Data Distribution and Archiving in Support of the Agricultural Ecosystems Program

Gail Steinhart
Cornell University
Data Distribution and Archiving in Support of the Agricultural Ecosystems Program

Gail Steinhart
Research Data & Environmental Sciences Librarian
Albert R. Mann Library
Cornell University
Collaborators

Cornell departments and units:
- Animal Science
- Biological and Environmental Engineering
- Crop and Soil Science
- Ecology and Evolutionary Biology
- Horticulture
- Natural Resources
- Mann Library

Other organizations:
- Cornell Cooperative Extension of Chemung County
- Institute of Ecosystem Studies
- Univ. Maryland Center for Environmental Science
- Univ. Nebraska-Lincoln School of Natural Resources
- Upper Susquehanna Coalition

Funding: USDA Cooperative State Research, Education, and Extension Service
Research context

- The Susquehanna River watershed contributes a large portion of the nutrients and sediments impacting the Chesapeake Bay, which suffers from excessive nutrient and sediment inputs.

- Improving nutrient and erosion controls in the watershed of the Susquehanna River, the largest river entering the Bay, is one way to improve the health of the Bay itself. New York is committed to reduce the impact of its part of the Susquehanna River watershed on the Bay.

- The research project is designed to better understand the sources and sinks of nutrients and sediments in the New York portion of the Susquehanna watershed, using this as a model for rural landscapes in general.
Goal

The AEP Upper Susquehanna River Basin project should serve as a source of valuable data and insight for people in county, state, and the Federal government, NGOs, and the public to improve water quality in the Upper Susquehanna Basin and the Chesapeake Bay.

Specific Objective: Provide an easy way for the public to learn about the scientific goals and results of the AEP Upper Susquehanna River Basin project.

Approach:
• Document data sets using Ecological Metadata Language (EML).
• Deposit and preserve data sets in Cornell’s digital repository, DSpace.
• Create a public research portal, with background information, research plans, results, and data.
Goal

The AEP Upper Susquehanna River Basin project should serve as a catalyst for innovative cross-disciplinary research at Cornell that will produce better scientific understanding of nitrogen, phosphorus, and sediment cycling in the Upper Susquehanna basin and the Chesapeake Bay.

Specific Objective: Provide an easy way for researchers to know what other project participants have already done, are doing, and are planning to do (who, what, when, where, why).

Approach: Create a wiki, with access restricted to project participants. The wiki can be used to share datasets, preliminary results, and any other information about the project as it progresses, before such information is made publicly available.
Metadata

- Metadata serve as documentation for data, describing the content, purpose, structure, format, and accessibility of datasets. Interdisciplinary and collaborative science creates an important demand for a set of “instructions” for researchers to make sensible judgments about whether and how they might use data provided by their colleagues.

- Metadata also serve a functional role in digital repositories, providing the raw material that makes it possible to display information about a dataset, and for search engines to index repositories and deliver results to users.
Avoid data entropy!

Michener et al., 1997
Ecological Metadata Language: EML

- Developed specifically for ecological data (ESA, LTER)
- Modular and extensible XML-based standard
- Accommodates information on methods, geographic coverage, temporal coverage, detailed descriptions of tabular data
- [http://knb.ecoinformatics.org/software/eml/](http://knb.ecoinformatics.org/software/eml/)
- Comes with tools!

Upstate Science Librarians, October 20, 2006
Morpho

- Easy to use, platform independent metadata editor.
- Allows users to upload metadata and data to a server.
- Allows users to search, view, and save public data and metadata. *(Interacts with Metacat)*
EML record

Data Package: gis:1.3.2

Title: Lake Ontario Embayments - temperature and dissolved oxygen profiles

Accession Number: gis:1.3.2

Keywords: Lake Ontario, Blind Sodus Bay, Little Sodus Bay, Sterling Point, Juniper Point, South Sandy Point, North Sandy Point.

Abstract: Temperature and dissolved oxygen profiles collected as part of the NSF funded project (OCE-0031316): Biodiversity, Physical, Biological, and Human interactions shaping the ecosystems of freshwater bays and lagoons.

License and Usage Rights: Protected Data. Data is freely shared with the research group. However, findings or conclusions made with using another individual’s data should be brought to that individual’s attention. Only the owner of a dataset may share it with individuals not affiliated with the research group. Acknowledgment of Support and Disclaimer: This research was supported by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
DSpace

- Open source
- Easy submission
- Indexable by search engines
- Communities, collections
- Customizable workflow
- Supports OAI-PMH
- Already implemented at Cornell

Upstate Science Librarians, October 20, 2006
Portal

- [http://www.usaep.mannlib.cornell.edu/](http://www.usaep.mannlib.cornell.edu/)
- Modeled after Vivo and CALS research portals
- Presents information on the project *in context*
- Links to data in DSpace
Wiki

- Share information within the group
- Gaining traction in the sciences:
  - [http://www.openwetware.org/](http://www.openwetware.org/)
  - “Science in the Web Age: Joint efforts”
Putting it all together

- Project meetings
- Metadata / Morpho workshop
- Documentation
- Individual consulting
Early findings...

- USDA proposal
- Scope creep
  - Data formats
  - Files sizes
  - Need to share large files, privately
- Portal maintenance
- Extensible model?
Thank you

Gail Steinhart
Research Data & Environmental Sciences Librarian
Albert R. Mann Library, Cornell University

GSS1@cornell.edu