Best Practices for Dataset Metadata in Ecological Metadata Language (EML)

Version 3

November 2017

This document is managed and distributed by the Environmental Data Initiative (EDI)

https://environmentaldatarepository.org

info@environmentaldatarepository.org

Please cite as:

Best Practices for Dataset Metadata in Ecological Metadata Language (EML Best Practices V3). 2017. Environmental Data Initiative.

ORGANIZATION AND CONVENTIONS OF THIS DOCUMENT

AUDIENCE

This document is intended for data managers. It assumes that readers are familiar with

- the basic structure of an XML document, and the ability to edit in an XML editor like OxygenXML or XMLSpy.
- the process for contributing data to a repository. If you reached this document from a repository's help-page, contact them for more information.

FONTS AND TYPEFACE

Numbered examples of EML nodes are in fixed-width font:

```
<?xml version="1.0" encoding="UTF-8"?>.
```

XML element and attribute names, XPath and references to element names in text are in bold face. Single element names are surrounded by angle brackets, as they appear in XML.

<dataTable>
/eml:eml/@packageId

Some recommendations have special context, e.g., an XML element or attribute may be requested by a community (e.g., LTER), or required by the EDI repository (but not by other repositories).

Context notes: Recommendations for EML usage in a specific context are called "context notes", and are placed in separate paragraphs, in italic.

DEFINITIONS

EML preparer: the person responsible for "building" the EML metadata record. Generally, this is a data manager working with a project or physical site that produces data.

Contributor: the research project contributing the data package, e.g., an LTER or OBFS site, or a Macrosystems project. Generally, the "EML preparer" works with or for the "Contributor."

Data package: the EML metadata together with its entity or entities. This is generally the unit housed in repositories. We use this term to avoid confusion with the EML element "**dataset**".

OTHER EML RESOURCES

Some sections refer to further information or tools. These can be found on the EDI website, under "Resources and How To...", at https://environmentaldatarepository.org

Table of Contents

I. INTRODUCTION

This document contains current "Best Practice" recommendations for EML content for metadata related to ecological and environmental data. It is intended to augment the EML schema documentation (normative documents) for a less-technical audience. The current version (v3, 2017) is one component of several resources available to EML preparers. The recommendations are directed towards the following goals:

- Provide guidance and clarification in the implementation of EML for data packages
- Minimize heterogeneity of EML documents to simplify development and re-use of software built to ingest it
- Maximize interoperability of EML documents to facilitate data synthesis

At time of this document's publication (late 2017), the version of EML currently in production was EML.2.1.1. EML 2.2.0 is anticipated within the next year. Contact EDI for more information.

HISTORY

EML Best Practice recommendations have evolved over time. The most active contributors have been members of the LTER Information Managers Committee in multiple working groups and workshops. EML has been widely used for several years with multiple applications written against it, and the community has had the opportunity to observe the consequences of many content patterns. As much as possible, recommendations have been aligned with those experiences, as well as with the capability of data contributors.

Timeline and Previous Revisions
- 2017 Best Practices for Dataset Metadata in EML v3 (this document)
- 2016 EDI inception, see http://environmentaldatainitiative.org
- 2011 EML Best Practices for LTER sites v2
- 2008 EML 2.1 release
- 2004 EML Best Practices for LTER sites
- 2003 LTER adopts EML as network exchange standard

Contributors, including LTER EML Best Practices Working Groups and workshops in 2003, 2004, 2010 (alphabetical order): Dan Bahauddin, Barbara Benson, Emery Boose, James Brunt, Duane Costa, Corinna Gries, Don Henshaw, Margaret O'Brien, Ken Ramsey, Inigo San Gil, Mark Servilla, Wade Sheldon, Philip Tarrant, Theresa Valentine, John Vande Castle, Kristin Vanderbilt, Jonathan Walsh, Yang Xia

GENERAL RECOMMENDATIONS

Following are general best practices for handling EML dataset metadata:

Metadata Distribution

Do not publicly distribute EML documents containing elements with incorrect information, e.g., as a workaround for missing metadata or to meet validation requirements. Pre-publication drafts, or EML produced for demonstration or testing purposes should be clearly identified as such and not contributed to public archives, because these are passed on to large-scale clearinghouses. For previews of drafts or handling test and demonstration data packages, consult your repository to learn about options.

Data Package Identifiers

Metadata and data set versioning are controlled by the contributor, and so identifiers are tied to local systems. Many repository systems that accept EML-described data support principles of immutable metadata and data entity versioning. EML has elements to contain package identifiers, although these may also be assigned externally. It is the responsibility of the submitters to understand the practices of their intended repository when using identifiers.

High-priority Elements

- To support locating data by time, geographic location, and taxonomically, metadata should provide as much information as possible for the data package, in the three **<coverage>**; elements of **<temporalCoverage>**; (when), **<geographicCoverage>**; (where) and **<taxonomicCoverage>** (what).
- For a potential user to evaluate the relevance and usability of the data package for their research study or synthesis projects, metadata should include detailed descriptions in the **<project>**,**<methods>**,**<protocols>**, and **<intellectualRights>** elements.

II. CONTENT RECOMMENDATIONS FOR ELEMENTS AND ATTRIBUTES

The root element: <eml:eml>

This element is the root element in all EML documents. The XPath notation is:
/eml:eml

The root element holds two important parts, both of which are optional, but recommended.

@schemaLocation (XML attribute)

This attribute is this location (XPath):

/eml:eml/@schemaLocation

The schemaLocation attribute tells a processor the name of the schema to which the EML document belongs and where to find it. Most repositories check schema compliance when data packages are deposited, but it is highly recommended that data managers know how and where to specify the schema that their metadata document should adhere to. This way, they can validate their own work in progress, e.g., through an XML editor like OxygenXML.

@packageId (XML attribute)
This attribute is found at this location (XPath):
/eml:eml/@packageId

As outlined elsewhere, EML preparers should manage unique identifiers and versioning at the local level (see **@system** discussion below). The **packageId** attribute can be used to contain the same identifier as is used by the repository.

See Section III for other information about EML documents in Metacat.

Context Note: The packageID attribute is required in all EML documents submitted to EDI. It is entered into the repository software, and the format is standardized to three parts: scope, package-number, revision. The scope should be "edi" unless another scope is justified by prior arrangement. See Example 1.

Top Level Elements
An EML **dataset** is composed of up to three elements under the root element (**<eml:eml>**):

<access>
<dataset>
<additionalMetadata>

id, system and scope (XML attribute-group)
This attribute group can be used on these EML elements:
<access>
<dataset>
<creator>
<associatedParty>

<contact>
<metadataProvider>
<publisher>
<coverage>
<geographicCoverage>
<temporalCoverage,>
<taxonomicCoverage>
<distribution>
<software>
<citation>
<protocol>
<project>
<dataTable>
<otherEntity>
<spatialRaster>
<spatialReference>
<spatialVector>
<storedProcedure>
<view>
<attribute>
<constraint>

These three attributes are found as a group and are usually optional. The primary use of the id attribute is as an internal reference, hence each id must be unique within one EML document. E.g.,. a **<creator>** must have a different id than a **<dataTable>**. And if the same person appears in seberal places (at **dataset/creator**, **protocol/creator** or **project/creator**, the same id cannot be repeated, so either the content of the id must be changed or a reference used for repeated instances.

This restriction can cause problems when content is drawn from a system with IDs (e.g. a personnel database), and is under consideration by the EML developers. Ideally the three attributes would work together. The **scope** attribute can have one of two values, "system" or "document". It is preferred that when the scope is set to "system", that the **system** attribute defines the ID-system, the **id** attribute content is (presumably) from that system.

Currently, a reasonable general practice should be to define a **system** on the **<eml:eml>** element and set it to the site (but not set the system attribute at any other level), and to set **scope**="document" on elements other than **<eml:eml>**.

Example 1: attributes packageId, id, system, and scope

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:ds="eml://ecoinformatics.org/dataset-2.1.0"
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns:eml="eml://ecoinformatics.org/eml-2.1.0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xmlns:stmml="http://www.xml-cml.org/schema/stmml"
    xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0
https://nis.lternet.edu/eml-2.1.0/eml.xsd"
```

```
packageId="knb-lter-fls.21.3"
system="FLS"
scope="system">
```

access

The data package title element is found at this location (XPath):

/eml:eml/access
/eml:eml/[entityType]/physical/distribution/access

<access> contains a list of rules defining permissions for this metadata record and its data entity.   Values must be applicable by the system where data is stored. Many repositories follow the KNB system of using access control format that conforms to the LDAP "distinguishedName (dn)" for an individual, as in "uid=FLS,o=LTER,dc=ecoinformatics,dc=org".

As of EML 2.1.0, **<access>** trees are allowed at two places: as the first child of the **<eml:eml>** root element (a sibling to **<dataset>**) for controlling access to the entire document, and in a **physical/distribution** tree for controlling access to the resource URL. With the exception of certain sensitive information, metadata should be publicly accessible. The **<access>** element is optional, and if omitted, the repository may presume that only the dataset submitter will be allowed access.

Example 2: access
```
<access authSystem="knb" order="allowFirst" scope="document">
    <allow>
        <principal>uid=FLS,o=lter,dc=ecoinformatics,dc=org</principal>
        <permission>all</permission>
    </allow>
    <allow>
        <principal>public</principal>
        <permission>read</permission>
    </allow>
</access>
```

dataset

This element is found at these locations (XPath):
/eml:eml/dataset

Under **<dataset>**, the following elements are available. Some are optional, but if they appear, this order is enforced by the schema. Generally, the recommendations are presented here in this order, with the exception of elements related to people and organizations which are grouped together so that the distinctions between the uses of

those elements are clear. Elements that can appear at different levels within an EML file are discussed at their first appearance, or highest level.

<alternateIdentifer>
<shortName>
<title>
<creator>
<metadataProvider>
<associatedParty>
<pubDate>
<language>
<series>
****
<keywordSet>
<additionalInfo>
<intellectualRights>
<distribution>
<coverage>
<purpose >
<maintenance>
<contact>
<publisher>
<pubPlace>
<project>

These elements are then followed by one or more elements for the data entity (or entities), designated by choosing:

[**dataTable** | **spatialRaster** | **spatialVector** | **storedProcedure** | **view** | **otherEntity**]

alternateIdentifier
The alternateIdentifier element is found at this location (XPath):
/eml:eml/ dataset/alternateIdentifier
/eml:eml/ dataset/[entity]/alternateIdentifier

The contributing organization's local data set identifier should be listed as the EML **<alternateIdentifier>**, particularly when it differs from the "**packageId**" attribute in the **<eml:eml>** element. The **<alternateIdentifier>** should also be used to denote that a package belongs to more than contributing organization by including each individual ID in

a separate **<alternateIdentifier>** tag. At the entity level, the **<alternateIdentifier>** should contain an alternate name for the data table (or other entity) itself (see additional comments under entities, below.)

title (dataset)

The dataset title element is found at this location (XPath):

/eml:eml/ dataset/title
/eml:eml/method/methodStep/protocol/title
/eml:eml/project/title

The dataset **<title>** should be descriptive and should mention the data collected, geographic context, research site, and time frame (what, where, and when).

Example 3: dataset, alternateIdentifier, shortName, title

```
<dataset id="FLS-1" system="FLS" scope = "system">
    <alternateIdentifier>FLS-1</alternateIdentifier>
    <shortName>Arthropods</shortName>
    <title>Long-term Ground Arthropod Monitoring Dataset at Ficity, USA from 1998
to 2003</title>
```

People and Organizations (Parties)

People and organizations are all described using a "ResponsibleParty" group of elements, which is found at these locations (XPath):

/eml:eml/dataset/creator
/eml:eml/dataset/contact
/eml:eml/dataset/metadataProvider
/eml:eml/dataset/associatedParty
/eml:eml/dataset/publisher
/eml:eml/dataset/project/creator
/eml:eml/dataset/method/methodStep/protocol/creator

General recommendations: When using **<individualName>** elements anywhere within an EML document, names should be constructed with English alphabetization in mind. Many sites have found that maintaining full contact information for every creator is impractical, however a few important contact information should be kept up to date (see below). If a name includes a suffix, it should be included in the **<surName>** element after the last name.

It is recommended to include complete contact information for a permanent role that is independent of the person holding that position. For example, for an information manager,

site contact, pay careful attention to phone number and use an e-mail alias that can be passed on. (see below, under**<contact>**).

With the advent of general identifiers such as ORCIDs, the text in the **<address>**, **<phone>**, and **<onlineURL>** elements may become unnecessary for individuals and so is optional if and an individual's ORCID is included. **<electronicMailAddress>** is recommended to simplify contacting responsible parties. See the **<userId>** field. ORCID identifiers are not yet available for organizations, so **<address>**, **<phone>**, and **<onlineURL>** elements should be included for them. In the examples, these elements are included for completeness.

<userId>
This element is found at this location (XPath):
/eml:eml/dataset/creator/userId
/eml:eml/dataset/contact/userId
/eml:eml/dataset/metadataProvider/userId
/eml:eml/dataset/associatedParty/userId
/eml:eml/dataset/publisher/userId
/eml:eml/dataset/project/creator/userId
/eml:eml/dataset/method/methodStep/protocol/creator/userId

The optional **<userId>** field holds identifiers for responsible parties from other systems. This element is repeatable so that multiple systems can be referenced. EML prepares should contact the system they plan to use to learn their preferences for inclusion in metadata. The examples here are for ORCID identifiers, and that organization has asked that its full URI be used as both the **system** attribute, and as the head of the identifier itself.

Example 4: creator
```
<creator id="org-1" system="FLS" scope="system">
    <organizationName>Fictitious LTER Site</organizationName>
    <address>
        <deliveryPoint>Department for Ecology</deliveryPoint>
        <deliveryPoint>Fictitious State University</deliveryPoint>
        <deliveryPoint>PO Box 111111</deliveryPoint>
        <city>Ficity</city>
        <administrativeArea>FI</administrativeArea>
        <postalCode>11111-1111</postalCode>
    </address>
    <phone phonetype="voice">(999) 999-9999</phone>
    <electronicMailAddress>fsu.contact@fi.univ.edu</electronicMailAddress>
    <onlineUrl>http://www.fsu.edu/</onlineUrl>
    <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>
</creator>
<creator id="pos-1" system="FLS" scope="system">
    <positionName>FLS Lead PI</positionName>
    <address>
        <deliveryPoint>Department for Ecology</deliveryPoint>
        <deliveryPoint>Fictitious State University</deliveryPoint>
```

```
            <deliveryPoint>PO Box 111111</deliveryPoint>
            <city>Ficity</city>
            <administrativeArea>FI</administrativeArea>
            <postalCode>11111-1111</postalCode>
        </address>
        <phone phonetype="voice">(999) 999-9999</phone>
        <electronicMailAddress>fsu.leadPI@fi.univ.edu</electronicMailAddress>
        <onlineUrl>http://www.fsu.edu/</onlineUrl>
        <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>

    </creator>
    <creator id="pers-1" system="FLS" scope="system">
        <individualName>
            <salutation>Dr.</salutation>
            <givenName>Joe</givenName>
            <givenName>T.</givenName>
            <surName>Ecologist Jr.</surName>
        </individualName>
        <organizationName>FSL LTER</organizationName>
        <address>
            <deliveryPoint>Department for Ecology</deliveryPoint>
            <deliveryPoint>Fictitious State University</deliveryPoint>
            <deliveryPoint>PO Box 111111</deliveryPoint>
            <city>Ficity</city>
            <administrativeArea>FI</administrativeArea>
            <postalCode>11111-1111</postalCode>
        </address>
        <phone phonetype="voice">(999) 999-9999</phone>
        <electronicMailAddress>jecologist@fi.univ.edu</electronicMailAddress>
        <onlineUrl>http://www.fsu.edu/~jecologist</onlineUrl>
        <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>

    </creator>
```

creator

This element is found at this location (XPath):

/eml:eml/dataset/creator

<creator> The creator is considered to be the author of the data package, i.e. the person(s) responsible for intellectual input into its creation. **<surName>** and **<givenName>** elements are used to build citations, so these should be completed fully for credit to be understandable. For long-term data, e.g., from an LTER Site, preparers should include the organization (using the **<organizationName>**) or current principal investigator (PI, using **<postitionName>**). It should be kept in mind that in the past, different approaches have lead to confusion over how to best search for long-term data, and searchers frequently default to searches using PI's last name. Therefore it is a reasonable practice to include more creators rather than fewer, even if it blurs the credit for long-term data.

metadataProvider

This element is found at this location (XPath):
/eml:eml/dataset/metadataProvider

The **<metadataProvider>** element lists the person or organization responsible for producing or providing the metadata content. For primary data sets generated by LTER sites, the LTER site should typically be listed under **<metadataProvider>** using the **<organizationName>** element. For acquired data sets, where the **<creator>** or **<associatedParty>** are not the same people who produced the metadata content, the actual metadata content provider should be listed instead (see Example below).

Example 5: metadataProvider

```
<metadataProvider>
    <organizationName>Fictitious LTER Site</organizationName>
    <address>
        <deliveryPoint>Department of Ecology</deliveryPoint>
        <deliveryPoint>Fictitious State University</deliveryPoint>
        <deliveryPoint>PO Box 111111</deliveryPoint>
        <city>Ficity</city>
        <administrativeArea>FI</administrativeArea>
        <postalCode>11111-1111</postalCode>
    </address>
    <phone phonetype="voice">(999) 999-9999</phone>
    <electronicMailAddress>fsu@fi.univ.edu</electronicMailAddress>
    <onlineUrl>http://www.fsu.edu/</onlineUrl>
    <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>
</metadataProvider>
```

associatedParty

This element is found at this location (XPath):
/eml:eml/dataset/associatedParty

List other people who were involved with the data in some way (field technicians, students assistants, etc.) as **<associatedParty>**. All **<associatedParty>** trees require a **<role>** element. The parent university, institution, or agency could also be listed as an **<associatedParty>** using **<role>** of "owner" when appropriate.

Example 6: associatedParty

```
<associatedParty id="12010" system="FLS" scope="system">
    <individualName>
        <givenName>Ima</givenName>
        <surName>Testuser</surName>
    </individualName>
    <organizationName>FSL LTER</organizationName>
    <address>
        <deliveryPoint>Department for Ecology</deliveryPoint>
        <deliveryPoint>Fictitious State University</deliveryPoint>
        <deliveryPoint>PO Box 111111</deliveryPoint>
        <city>Ficity</city>
        <administrativeArea>FI</administrativeArea>
```

```
        <postalCode>11111-1111</postalCode>
    </address>
    <phone phonetype="voice">(999) 999-9999</phone>
    <electronicMailAddress>itestuser@lternet.edu</electronicMailAddress>
<onlineUrl>http://search.lternet.edu/directory_view.php?personid=12010&amp;query=it
estuser</onlineUrl>
    <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>

    <role>Technician</role>
</associatedParty>
```

contact

This element is found at this location (XPath):

/eml:eml/dataset/contact

A **<contact>** element is required in all EML metadata records. Full contact information should be included for the position of data manager or other designated contact, and should be kept current and independent of personnel changes. If several contacts are listed (e.g. both a data and site manager) all should be kept current. Technicians who performed the work belong under **<associatedParty>** rather than **<contact>**. Complete the **<address>**, **<phone>**, **<electronicMailAddress>**, and **<onlineURL>** elements for the **<contact>** element.

Example 7: contact
```
<contact>
    <positionName id="pos-4">Information Manager</positionName>
    <address>
        <deliveryPoint>Department for Ecology</deliveryPoint>
        <deliveryPoint>Fictitious State University</deliveryPoint>
        <deliveryPoint>PO Box 111111</deliveryPoint>
        <city>Ficity</city>
        <administrativeArea>FI</administrativeArea>
        <postalCode>11111-1111</postalCode>
    </address>
    <phone phonetype="voice">(999) 999-9999</phone>
    <electronicMailAddress>fsu.data@fi.univ.edu</electronicMailAddress>
    <onlineUrl>http://www.fsu.edu/</onlineUrl>
    <userId system="https://orcid.org">https://orcid.org/0000-0000-0000-
0000</userId>
</contact>
```

publisher

This element is found at this location (XPath):

/eml:eml/dataset/publisher

The organization producing the EML metadata (e.g., an LTER site or field station) should be placed in the **<publisher>** element. Spell out the organization's name (**<organizationName>**). Complete the **<address>**, **<phone>**, **<electronicMailAddress>**,

and **<onlineURL>** elements for each publisher element. Some citation displays may use this element, although typically, the repository becomes the publisher in citations.

Example: publisher
```
<publisher>
      <organizationName>Fictitious LTER site</organizationName>
   </publisher>
```

---

pubDate

This element is found at this location (XPath):
/eml:eml/dataset/pubDate

The year of public release of data online should be listed as the **<pubDate>** element. Because this element may be used in constructing citations, the **pubDate** also should reflect the 'recentness' of a package, with **pubDate** updated along with significant revision or data additions (e.g., corrected data, or additions to an ongoing time series). There is an argument for **pubDate** referring to original date of release, but this is probably only useful for static data packages, or if the only metadata changes are to enhance discovery.

abstract

This element is found at these locations (XPath):
/eml:eml/dataset/abstract
/eml:eml/dataset/project/abstract

For a dataset, the abstract element can appear at the resource level or the project level. The **** element will be used for full-text searches, and it should be rich with descriptive text. In particular, descriptions should include information that does not fit into structured metadata, and focus on the "what", "when", and "where" information, general taxonomic information, as well as whether the dataset is ongoing or completed. Some general methods description is appropriate, and broad classes of measured parameters should also be included. For a large number of parameters, use categories instead of listing all parameters (e.g. use the term "nutrients" instead of nitrate, phosphate, calcium, etc.), in combination with the parameters that seem most relevant for searches.

keywordSet and keyword

This element is found at these locations (XPath):
/eml:eml/dataset/keywordSet

/eml:eml/dataset/project/keywordSet

It is recommended that meaningful sets of keywords each be contained within <**keywordSet**> tag. Use one <**keywordSet**> for a group of terms identifying the contributing organization(s), e.g., the LTER or OBFS site, LTREB or Macrosystems project , which is especially if data are co-funded or funding is leveraged. Meaningful geographic place names also are appropriate (e.g. state, city, county). If groups of keywords are from a specific vocabulary, its name belongs the optional tag <**keywordThesaurus**>.

Context: Communities sometimes have specific requests for keywords to assist in searches. E.g, the LTER requests that keywords should include a LTER core research area(s), the network acronym (LTER, ILTER, etc.), three-letter site acronym and site name. In addition to specific keywords, relevant conceptual keywords should also be included, e.g., from the LTER Controlled Vocabulary.

Example: pubDate, abstract,keywordSet, keyword
```
<pubDate>2014</pubDate>
<abstract>
    <para>Ground arthropods communities are monitored in different habitats in a
rapidly changing environment. The arthropods are collected in traps four times a
year in ten locations and determined as far as possible to family, genus or
species.</para>
</abstract>
<keywordSet>
    <keyword keywordType="place">City</keyword>
    <keyword keywordType="place">State</keyword>
    <keyword keywordType="place">Region</keyword>
    <keyword keywordType="place">County</keyword>
    <keyword keywordType="theme">FLS</keyword>
    <keyword keywordType="theme">Fictitious LTER Site</keyword>
    <keyword keywordType="theme">LTER</keyword>
    <keyword keywordType="theme">Arthropods</keyword>
    <keyword keywordType="theme">Richness</keyword>
    <keywordThesaurus>FLS site thesaurus</keywordThesaurus>
</keywordSet>
<keywordSet>
    <keyword keywordType="theme">ecology</keyword>
    <keyword keywordType="theme">biodiversity</keyword>
    <keyword keywordType="theme">population dynamics</keyword>
    <keyword keywordType="theme">terrestrial</keyword>
    <keyword keywordType="theme">arthropods</keyword>
    <keyword keywordType="theme">pitfall trap</keyword>
    <keyword keywordType="theme">monitoring</keyword>
    <keyword keywordType="theme">abundance</keyword>
    <keywordThesaurus>LTER controlled vocabulary</keywordThesaurus>
</keywordSet>
<keywordSet>
    <keyword keywordType="theme">populations</keyword>
    <keywordThesaurus>LTER core research areas</keywordThesaurus>
</keywordSet>
```

intellectualRights

This element is found at this location (XPath):
/eml:eml/dataset/intellectualRights

<intellectualRights> are controlled at the source, however it is recommended that data be released with as few restrictions as possible. Each data package should contain a data access policy, plus a description of any deviation from the general policy specific for this particular package (e.g. restricted-access packages). The timeframe for release should be included as well.

Context: If no **<intellectualRights>** *element is included EDI will insert text that releases data under "CC-0" (shown in example). The LTER Network-wide default policy is "CC-by". Please consult those organizations for more information and more details.*

Example: intellectualRights
```
<intellectualRights>
    <section>
        <title>Data Policy</title>
        <para> This data package is released to the "public domain" under Creative
Commons CC0 1.0 "No Rights Reserved" (see:
https://creativecommons.org/publicdomain/zero/1.0/). It is considered professional
etiquette to provide attribution of the original work if this data package is
shared in whole or by individual components. A generic citation is provided for
this data package on the website https://portal.edirepository.org (herein
"website") in the summary metadata page. Communication (and collaboration) with the
creators of this data package is recommended to prevent duplicate research or
publication. This data package (and its components) is made available "as is" and
with no warranty of accuracy or fitness for use. The creators of this data package
and the website shall not be liable for any damages resulting from
misinterpretation or misuse of the data package or its components. Periodic updates
of this data package may be available from the website. Thank you</para>
    </section>
</intellectualRights>
```

distribution

This element is found at these locations (XPath):
/eml:eml/dataset/distribution
/eml:eml/dataset/[entity]/physical/distribution

The **<distribution>** element can appear at both the dataset and entity levels. At the entity-level, it contains information on how that specific data entity can be accessed. The **<distribution>** element has one of three children for describing the location of the resource: **<online>**, **<offline>**, and **<inline>**.

Offline Data: Use the **<offline>** element to describe restricted access data or data that is not available online. The minimum that should be included is the **<mediumName>** tag, if using the **<offline>** element.

Inline Data: The **\<inline\>** element contains data that is stored directly within the EML document. Data included as text or string will be parsed. If data are not to be parsed, encode them as "CDATA sections," by surrounding them with "<![CDATA[" and "]]>" tags.

Online Data: The **\<online\>** element has two sub elements, **\<url\>**, and **\<onlineDescription\>** (optional). **\<url\>** tags may have an optional attribute named **function**, which may be set to either "download" or "information". If the "function" attribute is omitted, then "download" is implied.

@function="download": accessing the URL directly returns the data stream

@function="information": URL leads to a data catalog, intended-use page, or other page that provides information about downloading the object but does not directly return the data stream, then the "function" attribute should be set to "information".

Context: for am EML data package to be accepted into the EDI repository, it must include at least one URL; at the entity level (e.g., a dataTable at /eml:eml/dataset/dataTable/physical/distribution/url). The URL must include the function attribute with the value "download" (or empty, i.e., defaults to "download").

When used at the entity level, an alternative tag is available to **\<url\>**, called **\<connection\>**. This element is discussed under data entities, below.

As of EML 2.1, there is also an optional **\<access\>** element in a **\<distribution\>** tree at the data entity level **(/eml:eml/dataset/[entity]/physical/distribution/access**). This element is intended specifically for controlling access to the data entity itself. For more information on the **\<access\>** tree, see above, under the general access discussion.

Example: distribution

```
<distribution>
    <online>
        <onlineDescription>f1s-1 Data Web Page</onlineDescription>
        <url function="information">http://www.fsu.edu/lter/data/fls-1.htm</url>
    </online>
</distribution>

<dataTable>
    <physical>
…
        <distribution>
            <online>
                <onlineDescription>f1s-1 Data Web Page</onlineDescription>
                <url function="download">http://www.fsu.edu/lter/data/fls-1.csv</url>
            </online>
        </distribution>
    </physical>
</dataTable>
```

coverage

This element is found at these locations (XPath):
/eml:eml/dataset/coverage
/eml:eml/dataset/methods/sampling/studyExtent/coverage
/eml:eml/dataset/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/[entity]/coverage
/eml:eml/dataset/[entity]/methods/sampling/studyExtent/coverage
/eml:eml/dataset/[entity]/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/methods/sampling/studyExtent/coverage
/eml:eml/dataset/[entity]/attributeList/attribute/methods/sampling/spatialSamplingUnits/coverage
/eml:eml/dataset/project/studyAreaDescription/coverage

The **<coverage>** element can appear at the dataset, methods, entity and attribute levels, and contains three elements for describing the coverage in terms of space, taxonomy, and time, **<geographicCoverage>**, **<taxanomicCoverage>**, and **<temporalCoverage>**. Populating these elements as recommended enables advanced searches and understanding. Because they appear at many XPaths, there are many options for how coverage elements can be used.

geographicCoverage

General Information: The **<geographicCoverage>** element describes locations of research sites and areas related to the data, and is intended for general placement of points on a map. It is recommended to use the element at different levels for different types of information. The cardinality of the **<geographicCoverage>** element is one-to-many. The miminum requirement under **<geographicCoverage>** is two elements, a **<geographicDescription>** and **<boundingCoordinates>** with a bounding box containing N, S, E, W limits.

At the dataset level (**eml:eml/dataset/coverage**) one **<geographicCoverage>** element should be included, whose **<boundingCoordinates>** describe the extent of the data. As a default, this could be the nominal boundaries of a sampling area. A more accurate extent (recommended) would be the maximum extent of the data, for each of east, west, north and south.

Additional **<geographicCoverage>** elements should be included if there are significant distances between study sites and grouping them in one bounding box would be misleading or confusing. For example, a cross-site study should have bounding boxes for each site.

Example: geographicCoverage at the dataset level
```
<coverage>
    <geographicCoverage>
```

```
        <geographicDescription>Ficity, FI metropolitan area,
USA</geographicDescription>
        <boundingCoordinates>
            <westBoundingCoordinate>-112.373614</westBoundingCoordinate>
            <eastBoundingCoordinate>-111.612936</eastBoundingCoordinate>
            <northBoundingCoordinate>33.708829</northBoundingCoordinate>
            <southBoundingCoordinate>33.298975</southBoundingCoordinate>
            <boundingAltitudes>
                <altitudeMinimum>300</altitudeMinimum>
                <altitudeMaximum>600</altitudeMaximum>
                <altitudeUnits>meter</altitudeUnits>
            </boundingAltitudes>
        </boundingCoordinates>
    </geographicCoverage>
</coverage>
```

If sampling took place in discrete point location, those sites should also appear with or without a bounding box. Individual sampling sites may also be be entered under **<spatialSamplingUnits>**, each site in a separate coverage element (see below).

Example: geographicCoverage under spatialSamplingUnits
```
<spatialSamplingUnits>
    <coverage>
        <geographicDescription>sitenumber 1</geographicDescription>
        <boundingCoordinates>
            <westBoundingCoordinate>-112.2</westBoundingCoordinate>
            <eastBoundingCoordinate>-112.2</eastBoundingCoordinate>
            <northBoundingCoordinate>33.5</northBoundingCoordinate>
            <southBoundingCoordinate>33.5</southBoundingCoordinate>
        </boundingCoordinates>
    </coverage>
    <coverage>
        <geographicDescription>sitenumber 2</geographicDescription>
        <boundingCoordinates>
            <westBoundingCoordinate>-111.7</westBoundingCoordinate>
            <eastBoundingCoordinate>-111.7</eastBoundingCoordinate>
            <northBoundingCoordinate>33.6</northBoundingCoordinate>
            <southBoundingCoordinate>33.6</southBoundingCoordinate>
        </boundingCoordinates>
    </coverage>
    <coverage>
        <geographicDescription>sitenumber 3</geographicDescription>
        <boundingCoordinates>
            <westBoundingCoordinate>-112.1</westBoundingCoordinate>
            <eastBoundingCoordinate>-112.1</eastBoundingCoordinate>
            <northBoundingCoordinate>33.7</northBoundingCoordinate>
            <southBoundingCoordinate>33.7</southBoundingCoordinate>
        </boundingCoordinates>
    </coverage>
</spatialSamplingUnits>
```

Latitudes and longitudes should be in the same datum, commonly used (i.e., all values in WGS84 or NAD83) and expressed to at least six decimal places (the EML2.1 schema enforces decimal content). International convention dictates that longitudes east of the prime meridian and latitudes north of the equator be prefixed with a plus sign (+), or by the

absence of a minus sign (-), and that west longitudes and south latitudes be prefixed with minus sign (-). See Example below, and the EML specification for more information and other examples.

<geographicDescription> The description is a string. It should be comprehensive so that searches can be run against it, and include the country, state, county or province, city, general topography, landmarks, rivers and other relevant information. The method for determining **<boundingCoordinates>**, **<boundingAltitudes>**, coordinates, datums, etc., should be included with the **<geographicDescription>**, since those elements do not encode this information.

The **<datasetGPolygon>** element may be included when the required bounding box does not adequately describe the study location, for example, if an irregular polygon is necessary to describe the study area, or there is an area within the bounding box that is excluded. This element is optional, and has two subelements.

<datasetGPolygonOuterGRing>: This is the outer part of the polygon shape that encompasses the broadest area of coverage. It can be created either by a gRing (list of points) or 4 or more **<gRingPoint>**s. Documentation for an FGDC G-Ring states that four points are required to define a polygon, and the first and last should be identical. However this is not enforceable in XML Schema, and so in EML a minimum of three **<gRingPoint>**s is required to define the polygon, and it can be assumed that a since a polygon is closed, the last point can be joined to the first.

The **<datasetGPolygonExclusionGRing>** is the closed, nonintersecting boundary of a void area (or hole in an interior area). This could be the center of the doughnut shape created by the **<datasetGPolygon>**. It can be created either by a gRing (list of points) or one or more **<gRingPoint>**s. This is used if there is an internal polygon to be excluded from the outer polygon, e.g, a lake to be excluded from the broader geographic coverage.

There are alternative methods for including location information with EML, especially when it is intended for use in an external application. GIS shape files, Keyhole Markup Language (KML or KMZ), or EML spatial modules can be included as data entities (see additional resources for different data file types at EDI).

temporalCoverage
The **<temporalCoverage>** element represents the period of time the data were collected, not the year the study was conducted if it uses retrospective or historical data. Most commonly, **<singleDate>** or **<rangeOfDates>** elements are used. Sometimes an **<alternativeTimeScale>** is more appropriate, such as the use of "years before present", e.g., for long-term tree ring chronology dating back hundreds of years. Two formats are allowed, either a 4-digit year, or a date in ISO format: YYYY-MM-DD.

In some cases, a package may be considered "ongoing", i.e., data are planned to be added at intervals. It is not currently valid to leave an empty **<endDate>** tag in EML. Further, EML is intended to house "snapshots" of data which can be immutable (if the repository supports). So for a package which is planned to be ongoing, the best solution is to populate the **<endDate>** element with the end of the current data range and to update this metadata field along with data updates, so that the **<endDate>** tag reflects only the data that have already been included. It is better to state an end date that guarantees that data are present up to that date with more data possibly being available, than an end date in the future that includes a period of time for which no data are yet available. Use the **<maintanence>** tag (below) to describe the update frequency. The methods/sampling tree should be used to describe the ongoing nature of the data collection.

Example: temporalCoverage
```
<temporalCoverage>
    <rangeOfDates>
        <beginDate>
            <calendarDate>1998-11-12</calendarDate>
        </beginDate>
        <endDate>
            <calendarDate>2003-12-31</calendarDate>
        </endDate>
    </rangeOfDates>
</temporalCoverage>
```

---

taxonomicCoverage

The **<taxonomicCoverage>** element should be used to document taxonomic information for all organisms relevant to the study. The lowest available level, preferably the species binomial and common name should always be included, but higher-level taxa should also be included to support broader taxonomic searches. Blocks of **<taxonomicClassification>** elements should be hierarchically nested within a single **<taxonomicCoverage>** element rather than repeated at the same level. The **<generalTaxonomicCoverage>** element could include a) descriptions of the general procedure of how the taxonomy was determined (keys used, etc.), b) general textual description of all flora/fauna in the study (scope), and c) denote how finely grained the taxonomy is – for example to "family" or "genus and species."

Note that it is allowable to combine elements in the hierarchy under like **<taxonRankName>** entries to create a taxonomic "tree" (not illustrated), but this practice may impede combining and re-using **<taxonomicClassification>** information from multiple documents so should be considered carefully.

The optional **taxonomicCoverage/taxonomicSystem** trees may be used to detail the use of taxonomic identification resources and on the identification process. **<classificationSystem>** should be used to list authoritative taxonomic databases (such as

ITIS, IPNI, NCBI, Index Fungorum, or USDA Plants) or classification systems used for taxonomic identification. Documentation and relevant literature regarding, used authoritative sources, including URL's pointing to these sources, should be listed in **<classificationSystemCitation>**. Exceptions to, or deviation from, used authoritative sources should be explained in **<classificationSystemModification>**.

Methods and protocols used for taxonomic classification should be detailed using the **<identifierName>** and **<taxonomicProcedures>** tags. Examples of methods that should be listed in **<taxonomicProcedures>** are details of specimen processing, keys, and chemical or genetic analyses. **<taxonomicCompleteness>** may be used to document the status, estimated importance, and reason for incomplete identifications.

Example: taxonomicCoverage

```
<taxonomicCoverage>
    <taxonomicSystem>
        <classificationSystem>
            <classificationSystemCitation>
                <title>Integrated Taxonomic Information System (ITIS)</title>
                <creator>
                    <organizationName>Integrated Taxonomic Information
System</organizationName>
                    <onlineUrl>http://www.itis.gov/</onlineUrl>
                </creator>
                <generic>
                    <publisher>
                        <organizationName>Integrated Taxonomic Information
System</organizationName>
                        <onlineUrl>http://www.itis.gov/</onlineUrl>
                    </publisher>
                </generic>
            </classificationSystemCitation>
        </classificationSystem>
    <identifierName>
        <references>pers-1</references>
    </idnetifierName>
    <taxonomicProcedures>All individuals where identified and stored in alcohol,
except for one voucher specimen for each species which was tagged and
pinned.</taxonomicProcedures>
    </taxonomicSystem>
    <generalTaxonomicCoverage>Orthopteran insects (grasshoppers) were
 identified to species</generalTaxonomicCoverage>
    <taxonomicClassification>
        <taxonRankName>Kingdom</taxonRankName>
        <taxonRankValue>Animalia</taxonRankValue>
        <taxonomicClassification>
            <taxonRankName>Phylum</taxonRankName>
            <taxonRankValue>Mollusca</taxonRankValue>
            <taxonomicClassification>
                <taxonRankName>Class</taxonRankName>
                <taxonRankValue>Gastropoda</taxonRankValue>
                <taxonomicClassification>
                    <taxonRankName>Order</taxonRankName>
                    <taxonRankValue>Basommatophora</taxonRankValue>
                    <taxonomicClassification>
                        <taxonRankName>Genus</taxonRankName>
                        <taxonRankValue>Detracia</taxonRankValue>
                        <taxonomicClassification>
```

```
                    <taxonRankName>Species</taxonRankName>
                    <taxonRankValue>Detracia floridana</taxonRankValue>
                    <commonName>Florida Melampus</commonName>
                </taxonomicClassification>
            </taxonomicClassification>
        </taxonomicClassification>
      </taxonomicClassification>
    </taxonomicClassification>
  </taxonomicClassification>
  <taxonomicClassification>
      <taxonRankName>Kingdom</taxonRankName>
      <taxonRankValue>Animalia</taxonRankValue>
      <taxonomicClassification>
          <taxonRankName>Phylum</taxonRankName>
          <taxonRankValue>Mollusca</taxonRankValue>
          <taxonomicClassification>
              <taxonRankName>Class</taxonRankName>
              <taxonRankValue>Bivalvia</taxonRankValue>
              <taxonomicClassification>
                  <taxonRankName>Order</taxonRankName>
                  <taxonRankValue>Filibranchia</taxonRankValue>
                  <taxonomicClassification>
                      <taxonRankName>Genus</taxonRankName>
                      <taxonRankValue>Geukensia</taxonRankValue>
                      <taxonomicClassification>
                          <taxonRankName>Species</taxonRankName>
                          <taxonRankValue>Geukensia demissa</taxonRankValue>
                          <commonName>Ribbed Mussel</commonName>
                      </taxonomicClassification>
                  </taxonomicClassification>
              </taxonomicClassification>
          </taxonomicClassification>
      </taxonomicClassification>
  </taxonomicClassification>
</taxonomicCoverage>
```

maintenance

This element is found at these locations (XPath):

eml:eml/dataset/maintenance

The dataset/maintenance/description element should be used to document changes to the data tables or metadata, including update frequency. The change history can also be used to describe alterations in static documents. The description element (TextType) can contain both formatted and unformatted text blocks.

Example: maintenance
```
<maintenance>
    <description>
        <para>Data are updated annually at the end of the calendar year.</para>
    </description>
</maintenance>
```

methods

This element is found at these locations (XPath):
/eml:eml/dataset/methods
/eml:eml/dataset/[entity]/methods
/eml:eml/dataset/[entity]/attributeList/attribute/methods

General Information: In early EML versions, both "**<method>**" and "**<methods>**" elements were found, which caused confusion. In EML 2.1.0, the elements were standardized to "**<methods>**".

The **<methods>** tree appears at the dataset, entity, and attribute levels, and content is generally regarded as human readable, not machine-readable. As a 'rule of thumb', methods are *descriptive*, and protocols are *prescriptive*, i.e. the methods describe what was done when collecting data, and protocols are a set of procedures or prescribed actions. A method often includes or follows a particular protocol. As a minimum, a reference to an external protocol should be given at the dataset level. However, detailed, text methods at this are preferable so that their content can be perused in a browser or indexed for searching. If further refinement is needed, methods can be defined for individual data entities or even individual **<attribute>**, although these may not be not indexed. The scope of the method defined can be tailored to match the EML document level where it is applied. For example, methods at the dataset level describe the study, for a **<dataTable>** methods might include pre-/post-processing steps, and at the attribute level, quality control. The use of methods refinement varies and keeping all methods in one place and at one level (dataset) is simpler to manage. Since they are mostly for human consumption, one detailed description of all steps taken at the dataset level is frequently sufficient and more user friendly.

A description of methods contains the elements **<methodStep>**, **<sampling>**, and/or **<qualityControl>**.

methodStep

At least one **<methodStep>** is required under **<methods>**, and each step is a logical portion of the methods, for example, field, lab and statistical. All textual methods descriptions belong here, using **<description>** and TextType tags.

At a minimum, to describe an external document two tags can be used: **<citation>** for a referral to a published document or paper, or **<protocol>**. At a minimum, the **<protocol>** requires **<title>**, **<creator>** and **<distribution>** tags, where the **<distribution>** tree may be used to refer to an online document; see the recommendations above for using that tree. Alternatively, the entire protocol may be written into EML under protocol/methodStep.

instrumentation

The <**instrumentation**> tag should contain a full description of the instruments used, including manufacturer, model, calibration dates and accuracy. Changes in instrumentation and dates of changes should be mentioned earlier under the **<description>**.

dataSource
The optional <**dataSource**> tag is for nesting an EML dataset that is input to a <**methodStep**> of the data being described, e.g., calibration information for an instrument or input parameters for a model. It also may hold the source (provenance) data when describing a derived dataset.

*Context Note: The <**dataSource**> element is used by the EDI repository's provenance tracking system for linking between derived and source data packages. For more information, see additional data repository resources from EDI.*

sampling
This optional tree can contain valuable and very specific information about the study site, coverage and frequency in addition to that listed at other levels.

<**studyExtent**> Provides specific information about the temporal and geographic extent of the study such as domains of interest in addition to geographic, temporal, and taxonomic coverage of the study site. <**studyExtent**> can be a surrogate for the <**studyAreaDescription**> under <**project**>. Descriptions can be either as a simple text using **<description>** or by including detailed temporal or geographic <**coverage**> elements describing discrete time periods sampled or multiple sub-regions sampled within the overall geographic bounding box that was described at the dataset level.

*Context Note: In the past, LTER requested that individual sampling locations be listed here (under **studyExtent/spatialSamplingUnits**), and some LTER sites may have applications that specifically use that APath. However, in general use, the dataset-level geographicCoverage elements are more practical. See EDI "Other Resources", for more information about how indexers typically handle EML.*

<**samplingDescription**> a text based version, similar to the sampling methods section in a journal article.

qualityControl
Like other trees under <**methods**>, <**qualityControl**> can be used at the dataset, entity or attribute level, whichever is appropriate. At its most basic, use the <**description**> element. Tags are also available for a <**citation**> or <**protocol**>.

Example: methods

```
<methods>
    <methodStep>
        <description>
            <section>
                <title>Pitfall trap sampling for ground arthropod biodiversity
monitoring</title>
                <para>Supplies used: pitfall traps (P-16 plastic Solo cups with lids)
metal spades and large bulb planters (to dig holes in which to put traps) 70%
ethanol (to preserve specimens) Qorpak glass jars with lids from the VWR
Corporation, 120ml (4oz), cap size 58-400 (comes included), Qorpak no. 7743C, VWR
catalog no. 16195-703.</para>
                <para>Between 10 and 21 traps are placed at each site in siutable
location. </para>
                <para>All trapped taxa counted and measured (body length), most taxa
identified to Family, ants to Genus</para>
            </section>
        </description>
        <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
Electronics (model: 37-SM MicroCAT); parameter: Conductivity (accuracy: 0.0003 S/m,
readability: 0.00001 S/m, range: 0 to 7 S/m); last calibration: Feb 28,
2001</instrumentation>
        <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
Electronics (model: 37-SM MicroCAT); parameter: Pressure (water) (accuracy: 0.2m,
readability: 0.0004m, range: 0 to 20m); last calibration: Feb 28,
2001</instrumentation>
        <instrumentation>SBE MicroCAT 37-SM (S/N 1790); manufacturer: Sea-Bird
Electronics (model: 37-SM MicroCAT); parameter: Temperature (water) (accuracy:
0.002°C, readability: 0.0001°C, range: -5 to 35°C); last calibration: Feb 28,
2001</instrumentation>
    </methodStep>
    <sampling>
        <studyExtent>
            <description>
                <para> Arthropod pit fall traps are placed in three different
locations four times a year</para>
            </description>
        </studyExtent>
        <samplingDescription>
            <para>Six traps were set in a transect at each location.</para>
        </samplingDescription>
        <spatialSamplingUnits>
            <coverage>
                <geographicDescription>site number 1</geographicDescription>
                <boundingCoordinates>
                    <westBoundingCoordinate>-112.234566</westBoundingCoordinate>
                    <eastBoundingCoordinate>-112.234566</eastBoundingCoordinate>
                    <northBoundingCoordinate>33.534566</northBoundingCoordinate>
                    <southBoundingCoordinate>33.534566</southBoundingCoordinate>
                </boundingCoordinates>
            </coverage>
            <coverage>
                <geographicDescription>site number 2</geographicDescription>
                <boundingCoordinates>
                    <westBoundingCoordinate>-111.745677</westBoundingCoordinate>
                    <eastBoundingCoordinate>-111.745677</eastBoundingCoordinate>
                    <northBoundingCoordinate>33.64577</northBoundingCoordinate>
                    <southBoundingCoordinate>33.64577</southBoundingCoordinate>
                </boundingCoordinates>
            </coverage>
            <coverage>
                <geographicDescription>site number 3</geographicDescription>
                <boundingCoordinates>
                    <westBoundingCoordinate>-112.167899</westBoundingCoordinate>
```

```
                    <eastBoundingCoordinate>-112.16799</eastBoundingCoordinate>
                    <northBoundingCoordinate>33.76799</northBoundingCoordinate>
                    <southBoundingCoordinate>33.76799</southBoundingCoordinate>
                </boundingCoordinates>
            </coverage>
        </spatialSamplingUnits>
    </sampling>
    <qualityControl>
        <description>
            <para>All specimens are archived for future reference. Quality control
during data entry is achieved with standard database techniques of pulldowns that
prevent typos and constraints. Scientists inspect standard data summary statistics
after data entry.</para>
        </description>
    </qualityControl>
</methods>
```

Example: methods, with dataSource
```
<methods>
    <methodStep>
        <description>
            <section>
                <para> We utilize NPP data collected from 1906 to 2006 from the ONL
LTER site. The ONL NPP data unit definition is kg/m^2/yr. This unit does not
require conversion. </para>
            </section>
        </description>
        <dataSource>
            <title> NPP data from ONL 1906 to 2006 </title>
            <creator>
                <organizationName> ONL LTER </organizationName>
            </creator>
            <distribution>
                <online>
                    <url> http://metacat.lternet.edu/knb/metacat/knb-lter-onl.23.1
</url>
                </online>
            </distribution>
            <contact>
                <organizationName> ONL LTER </organizationName>
                <positionName> ONL Information Manager </positionName>
                <electronicMailAddress> im@onl.lternet.edu </electronicMailAddress>
            </contact>
        </dataSource>
    </methodStep>
</methods>
```

project
This element is found at this location (XPath):
/eml:eml/dataset/project

General information: EML is one of the few specifications with a detailed tree dedicated to
projects, and which can be nested, using **<relatedProject>** At its simplest, a **<project>**
tree can hold a general descriptions of the project sponsoring the data package and nested
if smaller sub-projects. A related project Minimally, the description of a project should
include **<title>**, **<personnel>** and ****, with the study area description and mission

statement. The **<distribution>** tree should link to the project's home page, or alternatively could link to a publication describing the project. As stated earlier, the description of elements that are reused (e.g., XML types) are discussed where they first appear, so the descriptions for these three elements (**<title>**, **<personnel>** and ****) can be found above, under **<dataset>**, above. Two elements are unique to the **<project>** tree, **<fundingSource>** and **<studyAreaDiscription>**.

<fundingSource> should contain the agency and grant number. It is not optional.

<studyAreaDiscription> tree and its accompanying **<citation>** tree are optional, and may be used to describe non-coverage characteristics of the study area such as climate, geology or disturbances or references to citable biological or geophysical classification systems such as the Bailey Ecoregions or the Holdridge Life Zones. The **studyAreaDiscription** tree also supports multiple **<coverage>** elements that can be used to describe the geographic boundaries of individual study sites within the larger area. These can be referenced by the **studyExtent/spatialSamplingUnits/referencedEntityId**. The sibling **<descriptor>** tag can be used for text descriptions of the site.

Example: project

```xml
<projec>
    <title>FSL basic monitoring program</title>
    <personnel id="pers-30" system="FLS">
        <individualName>
            <salutation>Dr.</salutation>
            <givenName>Eva</givenName>
            <givenName>M.</givenName>
            <surName>Scientist</surName>
        </individualName>
        <address>
            <deliveryPoint>Department of Ecology</deliveryPoint>
            <deliveryPoint>Fictitious State University</deliveryPoint>
            <deliveryPoint>PO Box 111111</deliveryPoint>
            <city>Ficity</city>
            <administrativeArea>FI</administrativeArea>
            <postalCode>11111-1111</postalCode>
        </address>
        <role>principalInvestigator</role>
    </personnel>
    <personnel id="pers-130" system="FLS">
        <individualName>
            <givenName>Monica</givenName>
            <givenName>D.</givenName>
            <surName>Techy</surName>
        </individualName>
        <address>
            <deliveryPoint>Department for Ecology</deliveryPoint>
            <deliveryPoint>Fictitious State University</deliveryPoint>
            <deliveryPoint>PO Box 111111</deliveryPoint>
            <city>Ficity</city>
            <administrativeArea>FI</administrativeArea>
            <postalCode>11111-1111</postalCode>
        </address>
        <role> principalInvestigator</role>
    </personnel>
```

```
    <abstract>
        <para>The FLS basic monitoring program consists of monitoring of arthropod
populations, plant net primary productivity, and bird populations. Monitoring takes
place at 3 locations, 4 times a year. Climate parameters a continuously measured at
all stations.</para>
    </abstract>
  </project>
```

[entity] = dataTable, spatialRaster, spatialVector, storedProcedure, view, otherEntity

This element is found at this location (XPath):

/eml:eml/dataset/dataTable
/eml:eml/dataset/spatialRaster
/eml:eml/dataset/spatialVector
/eml:eml/dataset/storedProcedure
/eml:eml/dataset/view
/eml:eml/dataset/otherEntity

General information: If at all possible, do not publish data in dated, proprietary, binary formats such as MS-Excel, and instead, export to plain text representations such as csv. The entity types **<dataTable>**, **<otherEntity>** and **<view>** cover many commonly encountered data structures and are covered here. **<spatialRaster>**, **<spatialVector>**, **<storedProcedure>**) will be addressed in more depth in a future version of this document. Table 1 gives the general features of EML's six entity types, to assist in selection.

Table 1. Summary of the six entities in EML 2, including the type of data entity typically described with that element, how they are created and a brief description of its metadata.

Element name	Used for	Created from	Metadata features
dataTable	Static ASCII tables	export from code, RDBMS or spreadsheets	columns/rows named and defined, e.g., measurement and storage typing
otherEntity	Binary files, images, maps, KML, KMZ, code	applications	type of entity
spatialRaster	grid, raster cell data, remote sensing data	applications, stylesheet conversions. See "Other Resources"	spatial organization of the raster cells, their data values, and if derived via imaging sensors, characteristics about the image and its individual bands

spatialVector	lines, points polygons, KML (if converted), ESRI shape files	applications, stylesheet conversions. See "Other Resources"	information about the vector's geometry type, count and topology level
view	Data returned from a database query	RDBMS	similar to dataTable, plus description of the query
storedProcedure	Data returned from a stored procedure in a database	RDBMS	similar to dataTable, plus procedure's parameters

Every EML data entity has a set of elements in common, called the **EntityGroup** tree, which describe general information about any data resource. Other elements are provided which are unique to each entity type. The elements in the **EntityGroup** appear first, and are

<alternateIdentifier>
 <entityName>
 <entityDescription>
<physical> (including optional **<access>**)
<coverage>
<methods>
<additionalInfo>

<alternateIdentifier> (optional): The primary identifier belongs in the id attribute of the entityName (e.g., **<dataTable id="xxx">** , but this tag can accommodate additional identifiers that might be used, possibly from different data management systems. It is used similarly to the **<alternateIdentifier>** element at the dataset level, above.

<entityName> (required): the name of the table, file or database table. In the early phases of EML adoption, this was often the original ascii file name. However, a better analogy is that the **<entityName>** is a class, e.g., "FLS time series of air temperature at field station", with its instantiation (filename) in the **<objectName>** element (see below).

*Context: The EDI repository requires that **<entityName>**s be unique within the entity.*

<entityDescription> This should be a longer, more descriptive explanation of the data in the entity. Like all descriptions, it is human-readable, and should help determine if it is appropriate for a particular use.

The **\<physical\>** tree (**/eml:eml/dataset/[entity]/physical**) further describes the physical format of the data.

\<objectName\> should be the name of the file when downloaded, or exported as text from a database. The **\<objectName\>** often is the filename of a file in a file system or that is accessible on the network.

\<externallyDefinedFormat\> For data entities in prescribed formats (e.g., NetCDF, KML, Excel), name that format here. It is recommended that where possible, formats follow mime type (e.g., "image/jpeg"). Descriptions that are software-specific should include manufacturer, program, and version, e.g. "Microsoft Excel 2003". A KML file of sampling locations can be declared here as either "KML" or "KMZ".

\<distribution\> provides information on how the resource is distributed, and the contents of this tree was generally covered at the dataset level. However, there are a few points which will be reiterated here.

The content of a **\<url\>** element at the entity level should deliver data, and not point to another application or use page. The **\<url\>**'s attribute, "function", should have the value "download". This is implied if the "function" attribute is omitted.

As of EML 2.1, there is also an optional **\<access\>** element in a **\<distribution\>** tree at the entity level. This element is intended specifically for controlling access to the data entity separately from the metadata. For more information on using the **\<access\>** tree, refer to the general access discussion above.

\<coverage\> provides information on the geographic, spatial and temporal coverages used in this [**entity**]. See the discussion at the dataset level for more information.

\<methods\> provides information on the specific methods used to collect information in this [**entity**]. Please see the discussion at the dataset level for more information.

\<additionalInfo\> is a text field for any material that cannot be characterized by the other elements for the data type.

Example: The elements in the EntityGroup, showing the \<dataTable\> entity.
```
<dataTable>
    <entityName>arthro_hab</entityName>
    <entityDescription> habitat description for the sampling
        locations</entityDescription>
    <physical>
        <objectName>fls-1.csv</objectName>
```

```xml
<dataFormat>
    <textFormat>
        <numHeaderLines>1</numHeaderLines>
        <numFooterLines>0</numFooterLines>
        <recordDelimiter>\r</recordDelimiter>
        <numPhysicalLinesPerRecord>1</numPhysicalLinesPerRecord>
        <recordDelimiter>#x0A</recordDelimiter>
        <attributeOrientation>column</attributeOrientation>
        <simpleDelimited>
            <fieldDelimiter>,</fieldDelimiter>
        </simpleDelimited>
    </textFormat>
</dataFormat>
<distribution>
    <online>
        <onlineDescription>fls-1 Data File</onlineDescription>
        <url function="download">http://www.fsu.edu/lter/data/fls-1.csv</url>
    </online>
</distribution>
</physical>
```

Each data type has a specific set of elements that follow the common elements. Table 2 shows the specific trees that are applied to each of the data type.

Table 2. Elements specific to each of the six entity types.

Entity Type	Typical Uses	Elements following EntityGroup
\<dataTable>	Static ASCII tables	**\<attributeList>** **\<constraint>** **\<caseSensitivity>** **\<numberOfRecords>**
\<view>	Data returned from a database query	**\<attributeList>** **\<constraint>** **\<queryStatement>**
\<storedProcedure>	Data returned from a stored procedure in a database	**\<attributeList>** **\<constraint>** **\<parameter>**
\<otherEntity>		**\<attributeList>** **\<constraint>** **\<entityType>**
\<spatialRaster>	Lines, points polygons, KML (if converted), ESRI shape files	**\<attributeList>** **\<constraint>** **\<spatialReference>** **\<georeferenceInfo>** **\<horizontalAccuracy>** **\<verticalAccuracy>**

		<cellSizeYDirection> <numberOfBands> <rasterOrigin> <rows> <columns> <verticals> <cellGeometry> <toneGradation> <scaleFactor> <offset> <imageDescription>
<spatialVector	Lines, points polygons, KML (if converted), ESRI shape files	<attributeList> <constraint> <geometry> <geometricObjectCount> <topolgyLevel> <spatialReference> <horizontalAccuracy> <vericalAccuracy>

attributeList

This element tree is found at (XPath):

/eml:eml/dataset/dataTable/attributeList
/eml:eml/dataset/view/attributeList
/eml:eml/dataset/storedProcedure/attributeList
/eml:eml/dataset/spatialRaster/attributeList
/eml:eml/dataset/spatialVector/attributeList
/eml:eml/dataset/otherEntity/attributeList

The **<attributeList>** tree is required for all data types except for **<otherEntity>**. It describes all variables in a data entity in individual **<attribute>** elements. The description includes the name and definition of each attribute, its domain, definitions of coded values, and other pertinent information.

<attributeName> is typically the name of a field in a data table. This is often short and/or cryptic. It is recommended that attributeNames be suitable for use as a variable, e.g., composed of ASCII characters, and that the **<attributeName>**s match the column headers of a CSV or other text table.

*Context: in the EDI repository, **<attributeName>***s must be unique within a data entity.*

<attributeLabel> (optional): is used to provide a less ambiguous or less cryptic alternative identification than what is provided in **<attributeName>**. **<attributeLabel>** is likely to be used as a column or row header in an HTML display.

<attributeDefinition> gives a precise and complete definition of attribute being documented. It explains the contents of the attribute fully so that a data user can interpret the attribute accurately.

<storageType> may be system specific, as for a RDBMS, i.e., A Microsoft SQL varchar, or Oracle datetime. This field represents a 'hint' to processing systems as to how the attribute might be represented in a system or language, but is distinct from the actual expression of the domain of the attribute. Non system-specific values include float, integer and string.

<measurementScale> indicates the type of scale from which values are drawn for the attribute. EML's attribute-unit model is described in detail; see "Other Resources".   One of the 5 scale types must be used: nominal, ordinal, interval, ratio, or dateTime, as follows:

Non-numeric types:
The <nominal> scale is used to represent named categories. Values are assigned to distinguish them from other observations. This would include a list of coded values (e.g. 1=male, 2=female), or plain text descriptions. Columns that contain strings or simple text are nominal. Example: plot1, plot2, plot3.

<ordinal> values are categories that have a logical or ordered relationship to one another, but the magnitude of the differences between the values is not defined or meaningful. Example: Low, Medium, High.

Both the nominal and ordinal scales are **<nonNumericDomain>** types, and can be either text or an enumerated list. The **<enumeratedDomain>** applies to coded values, and requires a **<codeDefinition>** or a referenced entity containing the code explanations. For **<textDomain>** an optional pattern may describe the text, e.g., a US telephone number can be described by the format "\d\d\d-\d\d\d-\d\d\d\d".

Numeric types:
<interval> measurements are ordinal, but in addition, use equal-sized units on a scale between values. Because the units are equal sized, these measurements are numeric. However, the starting point is arbitrary, so a value of zero is not meaningful. For example, the Celsius temperature scale uses degrees which are equally spaced, but where zero does not represent "absolute zero" (i.e., the temperature at which molecular motion stops), and 20 C is not "twice as hot" as 10 C.

<ratio> measurements have a meaningful zero point, and ratio comparisons between values are legitimate. For example, the Kelvin scale reflects the amount of kinetic energy of a substance (i.e., zero is the point where a substance transmits no thermal energy), and so temperature measured in kelvin units is a ratio measurement. Concentration is also a ratio measurement because a solution at 10 micromolePerLiter has twice as much substance as one at 5 micromolePerLiter.

The numeric types **<interval>** and **<ratio>** scales require additional tags describing the **<unit>**, **<numericDomain>**, and **<precision>**.

<unit> Units should be described in correct physical units. Terms which describe data but are not units should be used in **<attributeDefinition>**. For example, for data describing "milligrams of Carbon per square meter", "Carbon" belongs in the **<attributeDefinition>**, while the **<unit>** is "milligramPerMeterSquared".

<standardUnit> and **<customUnit>**: Unit names must be either **<standardUnit>**, from the unit dictionary included with EML (http://knb.ecoinformatics.org/software/eml/eml-2.1.0/eml-unitTypeDefinitions.html#StandardUnitDictionary) or **<customUnit>** and defined in the **<additionalMetadata>**.

For general purposes, the following guidelines (from ISO recommendations) apply to **<customUnits>**: Units should be written out, not abbreviated. Unit modifiers, such as "squared", should follow the unit being modified. For example, meterSquared is preferred, while squareMeter is improper. Units should be singular, such as "meter", and not plural, such as "meters".

*Context: EDI has adopted the LTER Unit Registry and recommends that **<customUnit>** element be used for all units with content pulled from the Unit Registry, even when the unit is already listed in the standard unit dictionary.*

<numericDomain> This tag includes elements specifying the **<numberType>** and the minimum and maximum allowable values of a numeric attribute. A measurement's **<numberType>** should be defined as real, natural, whole or integer as explained in EML handbook: (see "Other Resources"). The **<bounds>** are theoretical or allowable minimum and maximum values (prescriptive), rather than the actual observed range in a data set (descriptive). The **<bounds>** tree is optional.

<precision> describes the number of decimal places for the attribute. Currently, EML does not allow more than one precision value for a column. For example, a column containing lengths of fish may be measured to a precision of .01 meter for one species of fish (e.g., large), and .001 meters for a different species, but all the data on "fish length" are collected into one attribute and are measured using their appropriate precision values. For these

cases precision can be omitted, but the variable precision information should be described in detail in **method/methodStep**. Together, the information in **<numericDomain>** and **<precision>** are sufficient to decide upon an appropriate system-specific data type for representing a particular attribute. For example, an attribute with a numeric domain from 0-50,000 and a precision of 1 could be represented in the C language using a 'long' value, but if the precision is changed to '0.5' then a 'float' type would be needed.

The **<measurementType>** element, **<dateTime>**, is a date-time value from the Gregorian calendar and it is recommended that these be expressed in a format that conforms to the ISO 8601 standard. An example of an allowable ISO date-time is "YYYY-MM-DD", as in 2004-06-25, or, more fully, as "YYYY-MM-DDThh:mm:ssTZD" (eg 1997-07-16T19:20:30.45Z). The ISO standard is quite strict about the structure of date components. Since legacy data often contain non-standard dates, and existing equipment (e.g., sensors) may still be producing non-standard dates, the EML authors have provided additional allowable formats. See the EML documentation for a complete list. It is important to note that the dateTime field should not be used for recording time durations. In that case, use a unit such as seconds, nominalMinute or nominalDay, that defines the duration in terms of its relationship to SI second.

The **<missingValueCode>** is optional, but should be included to describe any missing value codes present in the data set (e.g. NA, NaN, ND, 9999). The missing value code is a string, not a value, which means that the content of this field must exactly match what appears in place of data values for it to be correctly interpreted. For example, if data are output with precision .01 and with missing values formatted to "-9999.00", then the content of the **<missingValueCode>** element must be "-9999.00" not "-9999".

The examples show two attribute trees. The first was generated from an SQL system with a defined storage type. The second **<attributeList>** includes tags for **<customUnits>**, with the Unit defined in the **<additionalMetadata>** tree.

Example: attributeList/attribute dataTable
```
<attributeList>
    <attribute id="soil_chemistry.site_id">
        <attributeName>site_id</attributeName>
        <attributeDefinition>Site id as used in sites table</attributeDefinition>
        <storageType typeSystem="http://www.w3.org/2001/XMLSchema-
datatypes">string</storageType>
        <measurementScale>
            <nominal>
                <nonNumericDomain>
                    <textDomain>
                        <definition>Site id as used in sites table</definition>
                    </textDomain>
                </nonNumericDomain>
            </nominal>
        </measurementScale>
    </attribute>
```

```xml
    <attribute id="soil_chemistry.pH">
        <attributeName>pH</attributeName>
        <attributeDefinition>ph of soil solution</attributeDefinition>
        <storageType typeSystem="http://www.w3.org/2001/XMLSchema-
datatypes">float</storageType>
        <measurementScale>
            <ratio>
                <unit>
                    <standardUnit>dimensionless</standardUnit>
                </unit>
                <precision>0.01</precision>
                <numericDomain>
                    <numberType>real</numberType>
                </numericDomain>
            </ratio>
        </measurementScale>
    </attribute>
    <attribute id="pass2001.q110">
        <attributeName>q110</attributeName>
        <attributeDefinition>Q110-Preference for front yard
landscape</attributeDefinition>
        <storageType typeSystem="http://www.w3.org/2001/XMLSchema-
datatypes">float</storageType>
        <measurementScale>
            <ordinal>
                <nonNumericDomain>
                    <enumeratedDomain>
                        <codeDefinition>
                            <code>1.00</code>
                            <definition>1-A desert landscape</definition>
                        </codeDefinition>
                        <codeDefinition>
                            <code>2.00</code>
                            <definition>2-Mostly lawn</definition>
                        </codeDefinition>
                        <codeDefinition>
                            <code>3.00</code>
                            <definition>3-Some lawn</definition>
                        </codeDefinition>
                    </enumeratedDomain>
                </nonNumericDomain>
            </ordinal>
        </measurementScale>
    </attribute>
    <attribute id="att.2">
        <attributeName>Year</attributeName>
        <attributeDefinition>Calendar year of the observation from years 1990 -
2010</attributeDefinition>
        <storageType>integer</storageType>
        <measurementScale>
            <dateTime>
                <formatString>YYYY</formatString>
                <dateTimePrecision>1</dateTimePrecision>
                <dateTimeDomain>
                    <bounds>
                        <minimum exclusive="false">1993</minimum>
                        <maximum exclusive="false">2003</maximum>
                    </bounds>
                </dateTimeDomain>
            </dateTime>
        </measurementScale>
    </attribute>
    <attribute id="att.7">
```

```
        <attributeName>Count</attributeName>
        <attributeDefinition>Number of individuals observed</attributeDefinition>
        <storageType>integer</storageType>
        <measurementScale>
            <interval>
                <unit>
                    <standardUnit>number</standardUnit>
                </unit>
                <precision>1</precision>
                <numericDomain>
                    <numberType>whole</numberType>
                    <bounds>
                        <minimum exclusive="false">0</minimum>
                    </bounds>
                </numericDomain>
            </interval>
        </measurementScale>
        <missingValueCode>
            <code>NaN</code>
            <codeExplanation>value not recorded or invalid</codeExplanation>
        </missingValueCode>
    </attribute>
    <attribute id="att.7">
        <attributeName>cond</attributeName>
        <attributeLabel>Conductivity</attributeLabel>
        <attributeDefinition>measured with SeaBird Elecronics
            CTD-911</attributeDefinition>
        <storageType>float</storageType>
        <measurementScale>
            <ratio>
                <unit>
                    <customUnit>siemensPerMeter</customUnit>
                </unit>
                <precision>0.0001</precision>
                <numericDomain>
                    <numberType>real</numberType>
                    <bounds>
                        <minimum exclusive="false">0</minimum>
                        <maximum exclusive="false">40</maximum>
                    </bounds>
                </numericDomain>
            </ratio>
        </measurementScale>
    </attribute>
</attributeList>
```

The examples below show complete entity trees for **\<spatialVector\>** and **\<spatialRaster\>** converted via XSLT (stylesheet) from ESRI metadata format. For details see "Other Resources".

Example: Entity and attribute information for spatialVector
```
<spatialVector id="Landuse for Ficity in 1955">
    <entityName>Landuse for Ficity in 1955</entityName>
    <entityDescription>This GIS layer represents a reconstructed generalized
landuse map for the area of current Ficity around the time period of
1955.</entityDescription>
    <physical>
        <objectName>fls-20.zip</objectName>
        <dataFormat>
```

```
        <externallyDefinedFormat>
            <formatName>Shapefile</formatName>
        </externallyDefinedFormat>
    </dataFormat>
    <distribution>
        <online>
            <onlineDescription>f1s-20 Zipped Shapefile File</onlineDescription>
            <url function="download"> http://www.fsu.edu/lter/data/fls-
20.zip</url>
        </online>
    </distribution>
</physical>
<attributeList id="Landuse for Ficity in 1955.attributeList">
    <attribute id="Landuse for Ficity in 1955.FID">
        <attributeName>FID</attributeName>
        <attributeDefinition>Internal feature number.</attributeDefinition>
        <storageType
typeSystem="http://www.esri.com/metadata/esriprof80.html">OID</storageType>
        <measurementScale>
            <nominal>
                <nonNumericDomain>
                    <textDomain>
                        <definition>Sequential unique whole numbers that are
automatically generated.</definition>
                    </textDomain>
                </nonNumericDomain>
            </nominal>
        </measurementScale>
    </attribute>
    <attribute id="Landuse for Ficity in 1955.Shape">
        <attributeName>Shape</attributeName>
        <attributeDefinition>Feature geometry.</attributeDefinition>
        <storageType
typeSystem=http://www.esri.com/metadata/esriprof80.html>Geometry</storageType>
        <measurementScale>
            <nominal>
                <nonNumericDomain>
                    <textDomain>
                        <definition>Coordinates defining the features.</definition>
                    </textDomain>
                </nonNumericDomain>
            </nominal>
        </measurementScale>
    </attribute>
    <attribute id="Landuse for Ficity in 1955.Z955">
        <attributeName>Z955</attributeName>
        <attributeDefinition>This field signifies the landuse value for each
polygon.</attributeDefinition>
        <storageType typeSystem="http://www.w3.org/2001/XMLSchema-
datatypes">string</storageType>
        <measurementScale>
            <nominal>
                <nonNumericDomain>
                    <enumeratedDomain>
                        <codeDefinition>
                            <code>Agriculture</code>
                            <definition>Agricultural land use</definition>
                        </codeDefinition>
                        <codeDefinition>
                            <code>Urban</code>
                                    <definition>Urbanized area</definition>
                        </codeDefinition>
                        <codeDefinition>
```

```
                                <code>Desert</code>
                                <definition>Unmodified area</definition>
                        </codeDefinition>
                        <codeDefinition>
                                <code>Recreation</code>
                                <definition>Recreational land use</definition>
                        </codeDefinition>
                    </enumeratedDomain>
                </nonNumericDomain>
            </nominal>
        </measurementScale>
    </attribute>
</attributeList>
<geometry>Polygon</geometry>
<geometricObjectCount>78</geometricObjectCount>
<spatialReference>
    <horizCoordSysName>NAD_1927_UTM_Zone_12N</horizCoordSysName>
</spatialReference>
</spatialVector>
```

Example: Entity and attribute information for spatialRaster

```
<spatialRaster id="fi_24k">
    <entityName>fi_24k</entityName>
    <entityDefinition>Ficiticiou State 7.5 Minute Digital Elevation
Model</entityDefinition>
    <physical>
        <objectName> fls-30.zip </objectName>
        <dataFormat>
            <externallyDefinedFormat>
                <formatName>Esri Grid</formatName>
            </externallyDefinedFormat>
        </dataFormat>
        <distribution>
            <online>
                <onlineDescription>f1s-30 zipped raster data File</onlineDescription>
                <url function="download"> http://www.fsu.edu/lter/data/fls-
30.zip</url>
            </online>
        </distribution>
    </physical>
    <attributeList id="fi_24k.attributeList">
        <attribute id="fi_24k.ObjectID">
            <attributeName>ObjectID</attributeName>
            <attributeDefinition>Internal feature number.</attributeDefinition>
            <storageType
typeSystem="http://www.esri.com/metadata/esriprof80.html">OID</storageType>
            <measurementScale>
                <nominal>
                    <nonNumericDomain>
                        <textDomain>
                            <definition>Sequential unique whole numbers that are
automatically generated.</definition>
                        </textDomain>
                    </nonNumericDomain>
                </nominal>
            </measurementScale>
        </attribute>
        <attribute id="fi_24k.Cell Value">
            <attributeName>Cell Value</attributeName>
```

```xml
            <attributeDefinition>Elevation Value</attributeDefinition>
            <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html"
                    >Integer</storageType>
            <measurementScale>
                <ratio>
                    <unit>
                        <standardUnit>meter</standardUnit>
                    </unit>
                    <precision/>
                    <numericDomain>
                        <numberType>integer</numberType>
                        <bounds>
                            <minimum exclusive="true">-5193.000000</minimum>
                            <maximum exclusive="true">14785.000000</maximum>
                        </bounds>
                    </numericDomain>
                </ratio>
            </measurementScale>
        </attribute>
        <attribute id="fi_24k.Count">
            <attributeName>Count</attributeName>
            <attributeDefinition>Count</attributeDefinition>
            <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html"
                    >Integer</storageType>
            <measurementScale>
                <ratio>
                    <unit>
                        <standardUnit>number</standardUnit>
                    </unit>
                    <precision/>
                    <numericDomain>
                        <numberType>whole</numberType>
                    </numericDomain>
                </ratio>
            </measurementScale>
        </attribute>
    </attributeList>
    <spatialReference>
        <horizCoordSysName>NAD_1927_UTM_Zone_12N</horizCoordSysName>
    </spatialReference>
    <horizontalAccuracy>not available</horizontalAccuracy>
    <verticalAccuracy>not available</verticalAccuracy>
    <cellSizeXDirection>30.0</cellSizeXDirection>
    <cellSizeYDirection>30.0</cellSizeYDirection>
    <numberOfBands>1</numberOfBands>
    <rasterOrigin>Upper Left</rasterOrigin>
    <rows>21092</rows>
    <columns>18136</columns>
    <verticals>1</verticals>
    <cellGeometry>matrix</cellGeometry>
</spatialRaster>
```

constraint

This element tree is found at (XPath):

/eml:eml/dataset/dataTable/constraint
/eml:eml/dataset/view/constraint
/eml:eml/dataset/spatialRaster/constraint
/eml:eml/dataset/spatialVector/constraint

/eml:eml/dataset/storedProcedure/constraint

The **<constraint>** tree is for describing any integrity constraints between entities (e.g. tables), as they would be maintained in a relational management system. Use of the **<constraint>** tree is encouraged when data elements contain integrity constraints from a relational database. Example TO-DO shows the constraints for the **<attributeList>** in Example TO-DO. If there are constraints in which several columns are involved, these should be described in methods/qualityControl, since EML is not currently equipped to handle keys defined by multiple columns. When the **<constraint>** tree is used, all of the entities that may be referenced should be in the same package. There are six child elements:

<primaryKey> is an element which declares the primary key in the entity to which the defined constraint pertains.

<uniqueKey> is an element which represents a unique key within the referenced entity. This is different from a primary key in that it does not form any implicit foreign key relationships to other entities; however it is required to be unique within the entity.

<nonNullConstraint> defines a constraint that indicates that no null values should be present for an attribute in this entity.

<checkConstraint> defines a constraint which checks a conditional clause within an entity.

<foreignKey> defines an SQL statement or other language implementation of the condition for a check constraint. Generally this provides a means for constraining the values within and among entities. It also provides the means to meaningfully link table for explanation of codes (de-normalization).

<joinCondition> defines a foreign key relationship among entities which relates this entity to another's primary key.

The **<primaryKey>**, **<uniqueKey>**, **< nonNullConstraint>** require an additional **<key>** tag defining the attribute to which this constraint applies, referenced by its id attribute (described in another area). All **<ConstraintType>** entities require additional **<constraintName>** and **<attributeReference>** tags.

Example: constraint
```
<constraint id="soil_chemistry.PRIMARY">
    <primaryKey>
        <constraintName>PRIMARY</constraintName>
        <key>
            <attributeReference>soil_chemistry.ID</attributeReference>
        </key>
            </primaryKey>
```

```
        </constraint>
        <constraint id="soil_chemistry.FK_soil_chemistry_sites">
            <foreignKey>
                <constraintName>FK_soil_chemistry_sites</constraintName>
                <key>
                    <attributeReference>soil_chemistry.site_id</attributeReference>
                </key>
                <entityReference>sites</entityReference>
            </foreignKey>
        </constraint>
```

additionalMetadata

This element tree is found at (XPath):
eml:eml/additionalMetadata

<additionalMetadata> is a flexible field for including any other relevant metadata that pertains to the resource being described. Its content must be valid XML. A unit as a **<customUnit>** must be described in this tree.

<describes> (optional) is a pointer to an "id" attribute on an EML element ("id" described in another area). This pointer must be identical to the attribute it is pointing at, so that automated processes are able to associate **<additionalMetadata>** to the described attribute. If the **<describes>** element is omitted, it is assumed that the **<additionalMetadata>** content applies to the entire EML document.

<metadata> contains the additional metadata to be included in the document. The contents can be any valid XML. This element should be used for extending EML to include metadata that is not already available in another part of the EML specification, or to include site- or system-specific extensions that are needed beyond the core metadata. The additional metadata contained in this field describes the element referenced in the **<describes>** element preceding it. If **<describes>** is not used, either **<metadata>** must contain sufficient information to define the association between **<additionalMetadata>** or the **<additionalMetadata>** can be presumed to apply to the entire data package.

An example of "sufficient information to define the association" is the definition of a **<customUnit>**. The EML Parser expects to find the description of a **<customUnit>** in the id attribute of a **<unit>** element in a **<unitList>**, i.e., at /eml:eml/additionalMetadata/metadata/unitList/unit. For example, "stmml:unit id="siemenPerMeter""points at the **<customUnit>** "simenPerMeter". The EML Parser is available from GitHub, with the EML project. For descriptions of custom units see "Other Resources".

Example: additionalMetadata custom unit
```
    <additionalMetadata>
        <metadata>
```

```
        <stmml:unitList>
            <stmml:unit id="siemensPerMeter" name="siemensPerMeter"
  abbreviation="S/m" unitType="conductance" parentSI="siemen" multiplierToSI="1"
  constantToSI="0">
                <stmml:description>conductivity unit</stmml:description>
            </stmml:unit>
        </stmml:unitList>
    </metadata>
</additionalMetadata>
```

III. DESCRIPTIONS OF EML SAMPLE FILES PROVIDED WITH THIS DOCUMENT

Example 1: Complete EML from which the examples in this document were derived can be found in EDI's GitHub repository https://github.com/EDIorg/dm-best-practices

INDEX
Tbd.