



Buen Metadata con EML

Mark Servilla

Network Information System

LTER Network Office

University of New Mexico





Topics

- Metadata – What is it?
- Ecological Metadata Language
- LTER EML Best Practices



What is "Metadata"?

Information about data

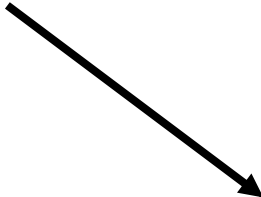
"higher level information that describe the **content**, **quality**, **structure**, and **accessibility** of a specific data set" –
Michener et al., 1997





A Simple Example

Data

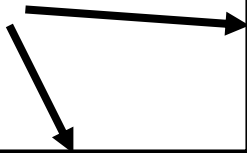


20040928	29.4	18.4
20040929	29.7	4.2
20040930	28.9	21.3



A Simple Example

Metadata



	Date (YYYYMMDD)	Temp (°C)	Precip. (mm)
Obs. #1	20040928	29.4	18.4
Obs. #2	20040929	29.7	4.2
Obs. #3	20040930	28.9	21.3



Data from the User's Perspective

- Who
 - Who collected these data?
 - Who is responsible for these data?
 - Who can access these data?
- What
 - What relevant data exist?
 - What is the structure and organization of these data?
- Where
 - Where were these data collected?
- Why
 - Why were these data collected and are they suitable for a particular use?
- When
 - When were these data collected?



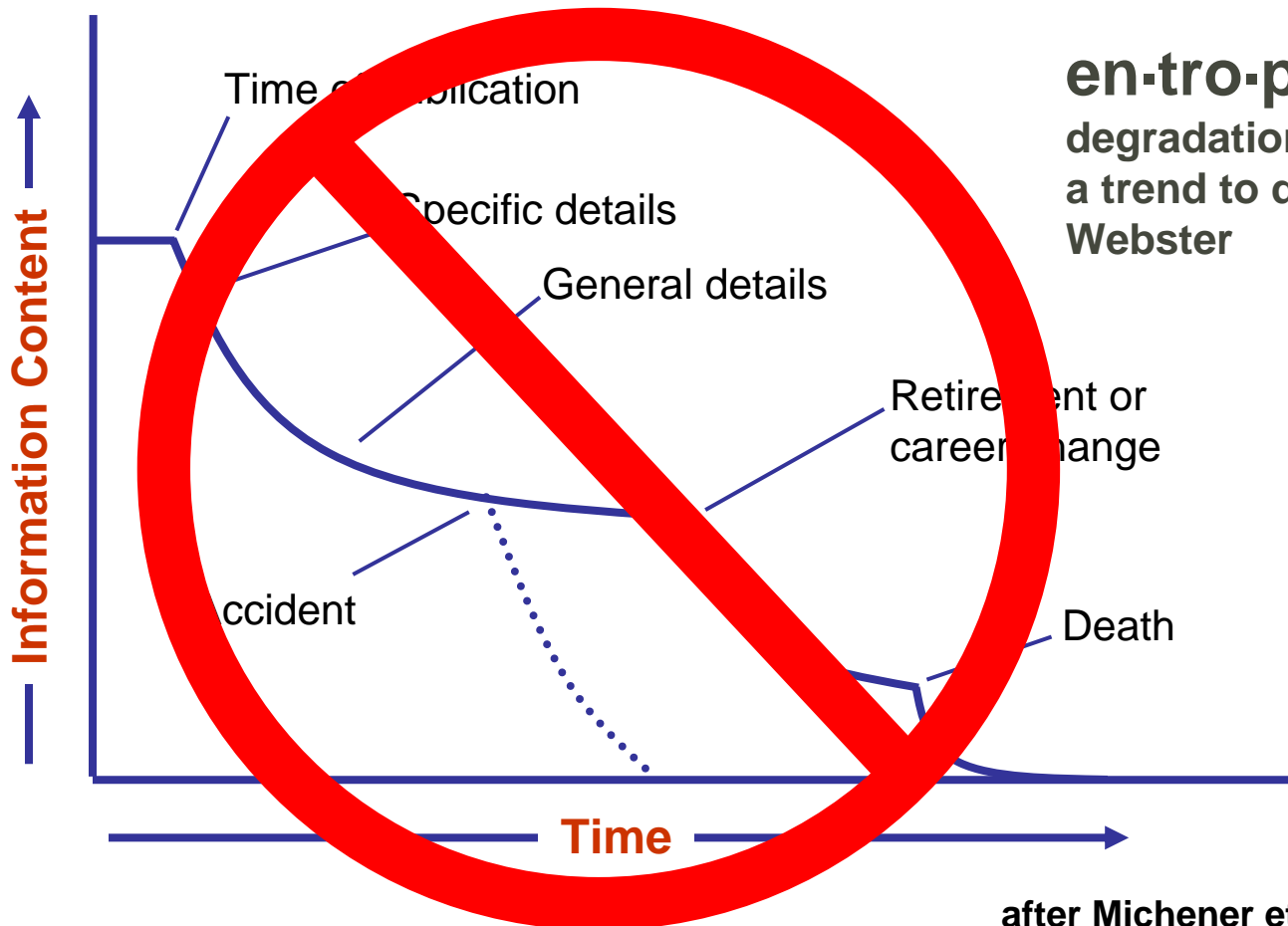
Data from the Collector's Perspective

- Data set
 - Title, abstract, researcher, ownership...
- Research origin
 - Hypothesis for research, site, methods...
- Data set status and accessibility
 - URL, URN, access rights...
- Data structural
 - Storage type, data attributes, data types...
- Supplemental
 - QA/QC, policies, other...



The purpose of Metadata is ...

... to reduce “information entropy”



en-tro-py : a process of degradation or running down or a trend to disorder – Merriam-Webster

after Michener et al., 1997

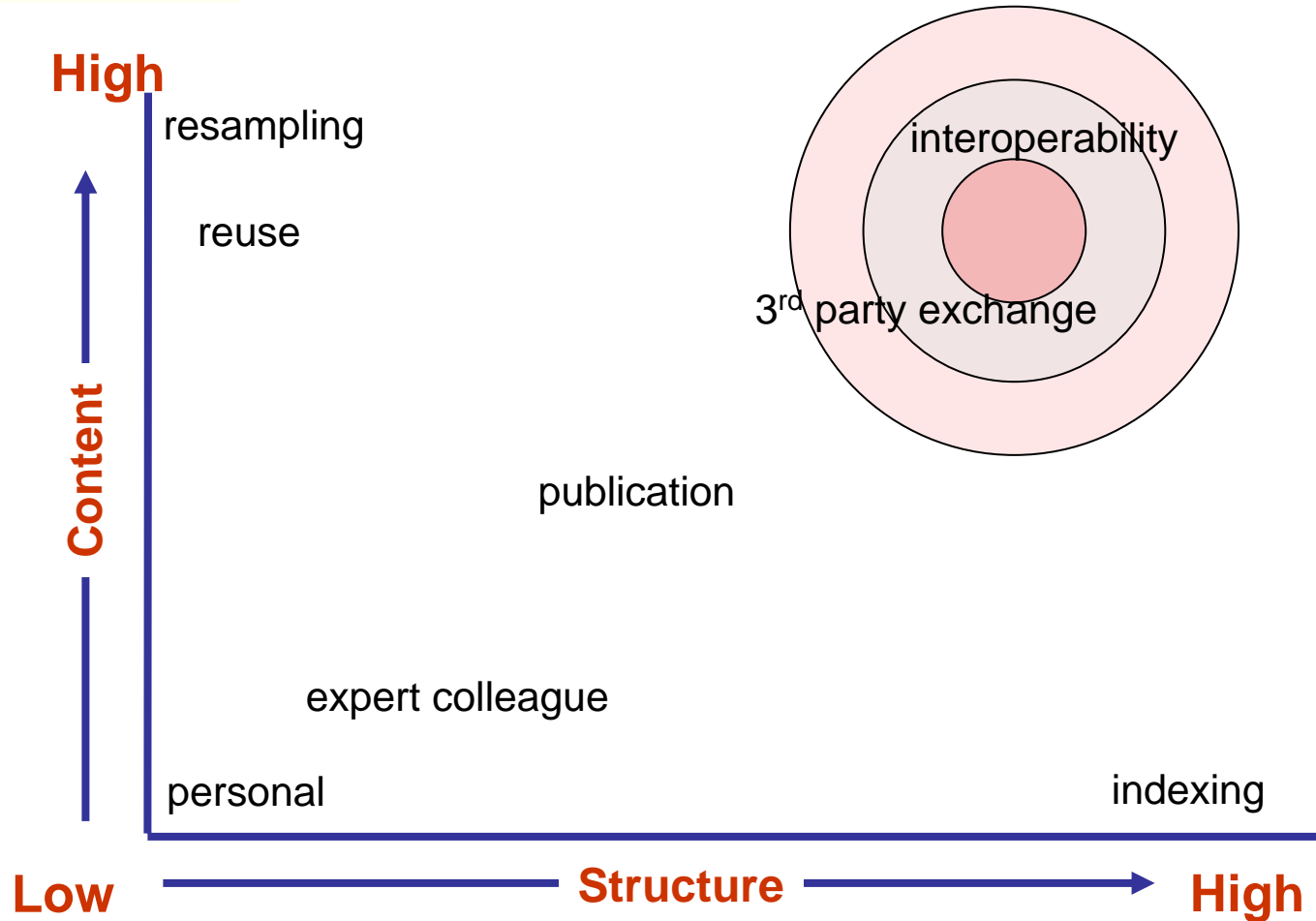


Open Metadata Standards

- Defines a common terminology
- Publicly available specifications for achieving a specific task
 - Supports a well defined structure
 - Promotes rich metadata
- Allows for system “cross-walks”; that is, mapping one metadata standard to another
- Allows for automation
 - Entry and validation
 - Discovery and access
 - Integration and synthesis
- Ensures complete and rich data documentation
- Foundation for higher-level knowledge



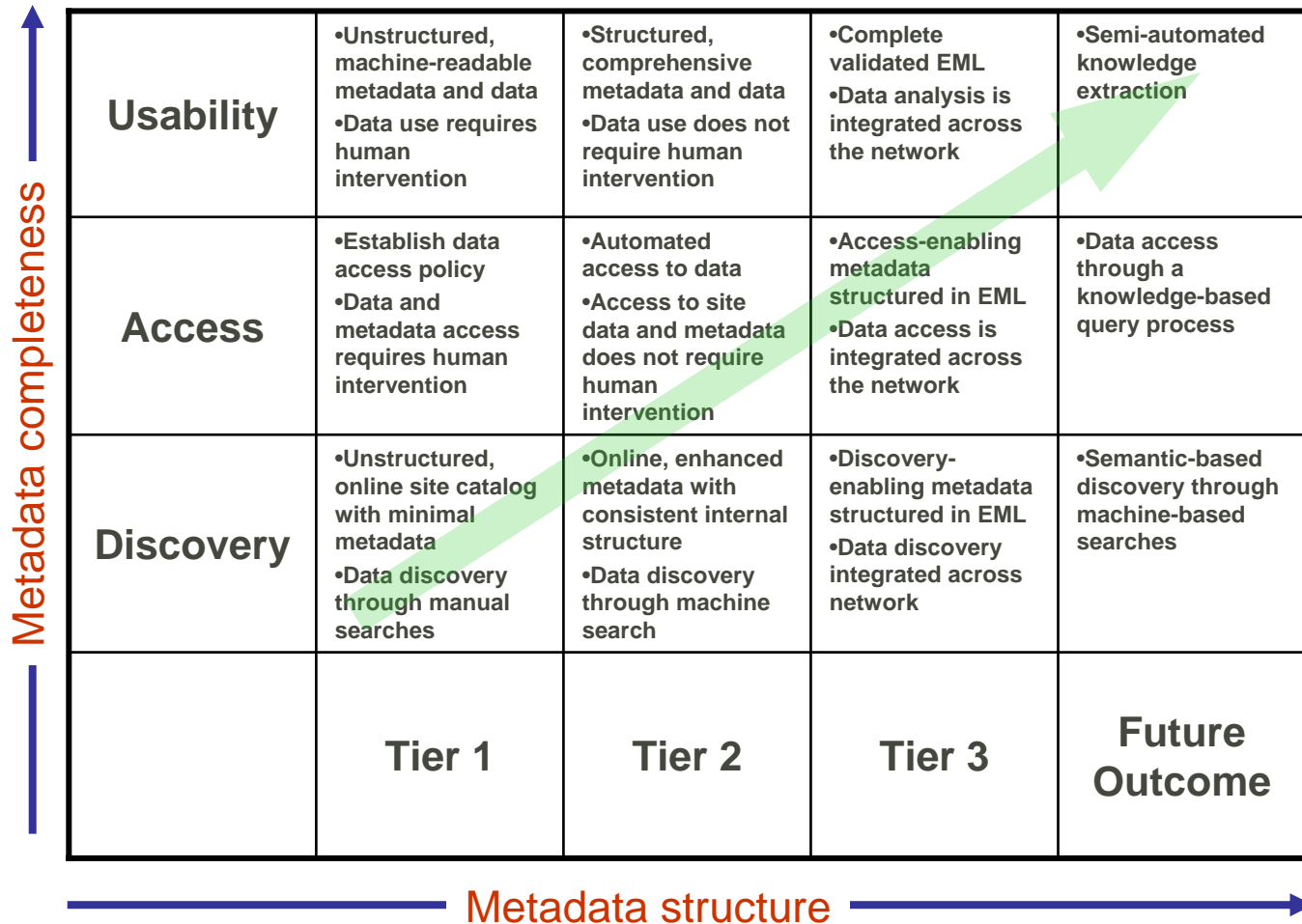
Benefits of Standardization



after Michener et al., 1997



LTER Tiered Trajectory





What is the Ecological Metadata Language?

EML is a metadata standard...

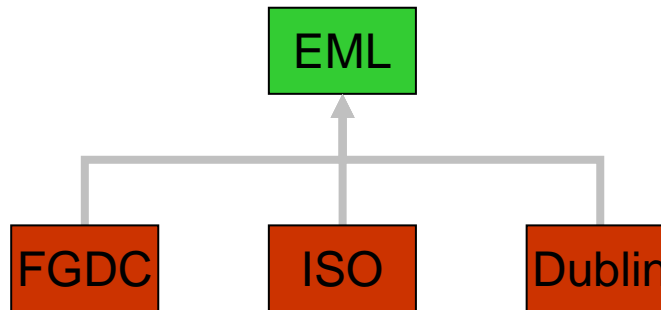
- Consists of four thematic modules
 - Dataset
 - Citation
 - Software
 - Protocol
- Defined by an XML Schema
- Supports a rich set of constructs to fully describe data, and much more...
- Based on other metadata standards compatibility



Conversion Among Standards

- EML represents a superset of:

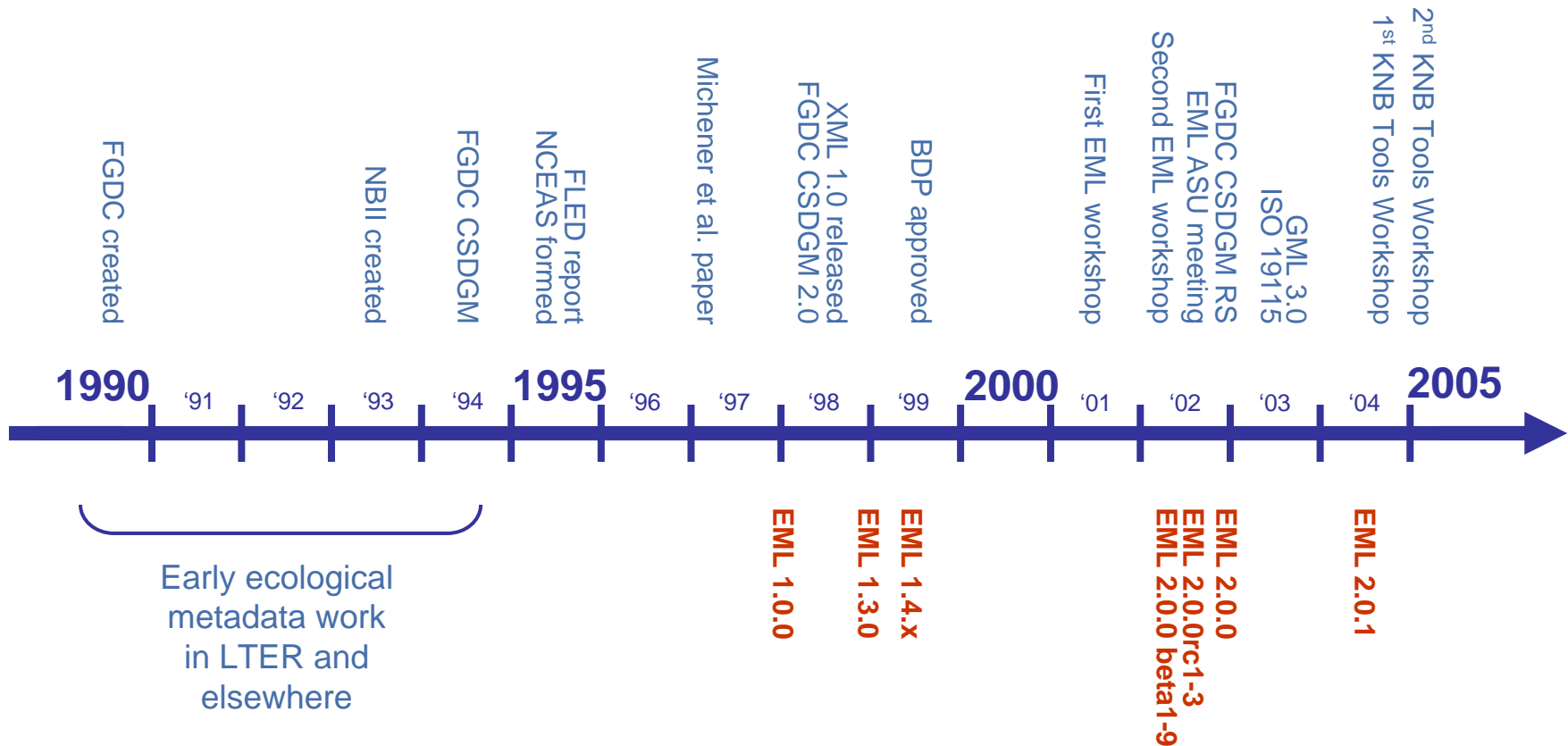
- FGDC
- ISO 19115
- Dublin Core
- and others...



- Theoretically can convert from EML to other standards
 - Extensive overlap among all of these standards
 - Practically, have conversion script for:
 - EML → BDP
 - Uses XSLT so can be used in a variety of software
 - Need to incorporate this into tools like Morpho and Metacat
 - Need other conversion scripts!
- More difficult going from other standards to EML
 - Fine grained content not supported in others

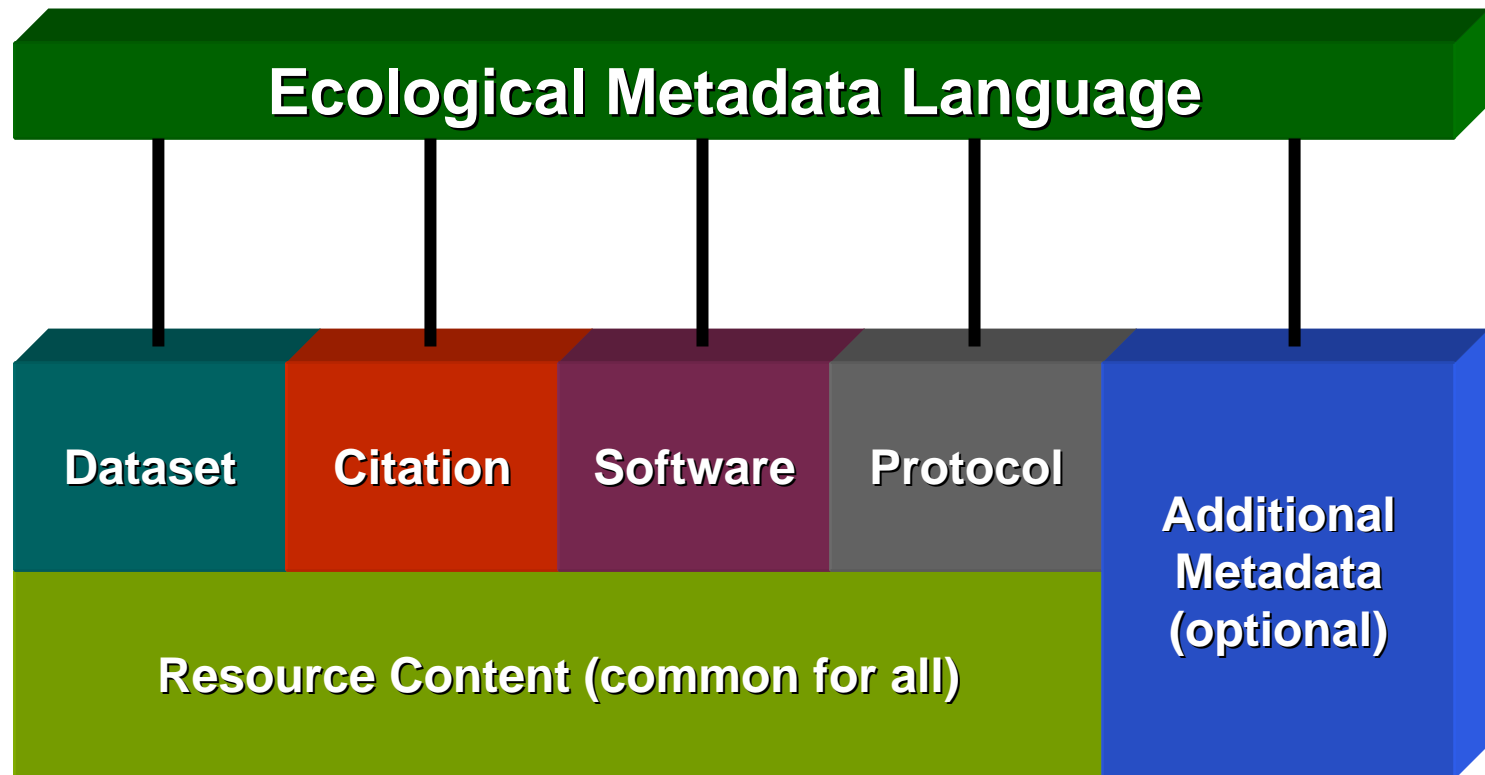


Abridged History of EML



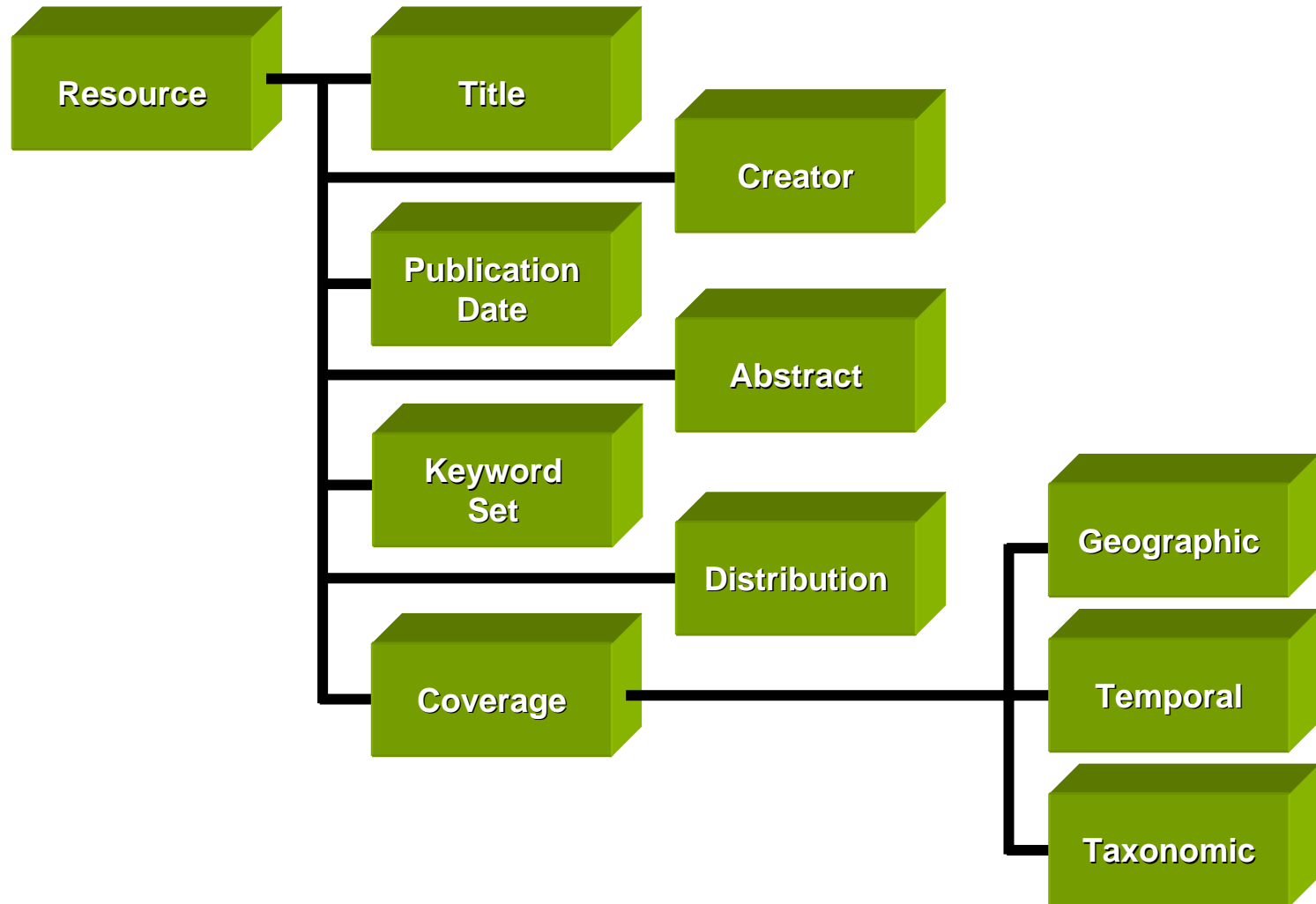


EML Module Organization



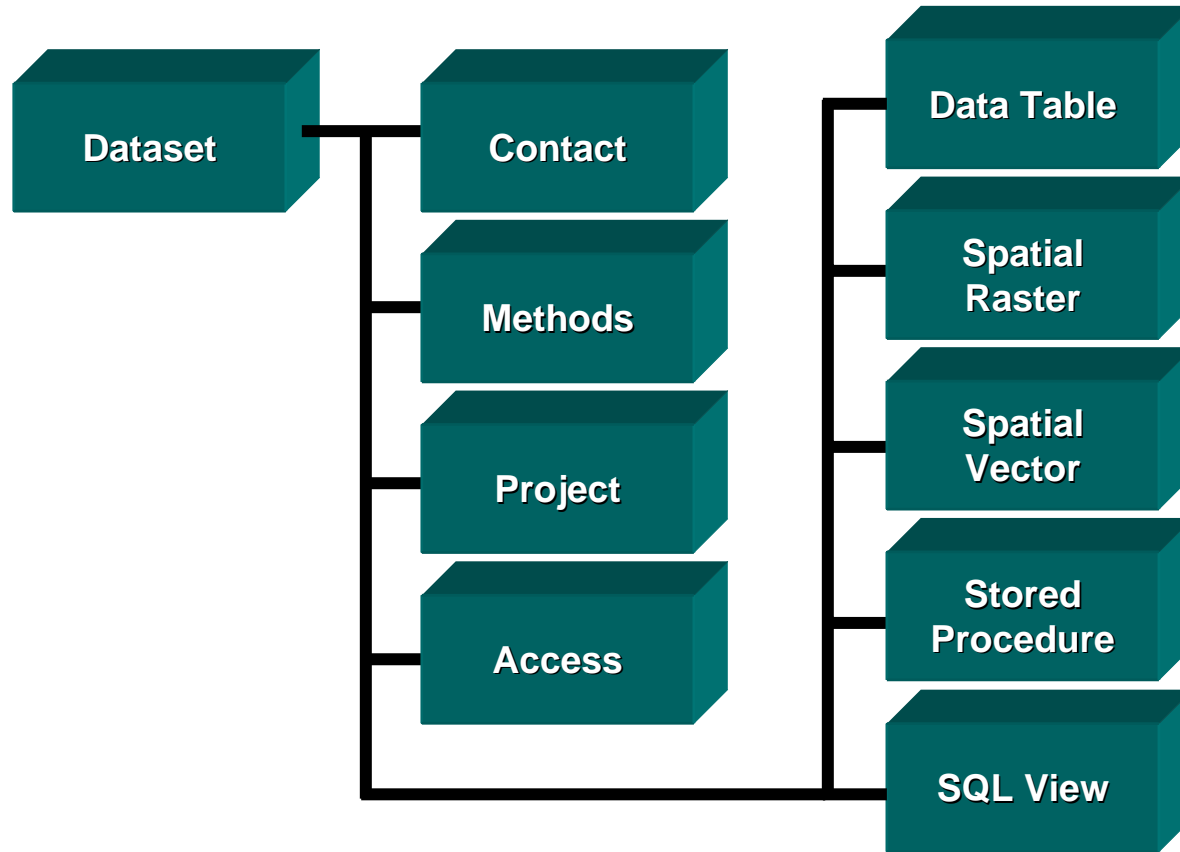


Resource Module



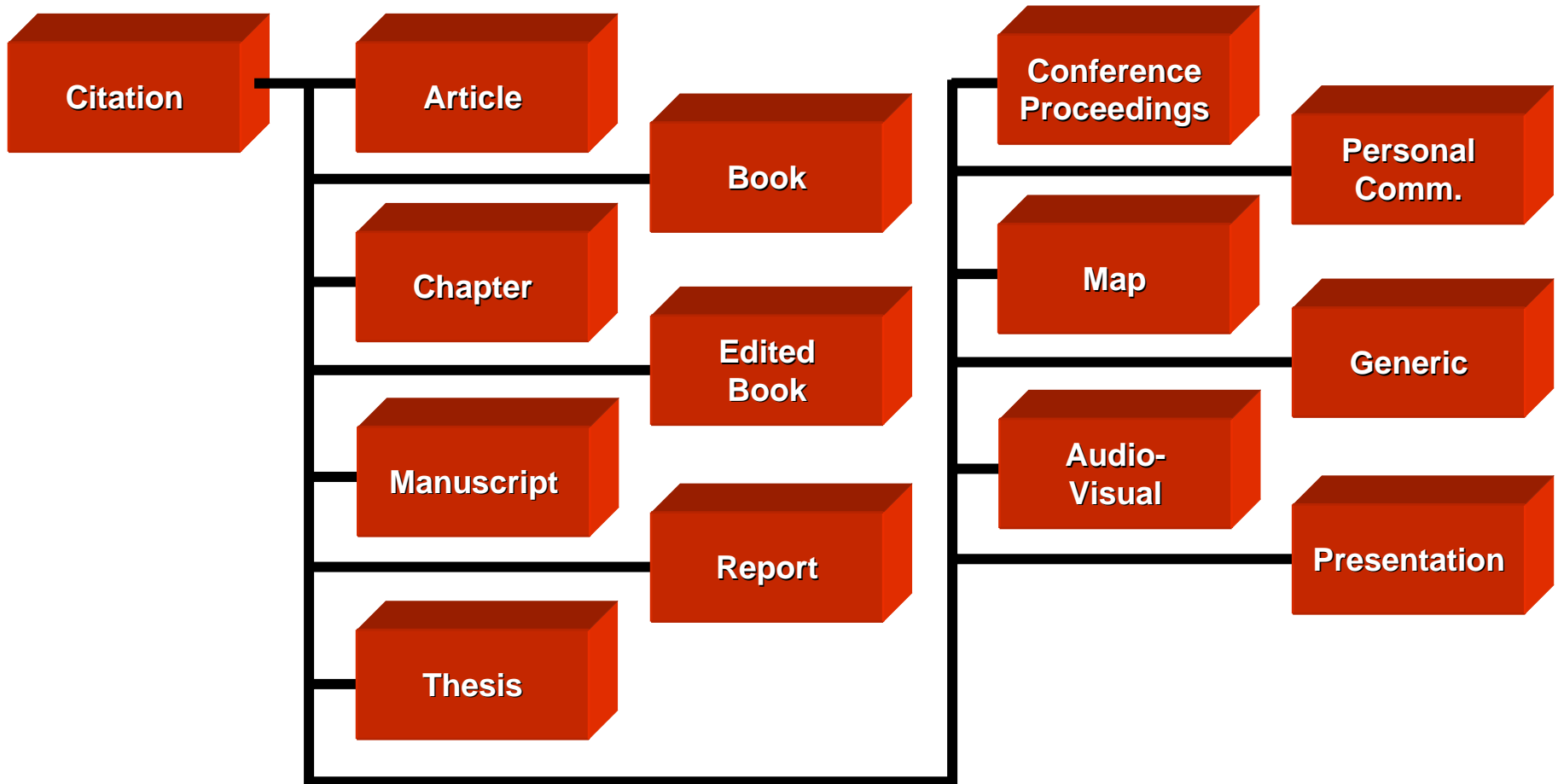


Dataset Module



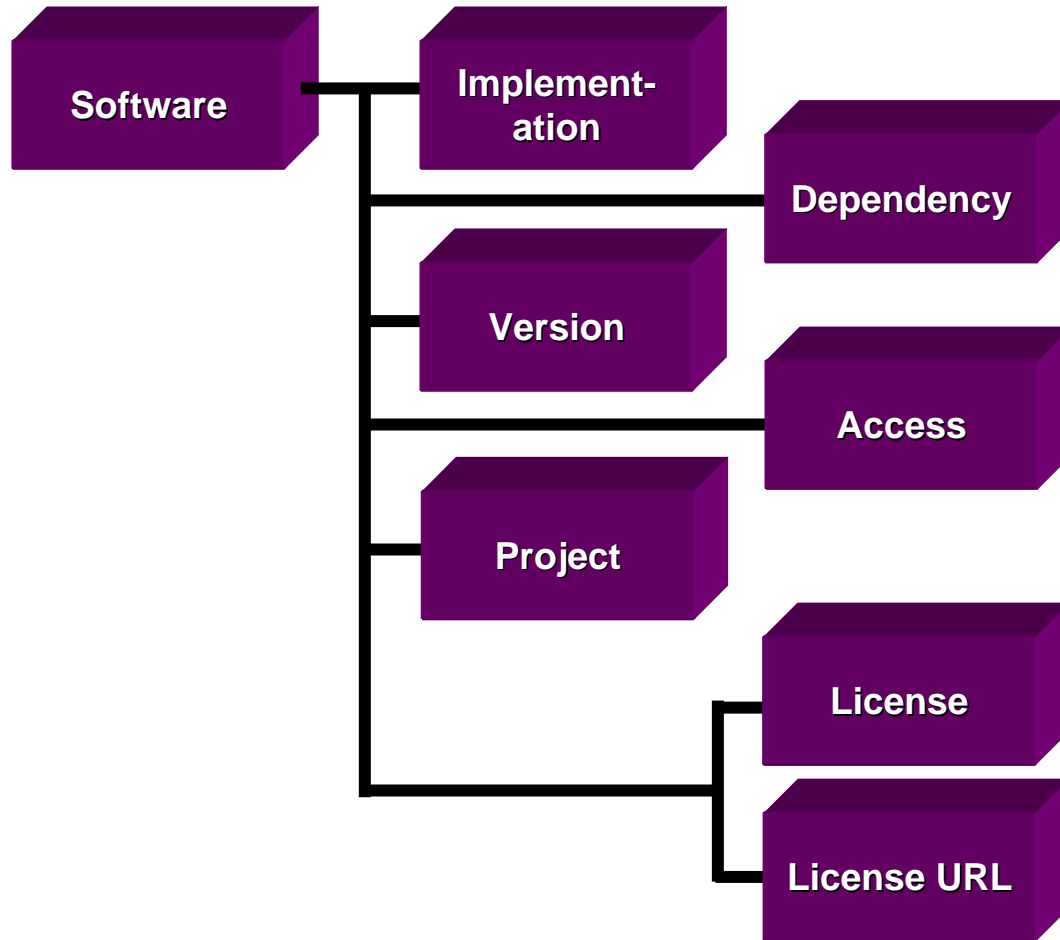


Citation Module



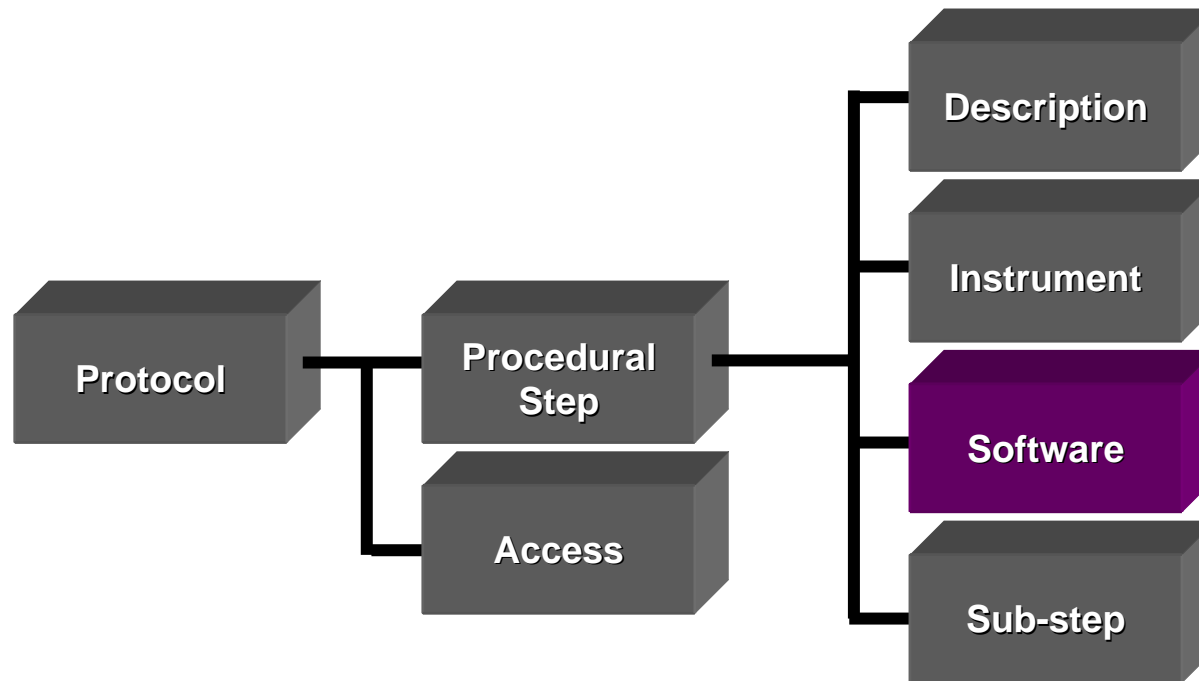


Software Module





Protocol Module





What is XML?

- eXtensible Markup Language
- How is it hierarchical?
 - Tree like
 - has a root
 - has inner leaves (inner nodes)
 - has outer leaves (end or outer nodes)
 - Parent/child relationship
 - Allows recursive nesting of nodes
 - All nodes have a beginning and ending marker ("tag")



Example XML Document

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

```
<BIB>
```

```
  <BOOK nickname="Dragon book">
```

```
    <AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
    <AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
    <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
    <TITLE>
```

```
      Compilers: Principles, Techniques, and Tools
```

```
    </TITLE>
```

```
    <PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
    <YEAR> 1985 </YEAR>
```

```
  </BOOK>
```

```
</BIB>
```

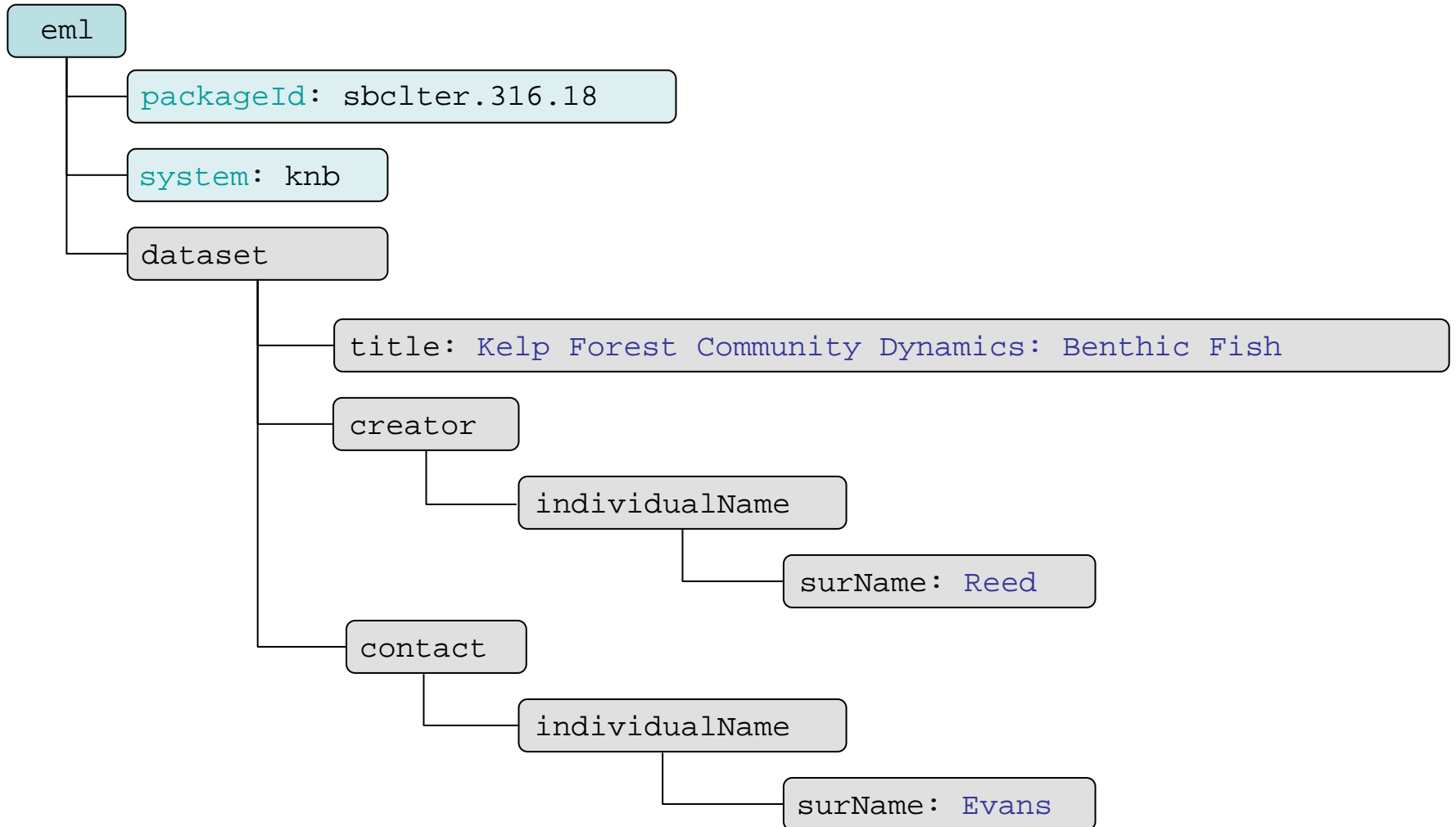


Benefits of XML

- Machine independent
- Can be read by many software applications
- Text based content (ascii, UTF-8, ISO...)
- Schemas support syntax validation
- Document object models support semantic validation
- Structure can easily be modified using transformation applications (e.g., xml to html; xml to text)



A simple EML example





Can be created in a text editor

```
eml-example.xml
1:0
<?xml version="1.0"?>

<eml:eml packageId="sbclter.316.18" system="knb"
  xmlns:eml="eml://ecoinformatics.org/eml-2.0.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.1 eml.xsd" >

  <dataset>
    <title>
      Kelp Forest Community Dynamics: Benthic Fish
    </title>
    <creator>
      <individualName>
        <surName>Reed</surName>
      </individualName>
    </creator>
    <contact>
      <individualName>
        <surName>Evans</surName>
      </individualName>
    </contact>
  </dataset>

</eml:eml>
```



...or in custom applications

The image displays two overlapping windows. The left window is a Mozilla Firefox browser showing the KNB Data Registry interface. The right window is a Microsoft Excel spreadsheet titled 'xls_eml_01_SAMPLE.xls'.

KNB Data Registry - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://knb.ecoinformatics.org/cgi-bin/register-dataset.c

ACMPD ACM TechNews Biohony Portal Celtic Sojourn Comcast LTER CVS

OGC

Bioco

Home

Data Registry F

Use this form to submit a

Please have a look at the
your browser's Reload/Re

If you have any question
help@nceas.ucsb.edu.

*Denotes a required field

BASIC INFORMATION

*First Name

*Last Name

*Data Set Title

*Organization Name

PRINCIPAL DATA SET

*First Name

*Last Name

Done

Morpho

File Edit Search Documentation Data Window Help

Current profile: servilla
(uid=mservilla,o=LTER,dc=ecoinformatics,dc=org)

Change profile: servilla

Create a new profile...

Network Status: Logged In

You are logged into the
files for which you have a

New Data Package Wizard

Welcome to the New

Logout from network

Work with your data...

Create a new data

Open an existing d

Search for an exist

This wizard creates a Data
themselves.
If you wish to improve you

An Introduction to Eco

which provides background
documentation is needed t
from the "Documentation".
Before beginning you sho

Step 1 of 15

• Title and abstr

• Keywords

• People and Or

Microsoft Excel - xls_eml_01_SAMPLE.xls

File Edit View Insert Format Tools Data Window Help Adobe PDF

Type a question for help

Arial 10 B U

B18 LTER site acronym

LTER EML Metadata Submission Template (Version 0.1 (October 2004))

(In compliance with the Ecological Metadata Language (EML) 2.0.1 Standards and Formats
and LTER EML Best Practices Tier 5 EXCEPT for the "Constraint" Element (September, 2004))

Documentation Instructions:

- Please fill in all blank, underlined rows as completely as possible unless directed otherwise by notes in the field. In some cases, field may not be applicable for data sets, such as 'Dataset Creator Organization Name'.
- Separate multiple entries using vertical line delimiters (|). (Applies to the Street Address and Research Project Temporal Coverage fields).
- Please do not insert or delete rows from the template (long entries will wrap within cells, except as noted below)
- Note that extremely long entries may not be completely displayed even when the cell is enlarged, but the full contents will be retained and will export properly.
- EML tags used for metadata field are listed in Column A of same metadata field row.
- Optional Metadata fields are magenta.
- Blue Metadata fields will most likely be completed by the Information Manager.
- If a row background is light blue, please enter multiple first names, addresses, keywords, etc. in new columns. Please note: Creator, Contact, Geographic description, Data Entity Taxon, and Research Project personnel sections must include all information for one person, taxon, or location in the same column.
- If user has trouble with Microsoft Excel adding erroneous hyperlinks to worksheet, user can delete all 'Hyperlinks' by running the 'DeleteAllHyperlinks' Excel Macro found under the tool section.

(NOTE: Move the mouse pointer over a field name to display instructions and comments)

(NOTE: Click +/- icon in column 1 to expand or hide EML category (i.e. Dataset Creator)).

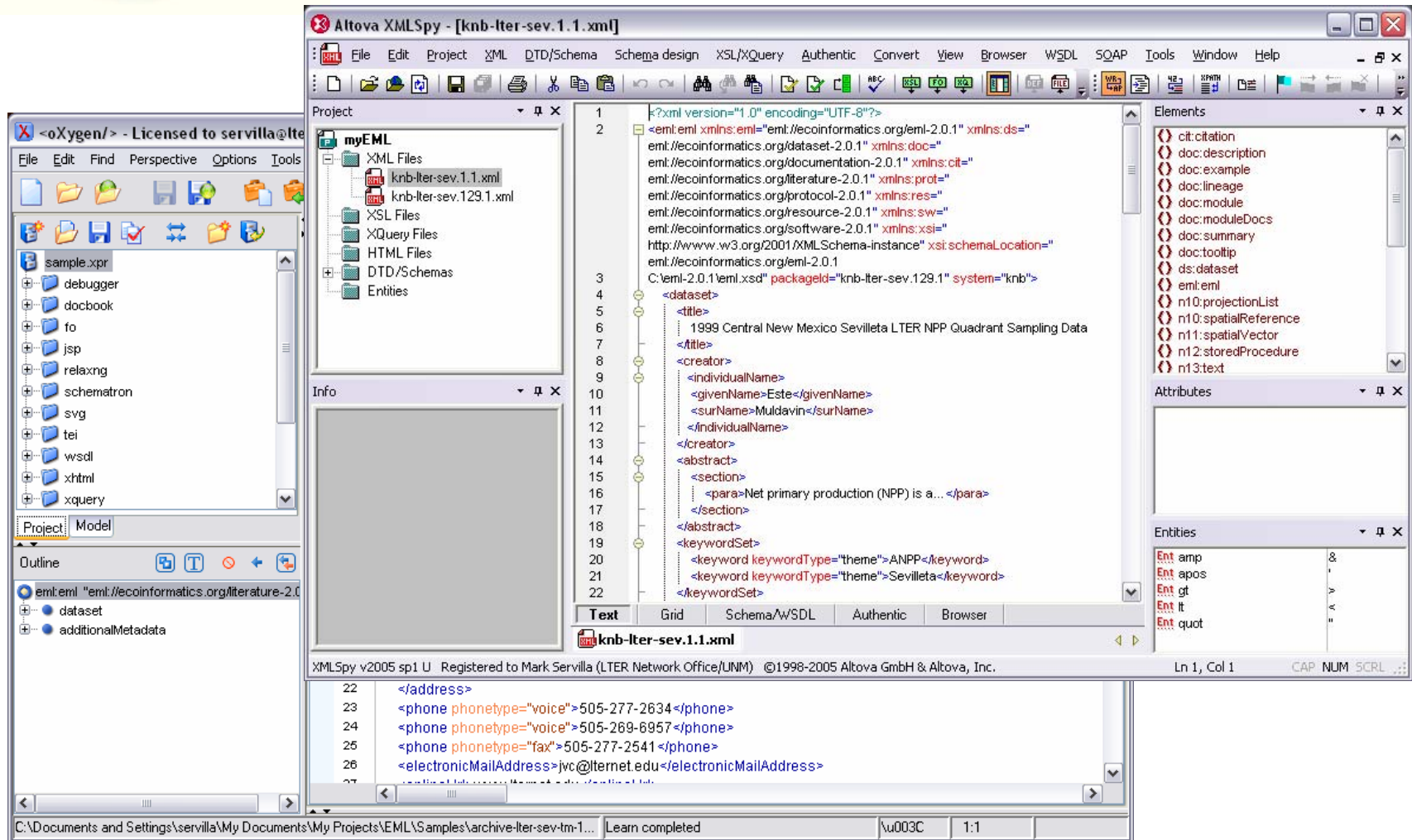
I. LTER Dataset Information

Field	Value
LTER site acronym	FCE
Metacat package ID	LT_ND_Grahl_001.0
Dataset LTER Identification Number	LT_ND_Grahl_001
Dataset Title	Water Quality Data (Extensive) from the Shark River

General Metadata / MethodsCitation / MethodsProtocol / ResearchProjects / DataTable / References / IM Use / NUM



...or in special
XML editors





EML Development & Communication

- Open Source project, welcomes contributions
- Developed by members of the community
- eml-dev@ecoinformatics.org
- irc.ecoinformatics.org, #eml channel for discussion
- Source code managed in cvs.ecoinformatics.org
- Documented specification found on the KNB website
- An EML validating service is available at:
<http://knb.ecoinformatics.org/emlparser>



Pete Taylor photo



EML Distribution

- Current release version eml-2.0.1
- Downloadable at:
<http://knb.ecoinformatics.org/software/eml>
- Development version available at:
<http://cvs.ecoinformatics.org/cvs/cvsweb.cgi/eml>





EML Best Practices





EML Best Practices

Why do we need “EML Best Practices”?

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



Level 1 Identification

- **Description** – Minimum content for adequate data set discovery
- **Major Elements Added :**
 - Title
 - Creator
 - Contact
 - Publisher
 - Publication Date
 - Keywords
 - Abstract
 - Dataset/distribution
(i.e. URL for dataset information)





Level 1 Code Example

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.0.1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.1
  http://someserver.fls.edu/eml-2.0.1/eml.xsd"
  packageId="knb-lter-fls.1.1" system="FLS" scope="system">

  <dataset id="FLS-1" system="FLS">
    <alternateIdentifier>FLS-1</alternateIdentifier>
    <shortName>Arthropods</shortName>
    <title>
      Long-term Ground Arthropod Monitoring Dataset at
      Silver City, NM USA from 1998 to 2004
    </title>
    . . .
  </dataset>
```



Level 1 Code Example

```
<creator id="pers-1" system="FLS">
  <individualName>
    <givenName>John</givenName>
    <surName>Ecologist</surName>
  </individualName>
  <organizationName>FLS LTER</organizationName>
  <address id="addr-1" system="FLS">
    <deliveryPoint>Department of Ecology</deliveryPoint>
    <deliveryPoint>University of New Mexico</deliveryPoint>
    <deliveryPoint>PO Box 1234</deliveryPoint>
    <city>Albuquerque</city>
    <administrativeArea>NM</administrativeArea>
    <postalCode>87131-1234</postalCode>
  </address>
  <phone phonetype="voice">(505) 999-9999</phone>
  <electronicMailAddress>jeco@unm.edu</electronicMailAddress>
  <onlineUrl>http://www.unm.edu/~jeco</onlineUrl>
</creator>
```



Level 2 Discovery

- **Description** – Level 1 content, plus coverage information to support targeted searches
- **Major Elements Added :**
 - Geographic Coverage
 - Taxonomic Coverage
 - Temporal Coverage





Level 2 Code Example

```
<coverage>
  <geographicCoverage>
    <geographicDescription>
      Silver City, NM USA
    </geographicDescription>
    <boundingCoordinates>
      <westBoundingCoordinate>-112.373634</westBoundingCoordinate>
      <eastBoundingCoordinate>-111.612936</eastBoundingCoordinate>
      <northBoundingCoordinate>+33.708829</northBoundingCoordinate>
      <southBoundingCoordinate>+33.298975</southBoundingCoordinate>
      <boundingAltitudes>
        <altitudeMinimum>304</altitudeMinimum>
        <altitudeMaximum>627</altitudeMaximum>
        <altitudeUnits>meter</altitudeUnits>
      </boundingAltitudes>
    </boundingCoordinates>
  </geographicCoverage>
  ...
```



Level 2 Code Example

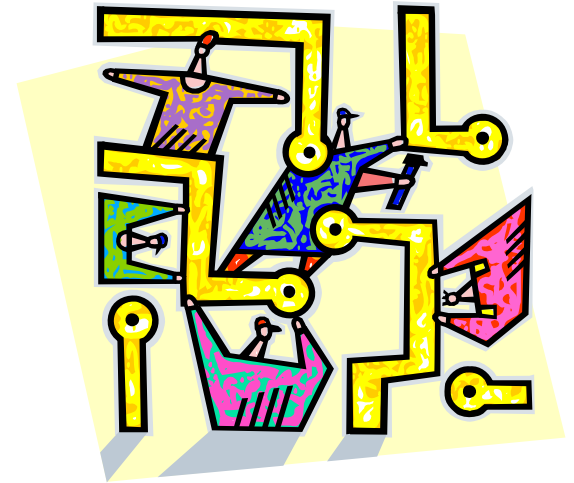
...

```
<temporalCoverage>
  <rangeOfDates>
    <beginDate>
      <calendarDate>1998-11-12</calendarDate>
    </beginDate>
    <endDate>
      <calendarDate>2003-12-31</calendarDate>
    </endDate>
  </rangeOfDates>
</temporalCoverage>
<taxonomicCoverage>
  <generalTaxonomicCoverage>
    Orthopteran insects (grasshoppers) were id using
    the 2004 BigKey to Orthoptera
  </generalTaxonomicCoverage>
  <taxonomicClassification>
    <taxonRankName>Kingdom</taxonRankName>
    <taxonRankValue>Animalia</taxonRankValue>
    <taxonomicClassification>
      <taxonRankName>Phylum</taxonRankName>
      <taxonRankValue>Arthropoda</taxonRankValue>
    </taxonomicClassification>
  </taxonomicClassification>
</taxonomicCoverage>
</coverage>
```



Level 3 Evaluation

- **Description** – Level 2 content, plus data set details to enable end-user evaluation of the methodology and data entities
- **Major Elements Added :**
 - Intellectual Rights
 - Project
 - Methods
 - Data Table/Entity Group
 - Data Table/Attributes (constrained by current version of EML)





Level 3 Code Example

```
<intellectualRights>
  <section>
    <para>
      The dataset is released to the public and
      may be used for academic or commercial purposes
      subject to the following restrictions:
    </para>
    <para>
      <itemizedlist>
        <listitem>
          <para>
            LTER will make every effort possible
            to control and document the quality of
            the data it publishes. Data are made
            available "as is"...
          </para>
        </listitem>
        ...
      </itemizedlist>
    </para>
  </section>
</intellectualRights>
```



Level 3 Code Example

```
...
<project>
  <title>Fictitious LTER Site (FLS) permanent monitoring program</title>
  <personnel id="pers-30" system="FLS">
    <individualName>
      <salutation>Dr.</salutation>
      <givenName>Eva</givenName>
      <surName>Scientist</surName>
    </individualName>
    <address>
      <reference>addr-1</reference>
    </address>
    <role>principalInvestigator</role>
  </personnel>
  <abstract>
    <para>
      The FLS basic monitoring program consists of monitoring of
      arthropod populations, plant net primary productivity, and bird
      populations. Monitoring takes place at 3 sites, 4 times a year.
      Climate parameters are continuously measured at all stations.
    </para>
  </abstract>
</project>
```



Level 3 Code Example

```
<methods>
  <methodStep>
    <description>
      <para>
        FSL Protocol for Surveying Ground Arthropods has been...
      </para>
    </description>
    <protocol>
      <title>
        FLS Protocol for Surveying Ground Arthropods
      </title>
      <creator>
        <references>pers-1</references>
      </creator>
      <pubDate>2000-02-23</pubDate>
      <abstract>
        <para>
          This protocol is being used by FLS arthropod...
        </para>
      </abstract>
      <keywordSet>
        <keyword keywordType="theme">Ecology</keyword>
        ...
      </keywordSet>
      <distribution>
        <online>
          <url>http://fls.univ.edu/protocols/arthro.html</url>
        </online>
      </distribution>
    </protocol>
  </methodStep>
  ...
```



Level 3 Code Example

```
<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: 0.002C, readability: 0.0001C, range: -5
    to 35C); last calibration: Feb 28, 2001
  </instrumentation>
</methodStep>
...
</methods>
```



Level 3 Code Example

```
...  
<dataTable>  
  <entityName>arthro_hab</entityName>  
  <entityDescription>  
    Habitat description for the sampling locations  
  </entityDescription>  
  <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>  
  ...
```



Level 3 Code Example

```
<attribute>
  <attributeName>cond</attributeName>
  <attributeLabel>Conductivity</attributeLabel>
  <attributeDefinition>
    measured with SeaBird Electronics 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>
...
```



Level 3 Code Example

```
...  
<additionalMetadata>  
  <unitList>  
    <unit id="siemensPerMeter" name="siemensPerMeter"  
      unitType="conductance" parentSI="siemen" multiplierToSI="1">  
      <description>  
        electrical conductance of a solution (conductivity)  
      </description>  
    </unit>  
  </unitList>  
</additionalMetadata>  
...
```



Level 4 Access

- **Description** – Level 3 content plus data access details to support automated data retrieval
- **Major Elements Added :**
 - Access
 - Physical





Level 4 Code Example

```
<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>
```



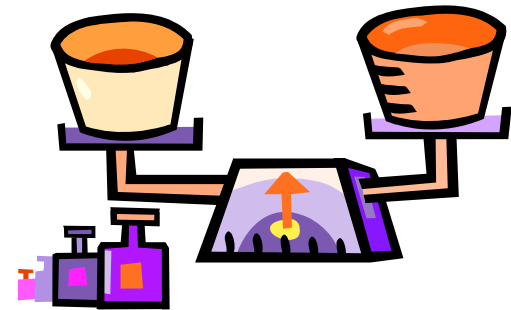
Level 4 Code Example

```
<dataTable>
...
<physical>
  <objectName>flslter.299.1</objectName>
  <size unit="bytes">59847</size>
  <dataFormat>
    <textFormat>
      <numHeaderLines>1</numHeaderLines>
      <attributeOrientation>column</attributeOrientation>
      <simpleDelimited>
        <fieldDelimiter>,</fieldDelimiter>
      </simpleDelimited>
    </textFormat>
  </dataFormat>
  <distribution>
    <online>
      <url>http://fls.unm.edu/flslter.296.1</url>
    </online>
  </distribution>
</physical>
...
```



Level 5 Integration

- **Description** – Level 4 content plus complete attribute and quality control details to support computer-assisted data integration and re-sampling; 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.
- **Major Elements Added :**
 - Attribute List (full descriptions)
 - Measurement Scale
 - Units
 - Constraint
 - Quality Control





Level 5 Code Example

```
...  
<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>  
...
```



Level 5 Code Example

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



Level 6 Semantic

- **Description** – Level 5 content plus semantic information (currently under development by SEEK, and may require extension to the EML schema)





Additional Recommendations

- packageID and Metacat document naming convention
Metacat and by extension the Metacat harvester rely on numerical data set ids and revision numbers for document management and synchronization - packageID attributes for EML contributed to the KNB Metacat should be formed as follows:

knb-lter-[site].[dataset number].[revision], e.g. knb-lter-sev.187.4
Scope UniqueID Revision#

- LDAP access control in Metacat
Metacat access control format conforms to the LDAP Distinguished Name concept:

<principal>uid=FLS,o=lter,dc=ecoinformatics,dc=org</principal>

- Organizational citation
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:

<organizationName>Sevilleta LTER</organizationName>



SEEK Acknowledgements

- This material is based upon work supported by:
- The National Science Foundation under Grant Numbers: 0129792 ,9980154, 0225676 and 0072909.
- Collaborators: University of New Mexico (Long Term Ecological Research Network Office), NCEAS (UC Santa Barbara), San Diego Supercomputer Center, University of Kansas (Center for Biodiversity Research), University of Virginia, University of California Berkeley (Hastings Biological Station), University of Wisconsin
- The Andrew W. Mellon Foundation



Addendum





eXtensible Markup Language

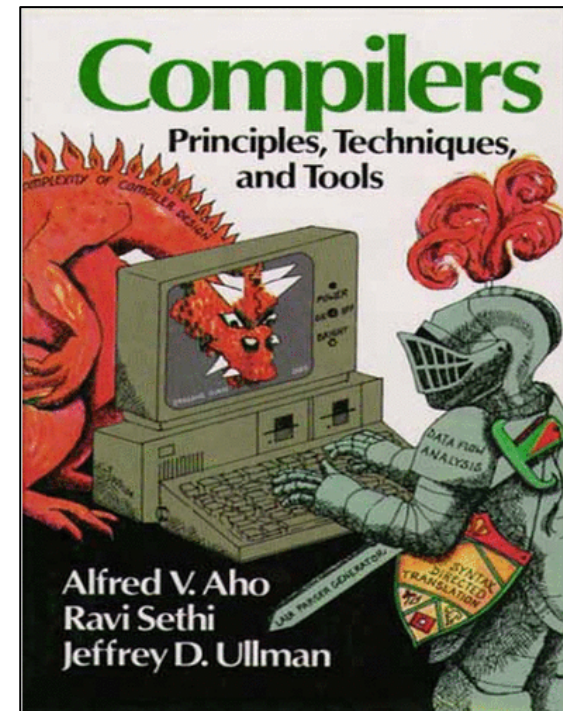
XML

- Development influenced by SGML and HTML – Version 1.0 in early 1998
- A semantic language that lets you more meaningfully annotate text (where HTML lets you define how text can be displayed, XML provides it with meaning).
- Support for presentation, exchange, and management of information (at the machine level)
- Tools include DTD, Schema, XSLT, and more...



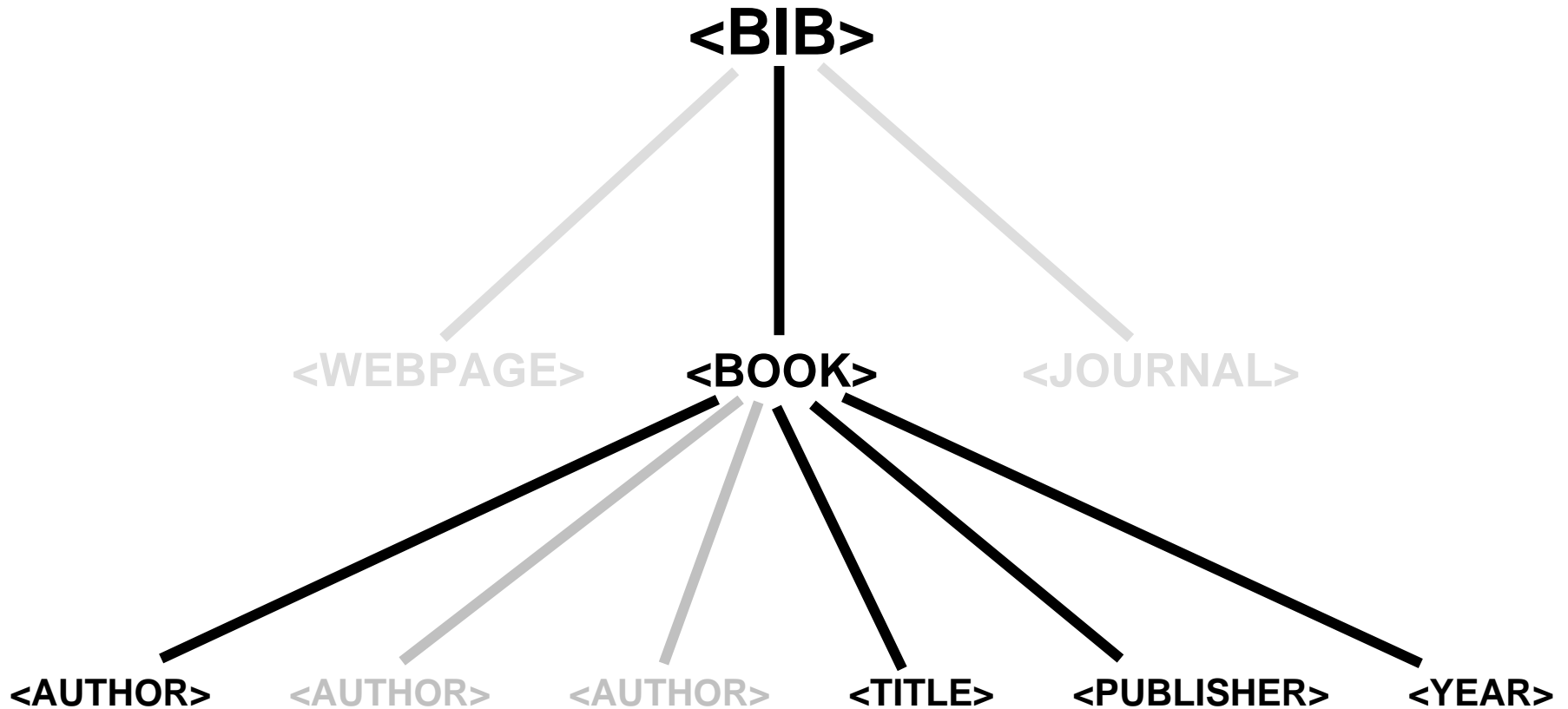
The Infamous Bibliography Example

Aho, A.V., Sethi, R., &
Ullman, J.D. (1985)
*Compilers: Principles,
Techniques, and
Tools*. Addison-
Wesley





XML is Hierarchical by Design





XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

```
<BIB>
```

```
  <BOOK nickname="Dragon book">
```

```
    <AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
    <AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
    <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
    <TITLE>
```

```
      Compilers: Principles, Techniques, and Tools
```

```
    </TITLE>
```

```
    <PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
    <YEAR> 1985 </YEAR>
```

```
  </BOOK>
```

```
</BIB>
```



XML Document Structure

<?xml version="1.0" standalone="yes">

<!-- This is an example bibliography. -->

Prolog

<BIB>

<BOOK nickname="Dragon book">

<AUTHOR id="aho"> Aho, A. V. </AUTHOR>

<AUTHOR id="sethi"> Sethi, R. </AUTHOR>

<AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>

<TITLE>

Compilers: Principles, Techniques, and Tools

</TITLE>

<PUBLISHER> Addison-Wesley </PUBLISHER>

<YEAR> 1985 </YEAR>

</BOOK>

</BIB>



XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

Comment

```
<BIB>
```

```
<BOOK nickname="Dragon book">
```

```
<AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
<AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
<AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
<TITLE>
```

```
    Compilers: Principles, Techniques, and Tools
```

```
</TITLE>
```

```
<PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
<YEAR> 1985 </YEAR>
```

```
</BOOK>
```

```
</BIB>
```



XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

<BIB>



Opening Tag
(Root Element)

```
<BOOK nickname="Dragon book">  
  <AUTHOR id="aho"> Aho, A. V. </AUTHOR>  
  <AUTHOR id="sethi"> Sethi, R. </AUTHOR>  
  <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>  
  <TITLE>  
    Compilers: Principles, Techniques, and Tools  
  </TITLE>  
  <PUBLISHER> Addison-Wesley </PUBLISHER>  
  <YEAR> 1985 </YEAR>  
</BOOK>
```

</BIB>



Closing Tag



XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

```
<BIB>
```

```
  <BOOK nickname="Dragon book">
```

```
    <AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
    <AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
    <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
    <TITLE>
```

```
      Compilers: Principles, Techniques, and Tools
```

```
    </TITLE>
```

```
    <PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
    <YEAR> 1985 </YEAR>
```

```
  </BOOK>
```

```
</BIB>
```

***Element
Content***





XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

```
<BIB>
```

```
  <BOOK nickname="Dragon book">
```

```
    <AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
    <AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
    <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
    <TITLE>
```

```
      Compilers: Principles, Techniques, and Tools
```

```
    </TITLE>
```

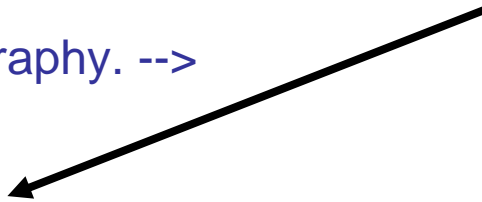
```
    <PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
    <YEAR> 1985 </YEAR>
```

```
  </BOOK>
```

```
</BIB>
```

Element
Attribute





XML Document Structure

```
<?xml version="1.0" standalone="yes">
```

```
<!-- This is an example bibliography. -->
```

```
<BIB>
```

```
  <BOOK nickname="Dragon book">
```

```
    <AUTHOR id="aho"> Aho, A. V. </AUTHOR>
```

```
    <AUTHOR id="sethi"> Sethi, R. </AUTHOR>
```

```
    <AUTHOR id="ullman"> Ullman, J. D. </AUTHOR>
```

```
    <TITLE>
```

```
      Compilers: Principles, Techniques, and Tools
```

```
    </TITLE>
```

```
    <PUBLISHER> Addison-Wesley </PUBLISHER>
```

```
    <YEAR> 1985 </YEAR>
```

```
  </BOOK>
```

```
</BIB>
```

***Complex
Element***



XML DTD and Schema

Defines the structure, content and to some extent, the semantics of XML documents

