The Environmental Data Initiative (EDI)

Make metadata with the EML assembly line

EDI is funded by the NSF DEB
5 Phases of Publishing Ecological Data

1. Assemble data and metadata
2. Format and QC data tables
3. Create EML metadata
4. Submit your data package (data and metadata) to repository
5. Cite your data package
5 Phases of Publishing Ecological Data

1. Assemble data and metadata
2. Format and QC data tables
3. Create EML metadata
4. Submit your data package (data and metadata) to repository
5. Cite your data package

www.environmentaldatainitiative.org
Quality metadata is VERY important!

- For understanding:
  - What the data are
  - When the data were collected
  - How the data were collected
  - Who collected the data
  - How the data have been modified
- For assessing fitness for use
- Critical to:
  - Open and reproducible science
  - Synthesis science
  - Extending the life and value of data
Structured metadata is VERY important!

- If not structured then:
  - Can’t search for the data in repositories or aggregators
  - Can’t search content at the dataset level
  - The speed of science and knowledge formation is slowed
  - Potential information loss
  - Can’t convert to other metadata standards
But, making quality EML is challenging!

Creating quality EML requires:
- Detailed info about: methods, data entities, personnel, keywords, etc.
- Detailed technical info about data entities
- Understanding of the EML schema (where content can go)
- Understanding EML best practices (where content should go)
- Construction of EML (how to create it)
But, making quality EML is challenging!

- Creating quality EML requires:
  - Detailed info about: methods, data entities, personnel, keywords, etc.
  - Detailed technical info about data entities
  - Understanding of the EML schema (where content can go)
  - Understanding EML best practices (where content should go)
  - Construction of EML (how to create it)
The EMLassemblyline R code package

- User:
  - Supplies core information in a familiar way (tables, text files)
- EMLassemblyline:
  - Provides meaningful error messages and guidance
  - Extracts technical info about data entities
  - Builds and validates the EML
  - Embodies EML best practices
  - Simplifies versioning
  - Facilitates automation
Step 1:
Step 1: Identify data type(s)
Step 1: Identify data type(s)
Step 1: Identify data type(s)
Step 1: Identify data type(s)
Step 1: Identify data type(s)
Step 1: Identify data type(s)
Step 2:
Step 2: Consult instructions
Step 2: Consult instructions

Instructions
- data table
- spatial vector
- spatial raster
- other
Step 2: Consult instructions

Instructions
- data table
- spatial vector
- spatial raster
- other
Step 2: Consult instructions

Instructions
- data table
- spatial vector
- spatial raster
- other
Step 2: Consult instructions
Step 3:
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates
Step 3: Import templates

abstract.txt  methods.txt  keywords.txt  attributes.txt
Step 4:
Step 4: Complete templates

abstract.txt
methods.txt
keywords.txt
attributes.txt
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 4: Complete templates
Step 5:
Step 5: Add custom details
Step 5: Add custom details

Instructions
- geographic
- taxonomic
- categorical
- other

- abstract.txt
- methods.txt
- keywords.txt
- attributes.txt
Step 5: Add custom details

Instructions
- geographic
- taxonomic
- categorical
- other

abstract.txt  methods.txt  keywords.txt  attributes.txt
Step 5: Add custom details

Instructions
- geographic
- taxonomic
- categorical
- other

files:
- abstract.txt
- methods.txt
- keywords.txt
- attributes.txt
Step 5: Add custom details

Instructions
- geographic
- taxonomic
- categorical
- other

- abstract.txt
- methods.txt
- keywords.txt
- attributes.txt
Step 5: Add custom details
Step 5: Add custom details
Step 5: Add custom details
Step 5: Add custom details
Step 6:
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Step 6: Translate to EML
Live demo!
Workflow:
Workflow: One-off publication
Workflow: One-off publication

Raw Data → cleaning → Processed Data → EMLassembline → edit templates → EML 1.0 → Upload → portal.edirepository.org → Storage
Workflow:
Workflow: Revision (same data format)
Workflow: Revision (same data format)
Workflow:
Workflow: Revision (new data format)
Workflow: Revision (new data format)

1. Reformat data
2. Data 1.2
3. EMLassembly
4. Edit templates
5. Data 1.2
6. Upload
7. portal.edirepository.org
8. Storage
Workflow:
Workflow: Automation (derived data)
Workflow: Automation (derived data)
Road map

- Where is the project going?
  - High-level functions
  - Auto-extracted metadata from data entities
  - Support for more data types
- How to get involved?
  - https://github.com/EDIorg/EMLassemblyline
  - colin.smith@wisc.edu
EDI Resources

- EDI website on “5 phases of data publishing”
  - environmentaldatainitiative.org/resources/assemble-data-and-metadata
- Contact EDI’s data curation team
  - info@environmentaldatainitiative.org
- Data Portal
  - portal.edirepository.org/nis/home.jsp
- GitHub
  - github.com/EDIorg
- Twitter
  - @EDIgotdata
- Slack
  - edi-got-data.slack.com