

# EML Best Practices for LTER Sites

## CONTENTS

- I. Introduction ..... 3
  - I.1. Changes from EML Best Practices Version 1 (2004) ..... 3
  - I.2 EML Management..... 5
    - I.2.1 Creating Datasets ..... 5
    - I.2.2 The Attribute – Value Data Model..... 6
    - I.2.3. EML produced from Geographic Information Systems (GIS) systems ..... 7
- II. Detailed content recommendations For Elements and Attributes ..... 7
  - II.1 General Recommendations..... 7
  - II.2 Recommendations by element..... 8
    - The root element: <eml:eml>..... 8
    - @packageId (XML attribute)..... 9
    - id, system and scope (XML attribute group) ..... 10
    - access ..... 11
    - dataset ..... 12
    - alternateIdentifier ..... 13
    - title (dataset) ..... 13
    - People and Organizations (Parties) ..... 13
      - creator ..... 14
      - metadataProvider ..... 15
      - associatedParty..... 16
      - contact ..... 16
      - publisher..... 17

pubDate..... 17

abstract..... 17

keywordSet and keyword..... 18

intellectualRights ..... 19

distribution..... 20

coverage..... 21

    geographicCoverage ..... 22

    temporalCoverage ..... 24

    taxonomicCoverage..... 25

maintenance..... 27

methods ..... 28

project..... 31

[entity] = dataTable, spatialRaster, spatialVector, storedProcedure, view, otherEntity ..... 32

attributeList ..... 36

constraint ..... 46

additionalMetadata ..... 48

III. Recommendations for Compatibility with External Applications..... 49

    III.1 Metacat ..... 49

    III.2 LTERMapS ..... 50

    III.3 Provenance Tracking in EML..... 50

    III.4 EML Congruence Checker ..... 52

IV. Descriptions of EML sample files provided with this document..... 53

V. Other resources..... 53

VI. Contributors to this Document..... 54

INDEX ..... 55

**Conventions used in this document:** Fonts and type faces: example XML nodes are in fixed-width font. XML element and attribute names, XPath, and references to element names in text are in bold face, and elements are surrounded by angle brackets. Recommendations or notes on possible changes to the EML schema are in italic.

## I. INTRODUCTION

The Ecological Metadata Language (EML) was adopted in 2003 as the exchange format for metadata contributed to the LTER network. As such, EML is one part of an LTER site's information management system. This document contains recommended current views for best practices for EML content for our network's use, and is also intended to augment the EML schema documentation (normative documents) for a less technical audience. Some notes on implementation are included as appropriate. This is one component of several Best Practice documents available to LTER sites, and related documents and resources are listed in the Additional Resources (Section V). The recommendations are directed towards achieving the following specific goals:

- a) Provide guidance and clarification in the implementation of EML for datasets
- b) Minimize heterogeneity of LTER EML documents to simplify development and reuse of software built for EML datasets
- c) Maximize interoperability of LTER EML documents to facilitate data synthesis

Please cite this document as:

Long Term Ecological Research Network. 2011. EML Best Practices for LTER Sites.

### I.1. CHANGES FROM EML BEST PRACTICES VERSION 1 (2004)

The EML Best Practice recommendations have evolved over time. They are the cumulative result of several working groups comprised of LTER information managers and Network Office representatives (Section VI). Each group aligned its recommendations with the current capability at most of the contributing LTER sites. As of 2011, 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. Following are the major changes between this document and Version 1 (2004).

EML Version: EML 2.1 is the current version of the specification, and all examples and recommendations refer to this. However, all recommendations for element content can be applied to EML 2.0.1.

Document Organization and EML “Metadata Levels”: Version 1 (2004) of this document discussed specific elements in sections of “levels of EML completeness” which corresponded to functionality tiers identified by NISAC in 2004. It used the completeness levels’ increasingly comprehensive descriptions of data resources so that EML could be implemented gradually. In this document (Version 2), the detailed content recommendations (Section II) are organized more simply and more or less in EML-document order. Other organizing concepts were considered, .e.g., distinguishing between ‘human-readable’ and ‘machine-readable’ elements. However, at present all information in EML is consumed by humans while only few elements are used by applications, (e.g. access control, geographic coverage, and the structural information in the entity tree). Few applications are able to make semantic use of TextType fields such as <abstract> and <methodStep>, although these fields can be machine-searched. Some fields with simple string content that is currently ad hoc (e.g., <attributeName>, <customUnit>) are currently undergoing a network wide standardization effort and will become machine-read in the near future.

Element Descriptions: Several important EML elements can be placed at multiple locations in a dataset, e.g., coverage and methods. To illustrate all possible locations, the allowable XPath locations are listed for every element. XPath is a representation of the document’s hierarchical structure, which is similar to fileserver directory paths. For each element, recommendations for content have been grouped together when the element first appears (often at the dataset level). This does not imply that these trees belong only at that level; in fact, locating trees at the most granular level is still advantageous and recommended.

Data Availability: The LTER Network policy on availability is that data products should be generally available. Therefore, there are limited recommendations here for how to describe a data entity which is unavailable to the public. More information on access can be found in the EML normative documents.

Data Types: Recommendations have been added in Section I.2 for describing certain dataset types or for common situations encountered in LTER datasets, specifically, conversions of spatial metadata to EML, and “string-of-pearls” or “attribute – value” data models.

External Applications: A new section (III) has been added to contain recommendations for EML which is to be used in specific external applications. These include Metacat, the LTERMaPs project, and PASTA (specifically, the EML Congruency Checker and its use in validating LTER datasets, and provenance for derived datasets).

## I.2 EML MANAGEMENT

The terms “dataset” and “data package” are somewhat interchangeable. The term “data package” is used here to mean the published unit of data and metadata together. “Dataset” has a special meaning within EML documents, because it is the top-level container for data objects (i.e., <dataset> as opposed to the other top-level elements, <citation>, <software> and <protocol>). However, a logical “dataset” may be something slightly different. For example, management needs for an ongoing collection may dictate that a dataset is published in increments, and the entire “Ongoing Climate Dataset” may be composed of multiple contributions, each of which could be an individual EML “dataset”. Subsequent data access applications may use the term dataset differently.

### *I.2.1 Creating Datasets*

Several approaches to creating datasets or data packages have emerged in the network and are all valid concepts. In general, this document does not recommend any one pattern. Following are several examples:

- (1) Data collected with defined beginning and end dates are published in logical units, with all ancillary data are described together in one EML file (i.e., <dataset>). EML accommodates descriptions of multiple data entities in one metadata document. For instance, a data table could be accompanied by a KML file or shapefile to describe the sampling locations.
- (2) Data from a large sampling campaign where many somewhat independent parameters are measured could either be combined into one EML dataset with multiple data entities, or broken up into several datasets, all accompanied by some of the same metadata, and each with other unique metadata. Both patterns are valid, and it is left up to the dataset designers to decide which style works best for their particular data and circumstances. It is wise to keep the data user in mind, and whether certain data entities naturally belong together.
- (3) For time-series observations (instrumented or not), some designers choose to update existing EML metadata when methods or other ancillary material remains constant. Others may choose to create new datasets for each logical unit (e.g., a summer sampling season). There are advantages and disadvantages to both approaches. The first approach does not allow for strict versioning of the data while the second approach forces the user to download and integrate many separate entities.
- (4) The dataset designer should consider some reasonable approaches to deciding which information is encoded directly in EML and which is better left in ancillary files (or tables).

For example, codes used in a column of the data table may either be encoded directly in EML or stored in a “look-up table” (described in EML as another entity) and linked via a **<constraint>** (foreignKey constraint) element or described directly in the attribute with the enumeratedDomain/entityCodeList/ tree. Another example is site locations, which may be encoded in EML under **<methods>** **<sampling>** **<spatialSamplingUnits>** or supplied as a KML or shape file and described as another entity. The decision between these two options probably depends mostly on the number of units to encode in EML. For example, 500 codes are better left in a second data entity, while ten code-definition pairs should be encoded directly in EML.

### ***1.2.2 The Attribute - Value Data Model***

The Attribute - Value or “string of pearls” data model is widely used for certain kinds of observational data where the more conventional matrix type model would cause many empty cells. This data model treats each point observation as a single record containing fields for location, time, variable (attribute) name, and value, plus various flags for methods, data quality etc. For ecologists, this data model is especially useful for (1) biodiversity and (2) sensor data that may otherwise require very wide tables in a matrix format, and/or have many empty cells e.g., when species are not regularly observed or sensors are not employed consistently.

The advantages of the Attribute - Value data model are its flexibility and efficiency. However datasets in this format are not easily described in the current version of EML. The basic problem is that the values in the value field do not necessarily share the same attributes (e.g. numerical type, collection methods, unit, and precision). So for now the best practices recommendations are as follows:

Biodiversity data: If all of the values in a given table share the same units (e.g., presence/absence, count, percent abundance), then the table should be represented in this Attribute - Value format and described accurately in EML. If not, then the data should be represented in matrix format, or possibly several matrices if the table is unreasonably wide.

Sensor data: The Attribute - Value format is especially efficient for sensor data and variations of this data model have been developed in communities handling large volumes of sensor data (e.g., ODM or Observation Data Model from CUAHSI). For a single site this data model is useful when sensors are frequently changed or redeployed or when several different sets of sensor deployments exist. For accurate description in EML, it is recommended that each data table contain only the Attribute - Value formatted values for a single sensor. If a single metadata document contains data from multiple sensors, values for each sensor should appear as separate entities (data tables) within one dataset. (Most

of these data models describe the concept of a single “data stream,” which may be used to achieve this recommended approach dynamically.)

### ***1.2.3. EML produced from Geographic Information Systems (GIS) systems***

Several established standards exist for documenting spatial datasets. The most common formats are Federal Geographic Data Committee (FGDC) geospatial standard, National Biological Information Infrastructure (NBII) biological profile, International Standards Organization (ISO) standard 19115, and ArcGIS metadata format. Tools are available at <http://www.fgdc.gov/metadata/geospatial-metadata-tools> for preparing FGDC compliant metadata, which can then be converted into EML using XSL transformation style sheets. The LTER network has the following stylesheets available:

esri2eml.xsl: converts XML documents from ArcCatalog metadata editor (up to ARCGIS version 9.3.1) <http://im.lternet.edu/project/Esri2Eml/docs>

bdp2eml.xsl: converts FGDC and NBII Biological Profile metadata to EML.  
<http://im.lternet.edu/project/bdp2eml>

Valid EML will be produced by esri2eml.xsl if the following guidelines are followed when preparing your FGDC, ISO, or ArcGIS compliant metadata:

- a) Personnel names: Last name, first name (i.e., use a comma to separate last name and first name) and for multiple names and contact elements (phone, fax, e-mail), use repeating fields.
- b) Attribute definition: Every attribute value needs a corresponding value definition. If the value definition is not available, record the value definition as “unknown”, “none”, or null”.

## **II. DETAILED CONTENT RECOMMENDATIONS FOR ELEMENTS AND ATTRIBUTES**

### **II.1 GENERAL RECOMMENDATIONS**

Following are general best practices for creating EML dataset metadata:

Metadata Distribution: Do not publicly distribute EML documents containing elements with incorrect information as dataset metadata (i.e. as a workaround for problems with metadata content availability or to meet EML validation requirements). EML produced as

draft, demonstration or for testing purposes should be clearly identified as such and not contributed to public metadata archives or clearinghouses.

Text Elements: Use EML text formatting tags whenever possible (e.g. **<section>**, **<para>**, **<orderedlist>**, etc.). Use **<literalLayout>** only when the field must contain HTML for formatting that is not available with Text Type. Note that the TextType elements were taken from docbook, and so do not use the “camelCase” notation that was generally adopted for EML.

Versioning: It is recommended that metadata and data set versioning be handled at the site level. For example, data entities and corresponding metadata for ongoing long term datasets can be archived annually with correct end dates and versions of the EML. Package IDs of the format {scope}.{identifier}.{revision} will be used in the LTER-NIS. Scope is of the format "knb-lter-fls". Data updates or revisions to data or metadata should retain the same identifier while incrementing the revision.

EML “id” Attribute: Many EML elements are allowed to have an “id” attribute. In EML 2.1.0, all ids in a document must be unique. Care should be exercised when using id attributes to reference and re-use EML content. It may be preferable to duplicate content without ids when generating EML dynamically from a relational database system to avoid potential id conflicts. See below for more information on the id attribute.

High-priority Elements: To support locating datasets by time, geographic location, and taxonomically, metadata should provide as much information as possible, in the three **<coverage>** elements of **<temporalCoverage>** (when), **<taxonomicCoverage>** (what), and **<geographicCoverage>** (where) for the dataset.

Metadata should include detailed descriptions in the **<project>**, **<methods>**, **<protocols>**, and **<intellectualRights>** elements in order for a potential user to evaluate the relevance of the data package for their research study or synthesis project.

The change history should be recorded in the **<maintenance>** element.

## II.2 RECOMMENDATIONS BY ELEMENT

### ***The root element: <eml:eml>***

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

After the XML declaration, all EML documents must have one root element (<**eml:eml**>). There is a benefit to including a schema declaration which points to an online resource. It is recommended that the following schema declaration be included in the root element:

```
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0  
http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
```

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

<**access**>

<**dataset**>

<**additionalMetadata**>

### **@packageId (XML attribute)**

This attribute is required in all EML2.1 documents, and is found at this location (XPath):  
**/eml:eml/@packageId**

As outlined in other sections (see versioning above, **@id** and <**alternateIdentifier**>, below) each site should manage unique identifiers and versioning at the local level (see **@system** discussion below). The LTER Network currently uses “Metacat” as its EML repository. Most sites synchronize the **@packageId** with the Metacat “docid” designated in the harvester or at document upload, however synchronization is not required. Metacat and its harvester rely on numerical data set ids and revision numbers for document management and synchronization. This may necessitate a workaround for sites that use non-numeric ids or don’t version data sets. Possible solutions include differentially generating EML optimized for Metacat (for sites capable of dynamic EML generation) and XSLT transformation.

Currently, the **@packageId** attribute is used to identify a site’s EML documents for searches in the LTER Data Catalog. A site’s **@packageId** attribute in EML contributed to the KNB Metacat should be standardized as follows, or another format agreed upon by the site and LNO Data Catalog managers.

knb-lter-[site].[dataset number].[revision], e.g. knb-lter-fls.187.4

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

***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, unless the **@id** attribute is used as a reference. Each **@id** must be unique in one EML document, i.e. a **<creator>** must have a different id than a **<dataTable>**. If the same person appears as dataset creator and protocol 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.

The current restrictions 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: 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
http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
  packageId="knb-lter-fls.21.3"
  system="FLS"
  scope="system">
```

---

**access**

The dataset 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 file and its data entity. Values must be applicable by the system where data is stored. Metacat access control format 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. Access elements for documents contributed to the KNB Metacat should be formed according to the Example below. With the exception of certain sensitive information, metadata should be publicly accessible. The **<access>** element is optional, and if omitted, presumably only the dataset submitter will be allowed access.

---

**Example: 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 (see also section I.1).

**<alternateIdentifier>**

**<shortName>**

**<title>**

**<creator>**

**<metadataProvider>**

**<associatedParty>**

**<pubDate>**

**<language>**

**<series>**

**<abstract>**

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

```
/eml:eml/ dataset/alternateIdentifier
/eml:eml/ dataset/[entity]/alternateIdentifier
```

The site's data set id 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 dataset belongs to more than one LTER site by including the each site's 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 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 mention the data collected, geographic context and research site (what, where). If the data will not be updated, the time frame (when).

---

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

### creator

This element is found at this location (XPath):

**/eml:eml/dataset/creator**

**<creator>** The creator is considered to be the author of the dataset, i.e. the person(s) responsible for intellectual input into its creation. **<surName>** and **<givenName>** elements are used to build dataset citations, so these should be completed fully for credit to be understandable. For long term data sets, include the name of the LTER Site (using the **<organizationName>**) or role of Site PI (using **<postitionName>**). It should be kept in mind that different approaches taken by sites have led to confusion over how to best search for long term datasets, 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 datasets.

---

### Example: 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>
</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>
</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>
</creator>

```

---

### **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). Complete the **<address>**, **<phone>**, **<electronicMailAddress>**, and **<onlineURL>** elements for each **<metadataProvider>** element.

---

### Example: 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>
</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: 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&query=itestuser</onlineUrl>
  <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 datasets. Full contact information should be included for the position of data manager or other designated site 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: contact**

```
<contact id="pos-4">
  <positionName> 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>
</contact>

```

## ***publisher***

This element is found at this location (XPath):

**/eml:eml/dataset/publisher**

The LTER site should be listed as the **<publisher>** of the data set. List the LTER site name, fully spelled out, in the **<organizationName>** element. Complete the **<address>**, **<phone>**, **<electronicMailAddress>**, and **<onlineURL>** elements for each publisher element.

*Recommendation for web display of LTER-EML: use **<publisher>** for the organization information*

Example: publisher using a reference to a creator from above

```

  <publisher>
    <references system="FLS">org-1</references>
  </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. The **<pubDate>** should be updated when data and/or metadata are updated or re released. The format can be either a 4-digit year, or an ISO date (yyyy-mm-dd).

## ***abstract***

This element is found at these locations (XPath):

**/eml:eml/dataset/abstract**

**/eml:eml/dataset/project/abstract**

In datasets, the abstract element can appear at the resource level or the project level. The **<abstract>** element will be used for full-text searches, and it should be rich with descriptive text. Extensive description should include what, when, and where information, some taxonomic information, as well as whether the dataset is ongoing or completed. Some general methods description is appropriate, and the 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>** tags. For example, use one **<keywordSet>** for a keyword identifying the LTER site, one for keywords from the LTER controlled vocabulary, one for the LTER core area keywords, etc.. Currently each **<keywordSet>** can include the name of a specific thesaurus (in the optional tag **<keywordThesaurus>**). Keywords that should be included are as many as possible from the LTER controlled vocabulary, at least one LTER core area (if appropriate), the three letter site acronym, some meaningful geographic place names (e.g. state, city, county), network acronym (LTER, ILTER, etc.), organizational affiliation, funding source (i.e. co-funded with other sources, non-LTER funding etc.). In addition to specific keywords, relevant conceptual keywords should also be included.

---

Example: pubDate, abstract, keywordSet, keyword

```
<pubDate>2000</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>** should contain site data access policy, plus a description of any deviation from the general access policy specific for this particular dataset (e.g. restricted-access datasets). The timeframe for release should be included as well. For example, LTER Network-wide data should be released on-line within 2-3 years, and if not, the reason needs to be documented in the metadata. (See also LTER Network Data Access Policy: <http://www.lternet.edu/data/netpolicy.html>)

---

### Example: intellectualRights

```

<intellectualRights>
  <section>
    <title>Copyright Notice</title>
    <para> Copyright Board of Regents, Fictitious State University. This dataset
is released to the public and may be used for academic, educational, or commercial
purposes subject to the following restrictions:</para>
    <para>
      <itemizedlist>
        <listitem>
          <para>While FLS LTER will make every effort possible to control and
document the quality of the data it publishes, the data are made available "as
is".</para>
        </listitem>
        <listitem>
          <para>FLS LTER cannot assume responsibility for damages resulting from
mis-use or mis-interpretation of datasets or from errors or omissions that may
exist in the data.</para>
        </listitem>
        <listitem>
          <para>It is considered a matter of professional ethics to acknowledge
the work of other scientists that has resulted in data used in subsequent
research.</para>
        </listitem>
        <listitem>
          <para>FLS LTER expects that any use of data from this server will be
accompanied with the appropriate citations and acknowledgments. </para>
        </listitem>
        <listitem>
          <para>FLS LTER encourages users to contact the original investigator
responsible for the data that they are accessing. Where appropriate, researchers
whose projects are integrally dependent on FLS LTER data are encouraged to consider
collaboration and/or co-authorship with original investigators. </para>
        </listitem>
      </itemizedlist>
    </para>
  </section>

```

```

    <para>FLS LTER requests that users submit to FLS LTER one copy of any
    publication resulting from the use of data obtained from this site. </para>
  </listitem>
  <listitem>
    <para>FLS LTER requests that users not redistribute data obtained from
    this site. However, links or references to this site may be freely posted.</para>
  </listitem>
</itemizedlist>
</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 appears at the dataset and entity levels and contains information on how the data described in the EML document 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 as XML. 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 URL provides only information about downloading the object but does not directly return the data stream, then the **function** attribute should be set to "information". If accessing the URL directly returns the data stream, then the **function** attribute should be set to "download". If the **function** attribute is omitted, then "download" is implied

An EML dataset should include at least one URL; at a minimum, this should be at the **<dataset>** level (XPath: **/eml:eml/dataset/distribution/url**), and may point to an application or website. This **<url>** **function** attribute can be set to either "information" or "download". However, a URL at the entity level (e.g, a dataTable at **/eml:eml/dataset/dataTable/physical/distribution/url**) should stream data to the

requesting application and should include an attribute **function** with the value “download”. In other words, at the entity level, the URL should lead directly to the data and not to a data catalog or intended-use page. For more information about describing a URL connection, see the EML documentation online.

When used at the entity level, an alternative tag is available to `<url>`, `<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.

Data access logging can be implemented by using the LTER Data Access Server (DAS) and its URL-proxy system. For more information about the DAS and how to use it, read the online documentation: [http://im.lternet.edu/im\\_practices/metadata/das](http://im.lternet.edu/im_practices/metadata/das)

---

Example: distribution

```

<distribution>
  <online>
    <onlineDescription>fls-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>fls-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 of a dataset in terms of space, taxonomy, and time, **<geographicCoverage>**, **<taxonomicCoverage>**, and **<temporalCoverage>**. Populating these elements as recommended allows for more advanced uses.

### **geographicCoverage**

**General Information:** The **<geographicCoverage>** element is used to describe geographic locations of research sites and areas related to the dataset. It is recommended to use the element at different levels for different types of information. Clearly there are many ways this element can be used and this document recommends one approach to make EML within LTER more consistent and easier to use in applications like the LTERMapS.

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, the bounding box could describe the entire LTER site as stored in SiteDB. However, it is strongly recommended that a more accurate extent of observations be provided, i.e., the maximum extent of the data, for each of east, west, north and south. Additional **<geographicCoverage>** elements may be entered at the dataset level if there are significant distances between study sites and it would be confusing if they were grouped into one bounding box. For example, cross-site studies where a bounding box would represent North American or more should have two bounding boxes, one 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 locations, those sites should appear at `eml:eml/dataset/methods/sampling/spatialSamplingUnits/coverage`. A list of individual sampling sites should be entered under **<spatialSamplingUnits>**, each site in a separate coverage element. More information about sampling sites can be found under **<methods>**,

---

Example: geographicCoverage for single sampling point locations

```
<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 commonly used datum (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 above. See the EML specification for more information and other examples.

**<geographicDescription>** The description is a string. It should be comprehensive 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 this information is not otherwise encoded under <**geographicCoverage**>.

The <**datasetGPolygon**> element may be included when the 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 gRingPoints. Documentation for an FGDC G-Ring states that 4 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 polygon is closed by joining the last point 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 gRingPoints. This is used if there is an internal polygon to be excluded from the outer polygon, e.g, a lake which is 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. These include GIS shapefiles, Keyhole Markup Language (KML or KMZ), or EML spatial modules. For more information, see Section III, Recommendations for Compatibility with External Applications.

### **temporalCoverage**

The <**temporalCoverage**> element of a dataset represents the period of time the data were collected, not the year the study was put together 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 dataset may be considered "ongoing", i.e., data are added continuously. It is not currently valid to leave an empty <**endDate**> tag in EML. For this type of dataset, the simplest solution is to populate the <**endDate**> element with the end of the current year

and update the metadata annually. Ideally, however, the **<endDate>** tag should reflect 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 **<maintenance>** 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 whenever possible 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 and is not generally recommended for data set documentation.

The optional **taxonomicCoverage/taxonomicSystem** trees may be used to detail the use of taxonomic identification resources as well as provide information 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 which 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>
    </identifierName>
    <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 florida</taxonRankValue>
  </taxonomicClassification>
</taxonomicCoverage>

```

```

        <commonName>Florida Melampus</commonName>
      </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

*Note: in earlier EML versions, both "<method>" and "<methods>" elements were found, which caused confusion. The elements have now been standardized to "<methods>".*

**General Information:** 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 methods at this level are preferable. If further refinement is needed, methods can be defined for individual data entities or even individual **<attribute>**s if necessary. The scope of the method defined should match the EML document level at which 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 among sites and some users prefer to have all methods in one place. 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>**.

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.

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>**.

The **<dataSource>** tag is for nesting an EML dataset which may be an input to a **<methodStep>** of the data being described, e.g., calibration information for an instrument or input parameters for a model. This element will also be used by the PASTA provenance tracking system for recording the source data when a derived product is created and described with EML. For more information, see Section III, Recommendations for Compatibility with External Applications.

The optional **<sampling>** 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. It 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. The example shows the location of three insect traps within the LTER site's overall sampling area, using the coverage tree. Note that single sampling site locations should be listed under **<spatialSamplingUnits>**.

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

**<spatialSamplingUnits>** This tree should contain individual study sites or point locations. In the example, stations within the city are listed as content in the structured coverage elements of the studyExtent tree. The LTERMapS project plans to use content or internal references from this tree; please see Section III for when creating metadata for that use.

Like other trees under **<methods>**, **<qualityControl>** can be used at the dataset, entity or attribute level, whichever is appropriate. As a minimum add a text description in 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 suitable
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>

```

---

## ***project***

This element is found at this location (XPath):

**/eml:eml/dataset/project**

Originally, it was recommended that a **<project>** tree be included which held a very general descriptions of the project sponsoring the data package, i.e., this LTER site. However, in practice it has turned out that if every dataset EML contains the same project description searches of the project tree become inefficient and bloated with inappropriate results. Instead, the current recommendation is to include trees for individual projects conducted at each LTER site, and nested if smaller sub-projects using **<relatedProject>**.

*NOTE: Research project management may also be done using the LTERProject schema. The supplementary XSL stylesheets for that schema include export of EML2.1 **<project>** trees. For details see online documentation: <http://im.lternet.edu/project/LTERProjectDatabase>).*

Minimally, the description of an LTER site-level project should include **<title>**, **<personnel>** and **<abstract>**, with the study area description and mission statement. The **<distribution>** tree should link to the project's home page, or alternatively could link to other publications describing the project. These elements are all used at the dataset level; see their individual sections for specific content recommendations.

Two elements are unique to the **<project>** tree:

**<fundingSource>** should contain the agency and grant number.

**<studyAreaDescription>** 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 **studyAreaDescription** tree also supports multiple **<coverage>** elements which 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

```

<project>
  <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>
  <funding>
    <para>National Science Foundation under Grant No. DEB 9999999, Fictitious Long-Term Ecological Research (FLS LTER)</para>
  </funding>
</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**

The entity types **<dataTable>**, **<otherEntity>** and **<view>** cover many commonly encountered data structures such as ASCII data tables, KML, and images. The other choices for entity (**<spatialRaster>**, **<spatialVector>**, **<storedProcedure>**) will be addressed in more depth in a future version of this document; however, the spatial data entities (**<spatialRaster>**, **<spatialVector>**) can be created via stylesheet conversions when there is ESRI or FGDC metadata available (see section I.2.3) and examples are given below. Spatial datasets that do not have FGDC or ESRI compliant metadata records will need to be created by some other mechanism (e.g., manually). Vector data spatial objects (lines, points, polygons) should use the **<spatialVector>** tree and include information about the vector's geometry type, count and topology level. Grid or raster cell data should be described with the **<spatialRaster>** tree and include the spatial organization of the raster cells, the cell data values, and if derived via imaging sensors, characteristics about the image and its individual bands.

The **<otherEntity>** element should be used for images and for KML files. KML files can also be converted to **<spatialVector>** types (see below).

Every EML data entity has certain 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. Often this is the original ascii file name.

**<entityDescription>** This should be a longer, more descriptive explanation of the data entity. It should describe the data contained within the entity. Like all descriptions, it is

human-readable, and should have enough information for a user to determine if it is appropriate for his/her 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, and might be the same as the **<entityName>**. However, a physical object (file) can have several entities, e.g. an Excel workbook with several sheets. If you are publishing Excel workbooks (not recommended practice), put the worksheet name in the **<entityName>**, and the filename in **<objectName>**.

**<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".

*Important: Whenever possible, sites should not publish data in dated, proprietary, binary formats such as MS-Excel workbooks.*

**<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. An exception is that an LTER-NIS DAS proxy URL may be used here so that data downloads may be logged by the Data Access Server. For more information about the DAS and how to use it, read the online documentation:

[http://im.lternet.edu/im\\_practices/metadata/das](http://im.lternet.edu/im_practices/metadata/das)

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.

*Recommendation: With the rare exception of sensitive information (for example, data covered by the Human Subjects Act), read-access to data entities should not be restricted.*

**<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>invertebrates</entityName>
  <entityDescription> invertebrate sampling data</entityDescription>
  <physical>
    <objectName>fls-1.csv</objectName>
    <size unit="rows">561</size>
    <characterEncoding>ASCII</characterEncoding>
    <dataFormat>
      <textFormat>
        <numHeaderLines>1</numHeaderLines>
        <numFooterLines>0</numFooterLines>
        <recordDelimiter>\r</recordDelimiter>
        <numPhysicalLinesPerRecord>1</numPhysicalLinesPerRecord>
        <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 which follow the common elements. Table 1 is a summary of the Entity types, with examples of typical uses and the elements used for that type.

Entity Type	Typical Uses	Elements following EntityGroup
<b>&lt;dataTable&gt;</b>	static ASCII table database export	<b>&lt;attributeList&gt;</b> <b>&lt;constraint&gt;</b>

		<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>	images KML maps	<attributeList> <constraint> <entityType>
<spatialRaster>	Remote sensing images	<attributeList> <constraint> <spatialReference> <georeferenceInfo> <horizontalAccuracy> <verticalAccuracy> <cellSizeYDirection> <numberOfBands> <rasterOrigin> <rows> <columns> <verticals> <cellGeometry> <toneGradation> <scaleFactor> <offset> <imageDescription>
<spatialVector>	ESRI shapfiles	<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 the official name of an attribute, typically the name of a field in a data table. This is often short and/or cryptic.

**<attributeLabel>** (optional): is used to provide a less ambiguous or cryptic alternative identification than what is provided in **<attributeName>**. This content may 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. One of the 5 scale types must be used: nominal, ordinal, interval, ratio, or dateTime, as follows:

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.

*Note: Both the nominal and ordinal scales are considered **<nonNumericDomain>** types, either text or an enumerated list. The **<enumeratedDomain>** applies to coded values, and 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”.*

**<interval>** These measurements are ordinal, but in addition, use equal-sized units on a scale between values. The starting point is arbitrary, so a value of zero is not meaningful. 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 Celsius is not “twice as hot” as 10 Celsius.

**<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.

*Note: The <interval> and <ratio> scales require additional tags describing <unit>, the <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 LTER EML documents, a registry and database of units is available <http://unit.lternet.edu/unitregistry/>. It is recommended that for LTER documents, the **<customUnit>** element be used for all units with content pulled from the Unit Registry, even in the case where a unit is listed in the standard unit dictionary. Units in LTER documents should be added to the Unit Registry.

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”.

**<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: <http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf>. 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” (e.g., 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 (Section V). It is important to note that the **dateTime** field should not be used for recording time durations. In that case, use a <**standardUnit**> such as seconds, nominalMinute or nominalDay, or a <**customUnit**> that defines the unit in terms of its relationship to SI second.

*Note: this measurement type was called <datetime> in previous versions of EML. It was changed from all lower case (<datetime>) to “camel case” (<dateTime>) for internal consistency.*

The <**missingValueCode**> is optional, but should be included to describe any missing value codes present in the data set (e.g. NaN, ND, 9999). Missing values should be explicitly coded, not just an empty field. 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="att.1">
    <attributeName>Site</attributeName>
    <attributeDefinition>FLS-LTER sampling site</attributeDefinition>
    <storageType>integer</storageType>
    <measurementScale>
      <ordinal>
        <nonNumericDomain>
          <enumeratedDomain>
            <codeDefinition>
              <code>1</code>
              <definition>FLS study site 1</definition>
            </codeDefinition>
            <codeDefinition>
              <code>2</code>
              <definition>FLS study site 2</definition>
            </codeDefinition>
            <codeDefinition>
              <code>3</code>
              <definition>FLS study site 3</definition>
            </codeDefinition>
            <codeDefinition>
              <code>4</code>
              <definition>FLS study site 4</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.3">

```

---

```

<attributeName>Month</attributeName>
<attributeDefinition>Calendar month of the observation</attributeDefinition>
<storageType>integer</storageType>
<measurementScale>
  <dateTime>
    <formatString>MM</formatString>
    <dateTimePrecision>1</dateTimePrecision>
    <dateTimeDomain>
      <bounds>
        <minimum exclusive="false">1</minimum>
        <maximum exclusive="false">12</maximum>
      </bounds>
    </dateTimeDomain>
  </dateTime>
</measurementScale>
</attribute>
<attribute id="att.4">
  <attributeName>Day</attributeName>
  <attributeDefinition>Calendar day of the observation</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <dateTime>
      <formatString>DD</formatString>
      <dateTimePrecision>1</dateTimePrecision>
      <dateTimeDomain>
        <bounds>
          <minimum exclusive="false">1</minimum>
          <maximum exclusive="false">31</maximum>
        </bounds>
      </dateTimeDomain>
    </dateTime>
  </measurementScale>
</attribute>
<attribute id="att.5">
  <attributeName>Transect</attributeName>
  <attributeDefinition>Transect number (randomly placed within the study
    site)</attributeDefinition>
  <storageType>integer</storageType>
  <measurementScale>
    <nominal>
      <nonNumericDomain>
        <textDomain>
          <definition>Transect number (randomly placed)</definition>
        </textDomain>
      </nonNumericDomain>
    </nominal>
  </measurementScale>
</attribute>
<attribute id="att.6">
  <attributeName>Species_Code</attributeName>
  <attributeDefinition>Coded species name of the insect
observed</attributeDefinition>
  <storageType>string</storageType>
  <measurementScale>
    <nominal>
      <nonNumericDomain>
        <enumeratedDomain>
          <codeDefinition>
            <code>G0</code>
            <definition>Other grasshoppers</definition>
          </codeDefinition>
          <codeDefinition>

```

---

```

        <code>G1</code>
        <definition>Orchelimum spp.</definition>
    </codeDefinition>
    <codeDefinition>
        <code>G2</code>
        <definition>Paroxya clavuliger</definition>
    </codeDefinition>
    <codeDefinition>
        <code>G3</code>
        <definition>Orphulella pelidna</definition>
    </codeDefinition>
    <codeDefinition>
        <code>G4</code>
        <definition>Mermiria intertexta</definition>
    </codeDefinition>
    <codeDefinition>
        <code>G5</code>
        <definition>Schistocerca obscura</definition>
    </codeDefinition>
    <codeDefinition>
        <code>G6</code>
        <definition>Romalea guttatus</definition>
    </codeDefinition>
</enumeratedDomain>
</nonNumericDomain>
</nominal>
</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.8">
    <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 section I.2.3.

---

#### 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>fls-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>
  <entityDescription>Ficiticiou State 7.5 Minute Digital Elevation
Model</entityDescription>
  <physical>
    <objectName> fls-30.zip </objectName>
  </physical>
</spatialRaster>

```

```

<dataFormat>
  <externallyDefinedFormat>
    <formatName>Esri Grid</formatName>
  </externallyDefinedFormat>
</dataFormat>
<distribution>
  <online>
    <onlineDescription>fls-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>
    <attributeDefinition>Elevation Value</attributeDefinition>
    <storageType typeSystem="http://www.esri.com/metadata/esriprof80.html"
>Integer</storageType>
    <measurementScale>
      <ratio>
        <unit>
          <standardUnit>meter</standardUnit>
        </unit>
        <precision>1</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>1</precision>
        <numericDomain>

```

```

        <numberType>whole</numberType>
      </numericDomain>
    </ratio>
  </measurementScale>
</attribute>
</attributeList>
<spatialReference>
  <horizCoordSysName>NAD_1927_UTM_Zone_12N</horizCoordSysName>
</spatialReference>
<horizontalAccuracy>
  <accuracyReport>not available</accuracyReport>
</horizontalAccuracy>
<verticalAccuracy>
  <accuracyReport>not available</accuracyReport>
</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. Constraints in which several columns are involved should be described as in the example below.

**<constraintType>** describes the relational integrity constraints of a relational database. This includes primary keys, foreign keys, unique keys, etc. When the **<constraint>** tree is used, all of the entities that may be referenced in the constraints (foreignKey) should be accessible within that same package. One of the 6 constraint types must be used: **primaryKey**, **uniqueKey**, **checkConstraint**, **foreignKey**, **joinCondition**, or **nonNullConstraint** as follows:

**<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.

*Note: the <primaryKey>, <uniqueKey>, <nonNullConstraint> require an additional <key> tag defining the set of attributes to which this constraint applies.*

**<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.

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

*Note: All <ConstraintType> entities require additional <constraintName> and <attributeReference> tags.*

---

Example: constraint (simple and compound primary key, foreign key)

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

<constraint id="invertebrate_pk">
  <primaryKey>
    <constraintName>invertebrate_pk</constraintName>
    <key>
      <attributeReference>att.1</attributeReference>
      <attributeReference>att.2</attributeReference>
      <attributeReference>att.3</attributeReference>
      <attributeReference>att.4</attributeReference>
    </key>
  </primaryKey>
</constraint>
```

```

    </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. An attribute unit which is a **<customUnit>** must be described in this tree (see below).

**<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 content 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.

The **<metadata>** element is a container for the additional metadata. 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, **<metadata>** must contain sufficient information to define the association between **<additionalMetadata>** and the attribute being described.

If the content is for definition of a **<customUnit>**, then it should be contained in valid STMML markup, starting with **<unitList>**. The **<unitList>** is the one exception to the use of a **<describes>** tag for associating **<additionalMetadata>** with specific EML elements. 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>** "siemenPerMeter". Content in STMML for the **<unit>** describing a **<customUnit>** should be retrieved in STMML format from the Unit Registry (<http://unit.lternet.edu/unitregistry/>).

---

Example: additionalMetadata custom unit

```

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

```

---

### III. RECOMMENDATIONS FOR COMPATIBILITY WITH EXTERNAL APPLICATIONS

#### III.1 METACAT

The KNB Metacat Server and related Morpho application, are generally compatible with all XML documents conforming to the EML schemas. However, the following recommendations will enhance management and display of EML documents using these tools and enable documents to be uploaded automatically using the LNO Metacat harvester application.

##### a) Document ids and revision numbers

packageId attributes for EML contributed to the KNB Metacat should be formed as follows:

knb-lter-[site].[dataset number].[revision], e.g. knb-lter-fls.187.4

Metacat (and by extension the Metacat harvester) rely on numerical data set ids and revision numbers for document management and synchronization. This may necessitate a workaround for sites that use non-numeric ids or do not version data sets. Possible solutions include differentially generating EML optimized for Metacat and XSLT transformation. Currently, the **packageId** attribute is used to identify a site's EML documents for searches in the LTER Data Catalog. A site's **packageId** attributes for EML contributed to the KNB Metacat should be standardized as above, or another format agreed upon by the site and LNO Data Catalog managers.

b) Metacat fully supports access control specified in EML documents. See section II.2 for more information about access control of EML documents

c) LTER's Metacat is configured to index certain paths for fastest searching. EML authors should be sure to include adequate content in these nodes:

title  
keyword  
abstract  
abstract/path  
abstract/section/path  
organizationName, surName or givenName (under associatedParty or creator)  
url  
taxonRankValue  
boundingCoordinates (under geographicCoverage)  
geographicDescription  
entityName

### III.2 LTERMAPS

**LTERMapS Application:** LTERMapS is an Internet mapping application for LTER site information to allow users to visualize, search, download, and explore site information and with spatial or tabular data at both the network and site level.

(<http://www.lternet.edu/map/>).

To interface EML datasets with the LTERMapS application, a dataset-level **<geographicCoverage>** tree must be included.

In addition, detailed sampling locations can also be listed in two ways, which are both acceptable for participation in the LTER MapS project.

**Content in <spatialSamplingUnits>:** Including station descriptions as content under the coverage elements of the **<spatialSamplingUnits>** may be the simplest method for a site that generates its EML 'on the fly'.

**Referenced data entity:** Sampling stations can be listed in a separate data entity, which should be referenced by methods/\*/spatialSamplingUnits/referencedEntityId. This may be the best option for a long list of stations, since these will create a very long display if listed by individual **<geographicCoverage>** elements. These data entities should be in one of the following formats: ESRI shape file, KML/KMZ (using **<otherEntity>**), or comma-delimited text (using **<dataTable>**). It should be kept in mind that if stations are listed in a comma-delimited table, it is not possible for a metadata search engine to find these locations. Instructions on how to create shapefiles, KML or KMZ files for study site locations are available at: [http://im.lternet.edu/im\\_practices/gis/locations](http://im.lternet.edu/im_practices/gis/locations)

### III.3 PROVENANCE TRACKING IN EML

Wiktionary defines "provenance" as the "place or source of origin" of an object (June 2010, <http://en.wiktionary.org/wiki/provenance>). EML has the capacity to manage provenance of a derived object (e.g., datasets) assuming that the source object(s) are also described by EML and can be referenced via a network protocol (e.g., through a metadata catalog such as Metacat).

The EML document for the derived object consists of three significant components:

- 1) the metadata of the derived object,
- 2) a description of any (if applicable) application process (e.g., unit conversion) applied to the source object to create the derived object, and
- 3) the metadata of the source object(s).

The metadata of the derived object should follow the same "best practices" provided here. The second two components (application processing and source metadata), are specifically related to the provenance of the derived object. The EML **<methods>** sub-tree is an ideal container to capture provenance information since its purpose is to document the descriptive information about the derivation. Each derivation step will be contained in a **<methodStep>** element.

The location of the **<methods>** tree in the new EML document will depend on where the derived object is being described - for example, the **<methods>** tree at the **<dataset>** level could be used to describe the provenance information for the entire derived object.

---

Example: provenance

```
<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>
```

```
</methodStep>  
</methods>
```

---

### III.4 EML CONGRUENCE CHECKER

*Note: this section anticipates features of the EML Congruence Checker currently under development at the LNO. This section will be further developed.*

The EML Congruence Checker (ECC) is designed to improve the quality of EML files and associated data files and to ensure that these files are ready for submission to the LTER Network Information System and ingestion by PASTA. The following tests are performed:

EML critical features: The EML file is examined to see if the file is

- (1) well-formed
- (2) complies with the XML schema
- (3) complies with EML rules for IDs and references

These functions are currently provided by the online EML ID and References Parser (<http://knb.ecoinformatics.org/emlparser/>). If the EML file fails any of these tests the ECC terminates with an error message indicating the cause of the failure.

EML recommended features: The EML file is examined to measure conformance with selected best practices from this document. The ECC returns detailed information for the following tests:

- (1) Completeness. The ECC checks to see that the recommended elements are present in the EML file and returns a list of recommended elements that are absent in the EML file.
- (2) URLs. The ECC checks all URLs in the EML file to see if they are active, and returns a list of URLs that are broken.
- (3) Keywords. The ECC will compare the keywords in the EML file to the current LTER Controlled Vocabulary list and returns a list indicating which keywords are present and which are absent in the Controlled Vocabulary list.
- (4) Units. The ECC compares units in the EML file to the current LTER Unit Dictionary and returns a list indicating which units are standard units, which are custom units, and which are not present in the Unit Dictionary.

EML and data congruency: The EML file and data file(s) are examined to see if the metadata description in the EML file is sufficiently accurate and complete to support ingestion of the data file into PASTA. For data tables, the ECC returns detailed information for the following tests:

(1) The ECC verifies that the information in the **<physical><distribution>** tree is sufficient to retrieve a data file.

(2) The ECC verifies that the information in the **<physical><dataFormat>** tree correctly describes the structure of the data file, including file size, compression and encoding method, number of header and footer lines, record delimiter, and field delimiter. The number of columns should match the number of attributes in the **<attributeList>** tree.

(3) For each attribute, the ECC verifies that the information in the **<measurementScale>** tree correctly describes the values for that attribute contained in the data file. For example, all **<enumeratedDomain>** values should be members of the specified set of coded values. **<numericDomain>** values should conform to the specified **<numberType>**. If the **<bounds>** element is present, all **<numericDomain>** values should also lie within the range given by **<minimum>** and **<maximum>**. **<dateTime>** values should match the specified **<formatString>**. Missing value codes (if present) should conform to the specified **<missingValueCode><code>**.

Similar tests are performed for other data types (spatialRaster, spatialVector, otherEntity).

## IV. DESCRIPTIONS OF EML SAMPLE FILES PROVIDED WITH THIS DOCUMENT

Example 1: Complete EML from which the examples in this document were derived

Example 2: EML for the provenance example

## V. OTHER RESOURCES

Documents are ordered by year. Applications are alphabetical by title.

### Documents:

EML 2.1.0 schema and normative documentation (2008):

<http://knb.ecoinformatics.org/software/eml/>

EML-2.1.0 FAQ (2008): <http://knb.ecoinformatics.org/software/eml/eml-2.1.0/eml-faq.html>

ESRI to EML conversion stylesheet documentation and recommendations (yyyy):  
<http://im.lternet.edu/project/Esri2Eml>

EML Best Practices for LTER Sites, Version 1 (2004):  
[http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices\\_oct2004\\_final.pdf](http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices_oct2004_final.pdf)

EML Handbook (yyyy): <http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf>

Report from the 2003 EML implementation workshop at SEV (2003):

Applications and Websites:

LNO Data Access Server: [http://im.lternet.edu/im\\_practices/metadata/das](http://im.lternet.edu/im_practices/metadata/das)

LNO Metacat Harvester info: <http://im.lternet.edu/node/418>

LTER Project Database: <http://im.lternet.edu/project/LTERProjectDatabase>

LTER Unit dictionary: <http://unit.lternet.edu>

Metacat and Morpho: <http://ecoinformatics.org/tools.html>

Observational Data Model (ODM) used by CUASHI hydrological information system at:  
<http://his.cuahsi.org/index.html>.

## VI. CONTRIBUTORS TO THIS DOCUMENT

**EML Best Practices Workshop, LTER Network Office, Albuquerque, NM (June 21-23, 2010)**

Best Practices Working Group:

- Dan Bahauddin (CDR)
- Emery Boose (HFR)
- Corinna Gries (NTL)
- Margaret O'Brien (SBC)
- Ken Ramsey (JRN)
- Mark Servilla (LNO)
- Theresa Valentine (AND)
- Jonathan Walsh (BES)
- M. Gastil Gastil-Buhl

Co-participants:

- James Brunt (LNO)
- Inigo San Gil (LNO)

- Philip Tarrant (CAP)
- Yang Xia (LNO)

**EML Best Practices Workshop, LTER Network Office, Albuquerque, NM (May 19-20, 2004)**

Best Practices Working Group:

- James Brunt (LNO)
- Corinna Gries (CAP)
- Jeanine McGann (LNO)
- Margaret O'Brien (SBC)
- Ken Ramsey (JRN)
- Wade Sheldon (GCE)

Co-participants:

- Duane Costa (LNO)
- Mark Servilla (LNO)

**LTER EML Implementation Workshop, Sevilleta Research Station, Sevilleta National Wildlife Refuge, NM (June 9-10, 2003)**

Best Practices Working Group:

- Barbara Benson (NTL)
- James Brunt (LNO)
- Don Henshaw (AND)
- John Vande Castle (LNO)
- Kristin Vanderbilt (SEV)

Working Group Support:

- Jeanine McGann (LNO)

## INDEX

abstract..... 17, 18, 31, 50  
access.....5, 11, 19, 20, 21, 34, 49  
additionalMetadata.....38, 40, 48

address .....14, 15, 16, 17  
alternateIdentifier ..... 13, 33  
alternativeTimeScale.....24

associatedParty .....	15, 16, 50	interval .....	37, 38
attribute.3, 6, 10, 13, 21, 22, 28, 29, 34, 37, 38, 39, 40, 43, 44, 47, 48, 49, 53		joinCondition .....	46, 47
attributeDefinition .....	37, 38	keyword .....	18, 50
attributeLabel .....	37	keywordSet.....	18
attributeList.....	21, 22, 28, 37, 40, 53	maintenance.....	25
attributeName.....	37	maintenance.....	8, 27
attributeReference .....	47	measurementScale.....	37, 53
boundingCoordinates .....	22, 23, 30	mediumName.....	20
bounds.....	39, 53	metadataProvider .....	15
checkConstraint .....	46, 47	methods....	5, 6, 8, 17, 21, 22, 25, 26, 28, 29, 32, 33, 35, 37, 50, 51
citation .....	5, 28, 29, 31	methodStep.....	28, 29, 39, 51
classificationSystem .....	25	missingValueCode.....	39, 53
classificationSystemCitation .....	26	nominal.....	37
constraint.....	6, 46, 47	nonNullConstraint .....	46, 47
constraintName.....	47	nonNumericDomain.....	37
constraintType.....	46	numberType.....	38, 53
ConstraintType.....	47	numericDomain .....	38, 39, 53
contact.....	14, 16, 17, 19	objectName .....	34, 35
coverage ....	8, 21, 22, 24, 27, 29, 30, 31, 35, 50	offline .....	20
creator .....	10, 14, 15, 17, 28, 50	online.....	17, 20, 21, 34, 35, 52
customUnit.....	38, 39, 48	onlineURL.....	15, 16, 17
dataset5, 8, 9, 10, 11, 13, 14, 17, 18, 19, 20, 21, 22, 24, 27, 28, 29, 31, 32, 36, 37, 49, 51		ordinal.....	37, 38
datasetGPolygon .....	24	organizationName .....	14, 15, 16, 17, 50
datasetGPolygonExclusionGRing.....	24	packageId.....	9, 11, 13, 49
datasetGPolygonOuterGRing.....	24	personnel.....	10, 16, 31
dataTable .....	10, 21, 28, 32, 34, 35	phone.....	14, 15, 16, 17
datetime.....	37, 39, 53	postitionName.....	14
descriptor.....	31	precision .....	6, 38, 39
distribution .....	11, 20, 21, 28, 31, 34, 35, 53	primaryKey.....	46, 47
electronicMailAddress.....	14, 15, 16, 17	project.....	8, 10, 17, 18, 22, 31, 50
endDate.....	24, 25	protocol.....	5, 10, 28, 29, 51
entityDescription.....	33, 35	pubDate.....	17, 18
entityName.....	33, 34, 35, 50	publisher.....	17
enumeratedDomain.....	37, 53	qualityControl.....	28
externallyDefinedFormat.....	34	role .....	14, 16
fundingSource.....	31	sampling.....	5, 6, 21, 22, 25, 28, 29, 35
generalTaxonomicCoverage.....	25	samplingDescription.....	29, 30
geographicCoverage.....	8, 22, 50	scope.....	11, 13, 14, 16, 25, 28
givenName.....	14, 15, 16, 50	spatialRaster .....	32, 53
identifierName.....	26	spatialSamplingUnits... 6, 21, 22, 29, 30, 31, 50	
individualName.....	13, 15, 16	spatialVector .....	32, 53
inline .....	20	standardUnit.....	38, 39
instrumentation .....	28, 29, 30	storageType.....	37
intellectualRights.....	19	storedProcedure.....	32
		studyAreaDiscription .....	31

studyExtent.....	21, 29, 30, 31	taxonRankName .....	25, 26
surName .....	14, 15, 16, 50	temporalCoverage.....	8, 22, 24, 25
system 10, 11, 13, 14, 16, 37, 39, 40, 46, 48, 54		textDomain .....	38
taxonomicClassification.....	25, 26	title .....	13, 19, 28, 29, 31, 50
taxonomicCompleteness.....	26	uniqueKey .....	46, 47
taxonomicCoverage .....	8, 25, 26	unit .....	5, 38, 39, 48, 51
taxonomicProcedures.....	26	view.....	16, 32