### Observation 1

LTER uses every concept in the recommendation. No other DataOne member node’s collection sample contained every concept.

### Conclusion 1

LTER has the most complete collection coverage because it is the only collection to contain all concepts in the recommendation.

### Observation 2

86%of the LTER sample are in the top 17% most complete signature groups

### Conclusion 2

LTER record more likely to be more complete than a record from any other member node. If we look at the entire collection of EML, there are many signature score groups. In fact, many of the signatures are unique within the collection, while other signatures are over 10% of the entire DataONE sample set. If you pair the signature score record groups with the member nodes from DataONE, you can see that most the LTER records in the metadata-set are distributed towards complete signature scores. In the top 17% of the entire sample set, 86% of the LTER sampled records occur. The following visualization shows the collection convergence towards the top of the most complete metadata records in the DataONE sample. LTER is the only EML collection with records that are complete for the Identification level. LTER and CLOEBIRD are the only collections using EML that contain records that are complete with respect to the Discovery level. GOA, GLEON, KNB, SANPARKS, PISCO, TFRI, and LTER have records that are Evaluation level complete. Over half of the LTER collection is Evaluation level complete. GOA, GLEON, KNB, IOE, SANPARKS, TFRI, USANPN, and LTER all have records that are complete with respect to the Access Level. LTER is the only collection with a record that is Integration level complete. LTER is the only collection that has a shining example of each LTER Recommendation Level.

### Observation 3

LTER contributes most of the Shining Examples.

### Conclusion 3

LTER more familiar with concepts and how to document.

### Observation 4

By level, LTER does not have a higher completeness percentage than all other member nodes LTER is not more complete on unweighted average either

### Conclusion 4

LTER is not favored as highly as a collection that contains few moderately complete records. LTER is more complete than the average of all DataONE member nodes that use EML including itself.

### Observation 5

Homogeneity leads to more complete concepts in a collection. Collections that have a high degree of homogeneity are also more likely to contain more unused concepts

### Conclusion 5

Homogeneity can be bad for completeness. CDL and TERN are examples of this.

### Questions

It appears CSDGM collections are more complete with respect to LTER. This case is only made more strongly when the dialect limitations are handicapped to dialect maximums for the levels. What are the common concepts between LTER and the FGDC recommendation that likely informed the creation of the CSDGM collections?

What effect does time have on record completeness? The LTER sample set may all be from 2005. Would new records from succeeding years be more complete? By improving the sampler to return a sample set published in a specific year it is possible to study this.

Why is LTER the only collection complete with respect to the Identification Level, but the concept occurrence percentage average of ESA is higher? LTER is made up of many sites itself. Perhaps this makes the sample set heterogeneous in that some sites have more experience with creating complete metadata than others, as there are a significant portion of the LTER sample that are complete or missing one or two concepts. (86%) so the rest must not be very complete. Are there multiple occurrences of metadata completeness evolution through time in the member node LTER that can be documented by creating a sample set with collections made up of records published in a specific time period?

Even though LTER has the most complete records in the DataONE sampling, the collection as a whole is not the most complete for any concept? Or recommendation level? Why are there a large number of LTER signature sums in the sample set that are less complete than the average for DataONE and EML at DataONE? It seems obvious that some records produced at LTER benefit from the recommendation. Do not all of the metadata record producers get the same level of guidance and use the same infrastructure to create their metadata?

This is not the case. Metadata is created at a number of sites. Can these sites be treated as member nodes in a new analysis and show a stronger case for collection evolution towards completeness through community usage of a recommendation?