

Sean Gordon (scgordon@hdfgroup.org)¹, Ted Habermann¹, Matthew B. Jones², Ben Leinfelder², Bryce Mecum², Lindsay A. Powers³, and Peter Slaughter²
 1. The HDF Group, 2. National Center for Ecological Analysis and Synthesis 3. United States Geological Society

Background

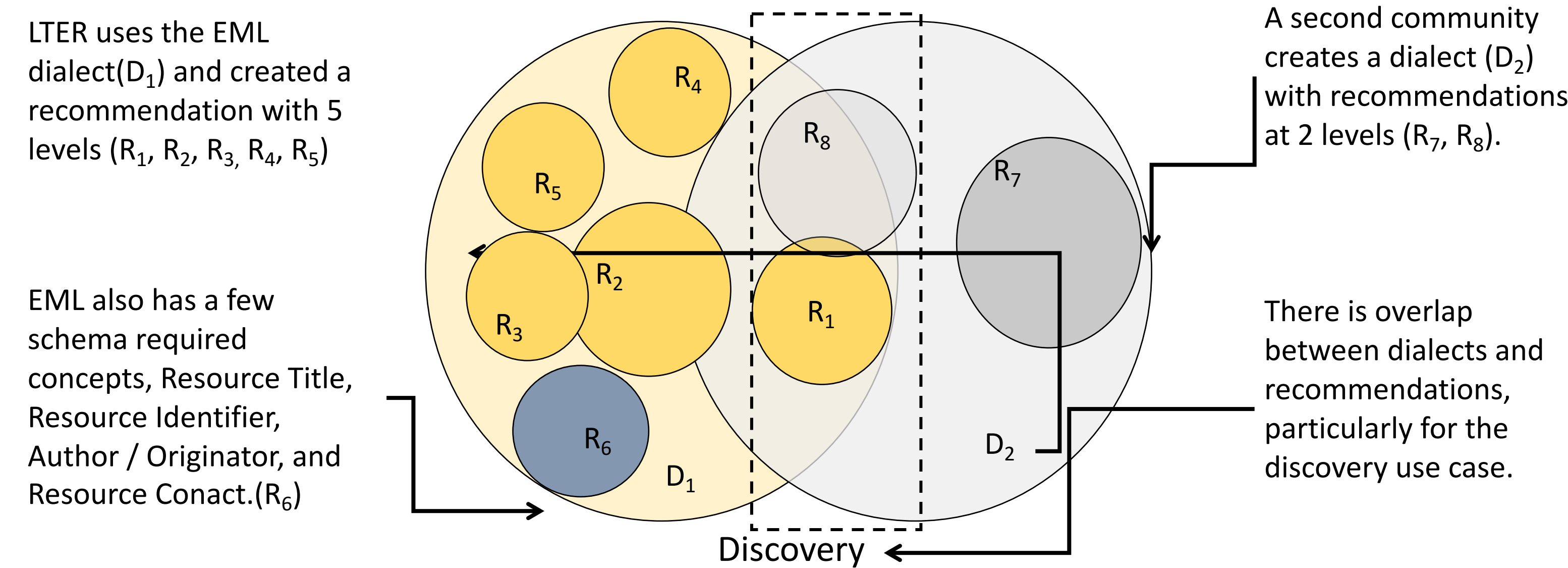
Many communities use the term "standard" when they describe their metadata and, as a result, there are many existing "standards". This approach focuses attention on differences between communities. We use the term "dialect" to focus attention on common concepts and goals.

Recommendations and Dialects:

Recommendations reflect community experiences and documentation needs. Communities have common documentation needs, so recommendations overlap, particularly for the discovery use case. Sharing recommendations is an important mechanism for sharing those experiences and community knowledge.

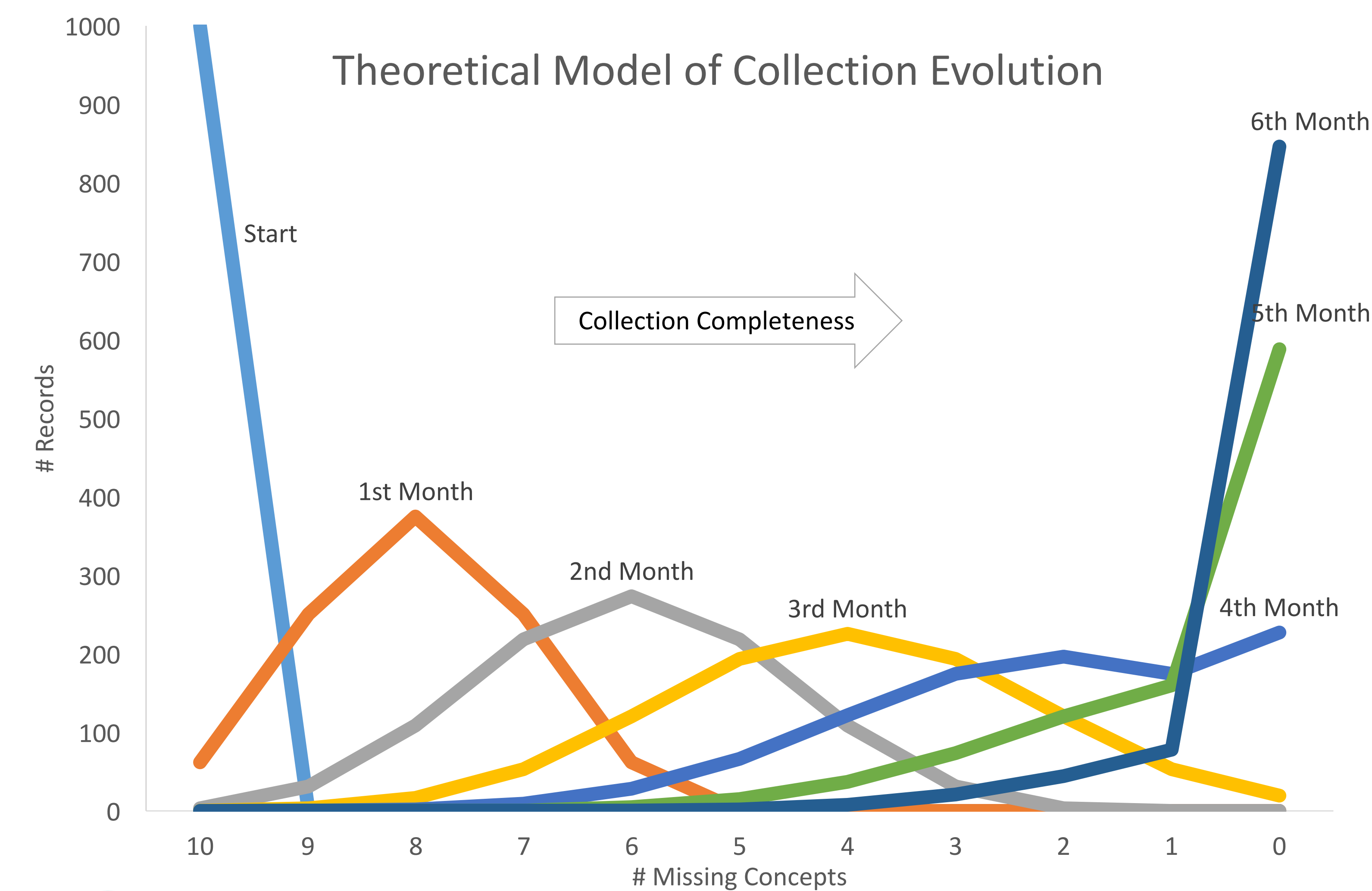
LTER and EML

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



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. The 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.



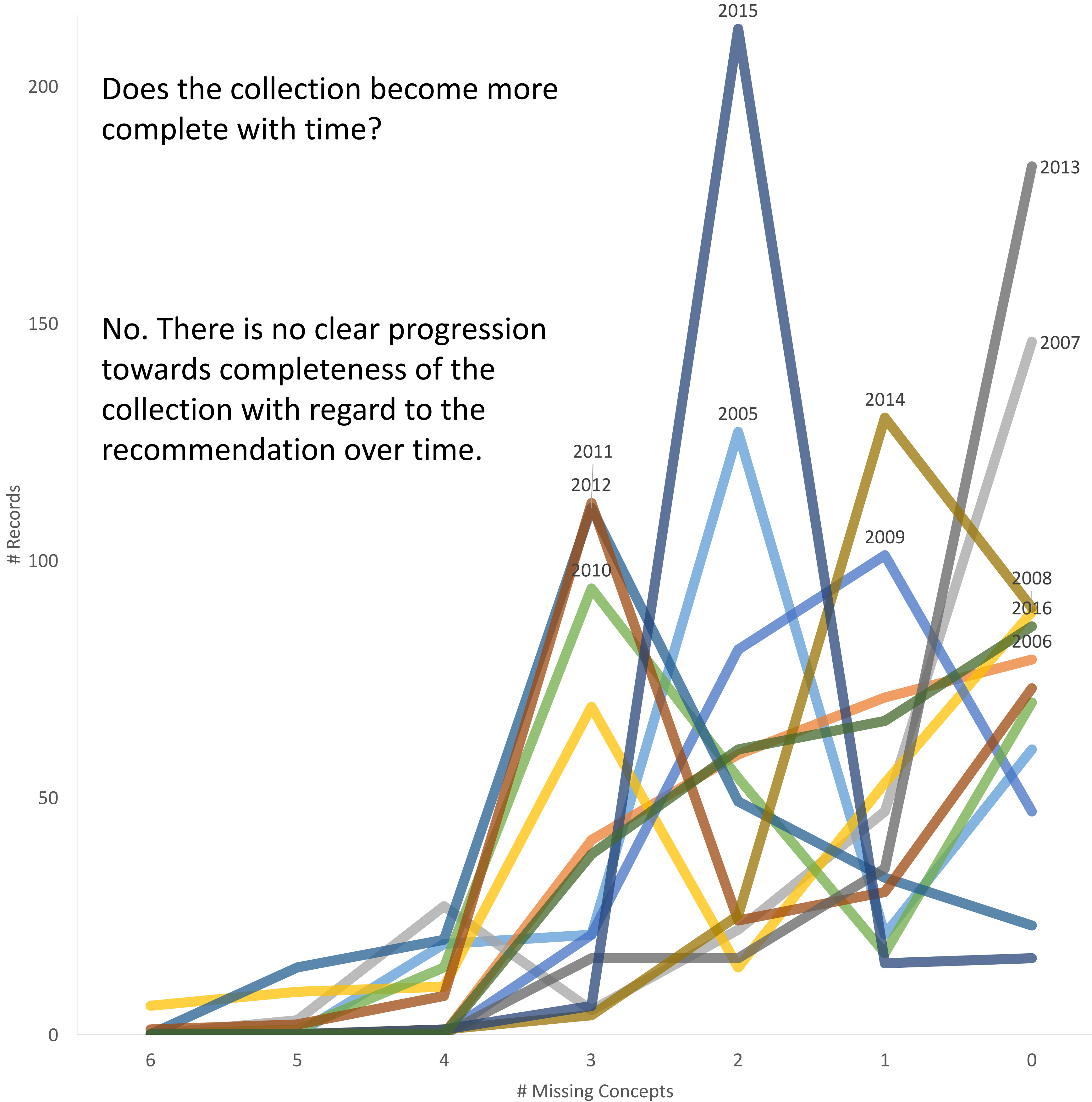
Process

- 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 Dashboard¹ for each years collection.
- Compared analyses across time periods using collection evolution² analysis and a variation that focuses on individual concept completeness.
- Compared heterogeneity of each collection to completeness using signature score groups¹ and another view of the distribution of completeness

Limitations

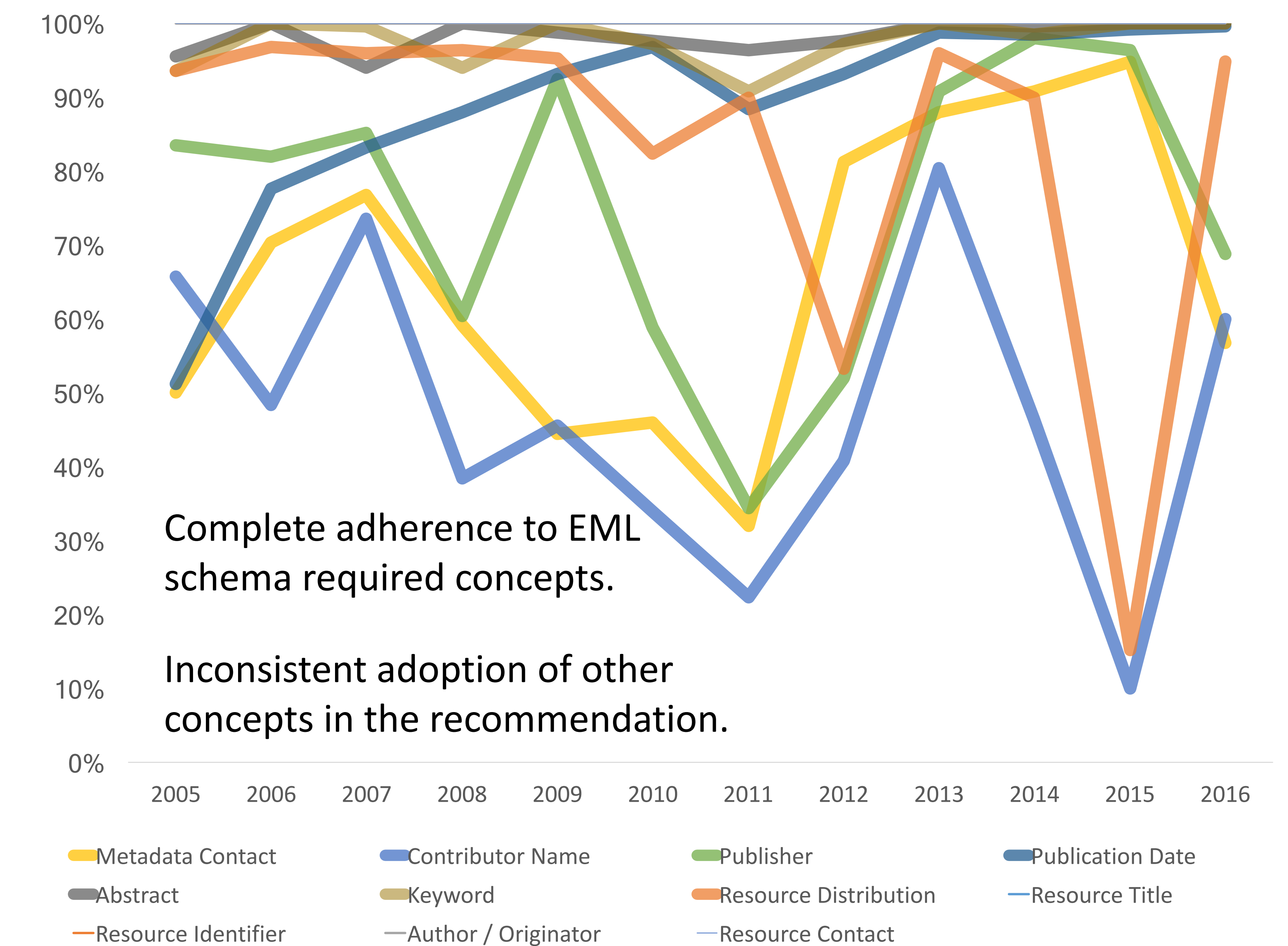
- Not a set of records through time.
- Sampling proportion vs sampling size.
- No ethnographic perspective.

LTER Collection Evolution of LTER Identification

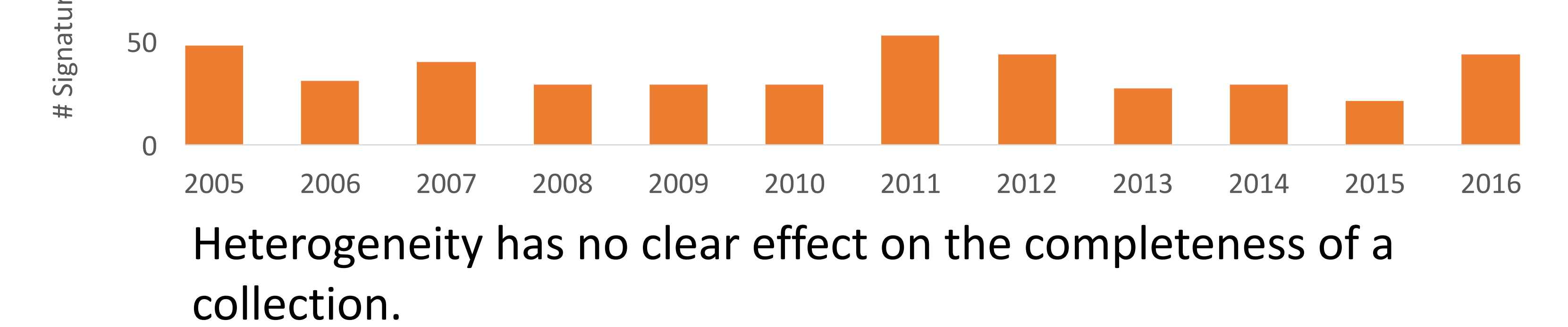


LTER Identification Concept Completeness

Do LTER metadata creators value a tacit subset of the concepts?



LTER Collection Heterogeneity



LTER Identification Completeness Distribution

