

# Metadictionary: Advocating for a Community-driven Metadata Vocabulary Application

## DataONE Preservation and Metadata Working Group (PAMWG)

Jane Greenberg, \*Angela Murillo, SILS/UNC-CH, Metadata Research Center

\*\*John Kunze, California Digital Library

\*\*Christopher Patton UC Davis

Sarah Callaghan, British Atmospheric Data Centre

Rob Guralnick, Univ. of Colorado at Boulder

Greg Janee, UC Santa Barbara

Nassib Nassar, HW Odum Institute UNC-CH

Karthik Ram, UC Berkeley



# Outline

---

1. Assumptions
2. Motivation
3. Introduce Sealce
4. Conclusions
5. Questions



# Assumptions

---

1. Prevailing methods for metadata development/generation has advantages and limitations
  - + Rich ...capture domain structure, scope,...
  - Costly, complex (process to implementation)
2. More than one way to skin a cat
  - Complementary, alternative approaches
  - Social technology
3. Ownership appeal
  - Empowerment and sustainability

# Motivation

---

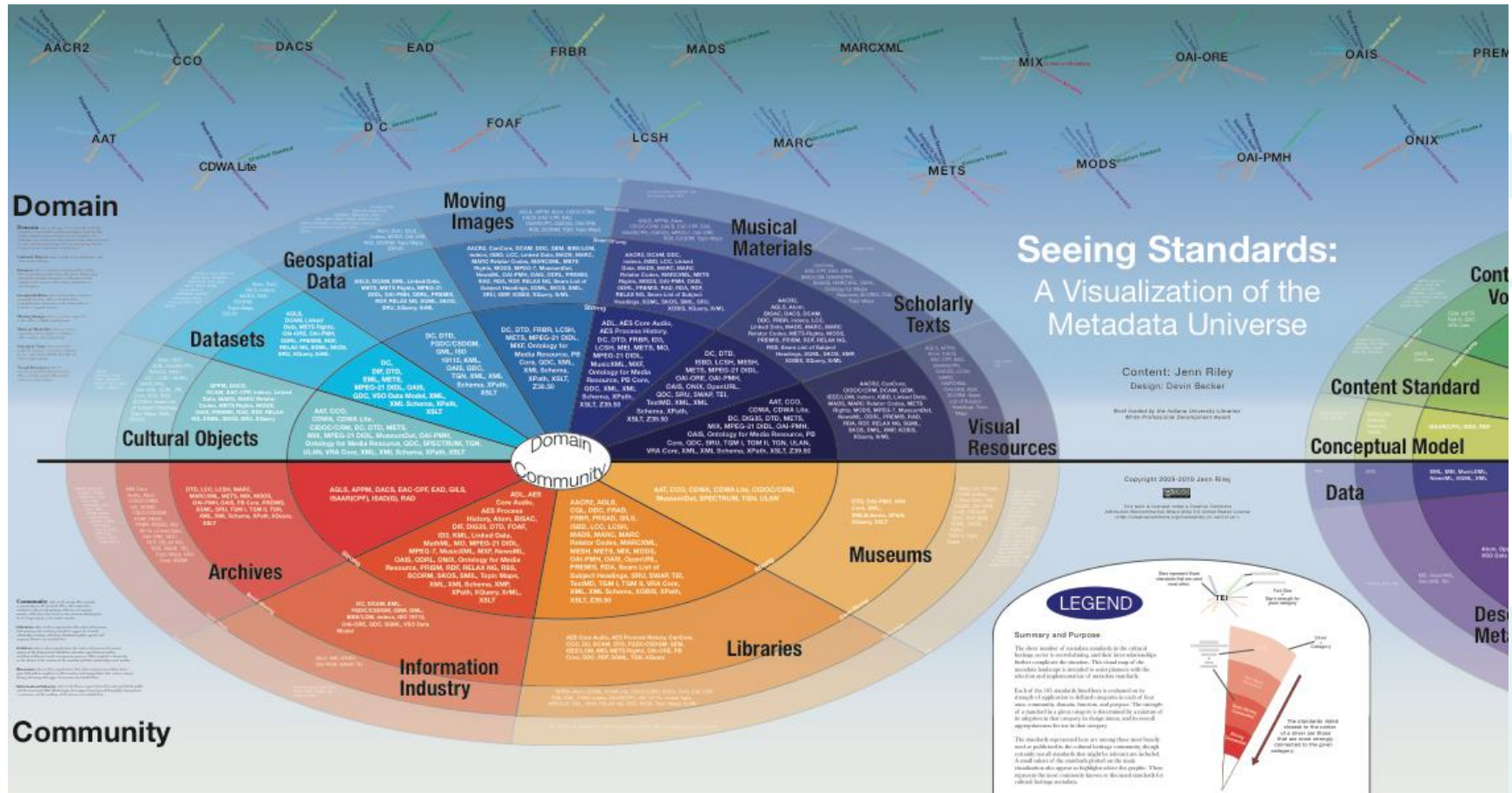
Toothbrush →



*Long tail*

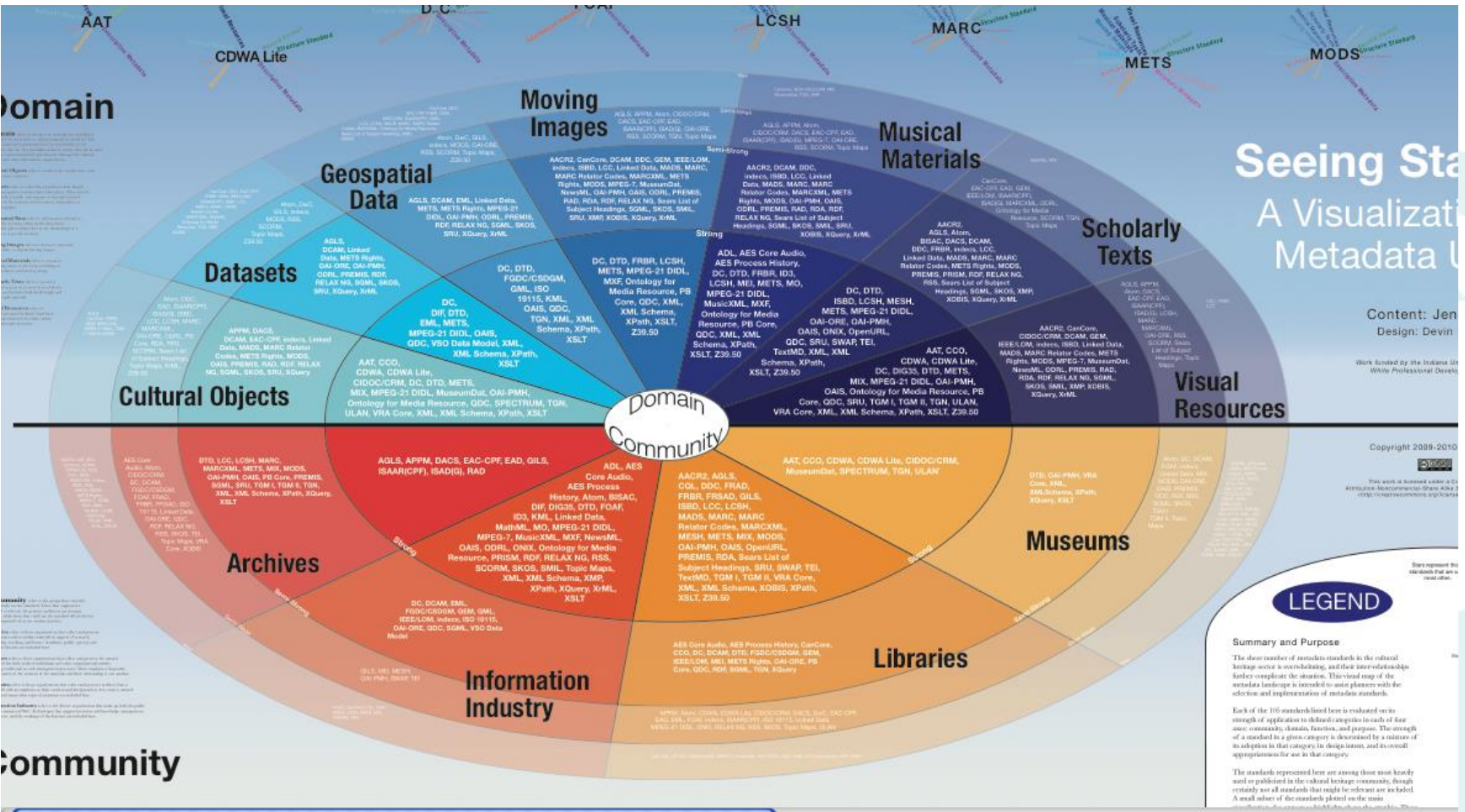


# The Metadata Universe



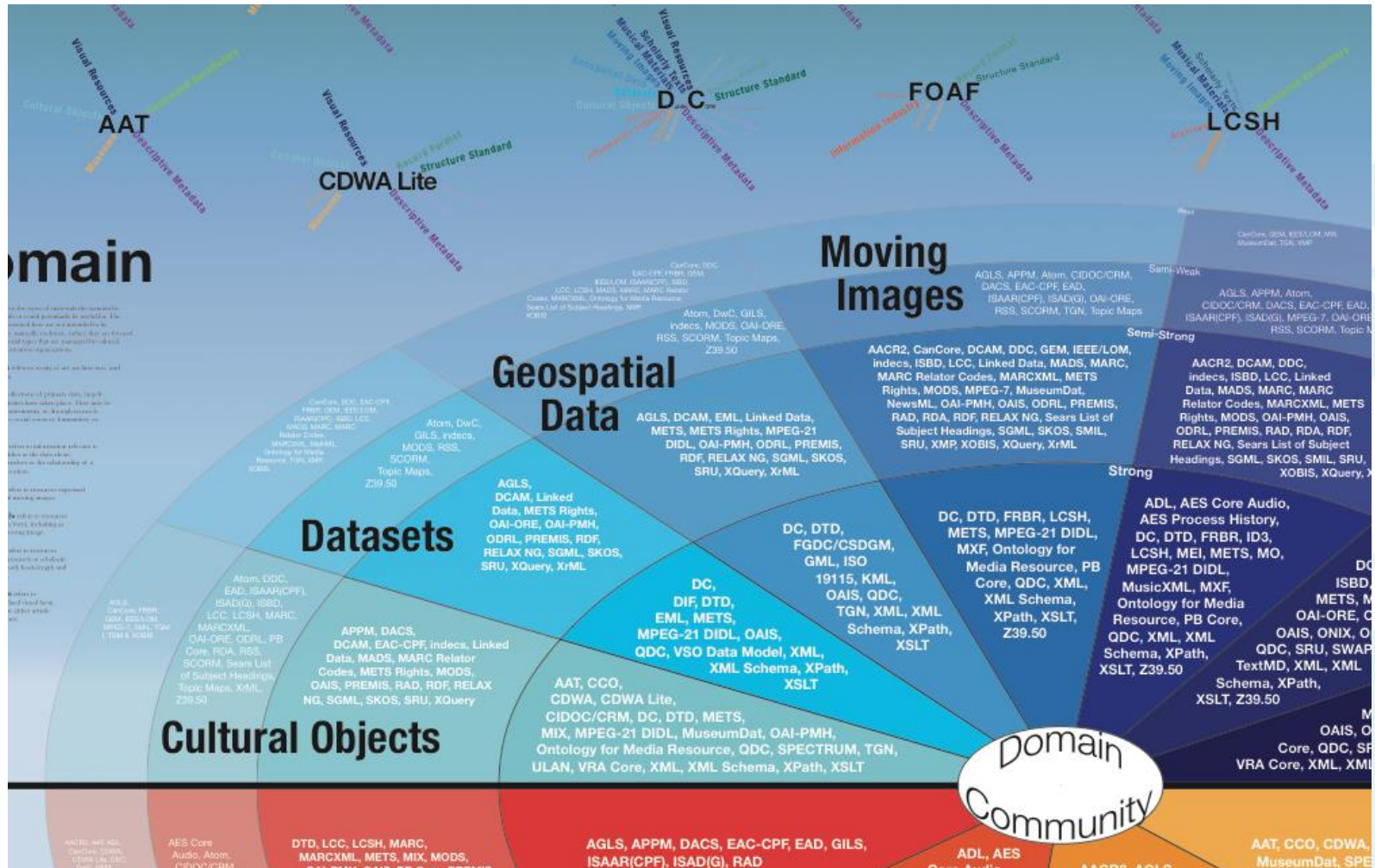
Jenn Riley, IU





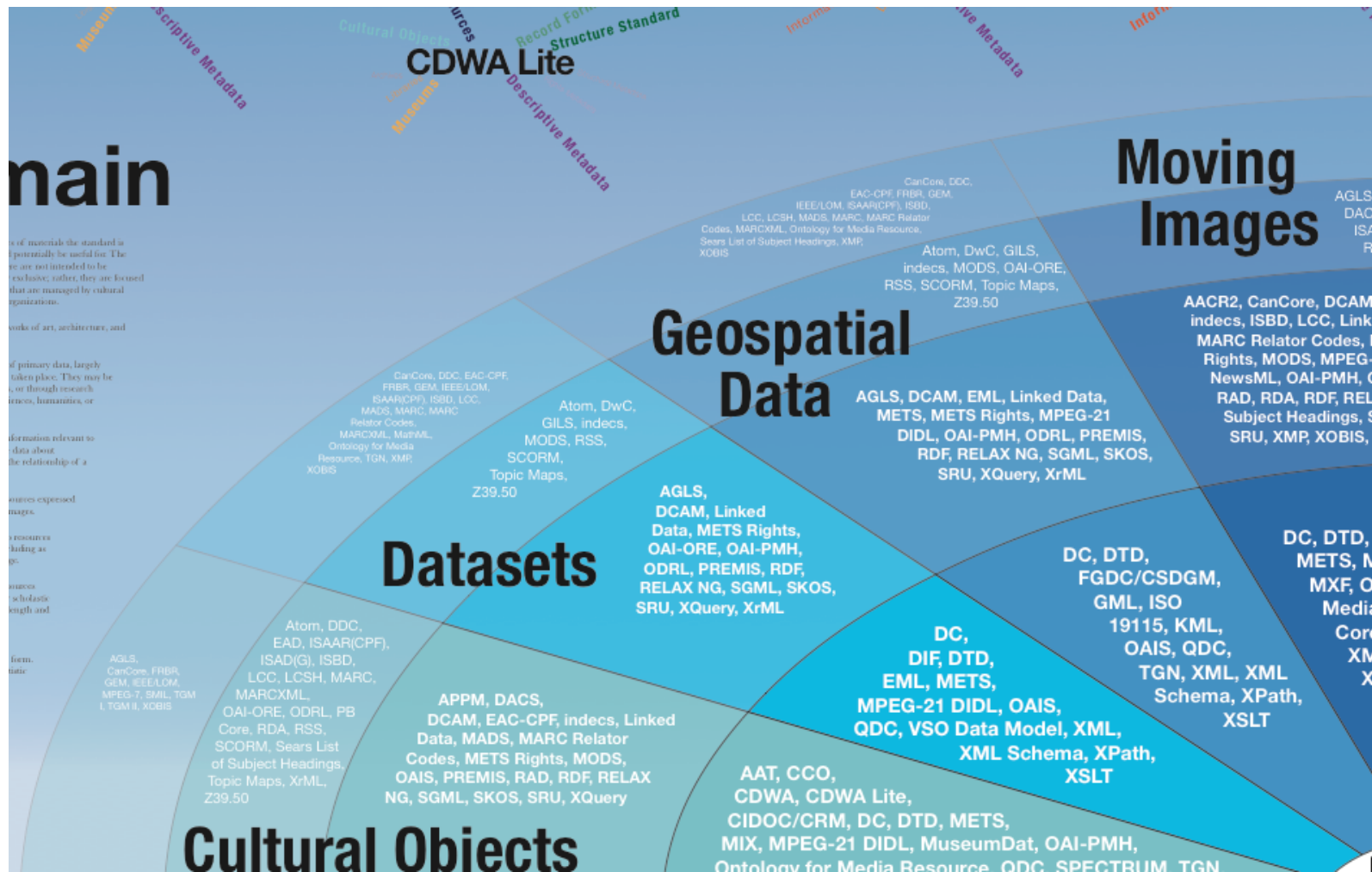
Jenn Riley, IU

# The Metadata Universe



Jenn Riley, IU

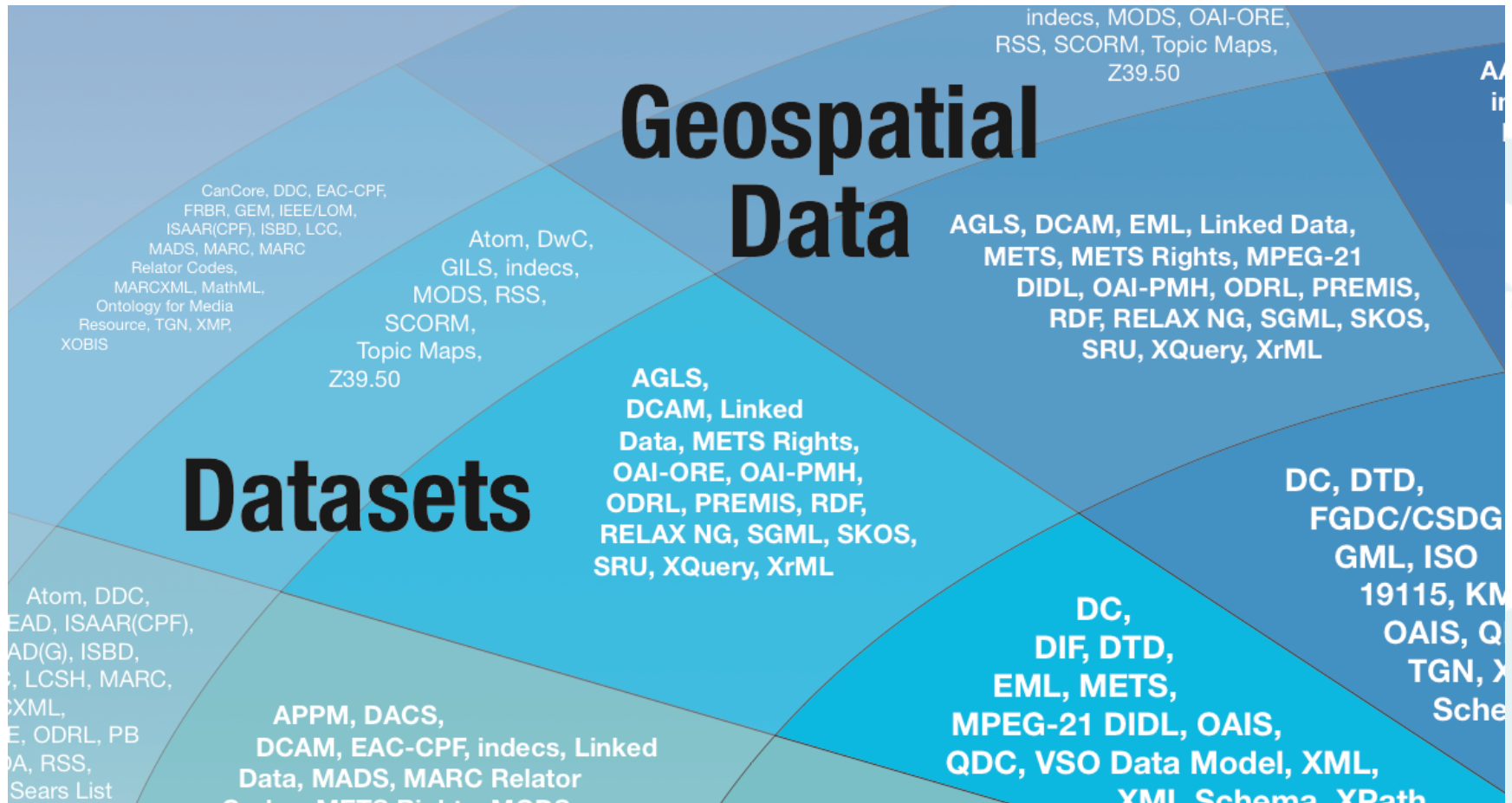
# The Metadata Universe



Jenn Riley, IU



# The Metadata Universe

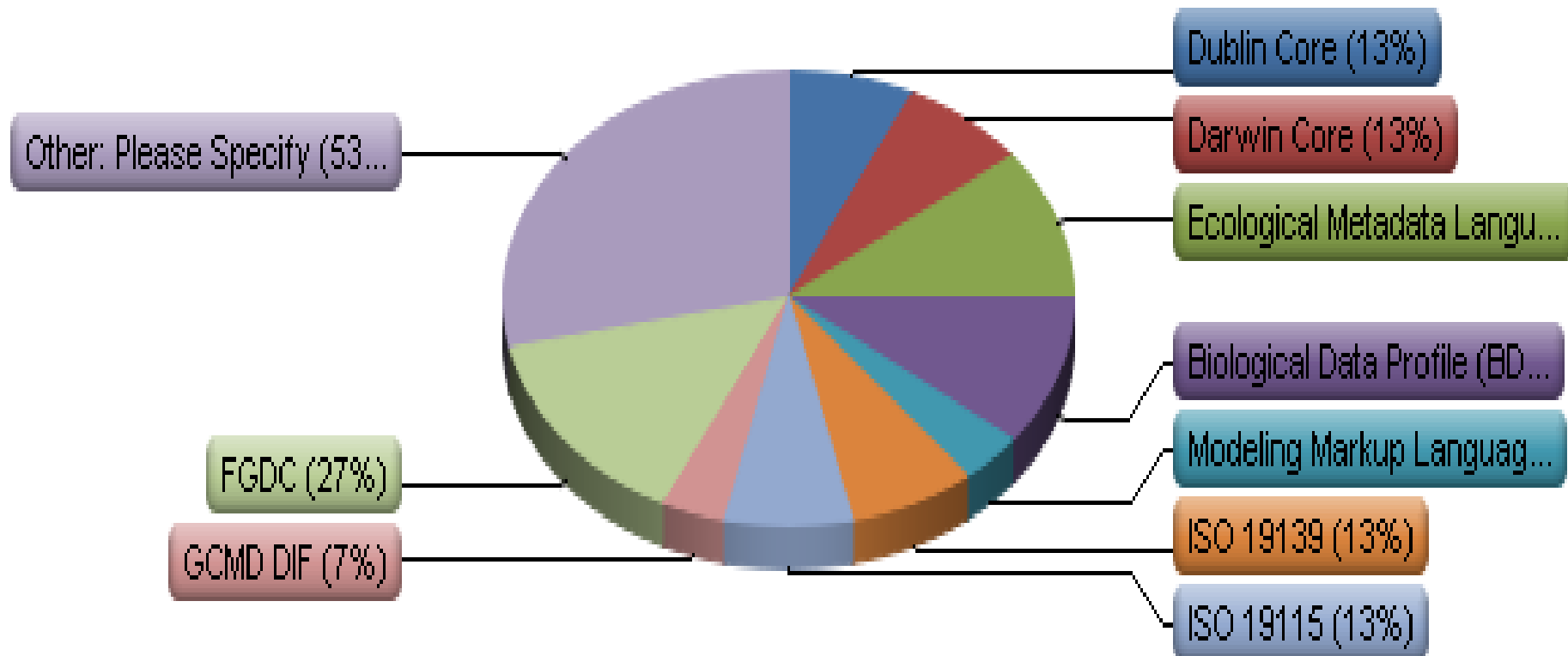


Jenn Riley, IU

## Components of Successful Metadata Registry Frameworks (A. Murillo, 2012)

---

14 standardized schemes used, lots of in-house  
n = 74 (biology, earth science, computer science, etc.)



# Metadata Vision

---

## *One dictionary, one namespace*

- Crowd sourced plus lightly supervised canon
- Anyone can look up terms
- Any domain, any part of “metadata speech”
  - Names, values, units, relationships, ...
- Anyone can propose and refine their terms
- Strong terms rise, weak terms decline

*What can we glean from Wikipedia, internet RFCs, and American Heritage Dictionary?*

# Translating a vision to principles

---

*Low barrier* for contributions.

*Transparency* in the review process.

*Collective review*, with rotating responsibilities among community members (scientists, developers, organizations, curators, etc.)

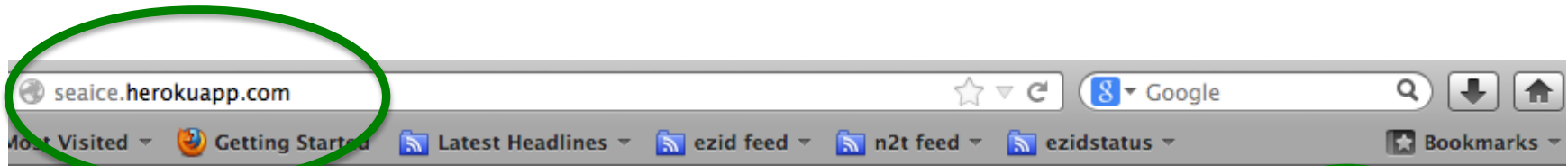
Consideration of elders (experts) to guide the review process and maintain thoughtful, balanced discussion.

Voting capacity of all users on the candidacy of terms submitted and their use.

*Collective ownership* of any user or organization.

Stakeholder engagement in the design and review process.





Sealce

[Browse](#)

[Add](#)

[About](#)

[Contact](#)

[Settings](#)

[Login](#)

# Metadictionary

A crowd sourced metadata dictionary. Search for terms, upvote useful ones.

**Tech stuff:** Python + off-the shelf tools; freely hosted on heroku with the evolving code on github; Project code name is Sealce.

## Notifications

TODO

## My terms

- [metadatum](#)
- [structured data](#)
- [metadata](#)
- [data](#)
- [structured datum](#)

## Starred

# 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:

Example(s):

# Browse dictionary

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

| Term                               | Score | Consensus | Class      | Contributed by | Last modified  |
|------------------------------------|-------|-----------|------------|----------------|----------------|
| <a href="#">data</a>               | 2     | 100%      | canonical  | John Kunze     | 1 day ago      |
| <a href="#">publisher</a>          | 2     | 100%      | canonical  | John Kunze     | 5 days ago     |
| <a href="#">creator</a>            | 1     | 100%      | canonical  | John Kunze     | 6 days ago     |
| <a href="#">datum</a>              | 1     | 100%      | canonical  | John Kunze     | 1 day ago      |
| <a href="#">description</a>        | 1     | 100%      | canonical  | John Kunze     | 6 days ago     |
| <a href="#">identifier</a>         | 1     | 100%      | canonical  | John Kunze     | 1 day ago      |
| <a href="#">metadata</a>           | 1     | 100%      | canonical  | John Kunze     | 6 days ago     |
| <a href="#">resource</a>           | 1     | 100%      | canonical  | Chris Patton   | 2 days ago     |
| <a href="#">identifier</a>         | 1     | 66%       | vernacular | John Kunze     | 5 days ago     |
| <a href="#">datum</a>              | 0     | 50%       | vernacular | Chris Patton   | 1 day ago      |
| <a href="#">hydraulic gradient</a> | 0     | 0%        | vernacular | Angela Murillo | 14 August 2013 |
| <a href="#">structured data</a>    | 0     | 0%        | deprecated | John Kunze     | 9 August 2013  |
| <a href="#">structured datum</a>   | 0     | 0%        | deprecated | John Kunze     | 5 days ago     |
| <a href="#">talus slope</a>        | 0     | 0%        | deprecated | Angela Murillo | 12 August 2013 |
| <a href="#">great</a>              | -1    | 12%       | deprecated | Nassib Nassar  | 6 days ago     |
| <a href="#">CHL</a>                | -1    | 0%        | vernacular | Greg Janée     | 6 days ago     |
| <a href="#">metadatum</a>          | -1    | 0%        | deprecated | John Kunze     | 12 August 2013 |
| <a href="#">token</a>              | -1    | 0%        | deprecated | John Kunze     | 1 day ago      |
| <a href="#">talus</a>              | -2    | 0%        | deprecated | Angela Murillo | 6 days ago     |



# Term Classes and Voting Impact

---

**Vernacular → canonical -- term is stable after two days and consensus is above 75%.**

**Vernacular → deprecated -- term is stable after two days and consensus is below 25%.**

**Canonical → vernacular -- term has been updated, restabilized, and consensus has dropped below 75%.**

**Deprecated → vernacular -- term has been updated, restabilized, and consensus has risen above 25%.**

**\* Nothing firmed about percentages, *just an illustrative consideration***

# Conclusions, Feedback welcome!

---

## Conclusions

- Complementary and alternative approaches
  - *More than one way to skin a cat...*
- Next steps...populate, test, engage
  - *User profiles*

Try out the prototype (Sealce) at

<http://seaice.herokuapp.com>

<https://github.com/cjpatton/seaice>