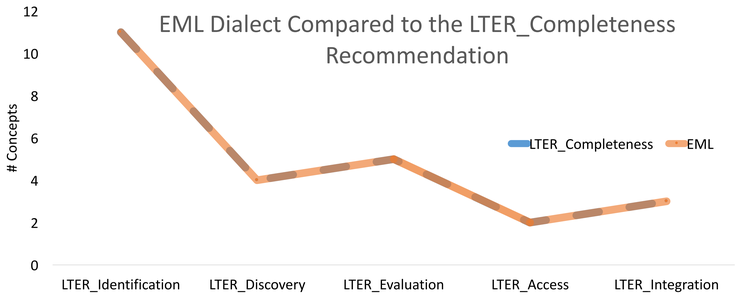
Do Community Recommendations Improve Metadata Completeness? (IN23C-1785)

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

Background

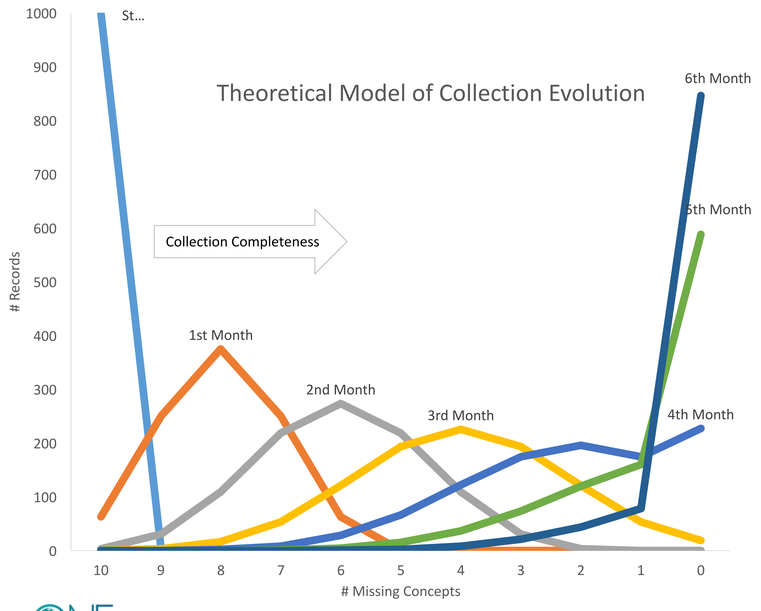
The Long Range Ecological Network created the LTER Recommendation for Completeness to help guide the creation of Ecological Markup Language records.

There are five documentation use cases in the LTER recommendation: Identification, Discovery, Evaluation, Access, and Integration. As shown below, EML and LTER have no concept gap.



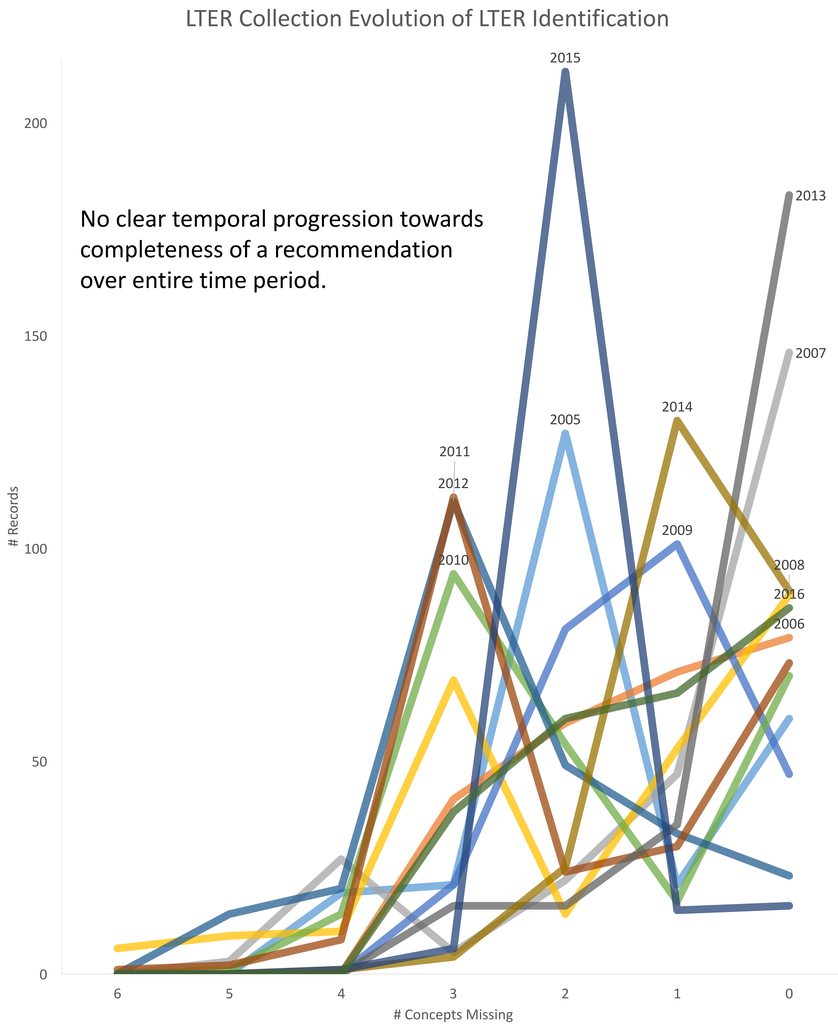
Premise

The LTER Completeness Recommendation includes concepts the LTER community considers important for creating quality metadata. Ideally the completeness of LTER metadata should improve over time. The graph below uses a theoretical model to illustrate how metadata can become more complete over time. This model output improves 500 out 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.

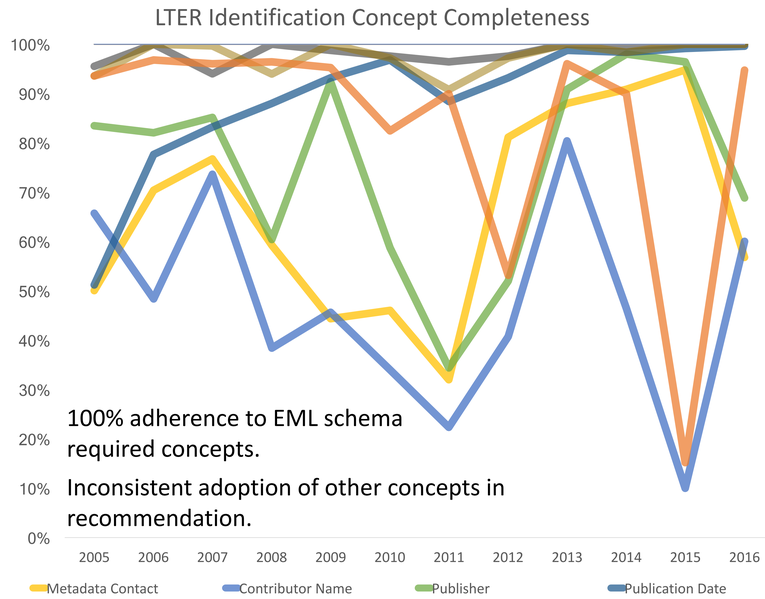


Process

- Utilized a python sampling tool that leveraged DataONE's SOLR index to identify and create XML collections of 250 LTER metadata records from each year 2005-2016.
- Used XSL rubrics to determine conceptual content in each record.
- Analyzed results for completeness of 25 concepts in the Recommendations Analysis Dashboard[1] for each years collection.
- Compared analyses across time periods using collection evolution[2] analysis variations.
- Compared heterogeneity of each collection to completeness using signature score groups.

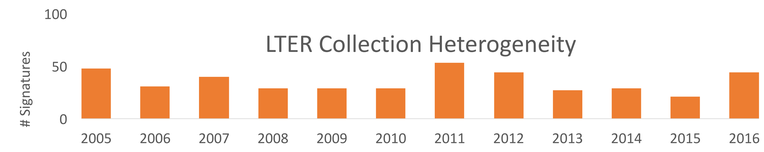


No clear temporal progression towards completeness of a recommendation over entire time period.

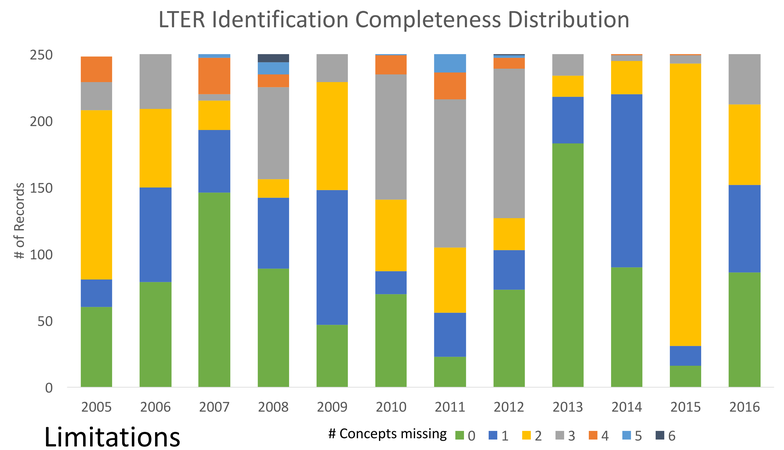


100% adherence to EML schema required concepts.

Inconsistent adoption of other concepts in recommendation.



Collection heterogeneity has no clear effect on completeness.



Limitations

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

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