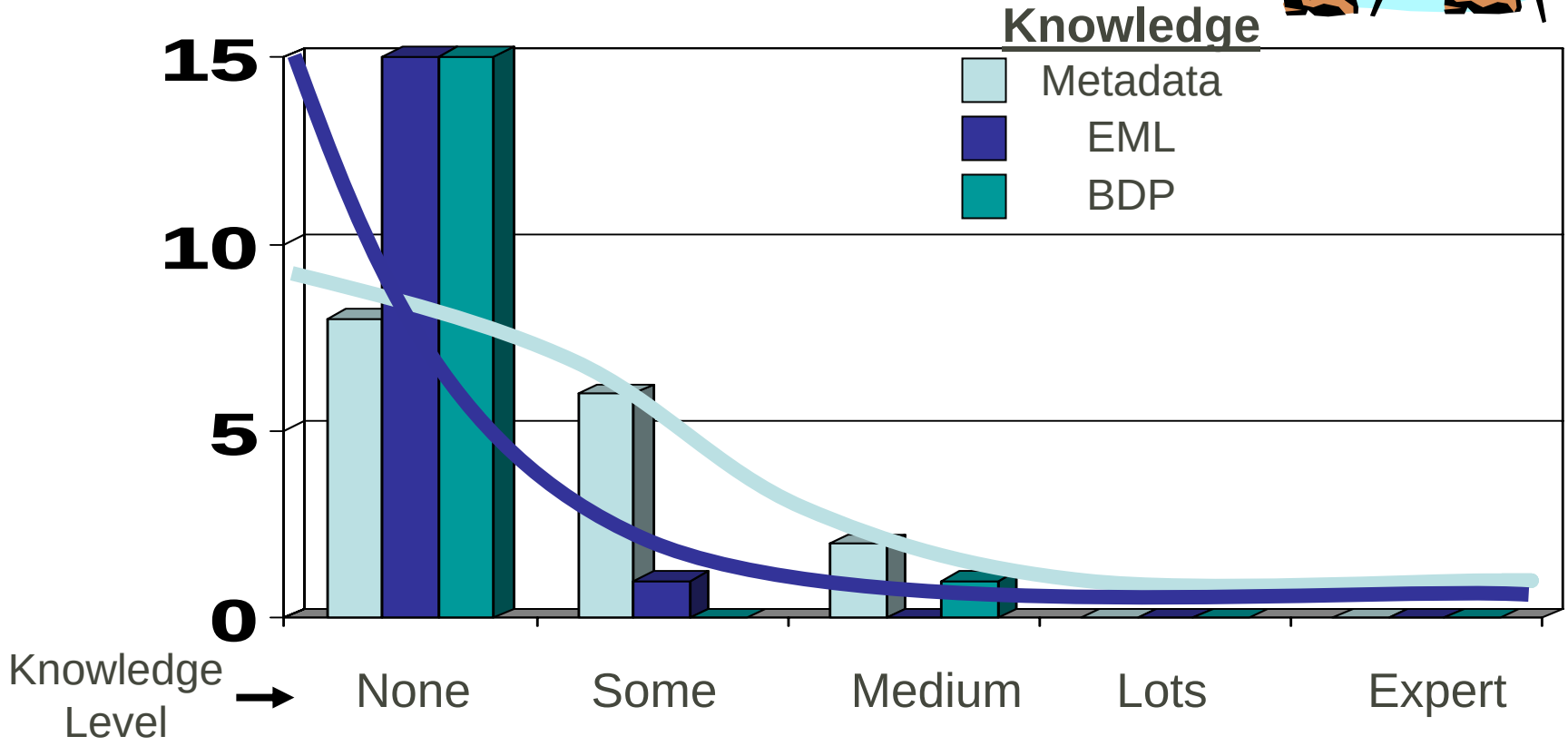
Enabling metadata discovery

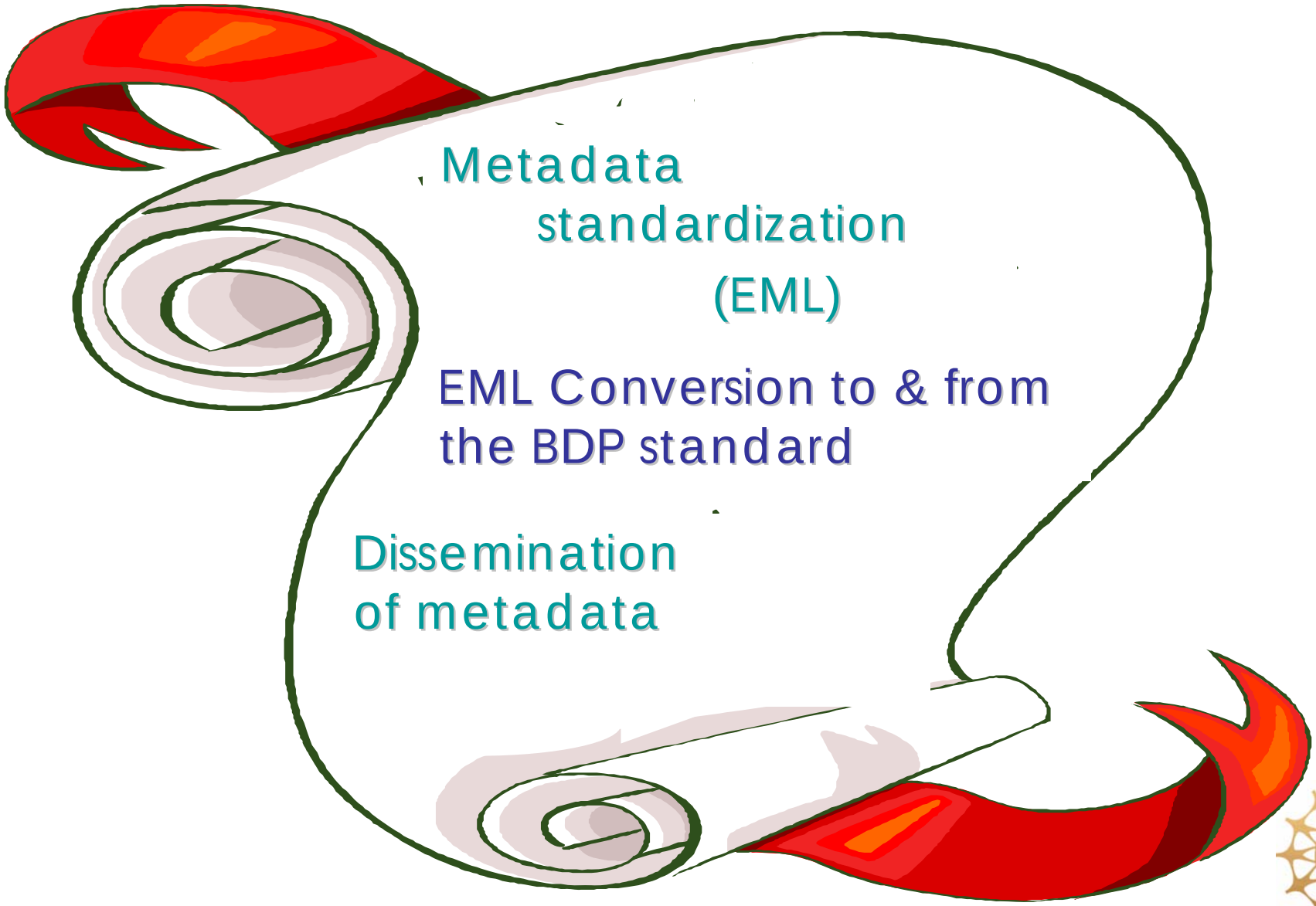
Practical metadata standardization tricks

Iñigo San Gil
LTER Network office
isangil@lternet.edu

How about YOU?



Talk Outline



Why enable metadata discovery?

Expand research

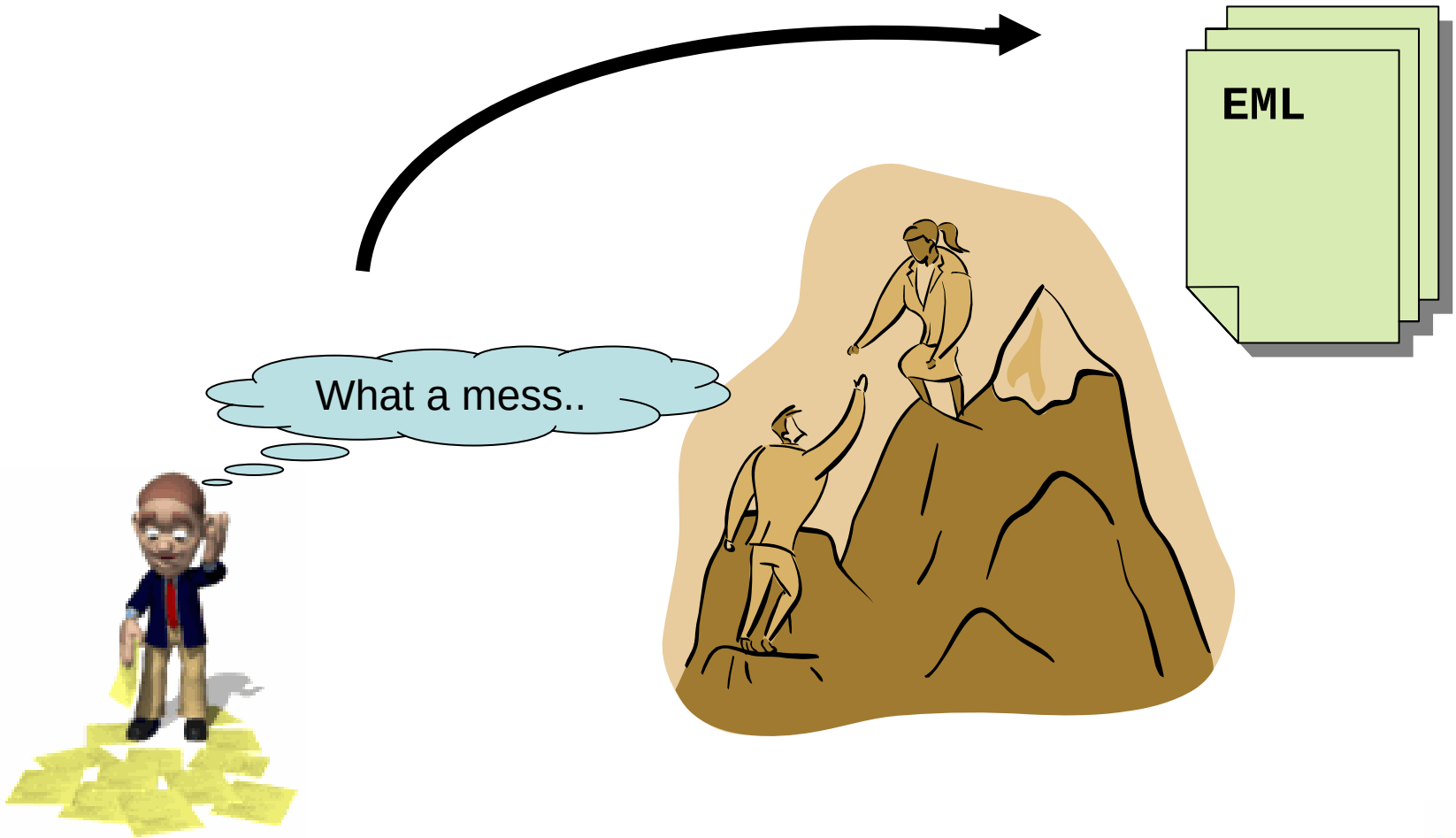
Forge alliances

Easier management

Common standard

Interoperability

Goal: standardize metadata



Metadata standardization topics:

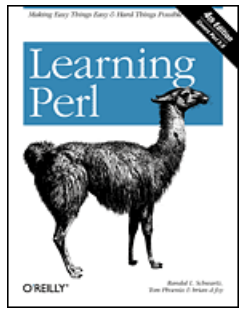
- How much, how rich and how is data stored
  - Just a few? Consider Morpho or an XML editor
  - A bunch? Databases: MS Access, SQLs, Paradox...
  - Flat text files: Structured Folders (or not)
- The LTER case. 26 sites, 26 forms of metadata
- Case study: transforming legacy metadata

Standardization. Questions about your metadata

How wide and deep is your metadata ? do we have it in a database? do we have flat text files?

Morpho  Morpho Data Management for Ecologists Next talk –Will Tyburczy

XML editor <oXygen/>® xmlspy



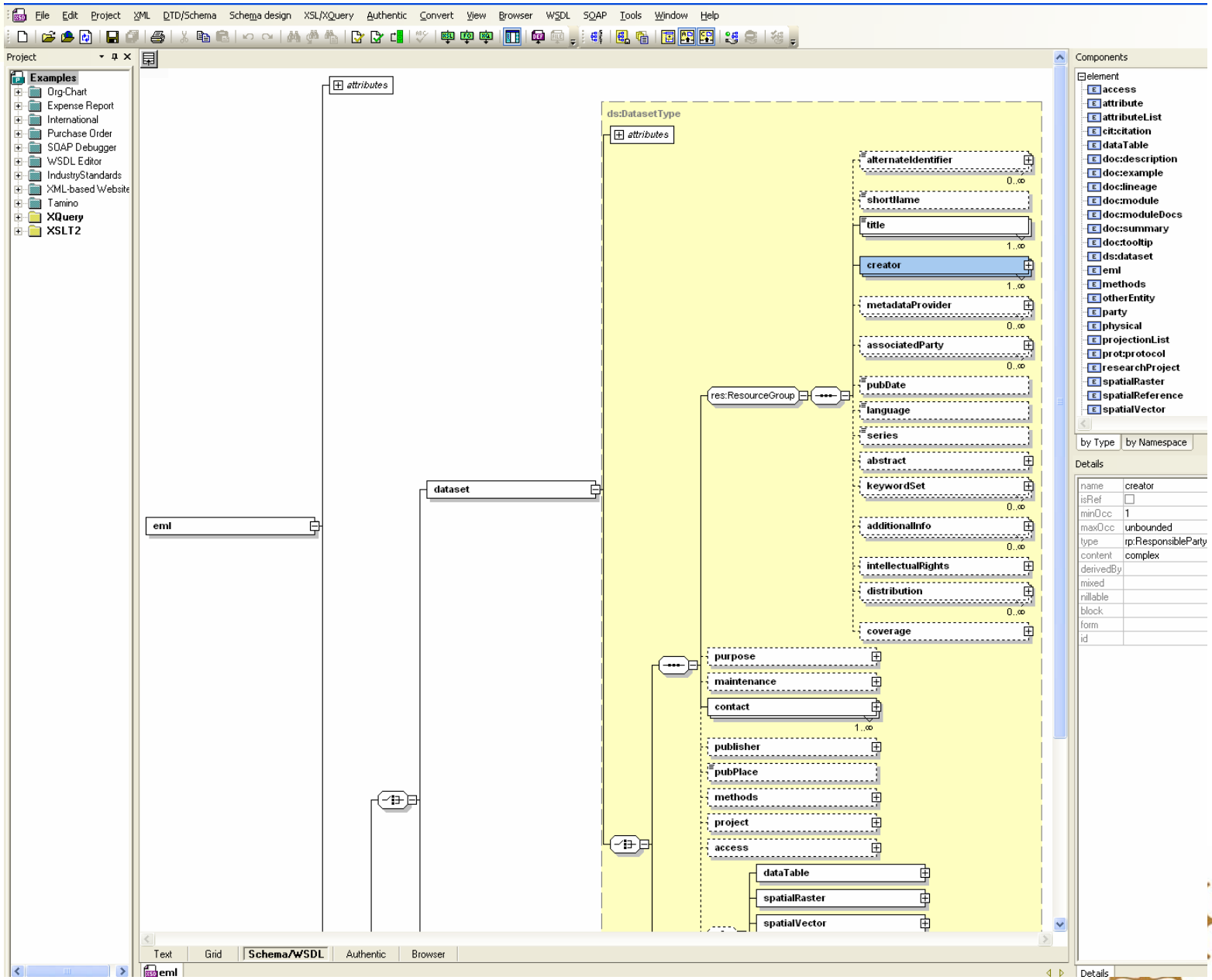
Perl (*Practical Extraction and Report Language*)

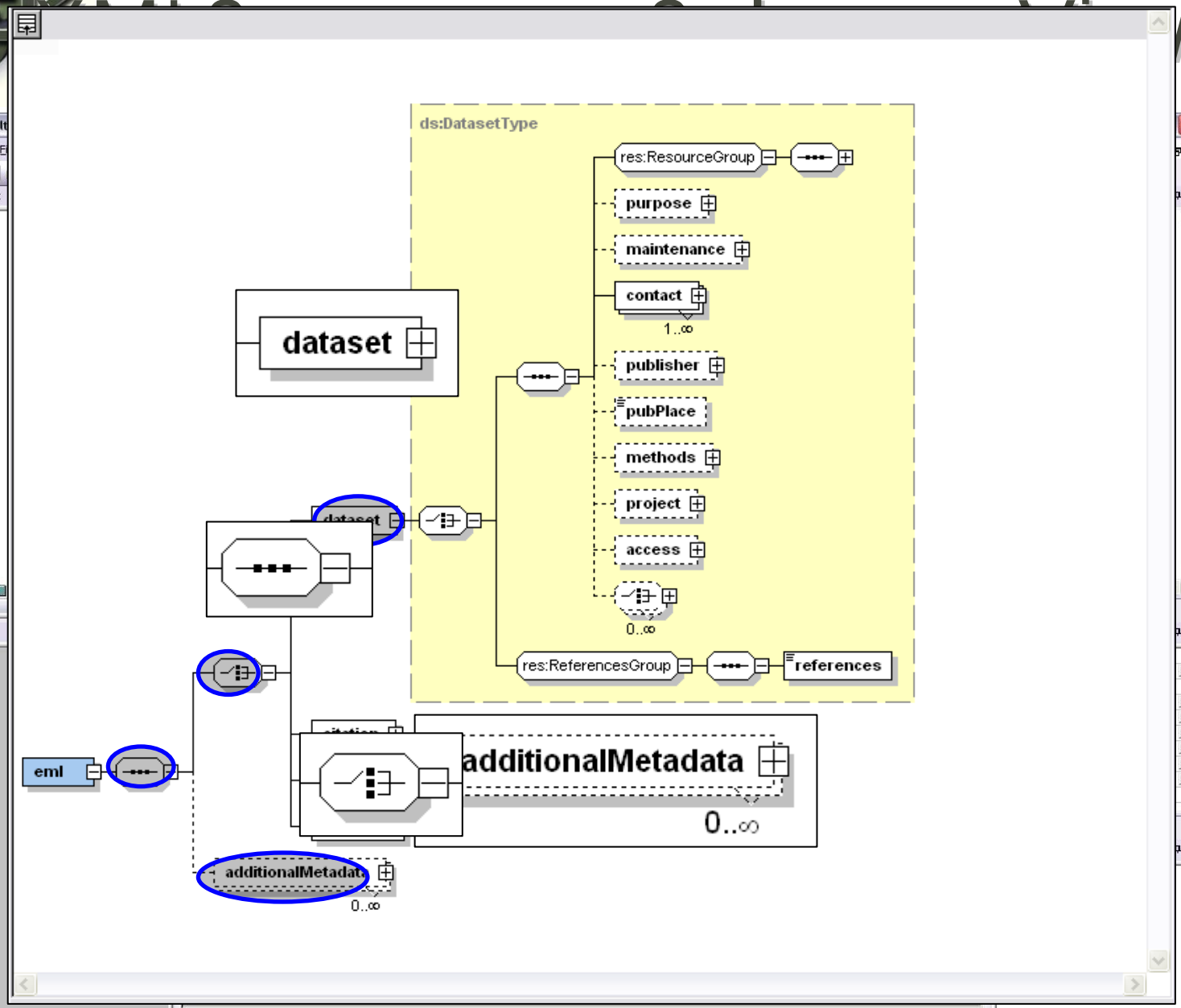
Other tools … *sorry, not time for the showcase here!*

Metadata Tool Matrix

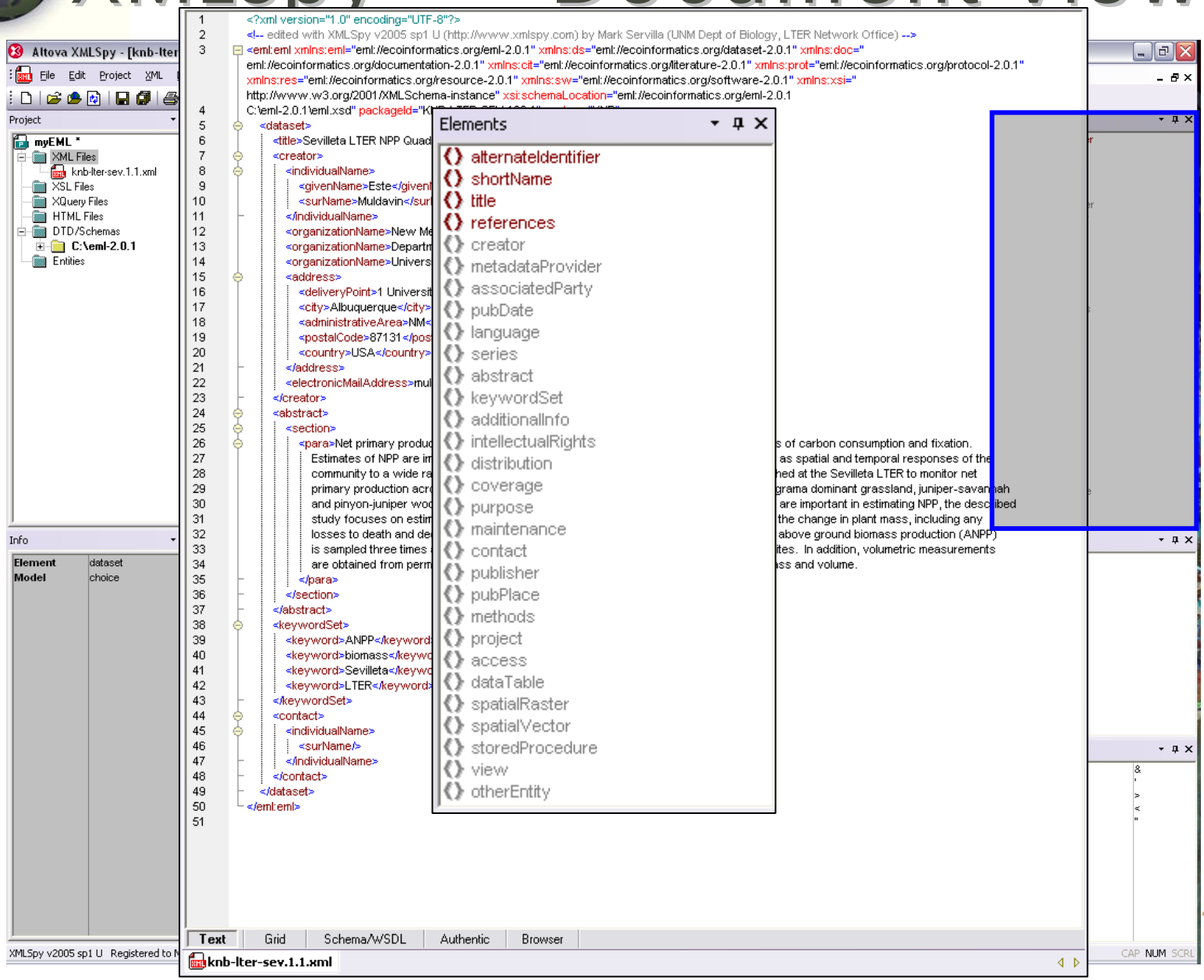
	Open Source Proprietary	On-line Stand Alone	Platform
XML Spy	Home ed - $0 Professional ~ $1000 Enterprise ~ $3000	Stand Alone	Windows
Morpho	Free	Stand Alone On-line*	Any
oXygen	Academic $50 Professional $200	Stand Alone	Any
Notepad, Word	Windows	Stand Alone	Windows
Emacs, vi, …	Free	Stand Alone	Unix
Databases	SQLs { $0 - $$$} Oracle {$50* - $1000+} MSAccess	Stand Alone On-line	Mixed
NBII tools	Free	Stand Alone On-line	Mixed

XML Editors -- XMLSpy

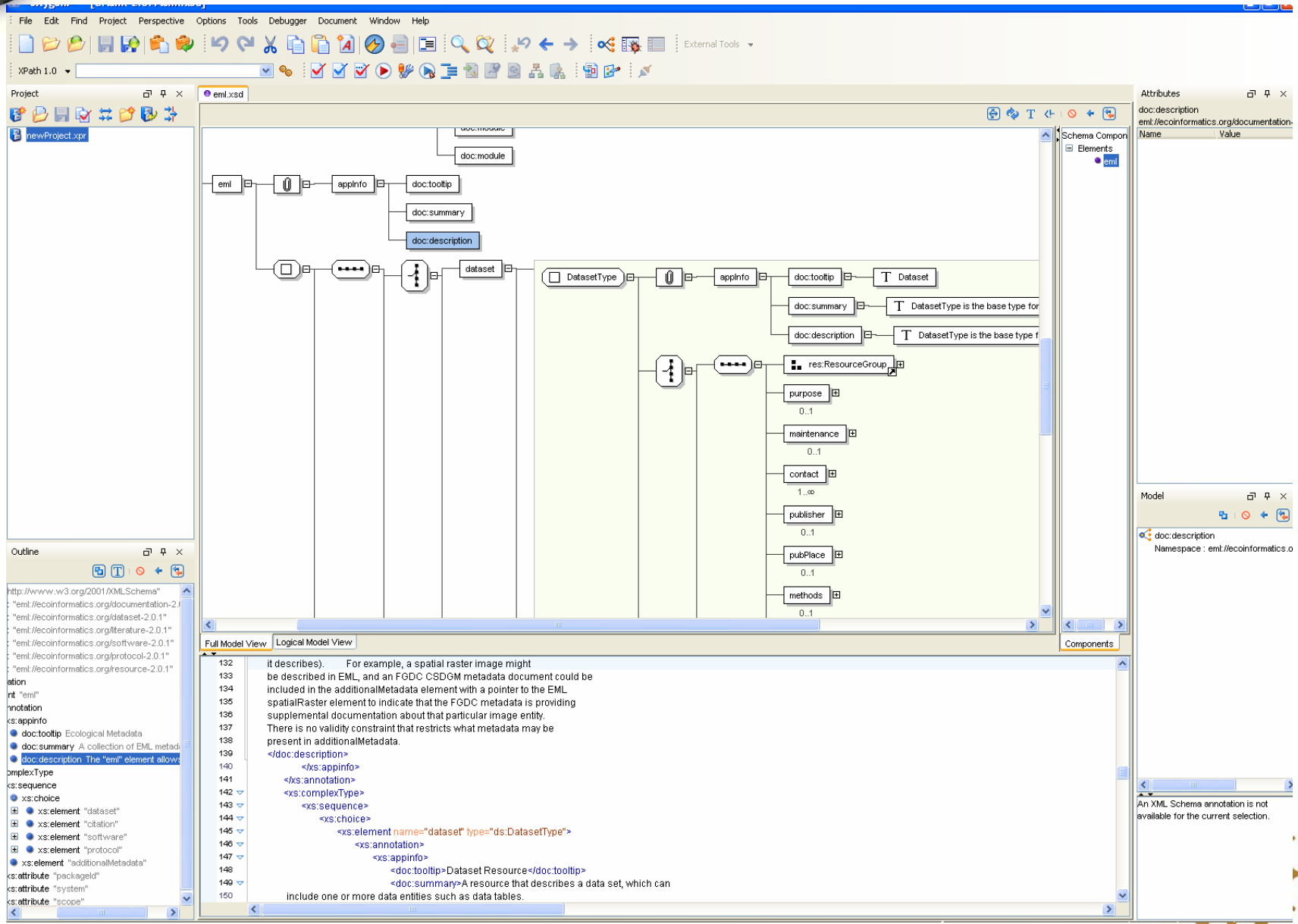



ds:DatasetType
res:ResourceGroup
purpose
maintenance
contact
1..∞
publisher
pubPlace
methods
project
access
0..∞
res:ReferencesGroup
references
dataset
eml
additionalMetadata
0..∞
Text
Grid
Schema/WSDL
Authentic
Browser
knb-lter-sev.1.1.xml
eml.xsd
Facets
XMLSpy v2005 sp1 U Registered to Mark Servilla (UNM Dept of Biology, LTER Network Office) ©1998-2005 Altova GmbH & Altova, Inc.
Ln 42, Col 41
CAP NUM SCRL

XMLSpy -- Document View



XML Editors – oXygen



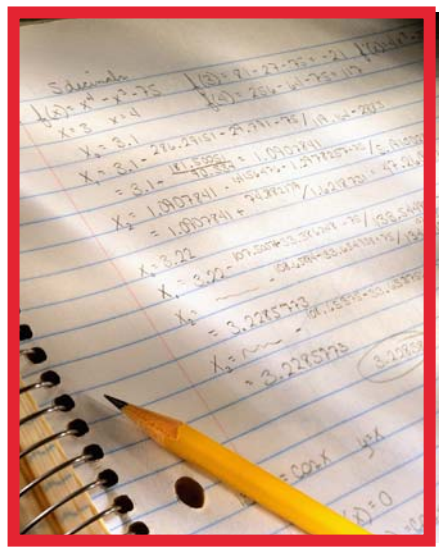
Data management models

Databases:

MySQL, MSAccess, Paradox, Metacat, etc

Spreadsheets:

Excel, etc

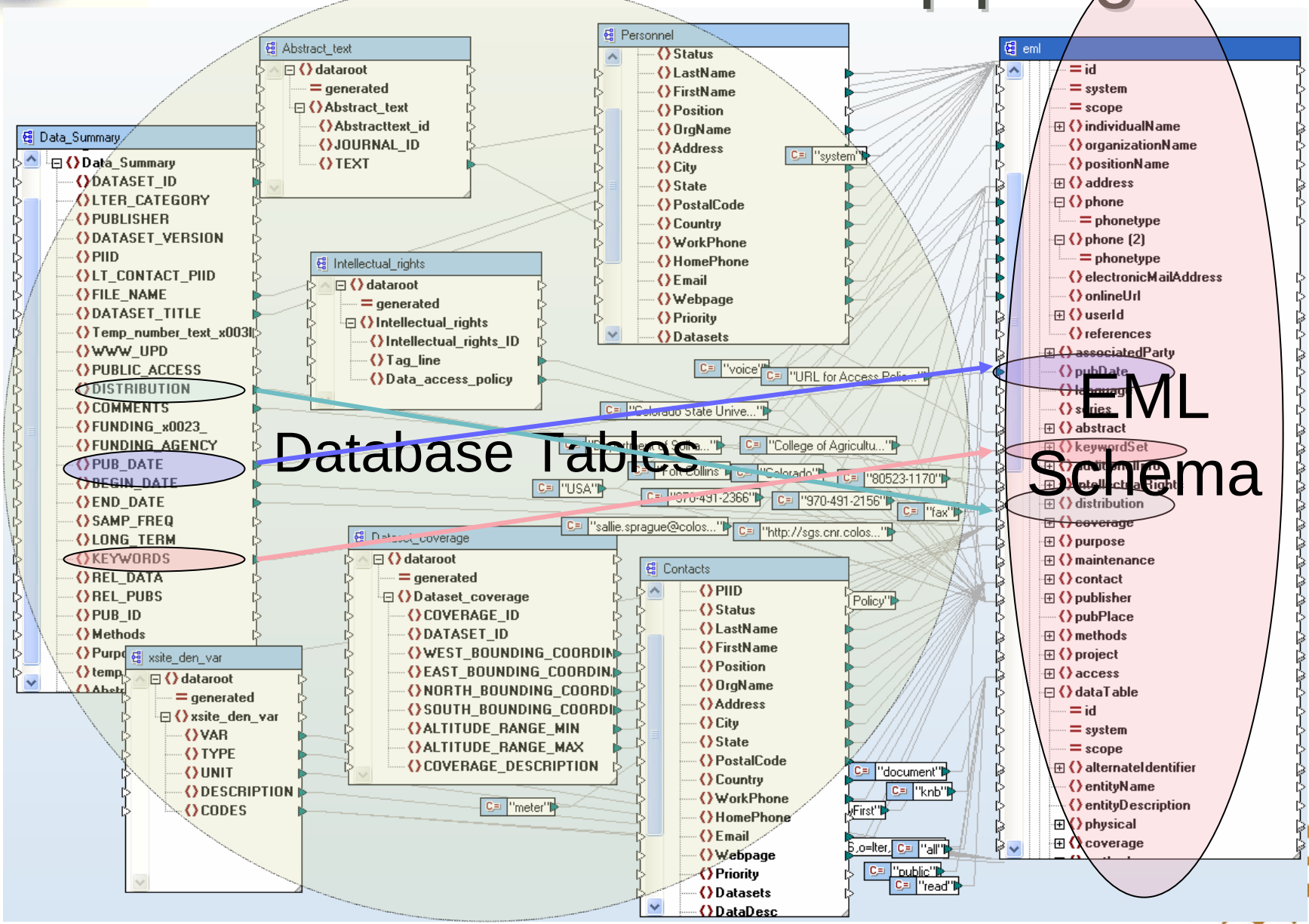


Text files:

Flat ASCII, Word documents.

Combos of the above

Database to EML mapping



Text to EML ma

Useful when text sources al

ARCTIC LTER DATABASE

(1) FILE NAME: 2004e5chlr.txt

(2) YEAR: 2004

(3) PI: Dr. Anne Giblin

(4) OTHERS: Chris Crockett, Lori Winters, Sam Kelse

(5) BRIEF DESCRIPTION OF DATA FILE: Corrected Chlorophyll a for lake E5 during the summer of 2004

(6) KEYWORDS: Chlorophyll a

(7) SITE TYPE: AQUATICS-LAKES

(8) RESEARCH LOCATION: Toolik Field Station, North Slope, Alaska.

(9) EXPERIMENTAL DESIGN AND METHODS:
Water samples are collected at a minimum of the epi, meta, and hypolimnion, samples at other depths are collected as needed dependent on lake

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2
C:\eml-2.0.1\eml.xsd" packageId="" system="">
  <dataset>
    <title> </title>
    <creator>
      <individualName>
        <surName/>
      </individualName>
    </contact>
  </dataset>
</eml:eml>
```

```
<title> </title>
<creator>
  <individualName>
<abstract>
  <section>
    <para>Corrected Chlorophyll a
    for lake E5 during the summer
    of 2004</para>
  </section>
</abstract>
```

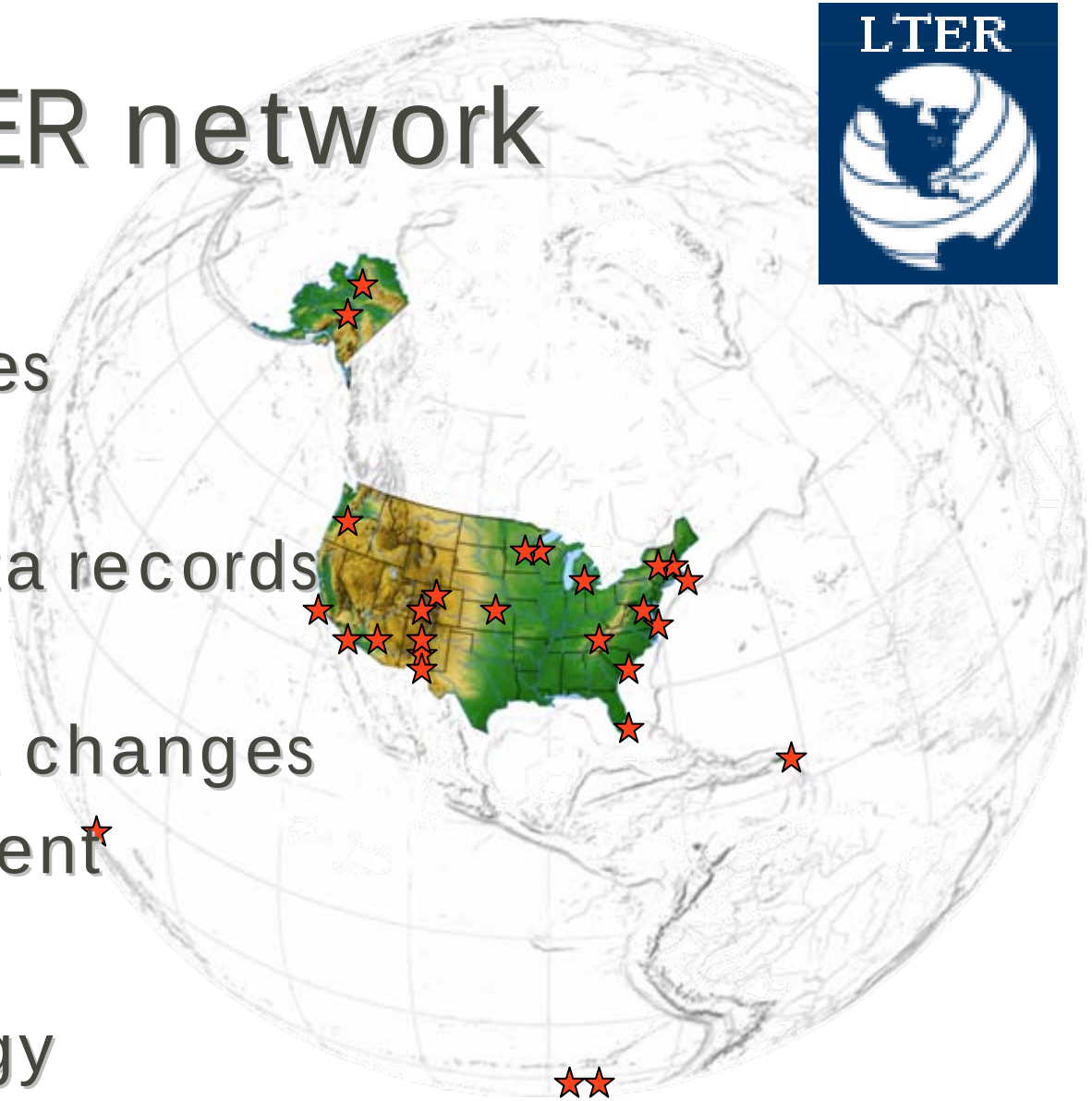
The LTER network

26 autonomous sites

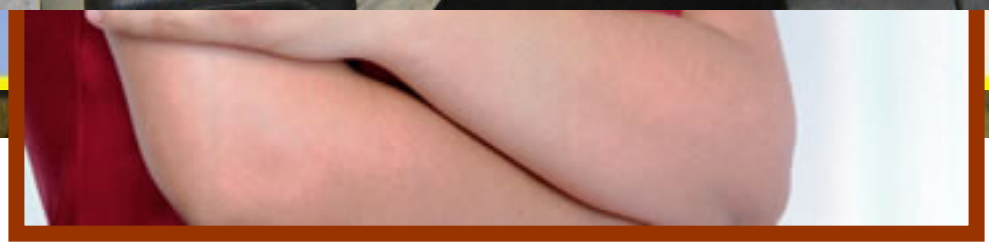
20 yr+ worth of data records

Relatively frequent changes in data management

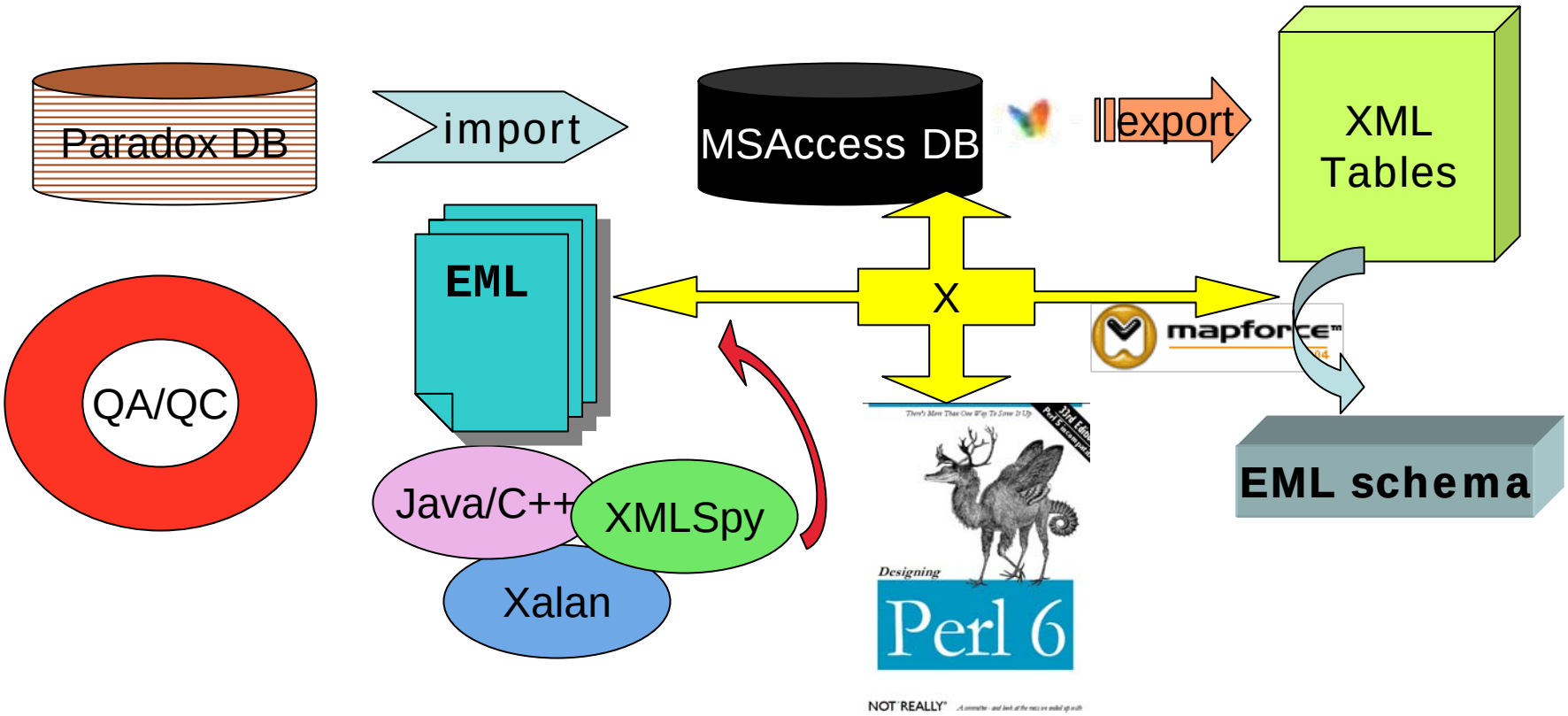
Evolving technology



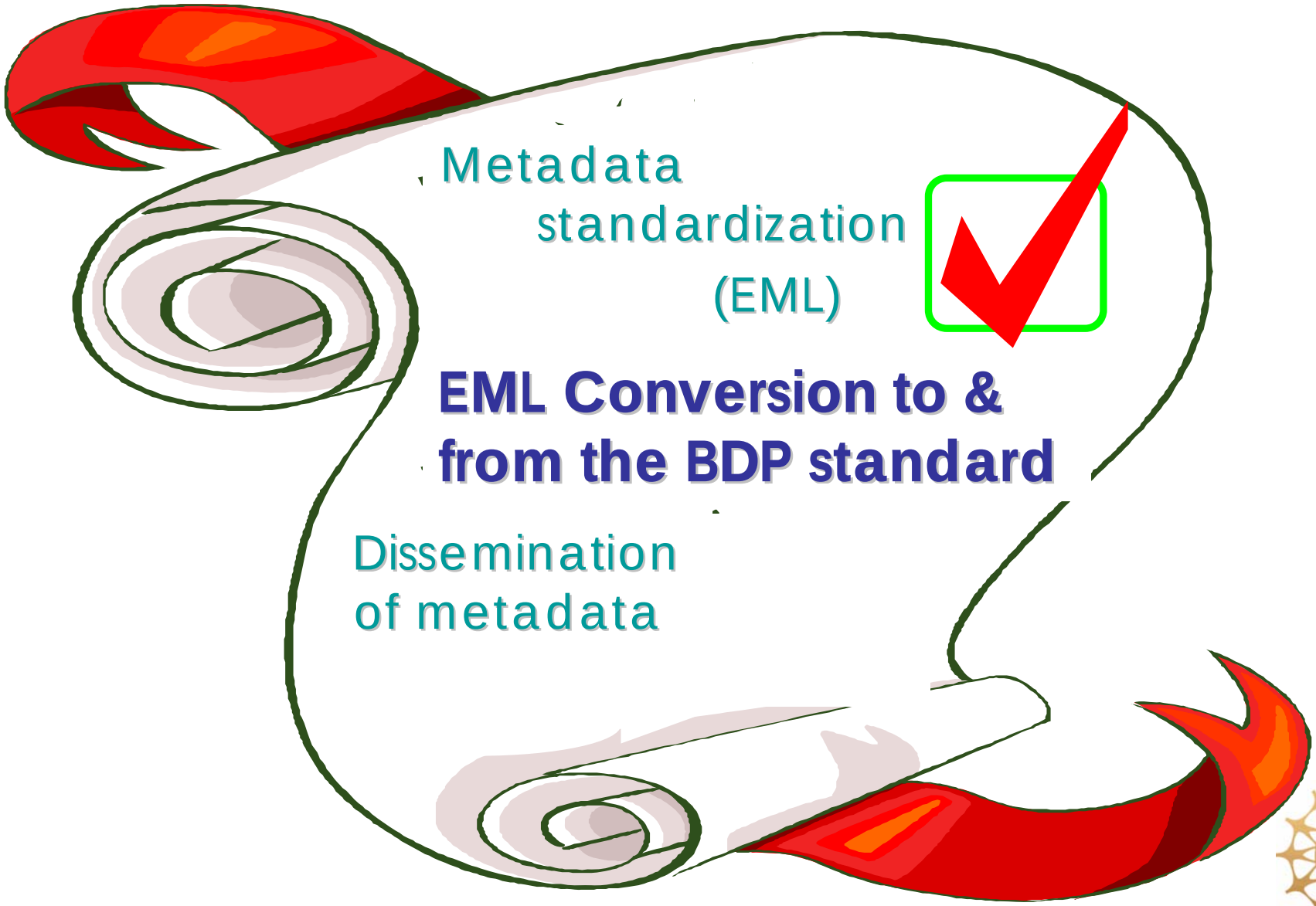
Case study – Luquillo



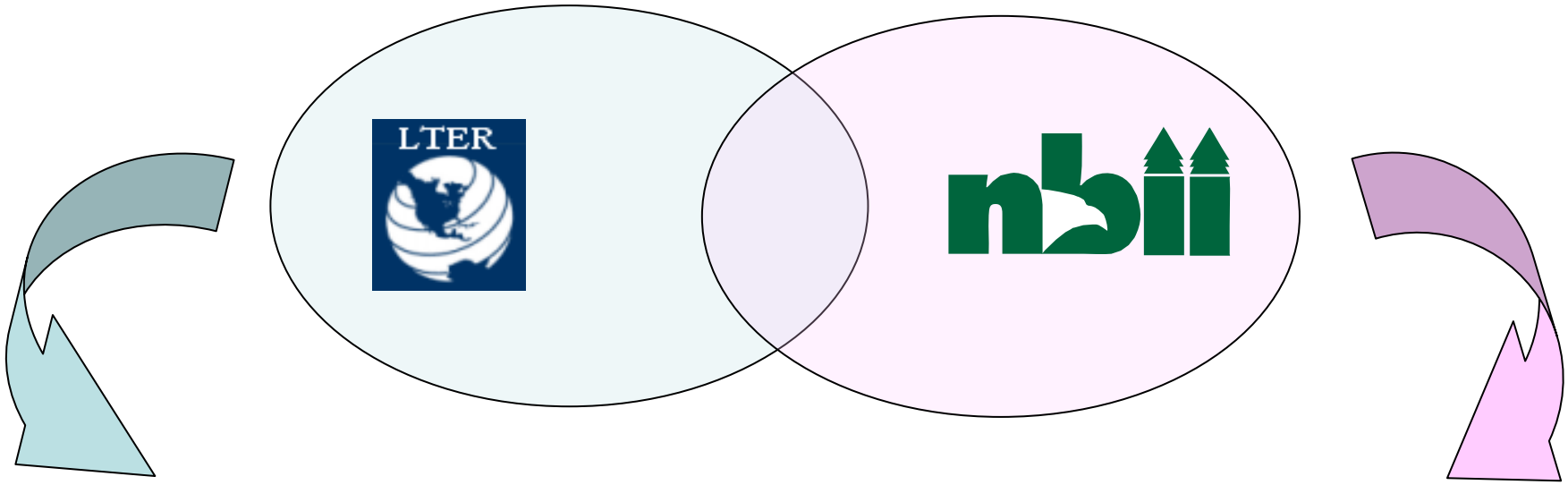
Case study – Luquillo



Talk Outline



Metadata Crosswalks



EML
Ecological Metadata Language

Metadata Crosswalk

BDP
Biological Data Profile

Metadata Crosswalk

Yeah, well, but HOW??

Web application at

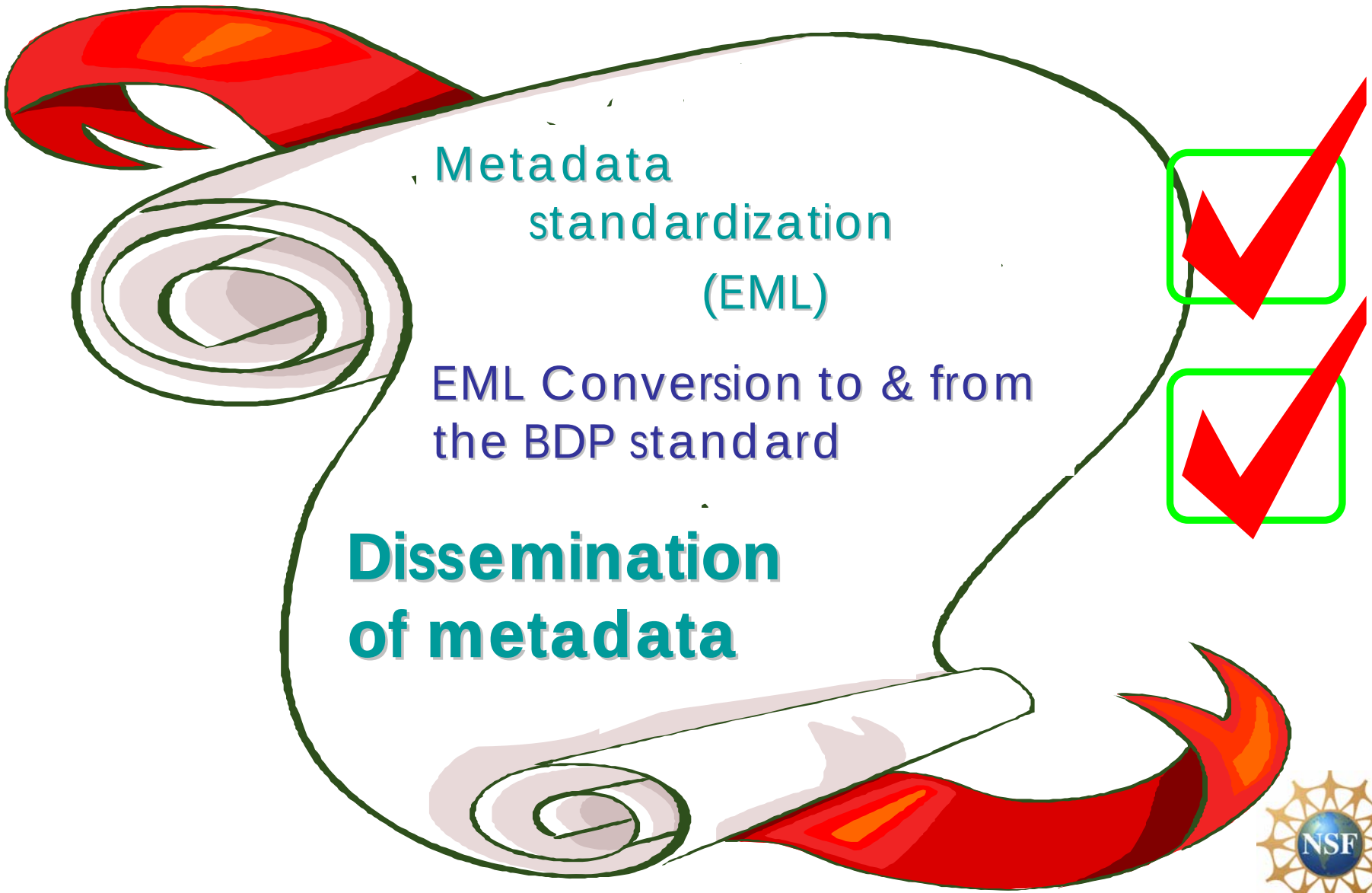
http://fire.lternet.edu/~isangil/Web_translation_application/EmlToBdp/eml2bdp.php

XSLT code available for free !

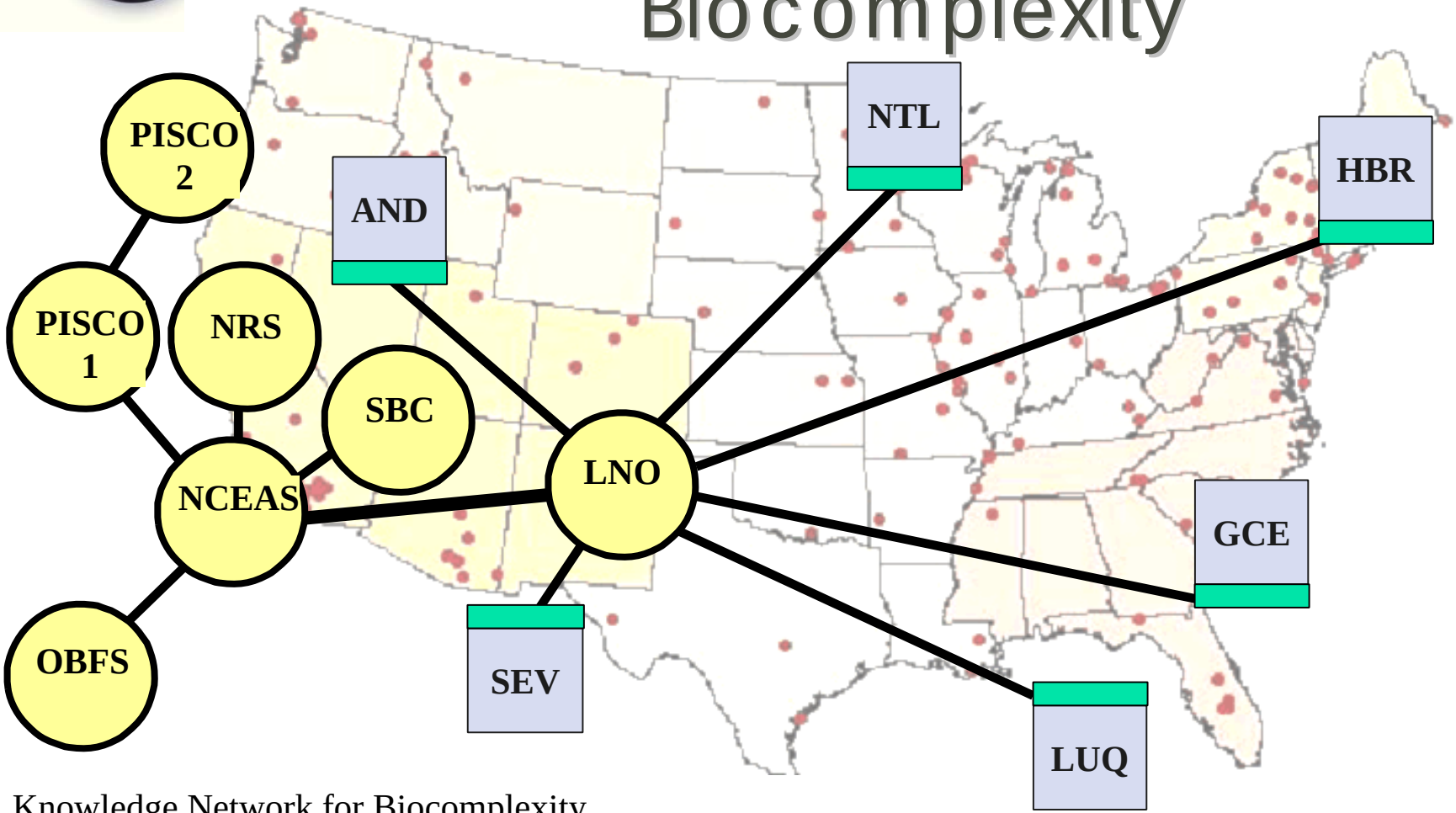
Web service to be released soon*

*Contact isangil@lternet.edu

Talk Outline



Knowledge Network for Biocomplexity



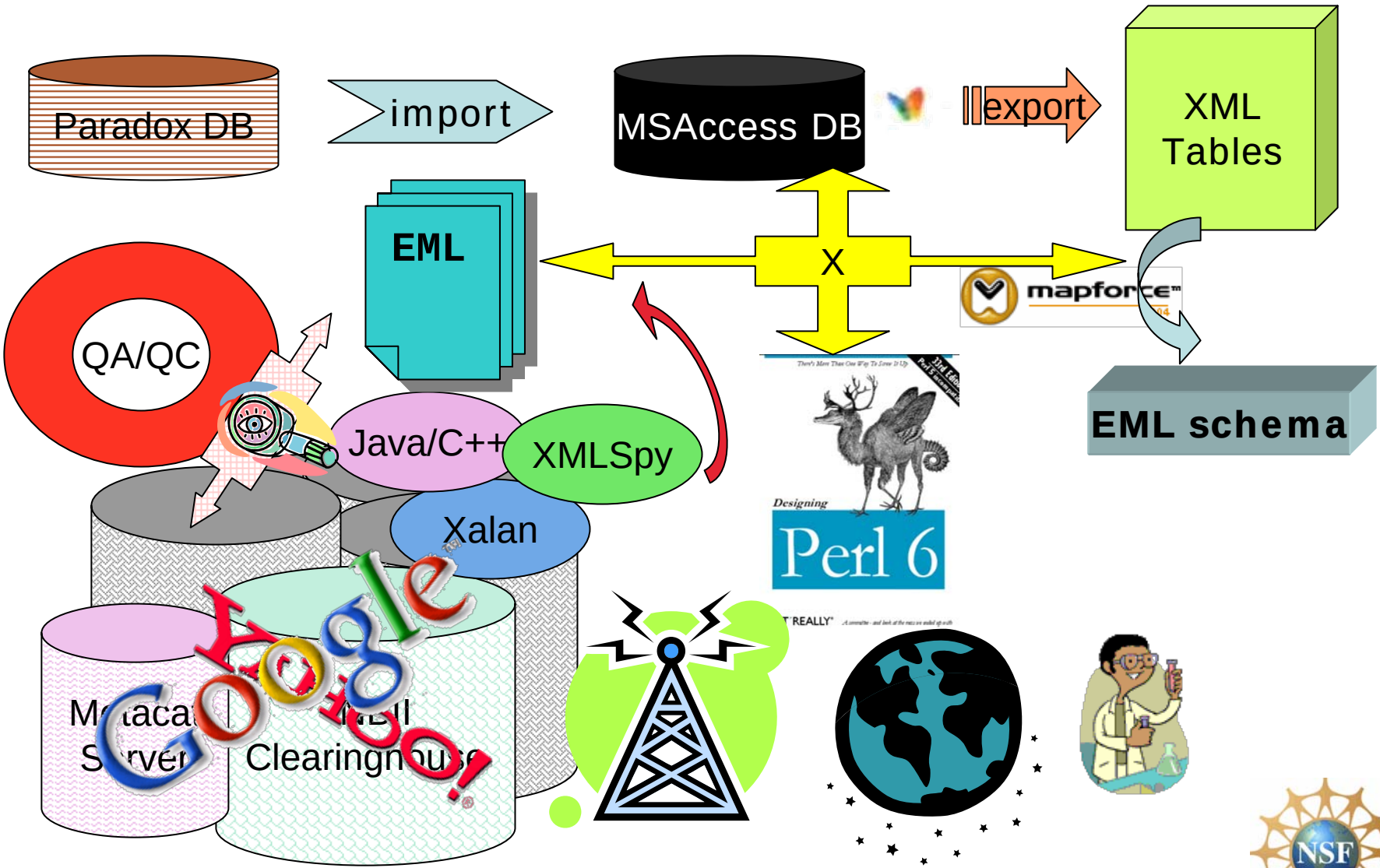
Knowledge Network for Biocomplexity

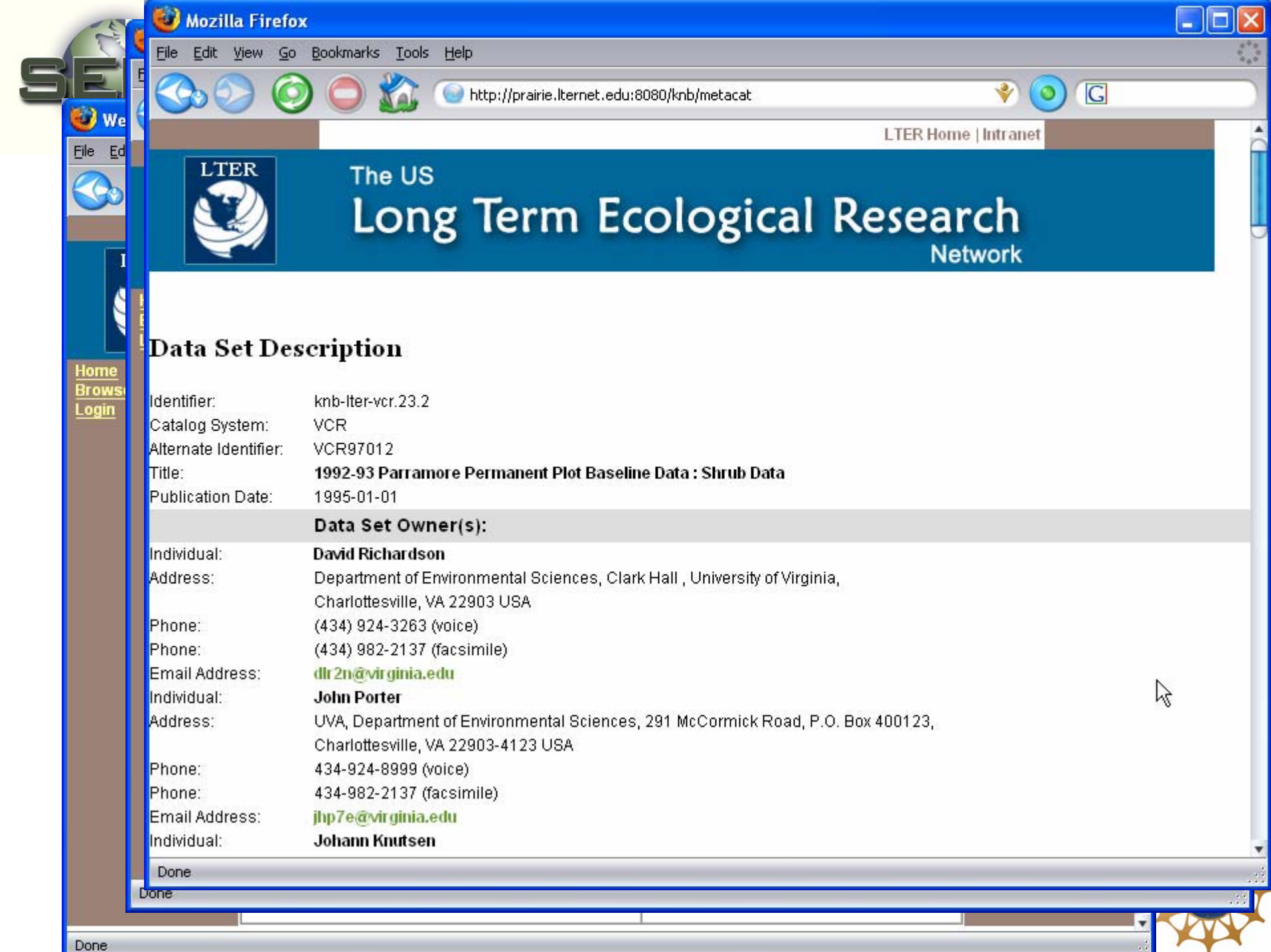
Metacat node

Site-specific node

LTER Network (27)
Organization of Biological Field Stations (180+)
UC Natural Reserve System (36)
Partnership for Interdisciplinary Studies of Coastal Oceans (4)
Multi-agency Rocky Intertidal Network (60)

Case study – Luquillo

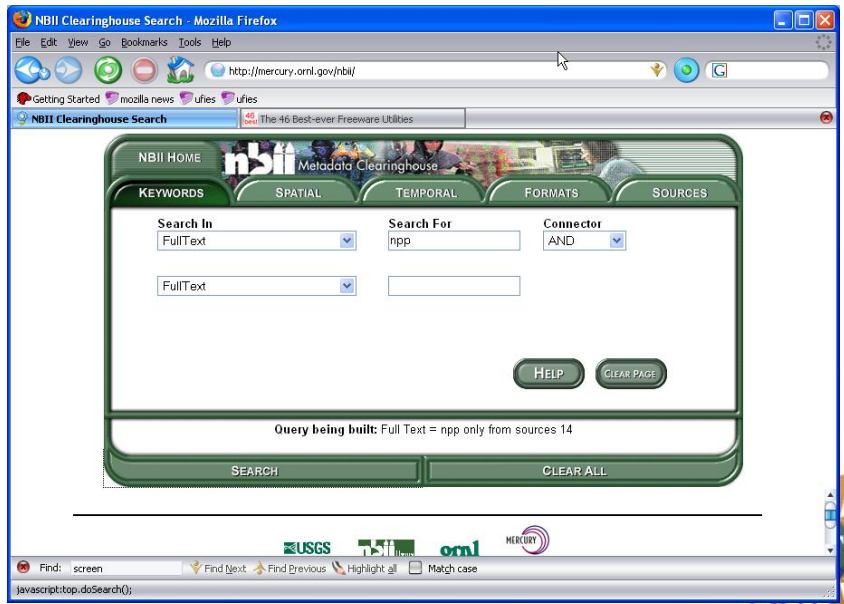



Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://prairie.lternet.edu:8080/knb/metacat
LTER Home | Intranet
LTER
The US
Long Term Ecological Research
Network
Data Set Description
Identifier: knb-lter-vcr.23.2
Catalog System: VCR
Alternate Identifier: VCR97012
Title: 1992-93 Parramore Permanent Plot Baseline Data : Shrub Data
Publication Date: 1995-01-01
Data Set Owner(s):
Individual: David Richardson
Address: Department of Environmental Sciences, Clark Hall , University of Virginia,
Charlottesville, VA 22903 USA
Phone: (434) 924-3263 (voice)
Phone: (434) 982-2137 (facsimile)
Email Address: dlr2n@virginia.edu
Individual: John Porter
Address: UVA, Department of Environmental Sciences, 291 McCormick Road, P.O. Box 400123,
Charlottesville, VA 22903-4123 USA
Phone: 434-924-8999 (voice)
Phone: 434-982-2137 (facsimile)
Email Address: jhp7e@virginia.edu
Individual: Johann Knutsen
Done
Done
Done

Metacat web user interface

Metadata Discovery: NBII

- URL: http://mercury.ornl.gov/nbii
- Searches can be as generic as you want, or really specific.
- 28 nodes, including the LTER node
- Metadata only



Links, people contacts

www.altova.com (XML Spy editor, MapForce)
www.oxygenxml.com (XML editor)
www.activestate.com (Free Perl runtime compiler)

Email addresses:
isangil@lternet.edu Inigo San Gil
servilla@lternet.edu Mark Servilla

http://Seek.ecoinformatics.org : these presentations!