EML Best Practices for LTER Sites

Table of Contents

I. Introduction

Background

This document contains a discussion of current views on best practices for EML metadata implementation by LTER sites, as outlined by two working groups comprised of LTER information managers and LNO representatives (see section V). These recommendations are directed towards achieving the following goals:

a) Identify useful subsets of the EML schema to support specific functionality tiers targeted by the LTER NIS Advisory Committee (NISAC)
b) Maximize interoperability of LTER EML documents to facilitate data synthesis
c) Minimize heterogeneity of LTER EML documents to simplify development and re-use of software tools and style sheets
d) Provide guidance to sites in their initial implementation of EML, and a roadmap for improving their implementation to achieve higher functionality

This document is also intended to augment the EML schema documentation and other resources listed below:

EML Handbook: http://intranet.lternet.edu/eml/EMLHandbook.pdf
EML FAQ: http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-faq.html
Report from the 2003 EML implementation workshop at SEV: http://intranet.lternet.edu/eml/emlimplementation.htm
EML 2.0.1 schema and documentation: http://knb.ecoinformatics.org/software/eml/

Overview

The following table summarizes the major levels of EML content "completeness", or tiers, identified by the two EML working groups. Each level adds more elements from the EML schema to provide a more comprehensive description of the data resources documented by the metadata, and thereby support higher functionality.

Completeness Level	Description and Major Elements Added
1: Identification	Minimum content for adequate data set discovery in a general cataloging system or repository (functionally equivalent to LTER DTOC): • title • creator • contact • publisher • pubDate • keywords • abstract (recommended) • dataset/distribution (i.e. url for general dataset information)
2: Discovery	Level 1 content, plus coverage information to support targeted searches, adding elements: • geographicCoverage • taxonomicCoverage • temporalCoverage
3: Evaluation	Level 2 content, plus data set details to enable end-user evaluation of the methodology and data entities, adding elements: • Intellectual Rights • project • methods • dataTable/entityGroup • dataTable/attributes (see issues outlined in the text)
4: Access	Level 3 content plus data access details to support automated data retrieval, adding elements: • access • physical
5: Integration	Level 4 content plus complete attribute and quality control details to support computer-assisted data integration and re-sampling, adding elements: • attributeList (full descriptions) • constraint • qualityControl
6: Semantic Use	Level 5 content plus semantic information (currently under development by SEEK, and may require extension to the EML schema)

II. EML Content Recommendations by Level

General Recommendations

The following are general best practices for creating EML metadata documents:

- Do not publicly distribute EML documents containing elements with incorrect information (i.e. included as a workaround for problems with metadata content availability or EML validation) as data set metadata. EML produced for demonstration or testing purposes should be clearly identified as such and not contributed to metadata archives or clearinghouses.
- For text type elements, use EML text formatting tags whenever possible (e.g. <section>, <para>, <orderedlist>). Only use <literalLayout> if HTML needs to be pasted into this field.
- Metadata and data set versioning are only relevant in the context of an archival- or repository-type information system. If a site does not have a local archival system that supports versioning (e.g. distributes data from ongoing collections via an RDBMS system), then versioning should only be applied to the metadata when EML is deposited in an external repository system such as Metacat (see Metacat interoperability notes below).
- Care should be exercised when using id attributes to reference and re-use EML content, because all ids in an EML document must be unique otherwise validation errors will occur. It may be preferable to duplicate content rather than use ids and references when generating EML dynamically from a relational database system to avoid potential id conflicts.

Level 1 – Identification:

Identification level EML is suitable for basic registration of datasets in a general cataloging system, similar to the current LTER Network Data Table of Contents (http://www.lternet.edu/DTOC). Identification level parameters include an alternative identifier (dataset ID used by site), dataset title, creator (researchers), metadata provider, other associated persons and organizations, date of public release, abstract, keywords, data distribution URL (if unrestricted dataset), contact (Information Manager), and publisher (LTER site). Listed below are the corresponding EML elements that should be completed to create a Level 1 EML document:

<**alternateIdentifier**> The site's data set id should be listed as the EML <alternateIdentifier> (see Example 1.1), particularly when it differs from the "packageId" attribute in the <eml:eml> element required by a given cataloging system.

<title> The dataset <title> (see Example 1.1) should be descriptive and should describe the data collected, geographic context, research site, and time frame (what, where, and when).

<creator> Full contact information for at least 1 <creator> (researcher) should be provided. It is important that the format be consistent among all site EML documents and that the contact information should be kept current as much as possible. When using <individualName> elements anywhere within an EML document, name suffixes should be included in the <surName> element after the last name (see Example 1.1). Complete the <address>, <phone>, <electronicMailAddress>, and <onlineURL> elements for each creator element.

Example 1.1. eml, dataset, creator tree:

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.0.0"
   packageId="1058377556406" system="FLS" scope="system"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.0
   http://someserver.fls.edu/eml.xsd">
   <dataset id="FLS-1" system="FLS">
      <alternateIdentifier>FLS-1</alternateIdentifier>
      <shortName>Arthropods</shortName>
      <title>Long-term Ground Arthropod Monitoring Dataset at Ficity,
         USA from 1998 to 2003</title>
      <creator id="pers-1" system="FLS">
         <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> The <metadataProvider> element lists the person or organization responsible for producing 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 associated party are not the same people who produced the metadata content, the actual metadata content provider should be listed instead (see Example 1.2).

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

<associatedParty> List people who were involved with the data in some way (field technicians, students assistants, etc.). The <address>, <phone>, <electronicMailAddress>, and <onlineURL> elements for each <associatedParty> element are optional, and if provided should be kept current. The parent University, institution, or agency could also be listed using the <owner> role when appropriate.

Example 1.2. metadata provider, associatedParty

```
...
<metadataProvider>
   <organizationName>Fictitious State University</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@fi.univ.edu</electronicMailAddress>
   <onlineUrl>http://www.fsu.edu/</onlineUrl>
</metadataProvider>
<associatedParty id="12010" system="FLS">
   <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>fsu@fi.univ.edu</electronicMailAddress>
   <onlineUrl>http://www.fsu.edu/~Imatestuser.htm</onlineUrl>
   <role>Technician</role>
</associatedParty>
...
```

<pubDate> The year of public release of data online should be listed as the <pubDate> element (see Example 1.3).

<**abstract>** The <abstract> element (see Example 1.3) will be useful for full-text searches, and it should be rich with descriptive text. The measured parameters should also be included. Extensive description should include what, when, and where information as well as whether the dataset is ongoing or completed, some taxonomic information, and some methods description (what, where, when, and why plus parameters). If there are too many parameters for a dataset, use categories of parameters instead of listing all parameters (ex. – use nutrients instead of nitrate, phosphate, calcium, etc.) in combination with the parameters that seem most relevant for searches.

<**keywordSet>** The <keywordSet> element (see Example 1.3) keyword listings should include the three letter site acronym, core research area(s), 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.). Multiple sets of key words can be included as illustrated in the example. In addition to specific keywords, relevant conceptual keywords should also be included. See the KNB keyword listing below (from http://knb.ecoinformatics.org/index.jsp) for some recommended keywords:

Taxonomy
Amphibian, Bird, Fish, Fungus, Invertebrate, Mammal, Microbe, Plant, Reptile, Virus
Measurements
Biomass, Carbon, Chlorophyll, GIS, Nitrate, Nutrients, Precipitation, Temperature, Radiation, Weather
Level of Organization
Molecule, Cell, Organism, Population, Community, Landscape, Ecosystem, Global
Evolution
Adaptation, Evolution, Extinction, Genetics, Mutation, Selection, Speciation, Survival
Ecology
Biodiversity, Competition, Decomposition, Disturbance, Endangered Species, Herbivory, Invasive Species, Nutrient Cycling, Parasitism, Population Dynamics, Predation, Productivity, Succession, Symbiosis, Trophic Dynamics
Habitat
Alpine, Freshwater, Benthic, Desert, Estuary, Forest, Grassland, Marine, Montane, Terrestrial, Tundra, Urban, Wetland

(Action item: An LTER Network-wide keyword thesaurus needs to be developed for site use)

Example 1.3. pubDate, abstract, keywords

```
...
<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>
</keywordSet>
<keywordSet>
   <keyword keywordType="theme">FLS</keyword>
   <keyword keywordType="theme">Fictitious LTER Site</keyword>
   <keyword keywordType="theme">LTER</keyword>
   <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">Richness</keyword>
   <keyword keywordType="theme">Abundance</keyword>
</keywordSet>
...
```

<**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 includes the <online>, <offline>, and <inline> elements. As a minimum (for level 1) the <online> element's <url> tag should be included at the dataset level and should point to a local data distribution application (example 1.4.). A URL listed at the table (entity) level, however, should stream data to the requesting application. In other words, if a distribution URL is provided at the entity level, the URL should lead directly to the data and NOT a data catalog or intended use page. For more information about describing a connection, see Example 1.8 and the online documentation. In most cases the <url> tag should be used. The <offline> element is used 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. The <inline> element contains data that is stored directly within the EML document.

Recommendation: data table access logging should be implemented by the cataloging system, e.g. Metacat, and relayed to the data provider when data is accessed directly via EML hosted in the system

---

Example 1.4. dataset distribution

```
...
<distribution>
   <online>
      <url>http://someserver.fls.edu/data/fls-1.htm</url>
   </online>
</distribution>
...
```

<contact> Full contact should be included for the Position of data manager (see Example 1.5) and should be kept current independently of personnel changes. If several contacts are listed (e.g. general site contact) all should be kept current. Technicians who performed the work should be listed as an <associatedParty> rather than contact. Complete the <address>, <phone>, <electronicMailAddress>, and <onlineURL> elements for the contact element (see Example 1.5.).

<publisher> The LTER site should be listed as the <publisher> (see Example 1.5) 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: Metacat should use <publisher> as the organization information for web display

---

Example 1.5. contact, publisher

```
...
<contact id="im">
   <positionName>Data 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">(111) 222-3333</phone>
      <phone phonetype="fax">(111) 222-3334</phone>
      <electronicMailAddress>fls.data@fi.univ.edu</electronicMailAddress>
      <onlineUrl>http://www.fsu.edu/lter</onlineUrl>
   </contact>
   <publisher>
      <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@fi.univ.edu</electronicMailAddress>
      <onlineUrl>http://www.fsu.edu/</onlineUrl>
   </publisher>
   ...
```

Level 2 – Discovery:

Discovery level metadata should provide as much information as possible to support locating datasets by time, taxa, and/or geographic location in addition to basic identification information. Discovery level EML should include the <coverage> elements of <temporalCoverage> (when), (taxonomicCoverage> (what), and <geographicCoverage> (where) for the dataset as well as the change history in the <maintenance> element.

<**coverage**> The <coverage> element appears at the dataset, methods, entity and attribute levels and contains 3 elements for describing the coverage of a dataset in terms of space, taxonomy, and time (see Example 2.1). The <geographicCoverage>, <taxanomicCoverage>, and <temporalCoverage> elements need to be populated at the discovery level of EML in order to allow for more advanced searches based on taxa, time, and geographic location than can be provided by the Identification level.

<**geographicCoverage**> The <geographicCoverage> element (see Example 2.1) is used to describe geographic locations of research sites and areas related to the dataset being documented. The method for determining <boundingCoordinates>, <boundingAltitudes>, coordinate datum, etc. can be included under <geographicDescription> since it is a simple text field. The description should be a comprehensive description of the location including country, county or province, city, state, general topography, landmarks, rivers, and other relevant information. The <boundingCoordinates> element should describe a bounding box surrounding the entire LTER site or full extent of observations (one

point for each extension to the east, west, north, south), with the latitude and longitude expressed to four decimal degrees in international convention (+/-). The <datasetGPolygon> element should be included when the bounding box does not adequately describe the study location (e.g. if there is an area within the bounding box that is excluded, or an irregular polygon is necessary to describe the study area). (Note: there is a possible error in the EML schema, and the gRing and gRingPolygon elements may need to be re-evaluated or their usage clarified.). If specific study site locations need to be listed within the bounding box the coverage element in <methods>/<sampling>/ <spatialSamplingUnits> should be used (see level 3). The <boundingAltitudes> element should be described in meters with a datum described in the <altitudeUnits> element.

Note: *geographicCoverage usage is currently under review by the LTER GIS committee*

<**temporalCoverage**> The <temporalCoverage> element (see Example 2.1) of a dataset represents the period of time the data was collected. Generally, the temporal coverage should be described as a <singleDate> or <rangeOfDates> element representing when the data were collected (not the year the study was put together if it uses retrospective or historical data). Sometimes an <alternativeTimeScale> is more appropriate, such as the use of "years before present" for something like long-term tree ring chronology dating back hundreds of years. The date format should be listed as described in the EML documentation (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. Use the <maintanence> tag (below) to describe the update frequency. The methods/sampling tree (described at Level 3) should be used to describe the ongoing nature of the data collection.

<**taxonomicCoverage**> The <taxonomicCoverage> element (see Example 2.1) should be used to document taxonomic information for all organisms relevant to the study. Genus, species name 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 as illustrated in Example 2.1 rather than repeated at the same level. The <generalTaxonomicCoverage> element should be included to describe the general procedure of how the taxonomy was determined (keys used, etc.), should include a general textual description of all flora/fauna in the study (scope), as well as how finely grained the taxonomy is broken down to – for example "family" or "genus and species."

Note that elements within common <taxonRankName> entries can be combined in the hierarchy 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.

<**maintenance>** The dataset/maintenance/description element should be used to document changes to the data tables or metadata, including update frequency (see Example 2.2). 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 2.1 coverage

```
...
<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>
   <temporalCoverage>
      <rangeOfDates>
         <beginDate>
            <calendarDate>1998-11-12</calendarDate>
         </beginDate>
         <endDate>
            <calendarDate>2003-12-31</calendarDate>
         </endDate>
      </rangeOfDates>
   </temporalCoverage>
   <taxonomicCoverage>
      <generalTaxonomicCoverage> Orthopteran insects (grasshoppers)
         were identified to species using the 2004 Great Big Key to
         Orthoptera (Great Big Key press, NY)</generalTaxonomicCoverage>
      <taxonomicClassification>
         <taxonRankName>Kingdom</taxonRankName>
         <taxonRankValue>Animalia</taxonRankValue>
         <taxonomicClassification>
            <taxonRankName>Phylum</taxonRankName>
            <taxonRankValue>Arthropoda</taxonRankValue>
            <taxonomicClassification>
               <taxonRankName>Class</taxonRankName>
               <taxonRankValue>Insecta</taxonRankValue>
               <taxonomicClassification>
                  <taxonRankName>Order</taxonRankName>
                  <taxonRankValue>Orthoptera</taxonRankValue>
                  <taxonomicClassification>
```

```
                        <taxonRankName>Family</taxonRankName>
                        <taxonRankValue>Acrididae</taxonRankValue>
                        <taxonomicClassification>
                           <taxonRankName>Genus</taxonRankName>
                           <taxonRankValue>Mermiria</taxonRankValue>
                           <taxonomicClassification>
                              <taxonRankName>Species</taxonRankName>
                              <taxonRankValue>Mermiria intertexta</taxonRankValue>
                           </taxonomicClassification>
                        </taxonomicClassification>
                     </taxonomicClassification>
                  </taxonomicClassification>
               </taxonomicClassification>
            </taxonomicClassification>
            <taxonomicClassification>
               <taxonRankName>Kingdom</taxonRankName>
               <taxonRankValue>Animalia</taxonRankValue>
               <taxonomicClassification>
                  <taxonRankName>Phylum</taxonRankName>
                  <taxonRankValue>Arthropoda</taxonRankValue>
                  <taxonomicClassification>
                     <taxonRankName>Class</taxonRankName>
                     <taxonRankValue>Insecta</taxonRankValue>
                     <taxonomicClassification>
                        <taxonRankName>Order</taxonRankName>
                        <taxonRankValue>Orthoptera</taxonRankValue>
                        <taxonomicClassification>
                           <taxonRankName>Family</taxonRankName>
                           <taxonRankValue>Tettigoniidae</taxonRankValue>
                           <taxonomicClassification>
                              <taxonRankName>Genus</taxonRankName>
                              <taxonRankValue>Orchelimum</taxonRankValue>
                              <taxonomicClassification>
                                 <taxonRankName>Species</taxonRankName>
                                 <taxonRankValue>Orchelimum fidicinum</taxonRankValue>
                                 <commonName>Salt Marsh Grasshopper</commonName>
                              </taxonomicClassification>
                           </taxonomicClassification>
                        </taxonomicClassification>
                     </taxonomicClassification>
                  </taxonomicClassification>
               </taxonomicClassification>
            </taxonomicClassification>
         </taxonomicCoverage>
      </coverage>
      ...
```

Example 2.2 maintenance

```
...
<maintenance>
   <description>
      <para>new data added monthly</para>
   </description>
</maintenance>
...
```

Level 3 – Evaluation:

Evaluation level metadata should include detailed descriptions of the project, methods, protocols, and intellectual rights in order for a potential user to evaluate the relevance of the data package for their research study or synthesis project. Ideally, evaluation-level metadata should also include at least basic descriptions of all data tables or other entities in the data set; however, the current version of EML (v2.0.1) requires that when the dataTable element is used, both entityGroup/entityName and the entire <attributeList> tree are also included.

The committee would prefer to recommend that only the dataTable/entityGroup/entityName be required at Level-3, but since both elements are required by EML2.0.1, this recommendation conflicts with the schema. If the EML requirement is changed in a future release (i.e., <attributes> becomes optional in the entityGroup tree), then basic descriptions of the data tables (the dataTable/entityGroup) will be required to meet Level 3, and inclusion of the attributeList will be moved to Level 4 or 5. Until that time, the committee chose to recommend the entire dataTable/attributeList/attribute tree be included at Level 3 in an effort to convey this critical entity information in metadata intended for evaluation purposes.

(Potential recommendation: the EML schema should be modified to make attributeList optional in dataTable for consistency with entity and to allow data tables to be named and described when attribute information is incomplete. Alternatively, measurementScale could be made optional in attribute to allow attribute names and definitions to be listed without full details.)

Level 3 therefore adds several major sections of metadata content, described in the text and examples below:

<intellectualRights> should contain site data access policy, plus a description of any deviation from the general access policy specific for this particular dataset (eg 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.

Example 3.1. intellectualRights

```
...
<intellectualRights>
   <section>
      <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>
         </itemizedlist>
      </para>
   </section>
</intellectualRights>
...
```

<**methods**> The <method(s)> tree appears at the dataset, entity, and attribute level. As a 'rule of thumb', methods are *descriptive*, and protocols are *prescriptive*. The minimum requirement for Level 3 is that a reference to an external protocol be given at the dataset level. If further refinement is needed, methods can be defined for individual tables or their attributes if necessary. The scope of the method defined should match the eml schema level at which it is applied. For example, methods at the dataset level describe the study, at the dataTable level methods might include pre-/post-processing steps, and at the attribute level, quality control. A description of methods contains the following elements:

<**methodStep**> each step is a logical portion of the methods (e.g. field, lab, statistical) <**proceduralStep**>, <**description**> and <**para**> tags are used for text descriptions. At a minimum,. Two tags can be used to include external documents: <**citation**> for referral to published procedures, or <**protocol**> to include a protocol. Note that at a minimum, the <**title**>, <**creator**> and <**distribution**> tags are required for a protocol, then the <**distribution**>/<**online**>/<**url**> may be used to refer to an online document or the entire protocol may be included under <**proceduralStep**>.

<methodStep><instrumentation> This 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 <**methods><description**>.

Example 3.2. methodStep/protocol and methodStep/instrumentation

```
<methods>
...
   <methodStep>
      <description>
         <para>FSL Protocol for Surveying Ground Arthropods has
            been used</para>
      </description>
      <protocol>
         <title>FSL Protocol for Surveying Ground Arthropods in school
            yards</title>
         <creator>
            <references>pers-1</references>
         </creator>
         <pubDate>2000</pubDate>
         <abstract>
            <para>This protocol is being used by FLS arthropod monitoring
               program. Ground arthropods are collected in pit fall traps buried
               into the soil and left for 72 hours. Sampling takes place four
               times a year in three traps per location. The trapped arthropods
               are determined and data entered into a database.</para>
         </abstract>
         <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">Richness</keyword>
            <keyword keywordType="theme">Abundance</keyword>
         </keywordSet>
         <distribution>
            <online>
               <url>http://fls.fi.univ.edu/protocol/arthropods/arthro.htm</url>
            </online>
         </distribution>
```

```
        </protocol>
      </methodStep>
      <methodStep>
        <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>
    ...
    </methods>
```

<methods><sampling> This optional tree can contain valuable and very specific information about the study site, and coverage in addition to that listed at Level 2. For example, whereas the level 2 geographicCoverage may encompass the entire study area, this tag may contain descriptions of specific sampling stations.

<**studyExtent**> provides specific information about the temporal and geographic extent of the study. This can be either as a simple text **<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 a city within the LTER site's overall sampling area, using the coverage/geographicCoverage/ tree. However, 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 can be used to optionally describe individual study sites. There are 3 ways to include site descriptions described here, and each has pros and cons. When the metadata is displayed in a browser window, all three of these methods will show their content under <spatialSamplingUnits>. In the example, stations within the city are listed as content in the structured coverage elements of the studyExtent tree.

a) As content in the 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'.

b) In the <project> tree: If a site has its project well described as a <project> tree with <relatedProject> trees for sub projects, each tree or subtree can contain the appropriate stations using a <geographicCoverage> element for each station, employing the id in the <geographicDescription> tag, and referencing the id value in methods/*/spatialSamplingUnits/referencedEntityId. However, by using the project tree, you may find you are including information in the metadata that is not directly pertinent to this dataset. One alternative would be to not include all the subtrees in every dataset. Another could be to put the project trees under additonalMetadata, where they might help a user identify other studies in the area. This might be the best choice for a site which creates fewer, large datasets. Remenber that all ids in an EML document must be unique. The <project> tree is more thoroughly described below.

c) In a table entity: Sampling stations can also be listed in a data table, which can 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. It should be kept in mind that if stations are listed in a table, it will not be possible for a metadata search engine to find these locations.

---

example 3.3. studyExtent, samplingDescription and spatialSamplingUnits

```
...
<studyExtent>
   <coverage>
      <temporalCoverage>
         <rangeOfDates>
            <beginDate>
               <calendarDate>1998-11-12</calendarDate>
            </beginDate>
            <endDate>
               <calendarDate>2003-12-31</calendarDate>
            </endDate>
         </rangeOfDates>
      </temporalCoverage>
      <geographicCoverage>
         <geographicDescription>Ficity, FI, 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>
   </studyExtent>
   <samplingDescription>
    <para>Arthropod pit fall traps are placed in three different locations at each
   site four times a year.</para>
   </samplingDescription>
   <spatialSamplingUnits>
      <coverage>
         <geographicDescription>site number 1</geographicDescription>
         <boundingCoordinates>
            <westBoundingCoordinate>-112.2</westBoundingCoordinate>
            <eastBoundingCoordinate>-112.2</eastBoundingCoordinate>
            <northBoundingCoordinate>33.5</northBoundingCoordinate>
            <southBoundingCoordinate>33.5</southBoundingCoordinate>
         </boundingCoordinates>
      </coverage>
      <coverage>
         <geographicDescription>site number 2</geographicDescription>
         <boundingCoordinates>
            <westBoundingCoordinate>-111.7</westBoundingCoordinate>
            <eastBoundingCoordinate>-111.7</eastBoundingCoordinate>
            <northBoundingCoordinate>33.6</northBoundingCoordinate>
            <southBoundingCoordinate>33.6</southBoundingCoordinate>
         </boundingCoordinates>
      </coverage>
      <coverage>
         <geographicDescription>site number 3</geographicDescription>
         <boundingCoordinates>
            <westBoundingCoordinate>-112.1</westBoundingCoordinate>
            <eastBoundingCoordinate>-112.1</eastBoundingCoordinate>
            <northBoundingCoordinate>33.7</northBoundingCoordinate>
            <southBoundingCoordinate>33.7</southBoundingCoordinate>
         </boundingCoordinates>
      </coverage>
   </spatialSamplingUnits>
   ...
```

The <**project**> tree should be included at Level 3 for more general descriptions of the project sponsoring the data package (i.e., this LTER site). Project trees may be nested, if smaller sub-projects are conducted under the auspices of the LTER site. In this case, the overall LTER site should be described first, and subprojects incuded as children using <**relatedProject**>.

Minimally, LTER site-level description should include:

<**title**> is the name of the LTER site

<**personnel**> list the lead PI and information manager
<**abstract**> containing study area description and mission statement
<**distribution**>, **online**/**url** minimally, should link to the project or site URL home page, but could link to other publications describing the site.
<**fundingSource**> agency and grant number

<**studyAreaDiscription**> tree and its accompanying <citation> tree may optionally be used to describe non-coverage characteristics of the study area such as climate, geology or disturbances or references to citable biological or geophysical classification systems such as the Bailey Ecoregions or the Holdridge Life Zones. The studyAreaDiscription tree also supports multiple <coverage> elements 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 3.4. project

```
...
<project>
    <title>Fictitious LTER Site permanent monitoring program(FLS(</title>
    <personnel id="pers-30" system="FLS">
      <individualName>
        <salutation>Dr.</salutation>
        <givenName>Eva</givenName>
        <givenName>M.</givenName>
        <surName>Scientist</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>
    <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>custodian</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 sites, 4 times a year.
            Climate parameters are continuously measured at all stations.</para>
      </abstract>
   </project>
   ...
```

Note: As stated earlier, the inclusion of the entire dataTable tree at Level 3 was not the preference of the committee, but was imposed by the EML schema and the logical progression of the Best Practices Levels. In the future, inclusion of some of the dataTable elements may be moved to Level 5.

For the **dataTable/entityGroup** tree:

<**alternateIdentifier**> (optional): The primary identifier belongs in the <dataTable id=”xxx”> , but this tag can accommodate additional identifiers that might be used, possibly from different data management systems.

<**entityName**> (required) this is often the table name, or could be the original ascii file name.

<**entityDescription**> This should be a longer more descriptive title, but does not need to reiterate the dataset title.

For the **dataTable/attributeList** tree:

<**attributeName**> usually the name of a field in a table, which is often short.

<**attributeLabel**> - if the attributeName is cryptic, consider using <attributeLabel> as well to provide a more intelligible name.

<**attributeDefinition**> – your last chance to be clear and unambiguous.

<**storageType**> - may be system specific, as for a RDBMS, ie A Microsoft SQL varchar, or Oracle datetime. Non system-specific values include float, integer and string.

<**measurementScale**> is required at this level. One of the 5 scale types must be used: nominal, ordinal, interval, ratio, or date-time, as follows:

<**nominal**> is used to represent named categories, a list of coded values, or plain text descriptions.

<**ordinal**> values are also named and in a set order with reference to one another, but the distance between them is not indicated (e. g., 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. The Celsius temperature scale represents part of the Kelvin scale that has been related to the physical properties of water, and so this measurement is interval.

<**ratio**> measurements have a meaningful zero point, and ratio comparisons between values are legitimate. For example, kelvin units reflect the amount of kinetic energy of a substance (ie, zero is the point where molecular motion stops), and so temperature measured in kelvin units is a ratio measurement. Concentration is also a ratio measurement because a solution at 10 micromolesPerLiter has twice as much substance as one at 5 micromolesPerLiter.

Note: The <**interval**> and <**ratio**> scales require additional tags describing <**unit>**, and the <**numericDomain>**. <**precision>** is optional at Level 3, now that the EML schema does not require this element as of version 2.0.1.

Unit names will be either <**standardUnit**> (from the unit dictionary) or <**customUnit**> (defined in the <**additionalMetadata**> section at end of document).

<**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://intranet.lternet.edu/eml/EMLHandbook.pdf p. 23. The <**bounds**> are theoretical or permitted minimum and maximum values (prescriptive), rather than the actual observed range in a data set (descriptive). 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.

<**datetime**> is a date-time value from the Gregorian calendar and it is recommended that these be expressed in a format that conforms to the ISO 8601 standard. An example of an allowable ISO date-time is "YYYY-MM-DD", as in 2004-06-25, or, more fully, as "YYYY-MM-DDThh:mm:ssTZD" (eg 1997-07-16T19:20:30.45Z). The ISO standard is quite strict about the structure of date components. Since legacy data often contain non-standard dates, and existing equipment (e.g., sensors) may still be producing non-standard dates, the EML authors have provided additional allowable formats. See the EML documentation for a complete list. (Note

that the datetime field should not be used for recording time durations. In that case, one should use a <**standardUnit**> such as seconds, nominalMinute or nominalDay, or a <**customUnit**> that defines the unit in terms of its relationship to SI second.)

The <**missingValueCode**> is optional, but should be included to describe any missing value codes present in the data set (e.g. NaN, ND, 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 3.5. attributeList tree (SQL system):

```
...
<attributeList id="arthro_taxa.attributeList">
     <attribute id="dbo.arthro_taxa.taxon">
        <attributeName>taxon</attributeName>
        <attributeLabel>taxon abbreviation</attributeLabel>
        <attributeDefinition>FSL specific six leter code for the taxon Name
        </attributeDefinition>
        <storageType typeSystem="Microsoft SQL Server">
           varchar</storageType>
        <measurementScale>
           <nominal>
              <nonNumericDomain>
                 <textDomain>
                    <definition>Unique species name
                       abbreviation</definition>
                 </textDomain>
              </nonNumericDomain>
           </nominal>
        </measurementScale>
     </attribute>
     <attribute id="dbo.arthro_taxa.taxon_id">
        <attributeName>taxon_id</attributeName>
        <attributeLabel>ITIS taxon ID</attributeLabel>
        <attributeDefinition>taxonomic code as used in ITIS (TSN)
        </attributeDefinition>
        <storageType typeSystem="Microsoft SQL Server">int</storageType>
        <measurementScale>
           <nominal>
              <nonNumericDomain>
                 <textDomain>
                    <definition>unique number as used in ITIS</definition>
```

```
                </textDomain>
              </nonNumericDomain>
            </nominal>
          </measurementScale>
        </attribute>
        <attribute id="dbo.arthro_taxa.taxon_name">
          <attributeName>taxon_name</attributeName>
          <attributeLabel>name of taxon</attributeLabel>
          <attributeDefinition>name of taxon as used in this
          study</attributeDefinition>
          <storageType typeSystem="Microsoft SQL Server">
            varchar</storageType>
          <measurementScale>
            <nominal>
              <nonNumericDomain>
                <textDomain>
                  <definition>taxon name</definition>
                </textDomain>
              </nonNumericDomain>
            </nominal>
          </measurementScale>
        </attribute>
        <attribute id="dbo.arthro_taxa.date_entered">
          <attributeName>date_entered</attributeName>
          <attributeDefinition>the date this taxon was first entered into the
           taxon list<attributeDefinition>
          <storageType typeSystem="Microsoft SQL Server">
            datetime</storageType>
          <measurementScale>
            <datetime>
              <formatString>MM/DD/YYYY</formatString>
              <dateTimePrecision>1</dateTimePrecision>
              <dateTimeDomain>
                <bounds>
                  <minimum exclusive="false">6/5/2000</minimum>
                  <maximum exclusive="false">12/12/2001</maximum>
                </bounds>
              </dateTimeDomain>
            </datetime>
          </measurementScale>
        </attribute>
      </attributeList>
...
```

Example 3.5. attributeList (numeric attributes with customUnits)

```
...
<attributeList>
   <attribute>
      <attributeName>temp</attributeName>
      <attributeDefinition>Water Temperature</attributeDefinition>
      <storageType>float</storageType>
      <measurementScale>
         <interval>
            <unit>
               <standardUnit>celsius</standardUnit>
            </unit>
            <precision>0.001</precision>
            <numericDomain>
               <numberType>real</numberType>
            </numericDomain>
         </interval>
      </measurementScale>
      <missingValueCode>
         <code>NaN</code>
         <codeExplanation>value not recorded or invalid</codeExplanation>
      </missingValueCode>
   </attribute>
   <attribute>
      <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>
   ...
   <addtionalMetadata>
      <unitList>
         <unit id="siemensPerMeter" name="siemensPerMeter"
            unitType="conductance" parentSI="siemen" multiplerToSI="1">
            <description>electrical conductance of a solution
               (conductivity)</description>
         </unit>
      </unitList>
   </additionalMetadata>
   ...
```

Level 4 – Access:

Access-level metadata should provide a user with all the information needed to access and download the data tables, even if the tables' attributes are not thoroughly described. The tags required at this level specify access control and the physical description of the table. The <**access**> tree describes the permission settings for the data package, and tags within the **dataTable/entityGroup** tree describe the physical structure for the data table.

<**access**> values must to be specific to the system where the data is stored. For Metacat, the format conforms to the LDAP distinguishedName for an individual, as in "uid=FLS,o=LTER,dc=ecoinformatics,dc=org". With the exception of sensitive information, metadata should be publicly accessible even if data tables are not.

Example 4.1. access

```
...
<access authSystem="FLS">
   <allow>
      <principal>PUBLIC</principal>
      <permission>read</permission>
   </allow>
   <allow>
      <principal>uid=fls,o=LTER,dc=ecoinformatics,dc=org</principal>
      <permission>all</permission
   </allow>
</access>
...
```

The **entityGroup/physical** tree further describes the physical aspects of the table:

<**objectName**> should represent the publicly available file with the specific file name (possibly exported as text from a database). This may be the same as the entityName but will be different if the object (ie, datafile) has several entities (e.g. as in a Excel workbook with several sheets, see EML Handbook).

<**externallyDefinedFormat**> descriptions of software should spell out manufacturer, program, and version, e.g. Microsoft Excel 2002. The NCEAS recommendation is following mime type.

<**distribution**> This may be system-specific, and not useful for outside users (example 4.2).

Example 4.2 physical (a Metacat ASCII table):

```
...
<dataTable>
   <entityName>arthro_hab</entityName>
   <entityDescription> habitat description for the sampling
      locations</entityDescription>
   <physical>
   <objectName>flslter.299.1</objectName>
   <size unit="bytes">59847</size>
   <dataFormat>
      <textFormat>
         <numHeaderLines>1</numHeaderLines>
         <recordDelimiter>#x0A</recordDelimiter>
         <attributeOrientation>column</attributeOrientation>
         <simpleDelimited>
            <fieldDelimiter>,</fieldDelimiter>
         </simpleDelimited>
      </textFormat>
   </dataFormat>
   <distribution>
      <online>
         <url>ecogrid://knb/flslter.296.1</url>
      </online>
   </distribution>
</physical>
...
```

Example 4.3. physical (a database table):

```
...
<dataTable id="dbo.arthro_taxa">
<entityName>dbo.arthro_taxa</entityName>
   <entityDescription/>
   <physical>
      <objectName>arthro_taxa</objectName>
      <dataFormat>
         <externallyDefinedFormat>
            <formatName>Microsoft SQL Server</formatName>
            <formatVersion>7.0.842</formatVersion>
         </externallyDefinedFormat>
      </dataFormat>
      <distribution>
         <online>
            <connection>
               <connectionDefinition>
                  <references>FLS1_SQL</references>
               </connectionDefinition>
               <parameter>
                  <name>catalog</name>
                  <value>monitoring</value>
               </parameter>
               <parameter>
                  <name>owner</name>
                  <value>dbo</value>
               </parameter>
            </connection>
         </online>
      </distribution>
   </physical>
</dataTable>
...
```

Level 5 – Integration:

Integration-level metadata should support computer-mediated access and processing of data, and therefore requires that all aspects of the data package be fully described. The additional trees to be included are those for **methods/qualityControl** and **dataTable/constraint**. The **dataTable/attributeList** must be fully described at this level, including the optional element <precision> which was omitted at Level 3.

<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 (eg, large), and .001 meters for a smaller species, but all the data on fish length are collected into one attribute and are measured using the closest precision values. For these cases precision can be omitted, but the variable precision information should be described in detail in **method/methodStep**.

The <**constraint**> tree is for describing any integrity constraints between entities (e.g. tables), as they would be maintained in a relational management system. Constraints include primary, foreign and unique key constraints, check constraints, and not-null constraints. The example shows the constraints for the attributeList described above. If there are constraints in which several columns are involved, these should be described in methods/QAQC, since EML is not currently equipped to handle multiple column Define keys.

Example 5.1 constraint (from table described in Examples 4.3 and 5.1)

```
...
<constraint id="pkarthro_taxa">
<primaryKey>
      <constraintName>pkarthro_taxa</constraintName>
      <key>
         <attributeReference>dbo.arthro_taxa.taxon</attributeReference>
      </key>
   </primaryKey>
</constraint>
<constraint id="arthro_taxa.taxonNotNull">
   <notNullConstraint>
      <constraintName>arthro_taxa.taxonNotNull</constraintName>
      <key>
         <attributeReference>dbo.arthro_taxa.taxon</attributeReference>
      </key>
   </notNullConstraint>
</constraint>
<constraint id="arthro_taxa.date_enteredNotNull">
   <notNullConstraint>
      <constraintName>arthro_taxa.date_enteredNotNull</constraintName>
      <key>
      <attributeReference>dbo.arthro_taxa.date_entered</attributeReference>
      </key>
   </notNullConstraint>
</constraint>
...
```

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

Example 5.2. a description of qualityControl at the attribute level

```
...
<attribute>
   <attributeName>kd_flag</attributeName>
   <attributeLabel>quality flag for diffuse attenuation
    coefficient</attributeLabel>
   <attributeDefinition>describes the uniformity of incident radiation during
    interval that kd was calculated. (0=uniform sfc radiation, ie, clean kc;
    14=most variable)</attributeDefinition>
   <storageType>float</storageType>
   <measurementScale>
      <ratio>
         <unit>
            <standardUnit>dimensionless</standardUnit>
         </unit>
      <precision>0.01</precision>
      <numericDomain>
         <numberType>real</numberType>
         <bounds>
            <minimum exclusive="false">0</minimum>
               <maximum exclusive="false">14</maximum>
            </bounds>
         </numericDomain>
      </ratio>
</measurementScale>
   <method>
      <qualityControl>
         <description>
            <para>Passage of clouds during a profile reduces the incident
            radiation, and leads to erroneous estimates of Kd. Variation of
            incident irradiance was described in two ways (before binning):
            1) the coefficient of variation (cv) over the 10m depth
            interval, and 2) difference in incident irradiance between two
            adjacent observations (first diff).  A low flags indicates that
            both cv &lt; 5% and first diff &lt; 0.01, and a high flag (up
            to a max of 14), that either or both of these conditions were
            not met. The flag was binned along with the data, so its
            magnitude indicates the relative proportion of possible errors
            for each Kd calculation in the binned profile.</para>
```

```
            </description>
          </qualityControl>
        </method>
      </attribute>
      ...
```

III. Recommendations for Metacat Compatibility

The NCEAS Metacat Server, and related Morpho client, are generally compatible with all XML documents conforming to the EML 2.0.0 and 2.0.1 schemas (see http://ecoinformatics.org). 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/KNB 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-gce.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 (for sites capable of dynamic EML generation) and XSLT transformation.

b) Access Control

Metacat access control format conforms to the LDAP distinguishedName for an individual, as in "uid=FLS,o=LTER,dc=ecoinformatics,dc=org". Access elements for documents contributed to the KNB Metacat should be formed as follows:

```
<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>
```

c) Display issues

The "Organization" field on the Metacat query results page is populated using the first eml:eml/dataset/creator/organizationName element in the document, so it is recommended that for LTER-contributed data sets the LTER site be included as the first creator (i.e. using <organizationName>)

IV. List of Example Documents

Completeness Level Examples

Level 1: emlbestpractices_level1.xml
Level 2: emlbestpractices_level2.xml
Level 3: emlbestpractices_level3a.xml (tabular), emlbestpractices_level3b.xml (RDBMS)
Level 4: emlbestpractices_level4a.xml (tabular), emlbestpractices_level4b.xml (RDBMS)
Level 5: emlbestpractices_level5a.xml (tabular), emlbestpractices_level5b.xml (RDBMS)

Each of these documents is also provided in template form (i.e. tags only, with all content removed), named according to the original file with "_template" appended to the base filename (e.g. emlbestpractices_level1_template.xml). These documents may be loaded into an XML or text editor and filled in to produce valid EML metadata documents at the desired level of completeness.

Example LTER Site Documents

*(**Note:** site documents are included to illustrate general patterns of valid EML currently being produced by LTER sites, and may deviate somewhat from best practice recommendations due to limitations of current site metadata content or other unresolved legacy issues)*

CAP-LTER:
example_cap1975.xml (spatial-vector; ESRI shapefile)
example_cap_stef_expert_max_phx_98.xml (spatial-raster; Landsat imagery)

GCE-LTER:
example_gce_21_6.xml (tabular dataset; river chemistry monitoring)
example_gce_129_5.xml (tabular dataset; aquatic invertebrate population survey)
example_gce_187_4.xml (tabular dataset, long-term physical oceanography mooring)

SBC-LTER:
example_sbclter.300.10.xml (edited Morpho tabular dataset; kelp abundance)
example_sbclter.379.26.xml (edited Morpho tabular dataset; stream chemistry)

V. List of Contributors to this Document

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

Participants:

- 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)