A Metadata Database and Usage Patterns in the LTER Network

ESIP Summer Meeting, July 16th, 2019
Agenda

**LTER-core-metabase**
- History.
- Design.
- How to start?

**EDI profile for EML**
- Why a profile?
- First findings.
Links
Group notes
http://tiny.cc/ESIPmetabase
1

Metadata databases
What’s a metadata DB

- Person or organization
- Bibliography
- Geographic location
- Taxonomy
- Temporal coverage
- Research project
- Search term
- Dataset
- Data table or file
- Method protocol
- Measurement
- Flag
- Definition
- Unit
- Data type
Why have a metadata DB

- Efficiency
  - Re-use
  - Update in one place to update all affected
- Consistency
- Flexibility in use (output via VIEWs)
“With redundancy comes the risk of inconsistency.”

"Complexity modeled is complication avoided"
You can use a metadata DB as...

...your dataset inventory
You can use a metadata DB as... the data source for making EML
You can use a metadata DB as...

...the source for a data catalog
LTER-core-metabase
History of metabases in LTER

Wade Sheldon wrote the first GCE Metabase in 2001.
Pre-release

LTER-core-metabase v.0.9

v0.9.20

atn38 released this 1 hour ago
But, there's more OneBigERD™
<table>
<thead>
<tr>
<th>Category</th>
<th>Frequent</th>
<th>Infrequent</th>
<th>Rare</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>title</td>
<td>New entries to Lists</td>
<td>Controlled vocabularies:</td>
</tr>
<tr>
<td></td>
<td>entity</td>
<td>- keyword</td>
<td>- file types</td>
</tr>
<tr>
<td></td>
<td>Xrefs</td>
<td>- missing code</td>
<td>- EML and network</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- people</td>
<td>controlled vocabs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- people ID (ie ORCiD)</td>
<td>- Unit dictionary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- protocols</td>
<td>- keyword thesauri</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- sites</td>
<td>- taxonomic providers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- taxa</td>
<td></td>
</tr>
</tbody>
</table>
Normalization

The DB models the 1-1, 1-many and many-many situations in EML.

For many-to-many relationships we use cross-reference tables.
Coming next version

- Additions for EML 2.2
  - Semantic annotation
  - Funding structure
  - taxonId
- Live migration path SQL
- Flexible methods view

github project board prioritizes planned enhancements
Cool! How do I start?

Get PostgreSQL Desktop or server

Download https://github.com/lter/LTER-core-metabase

Load The installation

Populate Your content

Export To EML
Use a tool that works for you

DBeaaver
pgAdmin
psql
R
oXygenXML Editor
Exposing your content

Your metadata content

- LTER-core-metabase
- VIEWs - Abstraction Layer
- EML

Exposed content

- SQL
- R

> library(MetaEgress)
Support

Comments peppered throughout the database (in-situ help)
Plenty of documentation on Github:
- Installation guide
- Content population
- What information goes where
- How to generate EML

And us site IMs, of course.
Discussion

Talking points:
● when not to do this?
● why migrate?
● ongoing development
schemaSpy: explore DB model

To explore in detail:
http://tiny.cc/metabaseSchema
Takeaways

• What did you get out of this
Credits to:

Wade Sheldon (GCE)
Mary Martin (HBR)
Sarah Elmendorf (NWT)
Thanks!

Join our channel at lter.slack.com/#metabase
Links and notes at tiny.cc/ESIPmetabase
Use these icons :)}