Do Community Recommendations Improve Metadata Completeness?

Sean Gordon (scgordon@hdfgroup.org)[1], Ted Habermann[1], Matthew B. Jones[2], Ben Leinfelder[2], Bryce Mecum[2], Lindsay A. Powers[3], and Peter Slaughter[2]

1. The HDF Group, 2. National Center for Ecological Analysis and Synthesis 3. United States Geological Society

IN23C-1785

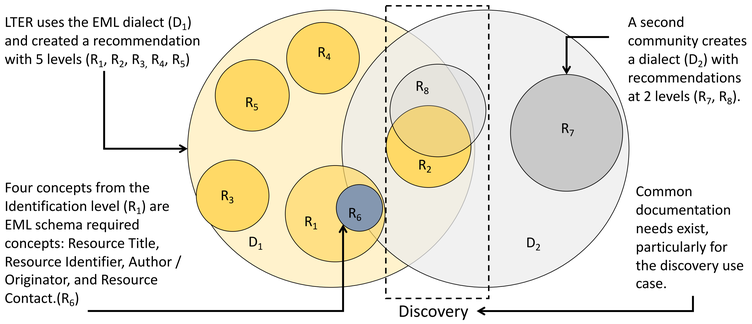
Background

Many communities use the term language when they describe their metadata. This approach focuses attention on differences between communities. Using the term dialect focuses attention on the common concepts and goals of scientific documentation.

Recommendations reflect the experiences and documentation needs of a community. Recommendations are an important mechanism for sharing those experiences and knowledge so that quality metadata is created throughout the community.

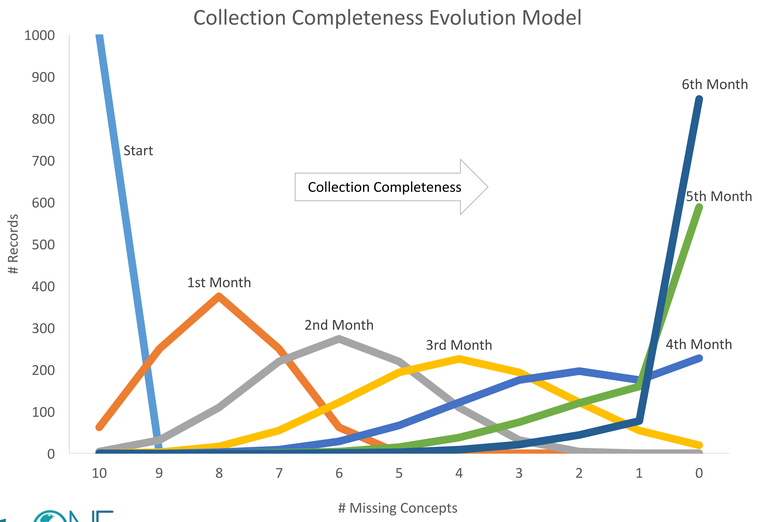
LTER and EML

The Long Range Ecological Network created the LTER Recommendation for Completeness to help guide the creation of Ecological Metadata Language records. There are five levels in the LTER recommendation: Identification, Discovery, Evaluation, Access, and Integration. All levels of the LTER recommendation are subsets of concepts in the EML dialect.



Premise

The LTER Completeness Recommendation includes documentation concepts the LTER community considers important for creating quality metadata. Ideally the completeness of LTER metadata would improve over time. The graph below illustrates how metadata collections evolve towards completeness. The model output improves 50% of 1000 records by one concept each time step. The visualization displays every fourth time step to simulate a 6 month period of collection development.



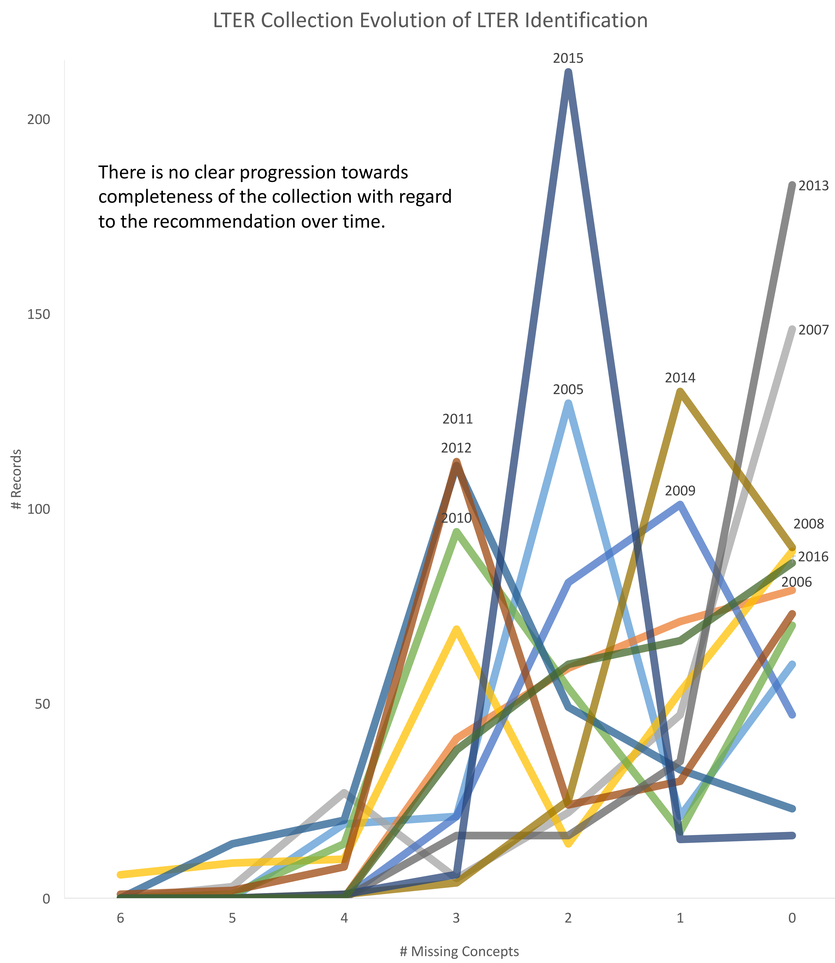
Process

- Sampled 250 LTER metadata records from DataONE to create collections for each year 2005-2016.
- Measured conceptual content existence in each record.
- Analyzed results for LTER Completeness in the Recommendations Analysis Dashboard[1] for each years collection.
- Compared analyses across time periods using collection evolution[2] analysis and a variation that focuses on individual concept completeness.
- Compared heterogeneity of each collection to completeness using signature score groups[1] and a distribution of completeness for each year.

Limitations

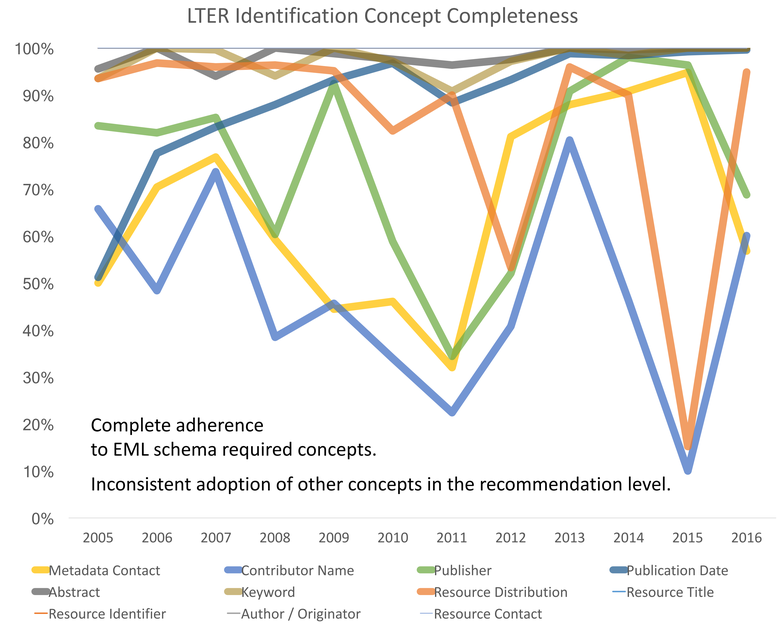
- Not a set of records through time.
- Sampling proportion vs sampling size.
- No ethnographic perspective.

Does the collection become more complete with time?



There is no clear progression towards completeness of the collection with regard to the recommendation over time.

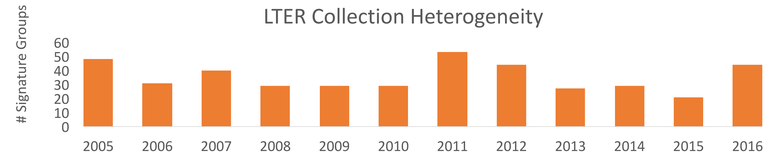
Are there recommendation concepts the community values more?



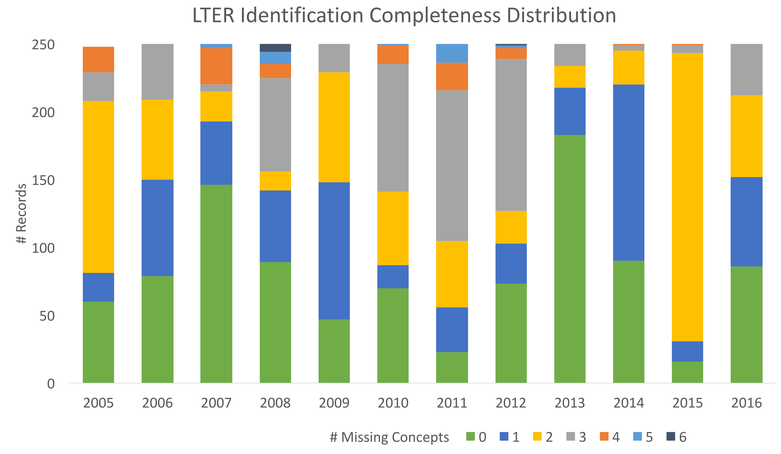
Complete adherence to EML schema required concepts.

Inconsistent adoption of other concepts in the recommendation level.

Do more heterogeneous collections have less complete metadata?



Heterogeneity has no noticeable effect on the completeness of a collection.



1. See bottom third of Evaluating and Evolving Metadata in Multiple Dialects, IN23C-1781 for a description 2. See top right of Evaluating and Evolving Metadata in Multiple Dialects, IN23C-1781 for a description

NSF-DIBBS Award 1443062