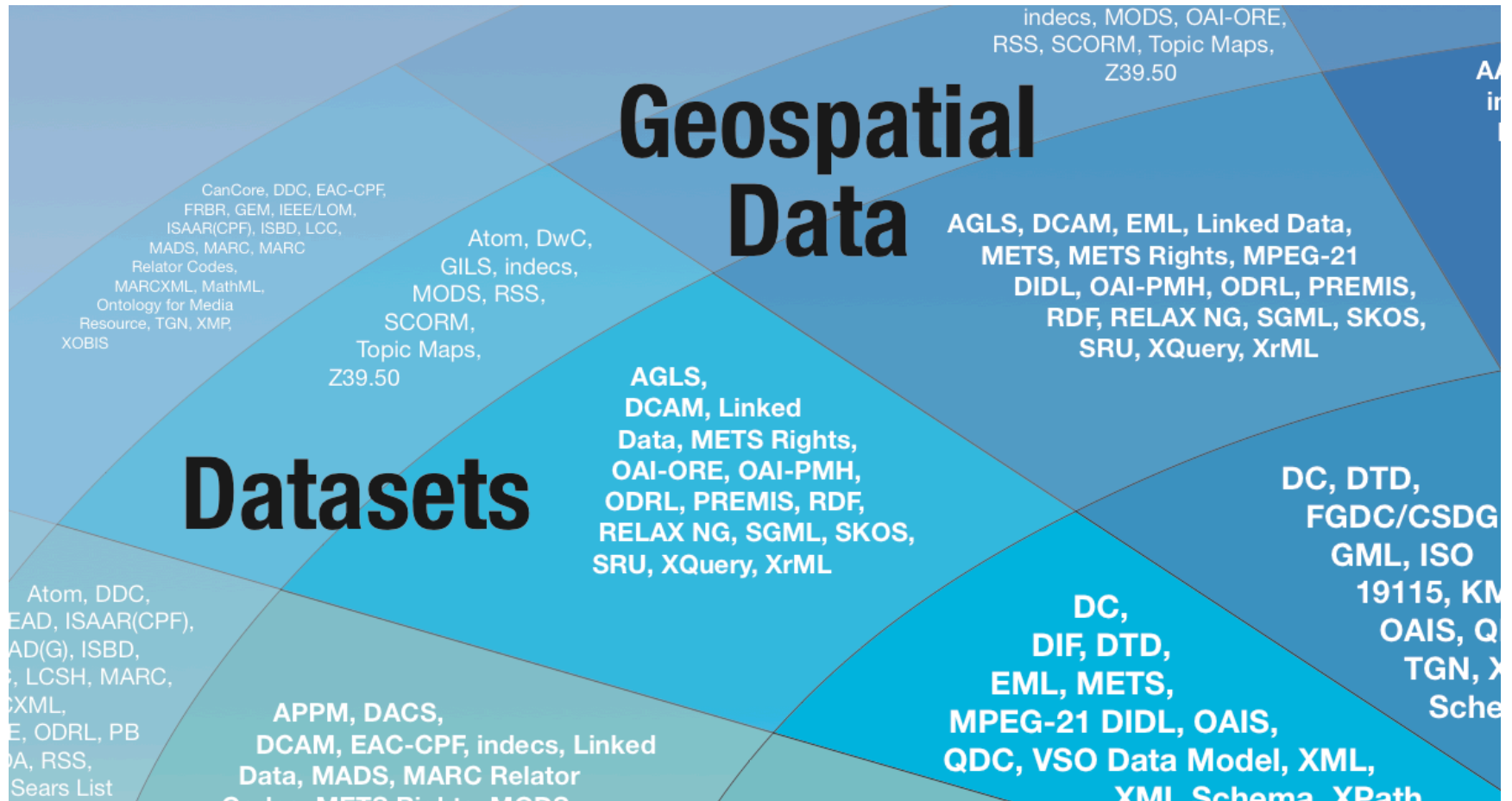
The Metadata Universe



The Metadata Universe



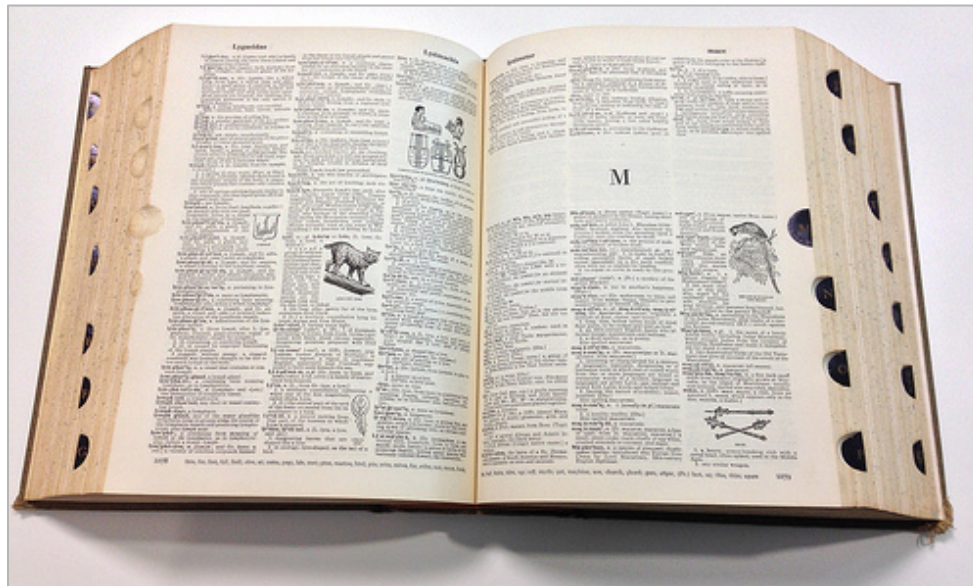
The Metadata Universe



Jenn Riley, IU

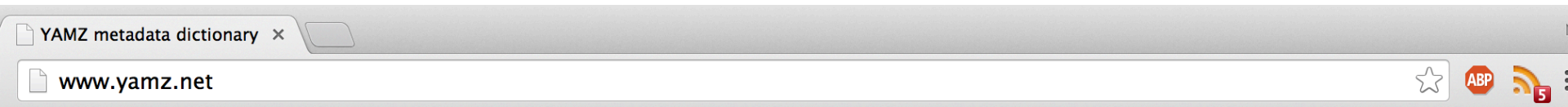
An alternate metadata universe

- 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

Crowdsourcing, but with voting

Each term in
1 of 3 classes

vernacular
canonical
deprecated

- all terms are born here
- these don't evolve
- terms available always

Each term gets a globally unique persistent identifier

term: oba

definition: other (Tagalog)

id: <http://n2t.net/ark:/99152/h1193>

Reputation-based voting resists “gaming”



Karunakar Rayker @flickr

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

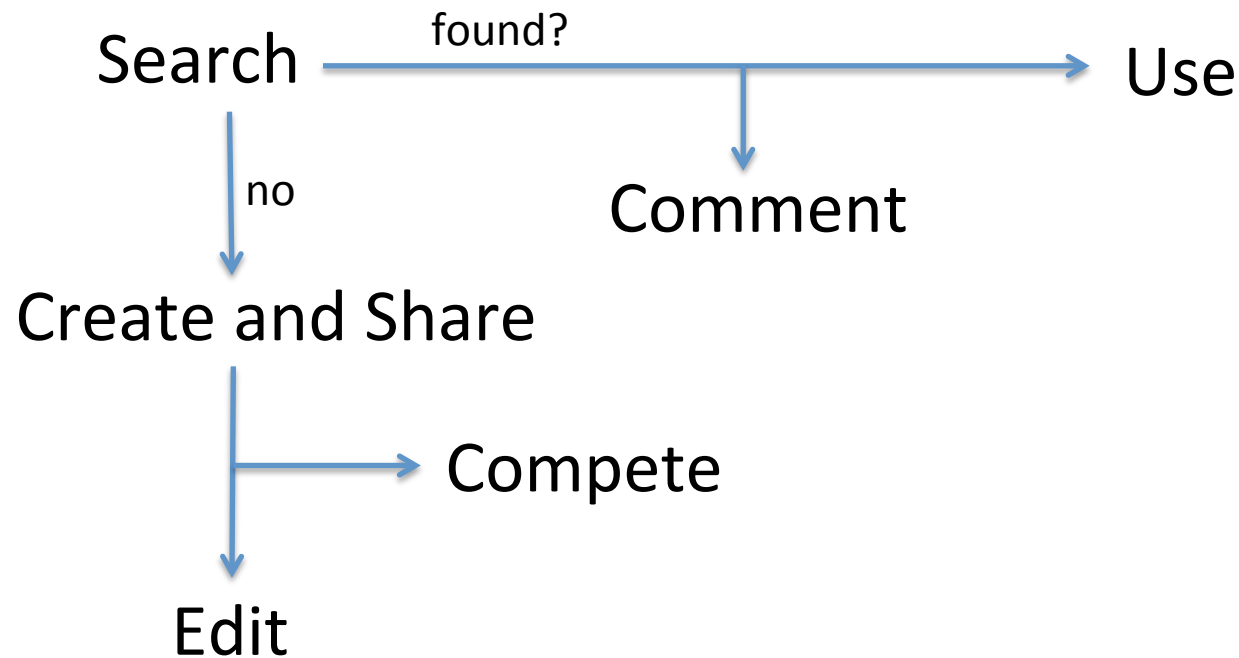
Browse dictionary

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

Term	Score	Consensus	Class	Contributed by	Last modified
oba	1	high	vernacular	John Kunze	13 October 2014
structured data	2	high	vernacular	John Kunze	7 October 2014
data	2	high	vernacular	John Kunze	7 October 2014
metadata	2	high	vernacular	John Kunze	7 October 2014
metadatum	2	high	vernacular	John Kunze	6 October 2014
datum	0	high	vernacular	John Kunze	6 October 2014
thing	1	high	vernacular	John Kunze	6 October 2014
datum	0	medium	vernacular	Chris Patton	4 October 2014
authority	1	high	vernacular	John Kunze	3 October 2014
identifier	2	high	vernacular	John Kunze	3 October 2014
association	1	high	vernacular	John Kunze	2 October 2014
manufactureDate	1	high	vernacular	Austin Mathews	29 May 2014
manufacturer	0	low	vernacular	Austin Mathews	29 May 2014
manufactureDate	-1	low	vernacular	Austin Mathews	29 May 2014
price	1	medium	vernacular	Josh Rice	29 May 2014
testscorepercentage	-1	low	vernacular	Silvana Goberdhan-Vigle	29 May 2014



Using dictionary terms



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 [guidelines](#) for best practice tips.

Term string:

Definition:

Term definition.

Example(s):

Examples of usage or related concepts.

Submit

Yamz use cases – macro level

1. develop a controlled vocabulary
 - borrow, invent, share, publish list of term references
2. create a pick list in software
 - same steps as in case 1, but refer to yamz online for user help (definitions and examples), then use the yamz API to add terms on behalf of users
3. add legacy vocabulary, and perhaps “correct” it
 - uploader only owns *their version* of terms

Finally

- Better, cheaper, faster metadata standardization
- Try it out at yamz.net
 - Code at <https://github.com/nassar/yamz>
 - Persistent identifiers (ARKs) from EZID
- Thanks to Jane Greenberg, Chris Patton, the NSF DataONE project, and the Digital Library Federation



THANK YOU!