Published on *LTER Information Management* (http://im.lternet.edu)

EML Best Practices for LTER Sites

N.B. This section is obsolescent. Info about EML and it's use is now maintained by the Environmental Data Initiative. See https://environmentaldatainitiative.org/resources [1]
See: "Five phases of data publishing" > "Phase 3"

EML Best Practices, V3
https://environmentaldatainitiative.org/resources/five-phases-of-data-pu... [2]
Previous version (1 and 2, 2004 and 2011 respectively) are below.

Earlier versions of the document are available from the network document archive.

EML Handbook, 2003
EML: Practical Application for Scientists [3] by D. Blankman and J. McGann, LNO. This describes EML use at a basic level. Note some important details such as <objectName> and <attributeLabel> are omitted. However, this is a good place to start an introduction to EML. Note that even back in 2003 it was advised that the data url be at the entity level.

Attachment	Size
2011: EML Best Practices for LTER Sites V 2 (PDF) [4]	880.28 KB
2011: example_full_a.xml [5]	33.94 KB
2011: example_provenance.xml [6]	6.65 KB
2011: example_lter_maps.xml [7]	61.52 KB

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 re-use 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/enitityCodeList/ 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.

I.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.)

I.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 [8] 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 [9]

bdp2eml.xsl: converts FGDC and NBII Biological Profile metadata to EML. http://im.lternet.edu/project/bdp2eml [10]
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

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.

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" [11]

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-1.1"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd" [11]
  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>
```

datasets

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).

<**alternateIdentifer**>
<**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**>)

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&amp;query=itestuser</onlineUrl>
  <role>Technician</role>
  </associatedParty>
```

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>
```

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 [12])

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>
      <listitem>
        <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: [13] http://im.lternet.edu/im_practices/metadata/das [14]

Example: distribution

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

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


Source URL: http://im.lternet.edu/node/910

Links:
[1] https://environmentaldatainitiative.org/resources
[2] https://environmentaldatainitiative.org/resources/five-phases-of-data-publishing/phase-3/metadata-best-practices/
[3] http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf
[4] http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices-2.0-FINAL-20110801_0.pdf
[5] http://im.lternet.edu/sites/im.lternet.edu/files/example_full_a.xml
[6] http://im.lternet.edu/sites/im.lternet.edu/files/example_provenance.xml
[7] http://im.lternet.edu/sites/im.lternet.edu/files/example_lter_maps.xml
[8] http://www.fgdc.gov/metadata/geospatial-metadata-tools
[9] http://intranet.lternet.edu/im/project/Esri2Eml/docs
[10] http://intranet.lternet.edu/im/project/bdp2eml
[11] http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
[12] http://www.lternet.edu/data/netpolicy.html
[13] http://www.google.com/url?q=http%3A%2F%2Fintranet.lternet.edu%2Fim%2Fim_practices%2Fmetadata%2Fdas&sa=D&sntz=1&usg=AFQjCNF5qpXdlir91CpUJAK7phLSymWTrQ
[14] http://im.lternet.edu/im_practices/metadata/das