Virtual Update Notes 11/3 and 11/4 2008 EcoTrends

Published on *LTER Information Management* (http://im.lternet.edu)

Home > Virtual Update Notes 11/3 and 11/4 2008 EcoTrends

Virtual Update Notes 11/3 and 11/4 2008 EcoTrends

Mon, 11/03/2008 - 4:29pm — sremillard

EcoTrends; the good, the bad, the ugly; lessons learned; Christine Laney
See attached PowerPoint

Participants:
Monday (11/3): Christine Laney (moderator), Suzanne Remillard (AND), Margaret O'Brien (SBC), Susan Stafford, Jason Downing (BNZ), Don Henshaw (AND), Sven Bohm (KBS), Dave Balsinger (NTL), Barbara Benson (NTL), John Porter (VCR), Karen Baker (PAL); James Conner (CCE), Mason Kortz (CCE), Kristin Vanderbilt (SEV), Duane Costa (LNO)
Tuesday (11/4): Christine Laney (moderator), Jonathan Walsh (BES), Linda Powell (FCE), Corinna Gries (CAP), Emery Boose (HFR), Wade Sheldon (GCE), Theresa Valentine (AND)

Monday notes:

(slide 1) Questions about how EcoTrends is progressing at meeting initiated this update. This will be an update and unresolved issue. How to make things

better.
Recap of project, goals, progress, where we're at, problems:

(slide 2) Cast of characters: LTER, USDA, USGS, technical advisors, editorial committee

(slide 3) Goals and Audience

(slide 4) Christine's challenges over the couple of years:
Asked for a variety of data then aggregated to monthly and annually for visual comparison.

Data in different formats. Specific scripts were developed to deal with all the different data sets and formats – very time consuming. Ended up with thousands of data sets. i.e. streamflow from AND resulted in multiple data sets over time (Ca, Ch, N, etc.). Needed to keep data and metadata connected, therefore developed an onsite database for source data products and metadata, personnel (from LNO), data derivation, scripts, relationship between products.

(slide 5) Communication

(slide 6) Book – lots of effort currently going into book. Planning to publish summer 2009.

Web site – will remain dynamic. It is currently up and running. Feel free to check it out. Can graph different sites and variables against each other.

(slide 7) PASTA – goal in future is to make this automated. PASTA will help automate it.

(slide 8) What are the strengths and weaknesses among and between all 26 LTER IM systems?
What can we learn from this process?
How can we make it better?
What are the most important implementation steps that can be taken right now to make thinks more efficient in the future?

(slide 9) The Good:

(slide 10) The Bad (or better opportunities):
EML is still incomplete, i.e., no units, no methods.
ClimDB metadata is very sparse.

(slide 11) The Ugly (or really great challenges):

(slide 12) Suggestions and Help Welcome:

There was lots of discussion. Some suggested an additional VTC to finish up. The last 2 slides will likely produce lots of questions.

Connection got bad towards the end of the call. We concluded with a very quick overview of the last two slides. It would be nice to have another update to allow for more questions. Christine will try to move through the first part of the

presentation quickly for the next group to allow for more time for discussion on the final slides.

Questions:
SS: What sort of anonymity is there with the socio-economic data? This is an issue that relates to urban site data.
CL: Socio-economic data from census bureau

JP: Will the site enable a skin? Can sites display EcoTrends within their own frameworks? Like a VCR-LTER skin?
CL: There is a generic data citation along with specific site and PI information.

MO: Seconds JP's idea about skinning because it makes EcoTrends more of a network product.
DC: DOT project and Bugzilla (ask Mark Sevilla)

MO: Perhaps need better integration between ClimDB and site metadata.

SR: ClimDB would really benefit from having metadata in EML and a metadata harvesting system.

BB: How serious is the problem with sites changing the format of their data?
CL: It is really serious and causes lots of problems with the R scripts. Other issues are changes in IMS, changes in URL, changes in

DH: (lots of static; will need to get Don's question in writing)

—

John Porter's notes:
IM VTC 11/3/2008

A. In attendance
1. Christine Laney
2. Susan Stafford
3. Jason Downing
4. Karen Baker, James (PAL/CCE)
5. Margret O'Brien
6. Sven Bohm
7. John Porter
8. Kristin
9. Suzanne Remillard
10. Duane Costa
11. Dave Balzinger
12. Barbara Benson
13. Don Henshaw
B. Christine Laney - Ecotrends Presentation
1. see powerpoint sent out by Suzanne
2. LTER 1/2 the sites, but 80% of the data

3. questions
a) any caveats on socioeconmic data regarding privacy
1) no, all data came from Census Bureau
b) Urban LTERs had some additional issues.... but ECOTRENDS
just dealt with publically available data from other sources
c) Plans for Ecotrends portals?
1) skins for different networks and sites would be valuable
2) viewing metadata does give credits....
d) we like the idea of "skinning" as well - allows parts of it
to be shared more easily
1) gives more of a network-wide feel
e) ClimDB and metadata - would like to have ClimDB do more with
EML metadata
f) Need to get it so we can generate code from EML for
processing data
g) issues of sites that store data by year in different formats
- will "break" R scripts....
1) moving data from one URL to another breaks the scripts...
2) move to query-based systems causes problems
4. Mark and Duane can give you a personal account for testing
purposes on Ecotrends site (note- site is for OLD data - not
updated data)
C. Action Items
1. John Porter and Margaret O'Brien will send better
description of skinning and potential use of Ecotrends
portal
a) put in bugzilla

—

Tuesday notes:

IM VTC 11/4/2008
E. Boose

Participants: Christine Laney (lead), Emery Boose, Corinna Gries, Linda Powell, Wade Sheldon, Theresa Valentine, Jonathan Walsh

See Christine's power point for details

EocTrends project includes 50 research sites, numerous agencies, an editorial committee (book), and a technical committee (website)

Goal is to make data available easily and quickly via the book and website

Contact Mark Servilla for access to website

Datasets on website have not been updated since last May. Some communications issues between JRN and LNO servers.

Considerable interest from Henry Gholz (NSF), especially regarding data accessibility

Intended audience includes not just researchers and educators, but also planners, managers, etc

LTER sites provided 85% of the data

Christine initially contacted PIs and asked for data at any resolution and level of documentation. The project will need to move away from that model to enable automation.

R scripts used to aggregate and analyze data

JRN database contains site details (e.g., met station coordinates), personnel info (PersonnelDB plus additions), data transformations, scripts utilized, relationships between personnel and datasets

Thousands of derived data files with specific file names. Copies of files are sent to LNO server. LNO creates EML for derived data files. Two to three days would be required to re-harvest and process all datasets.

Book includes standard time plots, maps of localities and spatial trends, summary tables

Book format is under development. Chapters written by subcommittees. Two-page site description, book introduction, and standardized plots sent to site PI and IM for review.

Site checking sometimes done by PI (or scientists chosen by PI) and sometimes by IM

Sheer bulk of information is a challenge. How to improve communications with sites?

Meeting of editorial committee and NISAC scheduled for late Feb or Mar 2009. Finalize book design and decide on next steps for website.

PASTA framework diagram (from Mark Servilla) shows goal for implementation. Not all pieces are available yet.

Collaboration and synthesis are increasing in LTER Network. EcoTrends can support synthesis.

What are next key implementation steps?

LTER datasets are by far the most available and best documented of the datasets used for this project

ClimDB has been very useful. Consistent format makes it easier to use.

Socio-economic data coordinated through CWT. 800 counties, 32 variables, over 20,000 records. Available now on two websites.

EcoTrends website is up and running. Many positive and useful suggestions have been received to date.

Data documentation is still incomplete and prone to error. It takes time to communicate problems back to sites.

Use ClimDB or data directly from sites? Site metadata may be more complete, but ClimDB data may be cleaner, more reliable, and more up to date (with better tools for analysis). Perhaps ClimDB should be augmented to read EML?

Site datasets sometimes have awkward formats and minor errors. Quality checking is sometimes lacking.

Data derivations not yet completed. Lots of auxiliary information. Sites sometimes change how data is stored and presented (e.g., shift to query-based system). Metadata does not always match the downloaded dataset.

Sites may be lumpers or splitters. Different strategies for creating data objects. These issues need to be addressed in best practices.

But metadata must match the data object received by the user. This issue not yet discussed at length by the IMC. EcoTrends is the first user on this scale.

A lag may result from time differences in data and metadata delivery systems

Migration of EML to derived products is not trivial. For example, methods for derived products may be different from methods for original dataset. How to nest methods and other information?

Need a better way to report problems back to the sites. Individual emails are time consuming.

Intra-site communications (PIs, IMs, other researchers) are often interesting

Researchers, teachers, and other users always want more capability. Import some land-use analysis capability into EcoTrends. Improve access to original source data. Researchers want to compare source and derived data. Teachers are interested in both long-term trends (derived data) and in how to work with data (source data).

Everything takes longer than planned. Data derivation errors, metadata interpretation, and removal of datasets all take time.

Non-LTER data is not as well documented, often not online, with less clear communication pathways and more personnel turnover

Biotic data are more complex. Often include different levels, different treatments. Current framework in Pasta and EcoTrends is not flexible enough to handle biotic data well. Current ad hoc solutions do not scale well.

EcoTrends has been criticized for lack of biotic data. Some biotic data have not been processed yet because it takes more time. Early focus on biogeochemical data. But also fewer long-term biotic datasets.

Need to find ways to automate the process. A person as the "black box" is not a feasible long-term solution. Harvest, data transformation, and EML documentation all need to be automated.

Opportunity to promote EML and data transparency among other agencies

Christine welcomes suggestions and comments. How to keep IMs better informed? How to improve EcoTrends project? Are there comparable projects at other agencies?

Check out the EcoTrends website. Send comments to Mark Servilla. John P. and Margaret suggested "skins" to customize look and feel to match site websites.

Set up Bugzilla for comments. Or use Drupal website. Drupal can support a project with issue tracking, suggestions, documentation downloads, etc.

See http://oregonexplorer.info [1] for a comparable project with ample funding

Recent work on EML and ontologies may prove helpful. Completion of level 5 EML across the LTER Network will also help.

EML syntax for units is difficult. Is this a problem for synthesis? No good metadata transformation tools at present. Christine uses Oxygen to read EML files manually. Automation of EcoTrends may require transforming units into a form more familiar to scientists.

Workflow tools such as Kepler might be used to document and automate manual steps

Problems with biotic datasets. Complex schemas, large numbers of variables, poorly defined species codes. Christine has mapped species to USDA plant codes for internal use. R script contains comments for mapping.

Methodology often not well documented. E.g., specific locations, changes in treatments over time, changes in codes over time. References to published papers are not sufficient. Resolving problems can be very time consuming.

Attachment	Size
EcoTrends_VTC_GoodBadUgly_Nov2008.ppt [2]	3.66 MB

- Virtual Updates [3]

This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Please contact us with questions, comments, or for technical assistance regarding this web site.

Source URL: http://im.lternet.edu/node/374

Links:
[1] http://oregonexplorer.info
[2] http://im.lternet.edu/sites/im.lternet.edu/files/EcoTrends_VTC_GoodBadUgly_Nov2008.ppt
[3] http://im.lternet.edu/taxonomy/term/169