Ecological Metadata Language (EML) and Morpho

Matthew B. Jones
jones@nceas.ucsb.edu

National Center for Ecological Analysis and Synthesis
University of California Santa Barbara

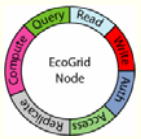
http://knb.ecoinformatics.org

KNB and SEEK Products

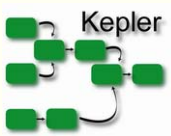
- Ecological Metadata Language (EML)

- Morpho
  - metadata and data management software

- Metacat
  - distributed metadata/data system
  - registries: KNB, UCNRS, OBFS, NCEAS, PISCO, LTSS



- EcoGrid
  - integrating distinct data systems and networks



- Kepler
  - grid-enabled scientific workflows

Ecological Metadata Language

- Metadata: a means to manage ecological data
  - Metadata is not an end, it's a means
- Many existing metadata standards
  - NBII Biological Data Profile (BDP), Dublin Core Element Set, ISO Geographic Metadata, GCMD Directory Interchange Format (DIF)
- EML initiative started in 1997
  - Common language for describing, archiving, and transporting data

EML In brief...

- Ecological Metadata Language
  - What is it? Documentation about data, aka metadata
  - Which data? Ecologically relevant data
    - Biodiversity surveys, hydrology, atmospheric chemistry, spatial data, behavioral experiments, ...
    - But really, any scientific data

Why do we need metadata?

- Ecologically relevant data are:
  - Heterogeneous
  - Dispersed
  - From many disciplines

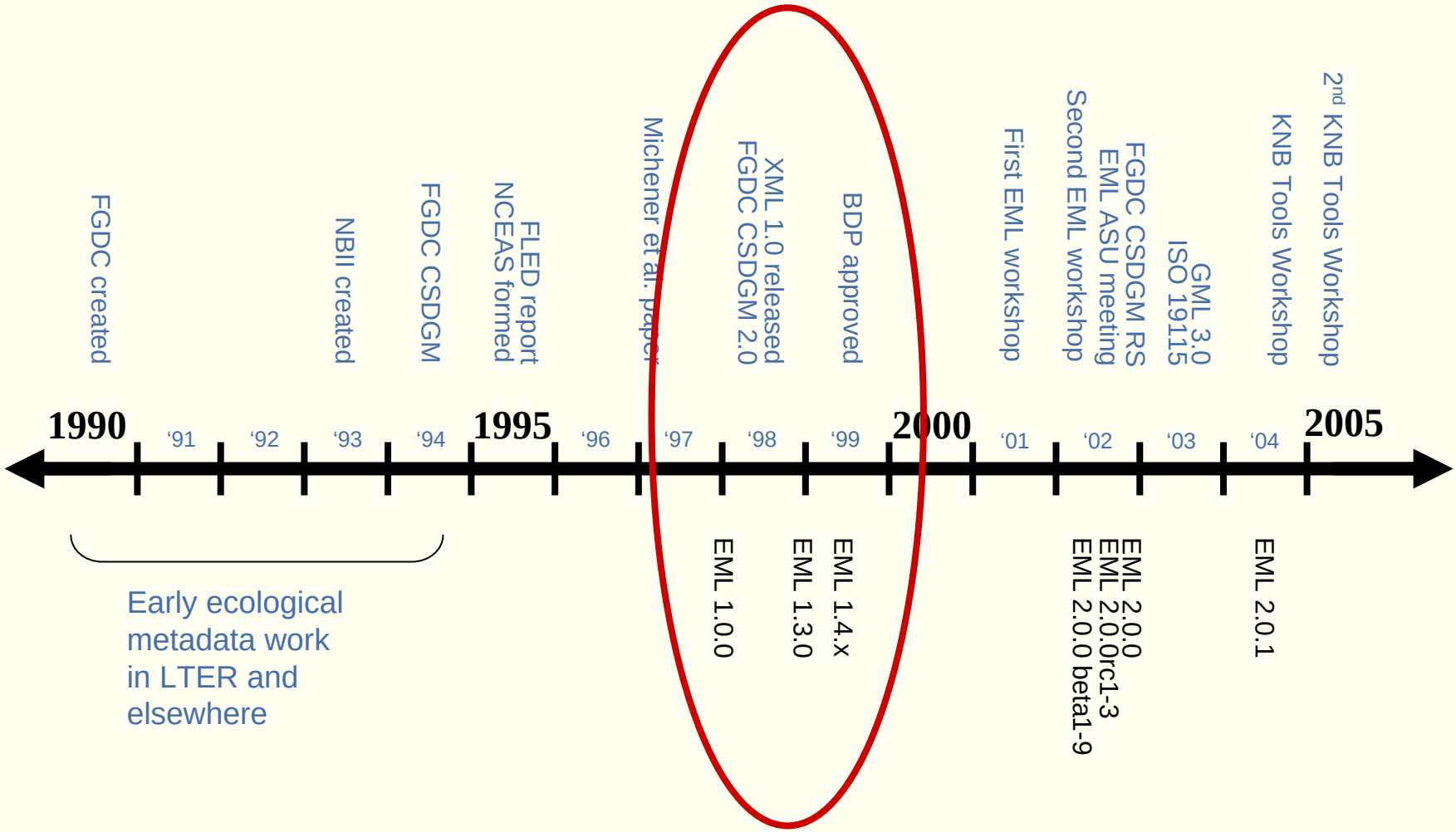
- Metadata: a means to manage ecological data
  - Accommodate heterogeneity and dispersion
    - There is no universal data model for ecology

- Data discovery
- Data interpretation
- Enable advanced analytical applications

Related metadata standards

- Dublin Core Element Set
  - Corresponds roughly to eml-resource
- Content Standard for Digital Geospatial Metadata (CSDGM)
  - Federal Geographic Data Committee (FGDC)
  - Corresponds to eml-spatialRaster, eml-spatialVector,eml-spatialReference
  - Overlaps in other modules (eml-resource)
- Biological Data Profile (BDP) of the CSDGM
  - Biological Data Working Group of the FGDC
  - Shares structure for taxonomicCoverage, geologicAge, and ascii table structures
- ISO 19115 Geographic information: Metadata
  - Incorporated in the eml-spatial* modules
  - Eml-party derived from ISO 19115
- Darwin Core
  - Partially overlaps with eml-coverage
- Geography Markup Language (GML)

Abridged History of EML



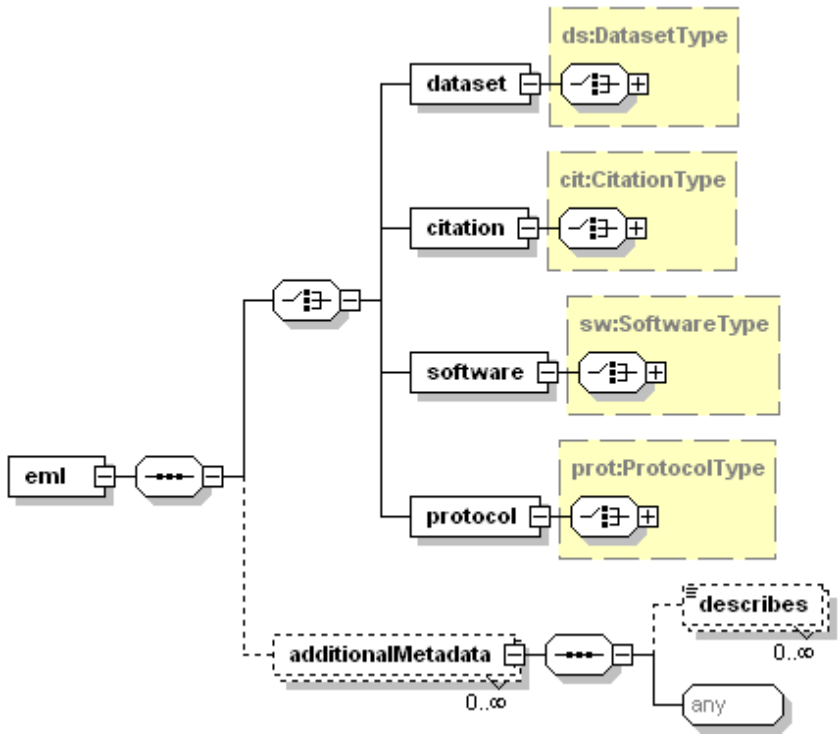
Conversion among standards

- Extensive overlap among all of these standards
- EML represents a superset of:
  - BDP
  - CSDGM
  - ISO 19115 (I think)
  - Dublin Core
- Can convert from EML to those
- Practically, have conversion script for:
  - EML → BDP
    - Uses XSLT so can be used in a variety of software

What does EML document?

- Scientific data sets
  - Tabular and relational data
  - Spatial images and GIS data
- Processing Software
- Literature citations
- Scientific Protocols

- Easily extensible
  - AdditionalMetadata



Information covered by EML

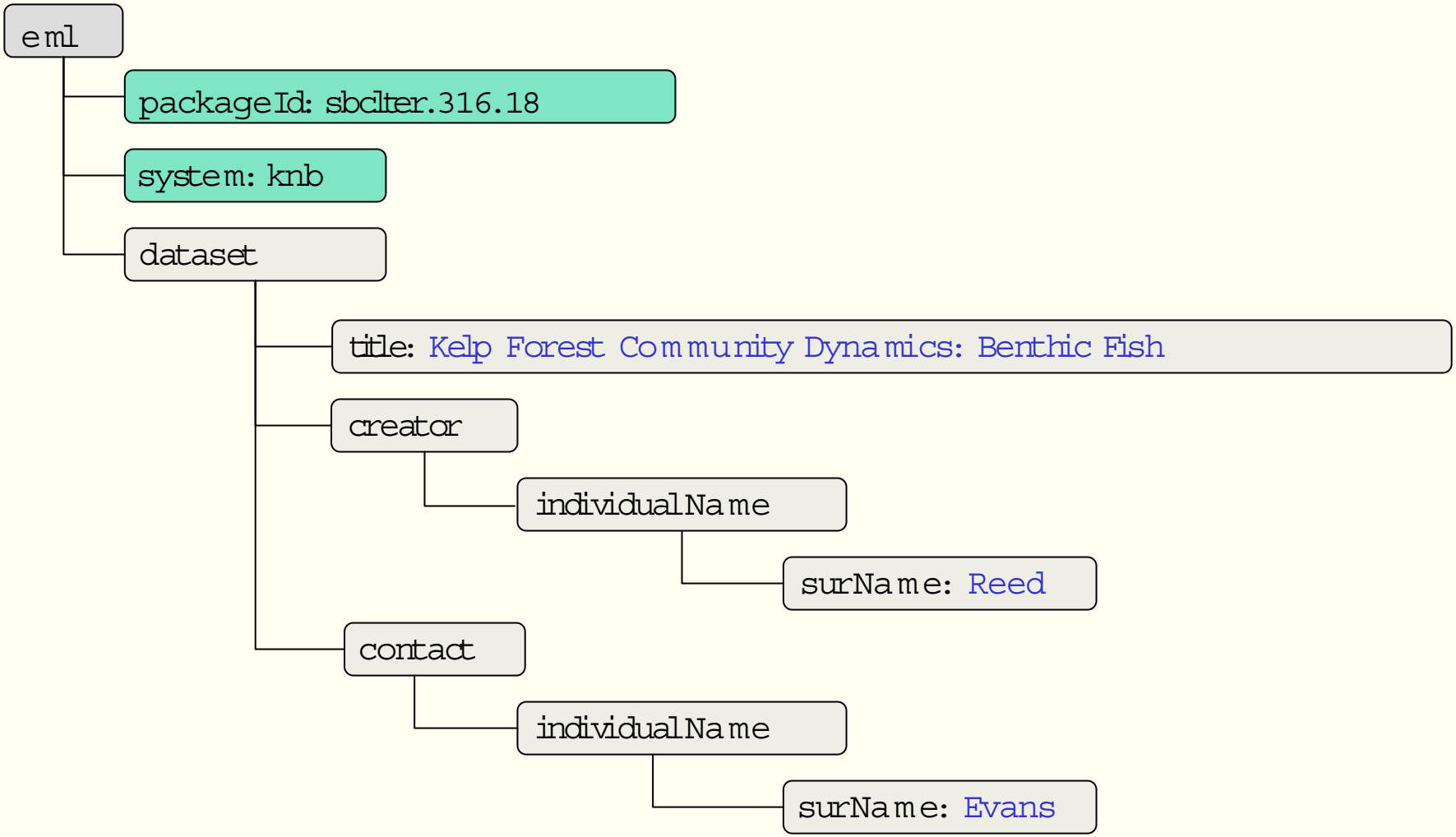
- Discovery information
  - Creator, Title, Abstract, Keyword, etc.
- Ownership and Citation information
- Intellectual property information
- Coverage
  - Geographic, temporal, and taxonomic extent
- Protocols and methods

Used in registries and catalogs

- Logical and physical data structure
  - Data semantics via unit definitions and typing

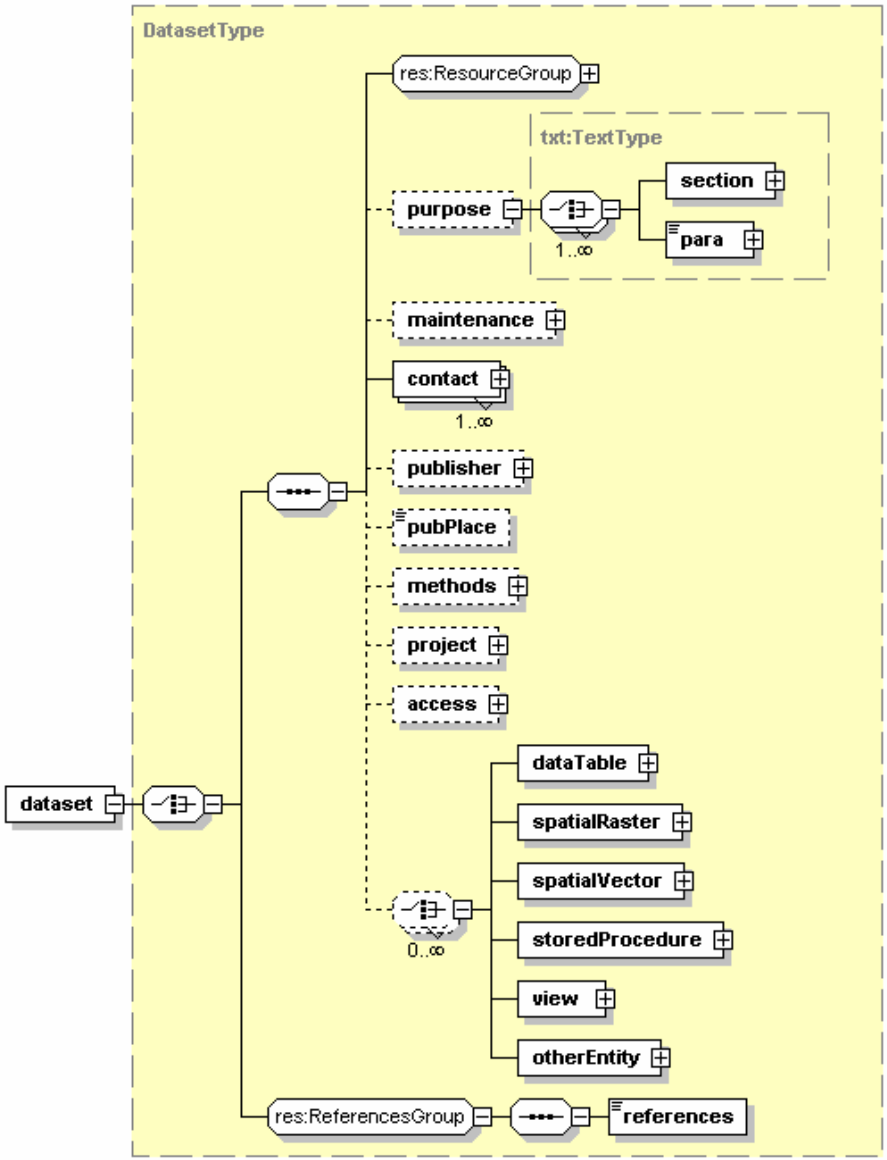
Used in Analytical applications

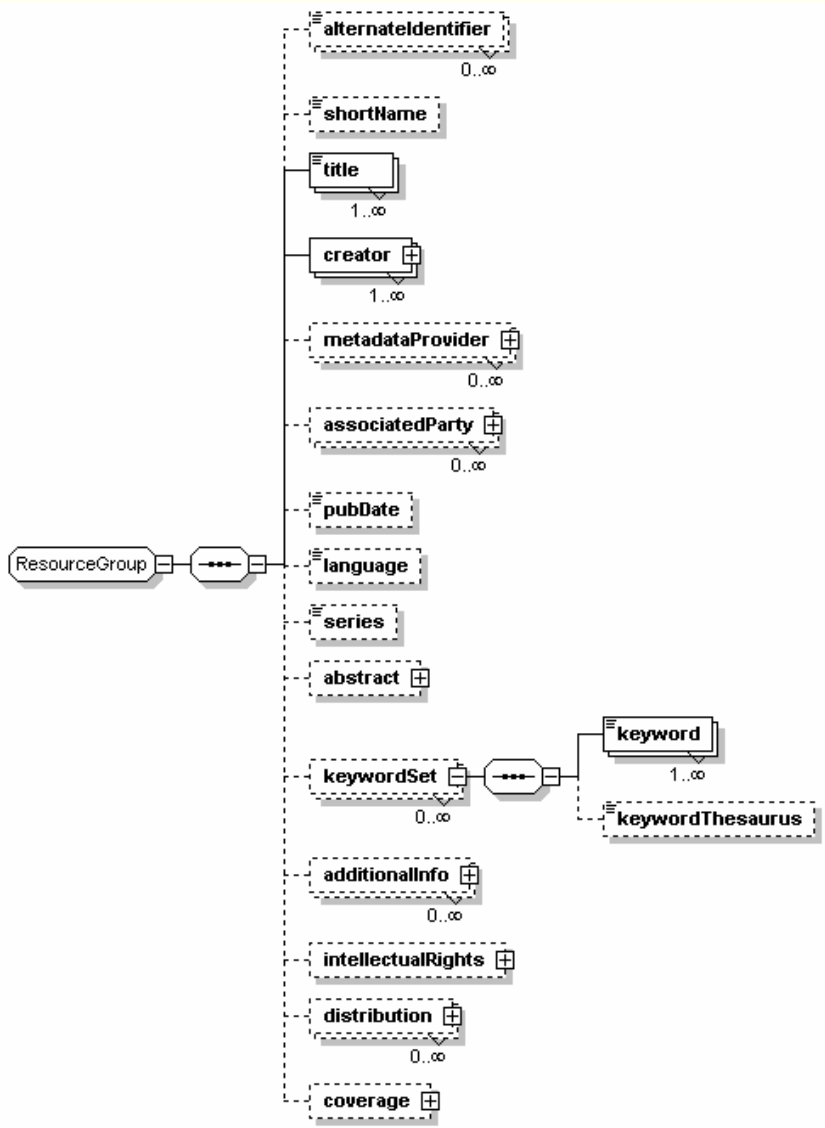
The simplest EML example



*barely enough information for 'identifying' the dataset

Identification: resource elements





EML Strengths

- EML is:
  - Designed for interoperability and portability
  - Supports validation and machine processing
- Modular
  - 24 modules for flexibility and reusability
- Extensible
  - Simple to add metadata from other standards or site-specific metadata
- Designed to support scientific analysis
  - Metadata details needed for analysis and models

EML uses open standards

- Uses Extensible Markup Language (XML)
  - Documents are cross-platform
  - EML documents support international languages
  - Excellent exchange lanuguage
- EML documents are human readable text
  - Therefore, archive friendly
- EML documents are fine-grained
  - Simpler conversion to other standards
  - More difficult to convert from coarse-grained standards like Dublin Core

An Example EML Document

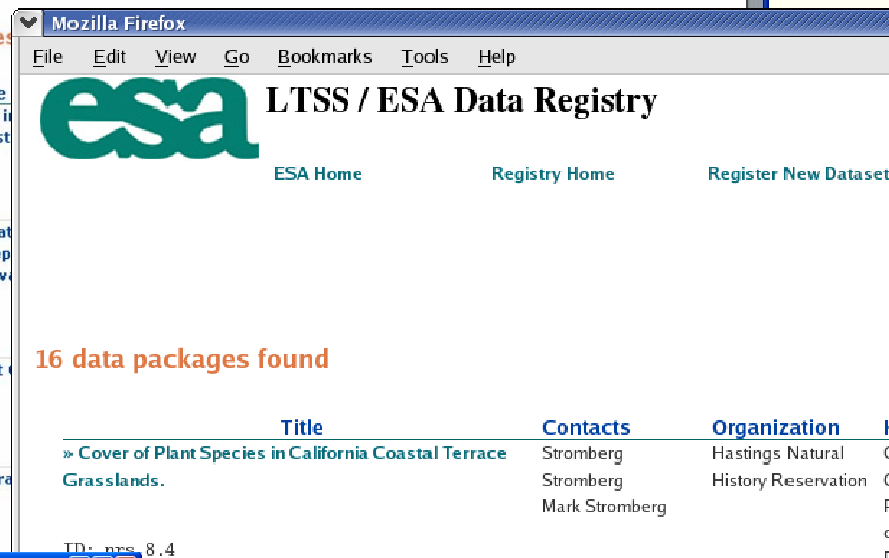
```
<?xml version="1.0"?>
<eml:eml packageId="piscoUCSB.5.20" system="knb"
xmlns:eml="eml://ecoinformatics.org/eml-2.0.0">
<dataset>
  <shortName>Alegria Temperatures</shortName>
  <title>PISCO: Intertidal Temperature Data:
      Alegria, California: 1996-1997</title>
  <creator id="C.Blanchette">
    <individualName>
      <givenName>Carol</givenName>
      <surName>Blanchette</surName>
    </individualName>
    <organizationName>PISCO</organizationName>
   <address>
    <deliveryPoint>UCSB Marine Science
      Institute</deliveryPoint>
    <city>Santa Barbara</city>
    <administrativeArea>CA</administrativeArea>
    <postalCode>93106</postalCode>
   </address>
  </creator>
  <abstract>
   <para>These temperature data were collected
     at Alegria Beach, California, and were ...
   </para>
  </abstract>
  <keywordSet>
   <keyword>OceanographicSensorData</keyword>
   <keyword>Thermistor</keyword>
  <keywordThesaurus>
     PISCO Categories
  </keywordThesaurus>
  </keywordSet>
  <intellectualRights><para>Please contact the
    authors for permission to use these data.
    Please also acknowledge the authors in any
    publications.</para>
  </intellectualRights>
  <contact>
    <references>C.Blanchette</references>
  </contact>
</dataset>
</eml:eml>
```

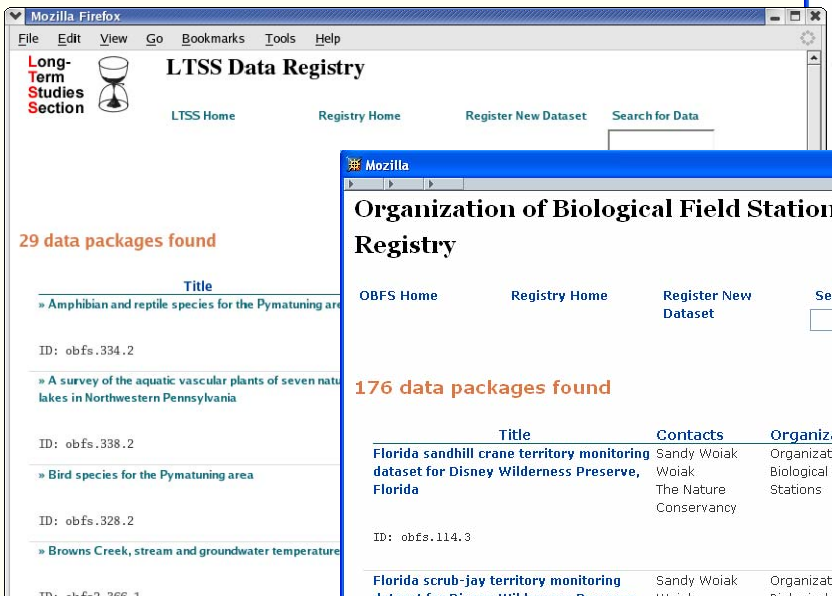
Transform

Data Registries

- Registries
  - UC NRS
  - OBFS
  - ESA LTSS
  - Spec Net
  - NCEAS

- Use metacat
- Web-based metadata entry





Morpho – Managing metadata and data

- Create and manage data and metadata using wizards
- Search and locate data
- Share data with colleagues

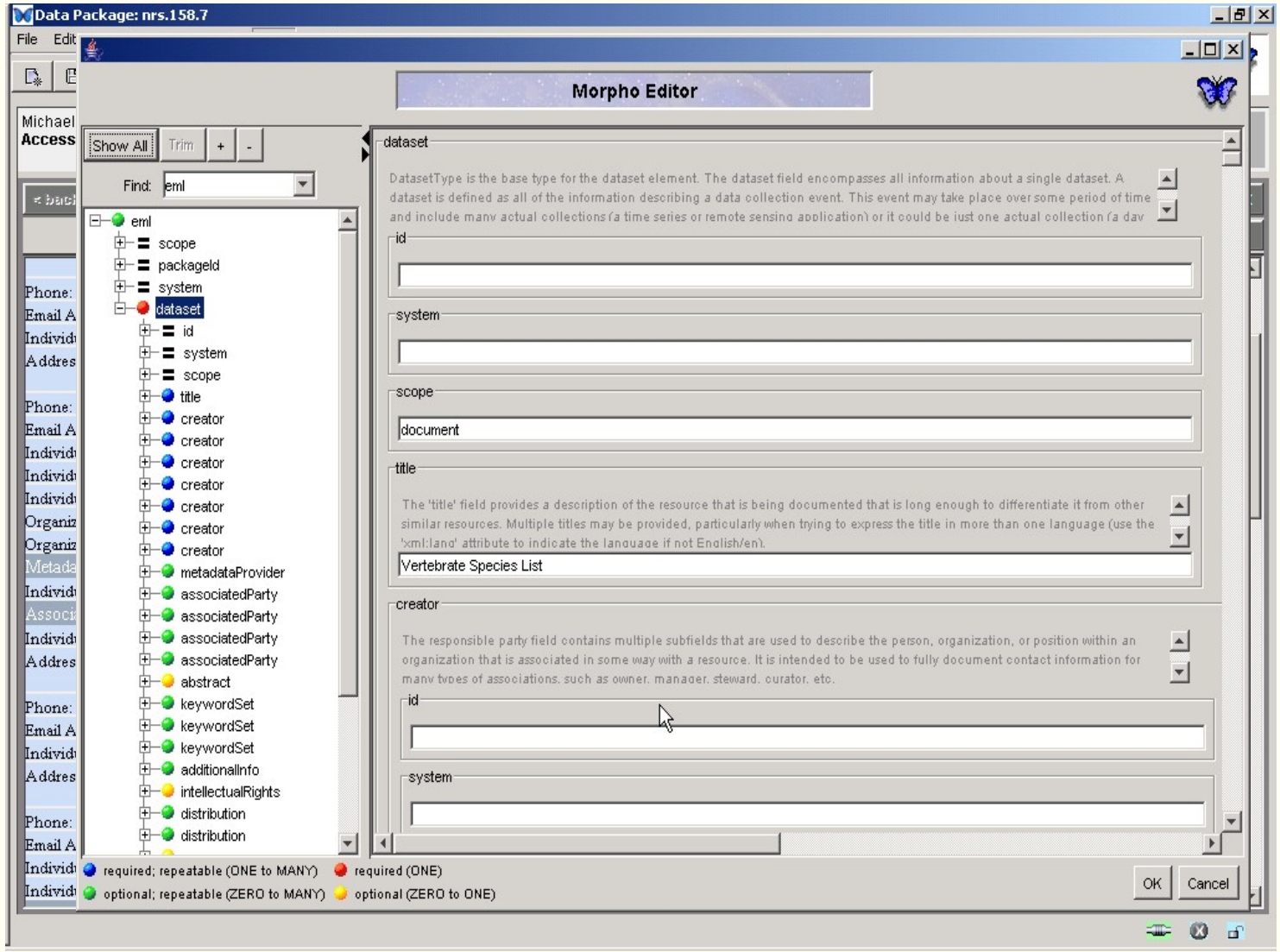
- Usable in field station environments
  - Can't assume a network connection is available always

- Can be run on Windows, Linux and Mac OS
- Is open source (GNU GPL)

What can Morpho do?

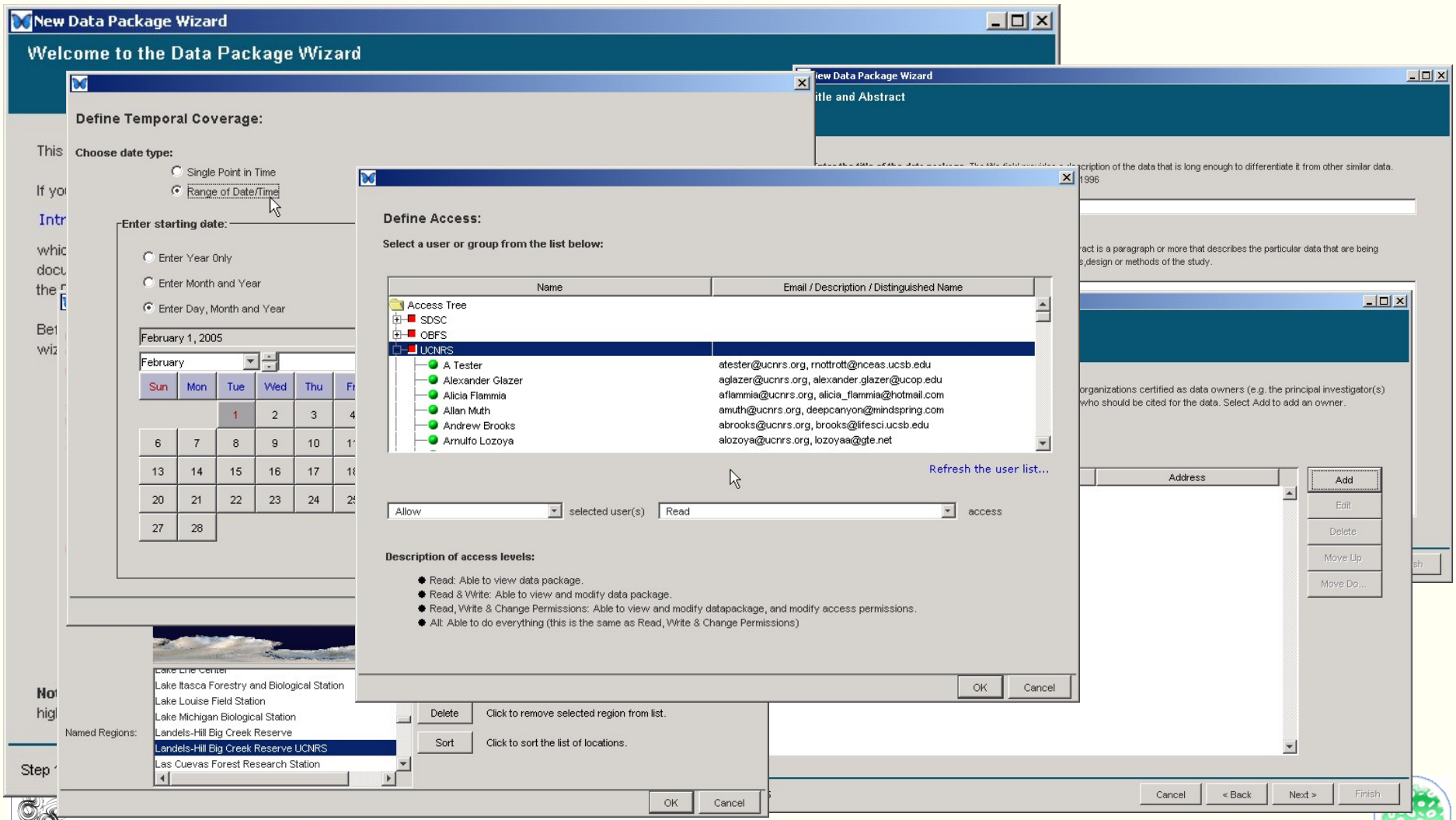
- Create EML data packages
  - Open and edit data packages, set access control
  - Import data into a data package
  - Save the EML and data
    - locally and to metacat servers
- Search for EML
  - locally and on metacat servers
- Export data packages

Morpho: Open and edit data packages



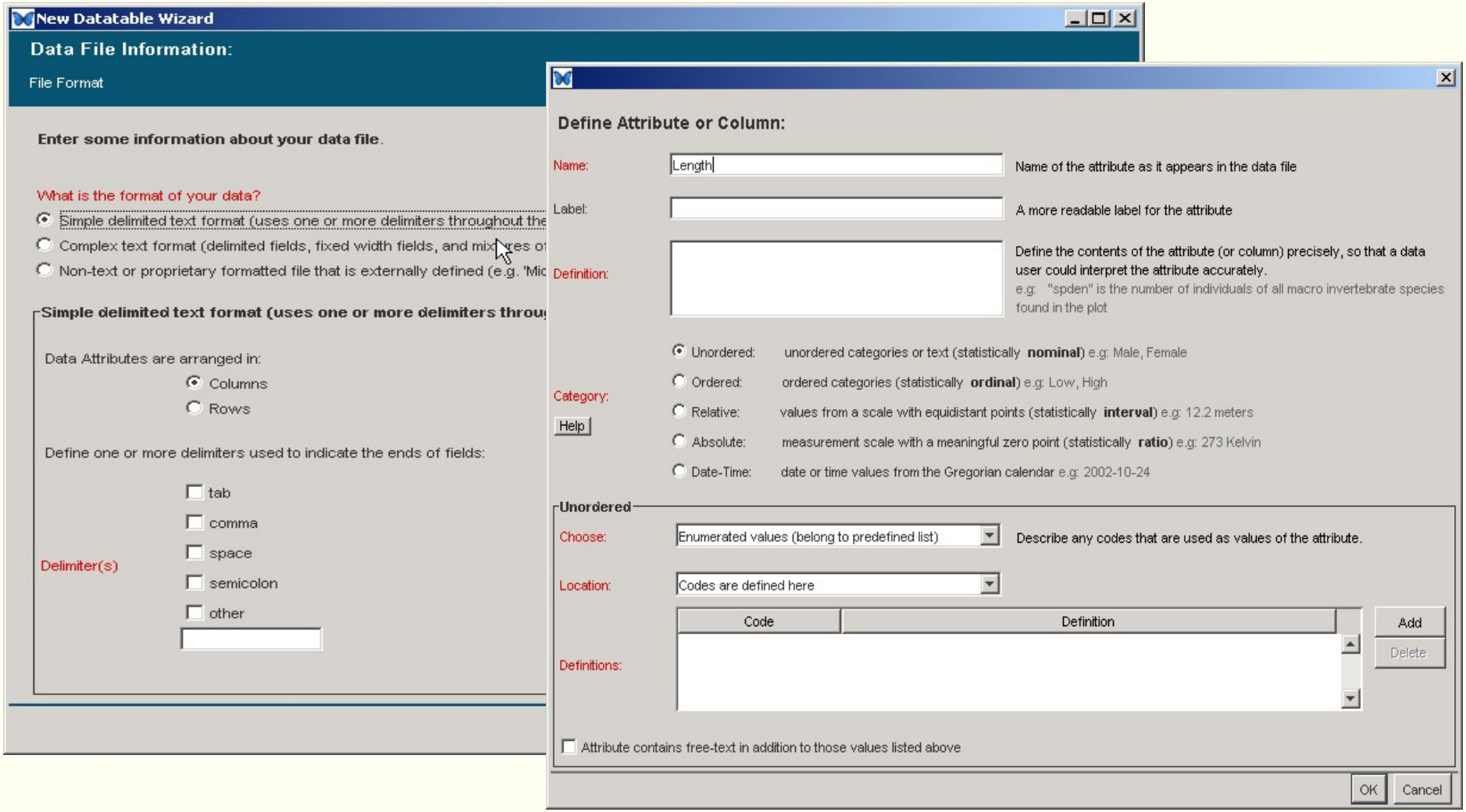
Morpho: Create data packages

- Data Package Wizard



Morpho: Import data

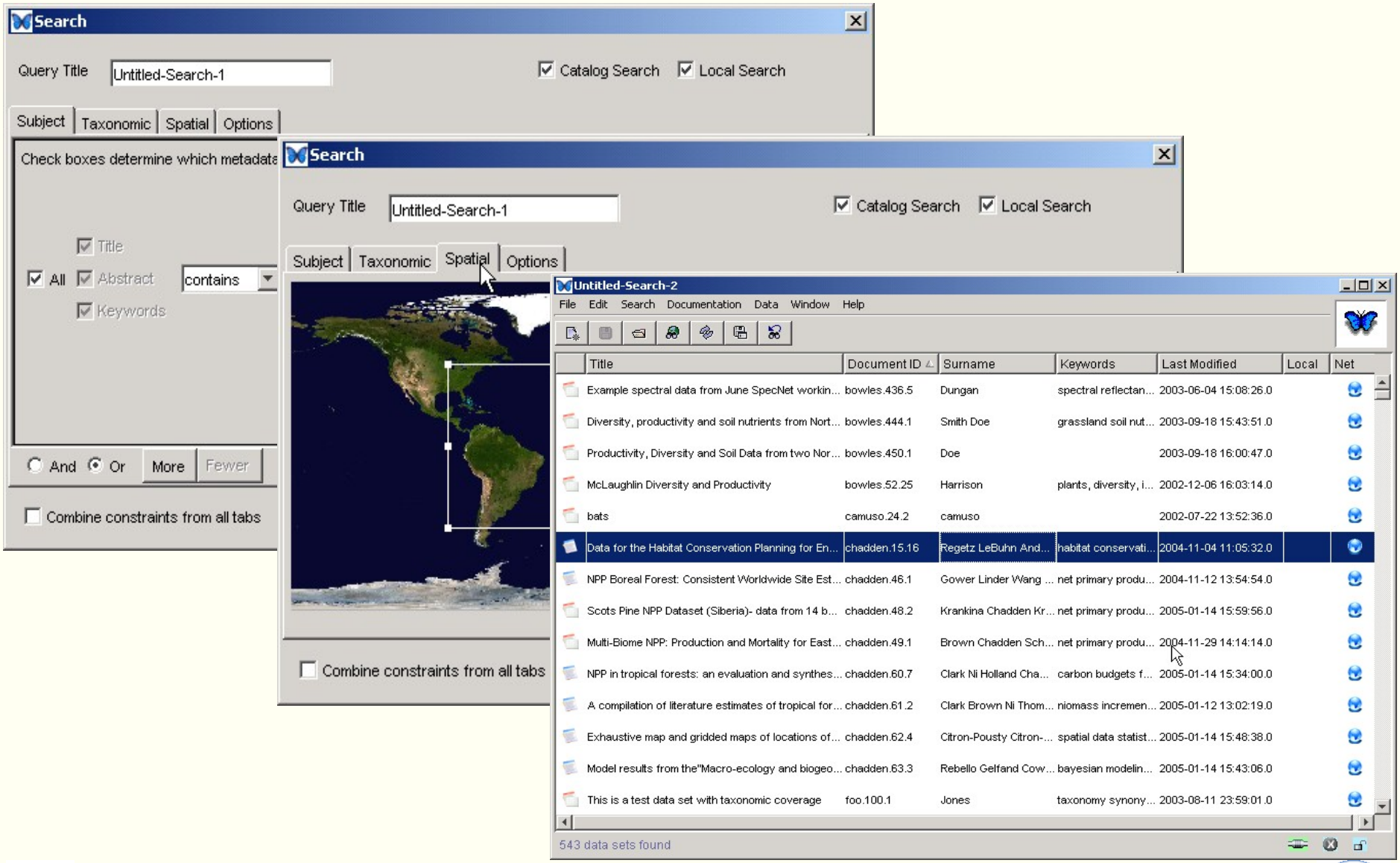
- New Data Table Wizard



Morpho: defining attributes

- Natural language description of each attribute
- Standardized measurement scales
  - Precisely define attribute domain
- Standardized measurement units
  - EML has unit dictionary with 140 pre-defined units
    - Relationship and conversion to base SI unit provided
  - Uses STMML to quantitatively define custom units

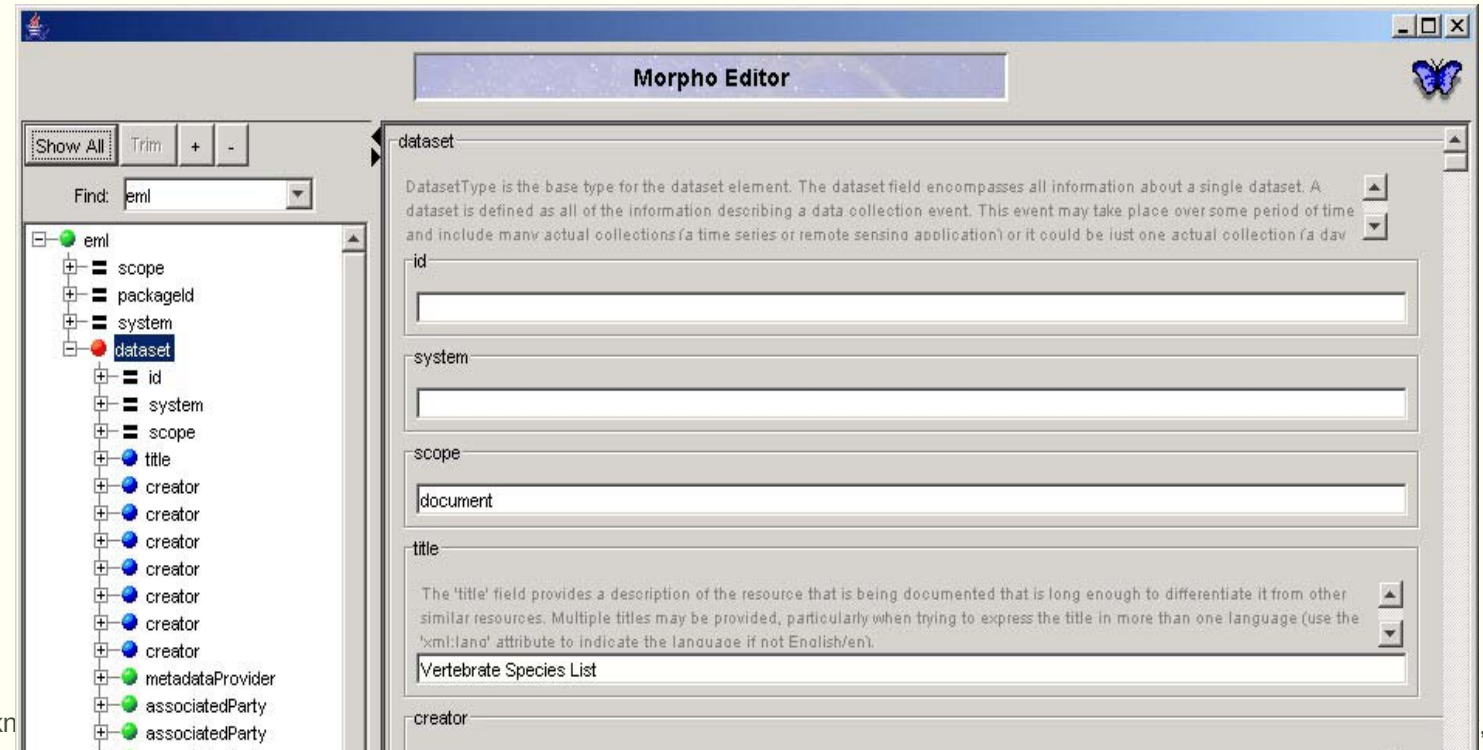
Morpho: Search for data packages



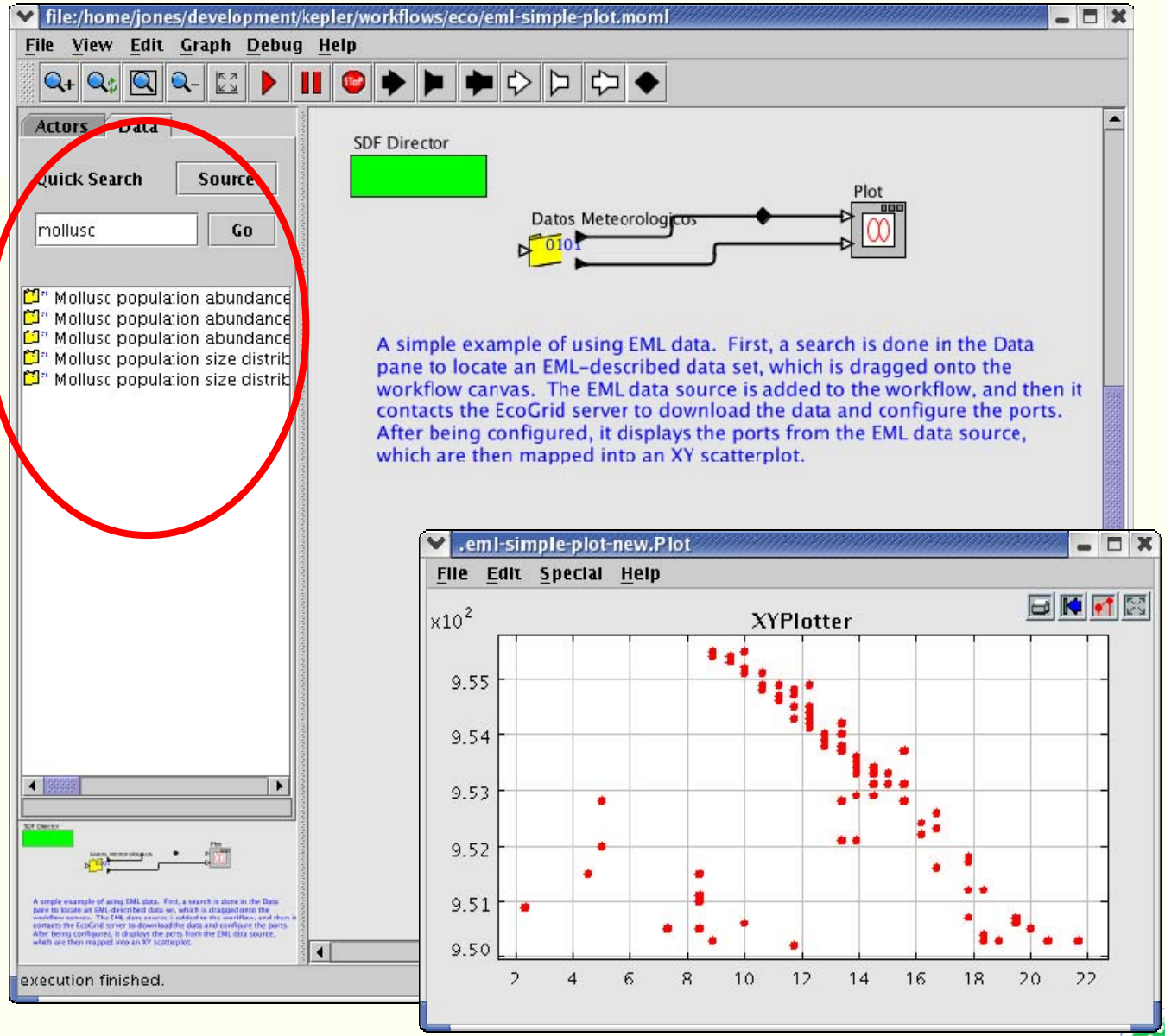
Morpho extensibility

- Originally designed to support multiple metadata standards
  - Later customized for ease-of-use for EML
  - Could be extended to support editing BDP or other standards

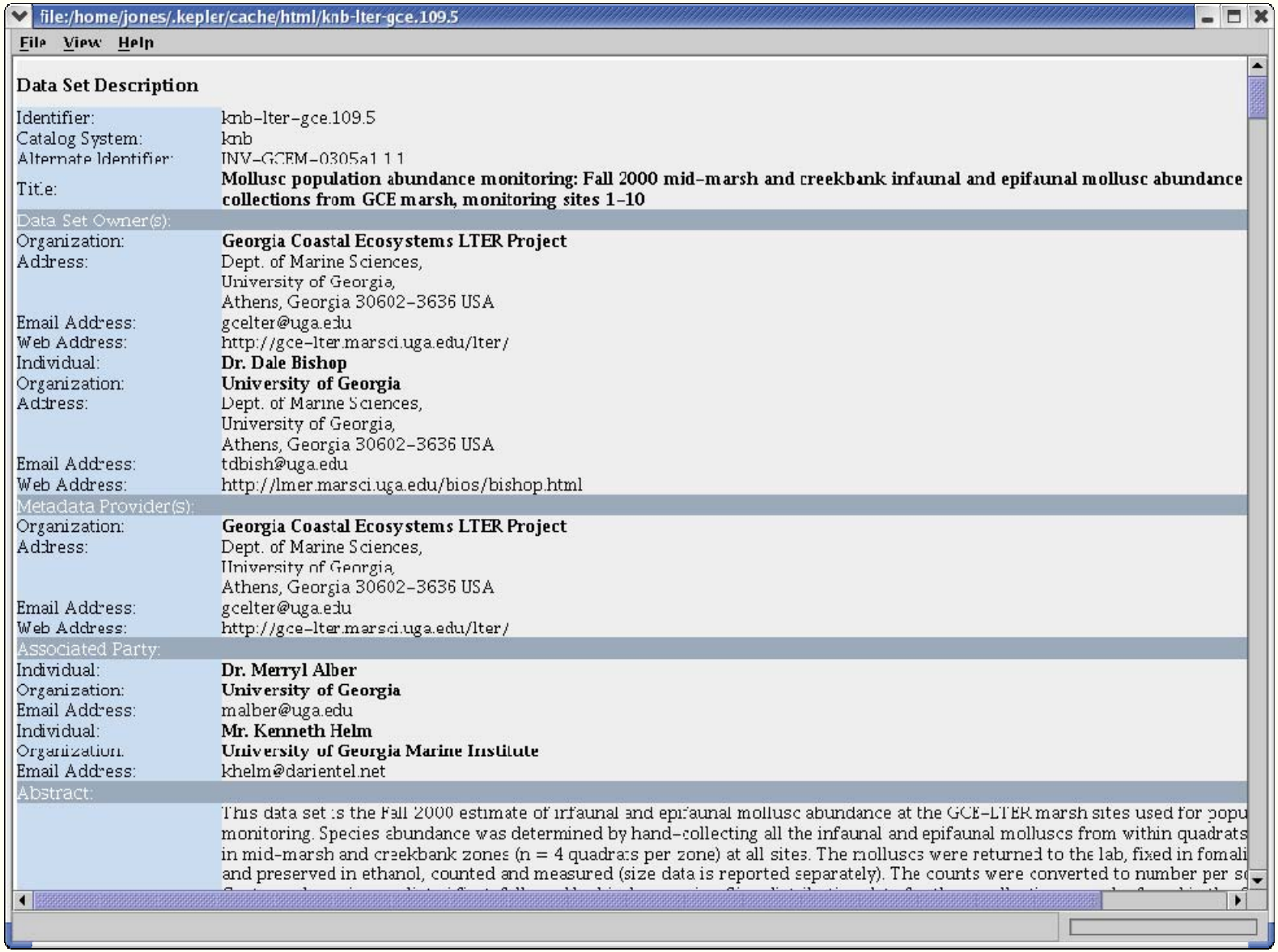
- Could add conversion capabilities to import/export BDP



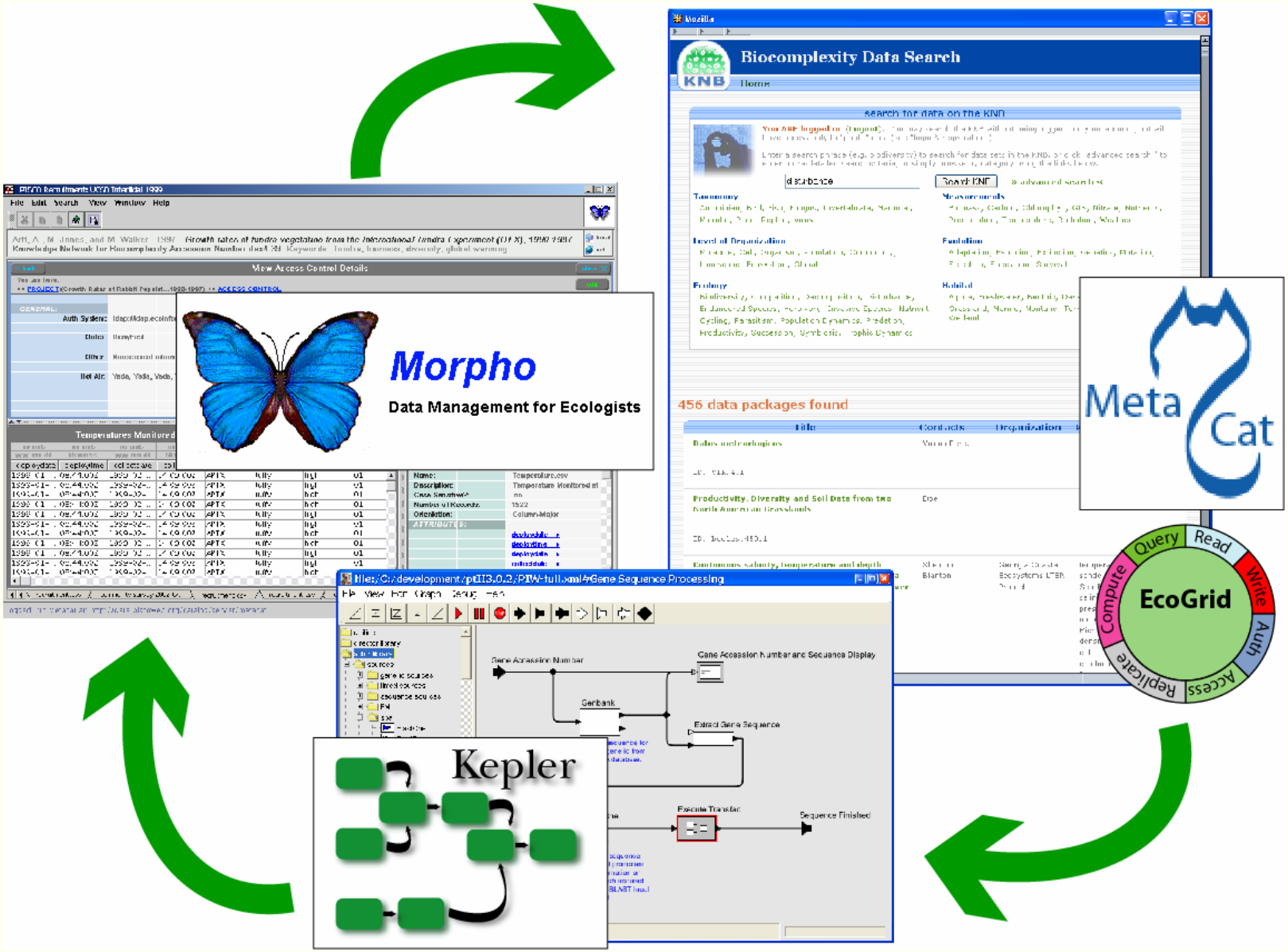
EML-driven analyses in Kepler



EML Metadata Display in Kepler



Metadata-driven analysis cycle



Acknowledgements

This material is based upon work supported by:

The National Science Foundation under Grant Numbers 9980154, 9904777, 0131178, 9905838, 0129792, and 0225676.

Collaborators: NCEAS (UC Santa Barbara), University of New Mexico (Long Term Ecological Research Network Office), San Diego Supercomputer Center, University of Kansas (Center for Biodiversity Research), University of Vermont, University of North Carolina, Napier University, Arizona State University, UC Davis

The National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant Number 0072909), the University of California, and the UC Santa Barbara campus.

The Andrew W. Mellon Foundation.

Kepler contributors: SEEK, Ptolemy II, SDM/SciDAC, GEON