## The LTER Recommendation

As the ecological research community gained experience with EML, it became clear that many metadata records were not complete or consistent enough to serve important community requirements. To address this problem, a group of LTER metadata experts developed a set of recommendations to help guide the creation and improvement of EML metadata records (EML Best Practices, n.d.). The LTER recommendation includes five levels: Identification, Discovery, Evaluation, Access, and Integration, each of which recommends specific elements designed to provide information about the dataset for a specific use case, or need.

The LTER recommendations were well publicized and supported in the LTER community, so we might expect that LTER metadata records are more complete with respect to these recommendations than other metadata collections. The DataONE Metadata Repository includes many metadata collections and thus provides an excellent opportunity for exploring the impact of the LTER recommendations across related communities.

We explore this impact in two ways. First, we compare the completeness of the LTER metadata collection in the DataOne metadata repository to collections from other ecological research groups that use the EML dialect. Second, we extend that comparison to metadata collections in DataONE documented in the CSDGM dialect. We accomplish both comparisons through a conceptual abstraction layer that provides a method of crosswalking dialect and recommendation specific elements. For example, the concept “Resource Title” is found in both the EML and CSDGM dialects at a specific location in the resource’s documentation. By connecting the structural locations, or dialect definitions in multiple dialects, conceptual recommendations can be measured across dialects. The dialect definitions for the LTER recommendation’s concepts in EML and CSDGM

The relationship between dialects and recommendations is illustrated in Figure 1. LTER uses the EML dialect (D1) and their recommendation had five levels (R1, R2, R3, R4, R5). All the concepts in the recommendation are included in the dialect. In some cases, the recommended concepts are required by the XML schema used to implement the dialect, illustrated as special case R6. There are many similar examples of metadata dialects and recommendations. When another community creates a second dialect (D2) with recommendations at two levels (R7, R8), their recommendations are all included in the second dialect.



We are interested in situations where documentation needs of different communities and dialects overlap. Figure one shows overlaps between D1 and D2 as well as R2 and R8. Such overlap is common in areas with clear common needs, such as data discovery, but can be less common as the metadata becomes more specialized. To identify these overlaps and do cross-dialect comparisons, the recommendations must be described in terms of fundamental documentation concepts that can be identified in multiple dialects.

The concepts included in the LTER Recommendation are listed in Table 1. The concepts that are included

Table 1 - Conceptual description of the LTER recommendations

|  |  |  |
| --- | --- | --- |
| Recommendation Level | # Concepts | Concept Titles |
| Identification | 11 | Resource Identifier, **Resource Title**, **Author / Originator,** **Metadata Contact**, **Contributor Name**, Publisher, **Publication Date**, **Resource Contact**, **Abstract**, Keyword, Resource Distribution |
| Discovery | 4 | Spatial Extent, Taxonomic Extent, **Temporal Extent,** Maintenance |
| Evaluation | 5 | **Resource Use Constraints**, Process Step, Project Description, Entity Type Definition, Attribute Definition |
| Access | 2 | **Resource Access Constraints**, Resource Format |
| Integration | 3 | Attribute List, Attribute Constraints, Resource Quality Description |

A second requirement for complete cross dialect comparisons is that the concepts must occur in both dialects. Of course, all the LTER recommendations are in the EML dialect, but they may not be included in other dialects. This is illustrated in Figure 1.

The conceptual design of the recommendation allows records in other dialects to be analyzed by the same recommendation. These concepts are The blue concept names are concepts that appear in the FGDC recommendation as well. 10 of the 25 concepts in the LTER recommendation are present in the FGDC recommendation. Many other concepts are closely related to concepts in the FGDC recommendation, such as Keyword and Spatial Extent. FGDC calls for Theme Keyword and Bounding Box.

Table 0 - Conceptual description of the recommendations

|  |  |  |
| --- | --- | --- |
| Recommendation Level | # Concepts | Concept Titles |
| Identification | 11 | Resource Identifier, Resource Title, Author / Originator, Metadata Contact, Contributor Name, Publisher, Publication Date, Resource Contact, Abstract, Keyword, Resource Distribution |
| Discovery | 4 | Spatial Extent, Taxonomic Extent, Temporal Extent, Maintenance |
| Evaluation | 5 | Resource Use Constraints, Process Step, Project Description, Entity Type Definition, Attribute Definition |
| Access | 2 | Resource Access Constraints, Resource Format |
| Integration | 3 | Attribute List, Attribute Constraints, Resource Quality Description |

## Comparison of DataONE dialects and the LTER Recommendation

Each level of the LTER recommendation contains metadata concepts needed for a specific documentation use case. As you can see in the chart below, EML contains every concept in each of these levels while CSDGM is missing one concept in each level except for Access. This means that a record at the CSDGM dialect maximum will never contain all the concepts in any of the levels except for access. CSDGM records can only be complete with respect to the CSDGM dialect maximum. CSDGM records will never be complete because there are concepts the dialect doesn’t contain. The *dialect maximum* is the number of concepts from a recommendation that a dialect contains. For example Mercury and BDP are other dialects in DataONE that extend CSDGM to contain taxonomic information in the case of BDP, or an identifier for the resource in Mercury’s case. In these cases, organizations have extended CSDGM when it did not contain the concepts they needed to describe in their metadata. The dialect maximum for BDP in the Discovery level of the LTER Recommendation is the same as the *recommendation maximum*, or count of concepts in a recommendation level.