YAMZ: better, faster, cheaper vocabulary standardization

John Kunze
California Digital Library
Problem: traditional standardization

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

Seeing Standards: A Visualization of the Metadata Universe

Jenn Riley, IU
The Metadata Universe

Seeing Stars
A Visualization of the Metadata Universe

Content: Jenn Riley
Design: Devon Diller

Work funded by the Indiana University Media Archival Design Lab (MAD Lab) and the Indiana University libraries.

Summary and Purpose
This first interactive metadata universe visualization is a result of collaborative efforts among institutions and community members. The project aims to create a comprehensive map of the metadata landscape, highlighting the various roles and relationships involved in metadata creation and use.

Key Features
- The universe is organized into domain communities, each representing a specific area of metadata focus.
- Each domain contains datasets, cultural objects, geospatial data, images, moving images, musical materials, scholarly texts, visual resources, and more.
- Interactive elements allow users to explore connections and relationships within the metadata universe.

Community Engagement
- This visualization serves as a platform for community engagement, encouraging collaboration and the sharing of metadata knowledge.
- It promotes awareness of the importance of metadata in preserving and sharing cultural heritage.

Acknowledgments
Many thanks to the institutions and community members who contributed to the creation of this visualization.

Copyright 2021-2022
This visualization is protected by copyright. No unauthorized use or reproduction is allowed without permission.

For more information, visit the Metadata Universe website at: metadatauniverse.com
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, ...
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.
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”

• Meritocracy: strong terms rise, weak terms decline
• Lessons from StackOverflow, Internet standards, and Wikipedia processes
<table>
<thead>
<tr>
<th>Term</th>
<th>Score</th>
<th>Consensus</th>
<th>Class</th>
<th>Contributed by</th>
<th>Last modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>oba</td>
<td>1</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>13 October 2014</td>
</tr>
<tr>
<td>structured data</td>
<td>2</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>7 October 2014</td>
</tr>
<tr>
<td>data</td>
<td>2</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>7 October 2014</td>
</tr>
<tr>
<td>metadata</td>
<td>2</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>7 October 2014</td>
</tr>
<tr>
<td>metadatum</td>
<td>2</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>6 October 2014</td>
</tr>
<tr>
<td>datum</td>
<td>0</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>6 October 2014</td>
</tr>
<tr>
<td>thing</td>
<td>1</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>6 October 2014</td>
</tr>
<tr>
<td>datum</td>
<td>0</td>
<td>medium</td>
<td>vernacular</td>
<td>Chris Patton</td>
<td>4 October 2014</td>
</tr>
<tr>
<td>authority</td>
<td>1</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>3 October 2014</td>
</tr>
<tr>
<td>identifier</td>
<td>2</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>3 October 2014</td>
</tr>
<tr>
<td>association</td>
<td>1</td>
<td>high</td>
<td>vernacular</td>
<td>John Kunze</td>
<td>2 October 2014</td>
</tr>
<tr>
<td>manufactureDate</td>
<td>1</td>
<td>high</td>
<td>vernacular</td>
<td>Austin Mathews</td>
<td>29 May 2014</td>
</tr>
<tr>
<td>manufacturer</td>
<td>0</td>
<td>low</td>
<td>vernacular</td>
<td>Austin Mathews</td>
<td>29 May 2014</td>
</tr>
<tr>
<td>manufactureDate</td>
<td>-1</td>
<td>low</td>
<td>vernacular</td>
<td>Austin Mathews</td>
<td>29 May 2014</td>
</tr>
<tr>
<td>price</td>
<td>1</td>
<td>medium</td>
<td>vernacular</td>
<td>Josh Rice</td>
<td>29 May 2014</td>
</tr>
<tr>
<td>testscorepercentage</td>
<td>-1</td>
<td>low</td>
<td>vernacular</td>
<td>Silvana Goberdhan-Vigle</td>
<td>29 May 2014</td>
</tr>
</tbody>
</table>
Using dictionary terms

- Search
  - found?
  - no
    - Create and Share
      - Edit
    - Compete
  - Use
    - Comment
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):

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](https://yamz.net)
  - Code at [https://github.com/nassar/yamz](https://github.com/nassar/yamz)
  - Persistent identifiers (ARKh) from EZID
- Thanks to Jane Greenberg, Chris Patton, the NSF DataONE project, and the Digital Library Federation

THANK YOU!