Ecosystem Informatics Working Group
LTER All-Scientists' meeting
September 21, 2006

Attendees

Name	LTER site/affiliation	e-mail
Julia Jones	AND	jonesj@geo.oregonstate.edu
Stephen Mitchell	AND	Stephen.Mitchell@orst.edu
Chris Graham	AND	Chris.Graham@oregonstate.edu
Aaron Ellison	HFR	aellison@fas.harvard.edu
Jeannine Cavender-Bares	CDR	cavender@umn.edu
Judy Cushing	Evergreen State College	judyc@evergreen.edu
Weigi Zhou	BES/UVM	Wzhou1@uvm.edu
William Michnere	LNO	wmichener@lternet.edu
Barbara Kimbell	UNM/LNO	bkimbell@unm/edu
Kate Buenau	MCR	buenau@lifesci.ucsb.edu
Steph Lyon	CDR	Pimmx001@umn.edu
Zac Kayler	AND	Zachary.kayler@oregonstate.edu
Alan Tepley	AND	tepleya@geo.oregonstate.edu
Atzimba Lopez	Mexican LTER	atzimba@mexlter.org.mx
Kristin Vanderbilt	SEV	vanderbi@sevilleta.unm.edu
Karen Baker	PAL, CCE	kbaker@ucsd.edu
R. Gil Pontius, Jr.	PIE	rpontius@clarku.edu
Rebecaa Koskela	BNZ	fnrjk@uaf.edu
Alan Covich	LUQ	alanc@uga.edu
Steven Paton	STRI	patons@si.edu
Kathi Schleidt	ALTER-net	Katharina.Schleidt@umweldbundesamt.at
Peter McCartney	NSF	pmccartn@nsf.gov

Summary: The workshop consisted of 4 presentations (Julia Jones, Judy Cushing, Aaron Ellison, Bill Michener/KristinVanderbilt) and comments from 4 PhD students in the Ecosystem Informatics IGERT. The presentations covered Ecosystem Informatics education at the undergraduate, PhD, post-doc and beyond levels, as well as research projects. Detailed notes from the presentations and discussion are included below.

Julia Jones, Oregon State University

Workshop Objectives:

- To consider multiple diverse approaches to "ecosystem informatics"
- What activities within LTER involve ecosystem informatics?
    - Ensuring quality of and extracting information from large datasets that reveal underlying structures and trends in ecosystems
    - Expressing ecosystem questions in models that allow hypothesis testing and provide general insights into ecosystem structure and function

- By furthering the development of ecosystem informatics, how might the LTER network contribute to ecosystem science beyond the LTER network?
    - Observatories and sensor arrays will be producing larger datasets with more need for automated analysis and quantification of sources of uncertainty
    - Increased appreciation for hierarchical structure of ecosystems and approaches that integrate data and models
    - Ongoing technology revolution requiring ecologists to collaborate with computer scientists
    - The need for ecologists to catch up with other branches of science and engineering in its use of mathematics for formulating questions and providing overarching insights about system function

1. What is Ecosystem informatics?

   Many different uses of "informatic." Try for the most inclusive. EI is a collection of opportunities for collaborative research and education, arising from:
   1. Ecosystem science, bringing diverse science questions, policy concerns, and datasets
   2. Computer science, Bringing algorithms for storing, retrieving, searching, analyzing, and displaying information about ecosystems
   3. Mathematics, brining coherent framinf of questions, clarity from analytic solutions, and general insights across disparate problems in ecosystem science
2. What are key areas of research in EI?

   EI research collaborations can be divided into 2 areas: (1) ecosystem science and policy-driven research (top-down), and (2) technologically-driven research (bottom-up).
   Four key areas of research provide opportunities for collaboration in EI:
   a) carbon cycling and climate change (nutrient cycling and environmental change); example question/project: What are the long-term carbon storage levels in forests of varying growth functions, decomposition rates, and disturbance regimes (random vs. regular)?
   b) Landscape disturbance patterns and processes; example question/project: What orders and magnitudes of historical disturbance events give rise to particular forest stand structures and specices compositions?
   c) Network structure and function; example questions/project: What configurations of stream networks and disturbances permit persistence of populations of aquatic species?
   d) Stability and resilience in ecological communities: example question/problem: what combinations of qualitative species interactions produce stable communities?

   Some EI research collaborations are motivated by technological challenges, for example
   e) automated outlier detection data cleaning from large sensor arrays (e.g. Aaron's talk on flux tower data)
   f) information extraction from large datasets (supervised, unsupervised machine learning)

g) integrating disparate datasets (ontologies, semantics) (e.g. Judy Cushing's presenation?)
h) interpretation of content by displaying large datasets (scintifice viusualiztion)

3. What are key elements of education in EI?

Ei education occurs at the graduate (PhD) level (IGERT) and at the undergraduate level (BBSI). Key elements of successful interdisciplinary education in EI involve:

a) shared questions (ecosystem science and policy-driven research above)
b) common vocabulary, or at least mutually understood, differences in meaning of common terms, e.g. use of "stability, model, uncertainty, scaling
c) effective collaborative process, including (i) who leads, who follows, and when; (ii) balanced use of constructive and critical impulses
d) appropriate matching of technique to problem in teaching examples

4. Speakers for today's session

Nine speakers (4 researchers, 5 graduate students) will share their impressions about research and education in EI, drawing upon their own experiences:

a) Judy Cushing, Evergreen State University. Biodiversity and Ecosystem Informatics conferences. Canopy database and associated projects.
b) Aaron Ellison, Harvard Forest LTER. Information technology for automated calculation of C exchange from flux towers. Bayesian inference.
c) Kristin Vanderbilt, Sevilleta LTER.
d) Bill Michener, LTER Network Office. The SEEK project. Kepler workflow analyses.
e) Chris Graham, PhD student, Forest Engineering, Oregon State University, and Andrews LTER. Integrating mathematical modeling and hydrology experiments across sites.
f) Dan Sobota, PhD student, Fisheries & Wildlife, Oregon State University and Andrews LTER. Approaches to data management and analysis from intersite lotic nitrogen experiments.
g) Alan Tepley, PhD student, Geosciences and Forest Sciences departments, Oregon State University, and Andrews LTER. Inferring fire disturbance history forest stand structure and composition.
h) Zac Kayler, PhD student, Forest Science department, Oregon State University and Andrews LTER. Carbon exchange from soils.
i) Steve Mitchell, PhD student, Forest Science Department, Oregon State University and Andrews LTER. Generalized likelihood uncertainty estimation for carbon storage in forest ecosystems.

Judy Cushing, Evergreen State College
Canopy database project
Thousand year chronosequence – within-tree and stand structure. Nalini Nadkarni. 6 years. First 3 years collecting data on what represents structure of the forest, now people are doing function studies. Will be featured on National Geographic next week (week of Sept 25).
Canopy DB vision – how to help scientist better manage data and contribute to archives.
Building prototypes:
database generator,
palm-top hands-free data acquisition,
data visualization
ecology database repository
research referenced tool – BCD
LTER grasslands biomass data synthesis
Which trees to leave?

Evergreen – interdisciplinary education

2. interdisciplinary education needed
3. undergraduates can do research
4. problem-based, interdisciplinary, team-taught, full-time yearlong education
5. no departments, no tenure, no grades, no departmental requirements
6. now 4000 students, 200_ faculty,

data and information: quantitative ecology at Evergreen

1. explore ecology research process via a case study
2. generalize from that to ecology and other sciences
3. provide the next generation ecologusts, computer scientists, and mathematicians with the education to work together as professionals to solve complex scientific problems
4. create a supporteive, effective, and enjoyable learning environment

program components (courses)

a. ecology case study
b. statistics – python and R
c. programming in python
d. seminar – history and philosphy of data-driven science

research opportunities for undergrads

a. in the field
b. in the lab
c. senior capstone project
d. collaborators

other research at Evergreen
research ambassodor prograj
EEON – compaus a s observatory

Lessons learned:

1) think outside the box
2) research can inform and enrich teaching, vice-versa
3) small colleges can contribute to eco-informatics, use them!
4) If you are at a small college, find a local collaborator

NSF CISE ICER initiative
Integrative computing education and research

- CS involves other fields
- No agreement on curriculum in CS
- CS Graduates lack "systems approach"
- Swindling student pipeline in CS
- US IT competitiveness threatened

ICER vision

- integrate IT educationand research – campus wide
- design currucula ro reflect the integrateive nature of computing
- four regional workshops
- new NSF program CPATH – announced yesterday: a solicitation for education programs

Aaron Ellison, Harvard Forest

Analytic web "ecologists don't like to program"
HFR Boose, Ellison, Foster, Hadley
U Mass Lori Clarke, Leon Osterweil, Alexander Wise

Research foci

- document analytical processes that scientists carry out, using formalisms from software engineering
- allow for precise reproduction of results and sunthesis of multiple datasets
- enable modeling of archival and real-time data streams

An analytic web integrates threei types of graph:

- data flow graph – scientists understand
- data derivation graph – tracing backward to source of data
- process definition graph:
    - graphs using little-JIL, designed for coordinating agents
    - accomdates automated and human intervention
    - handles exceptions
    - processing sequential, parallel, either

Applications of the analytic web: modeling carbon flux using eddy-covariance
50-80% of data are thrown away. User-dependent tossing of data.

Ellison et al 2006. Analytic webs support the synthesis of ecological datasets. Ecology 87: 1345-1358.

Modeling water budgets in real time
How do scientists model key components of the water budget

Sources: met station, flux tower, sream gauge
Processiong

1. realtime data – every 30 minutes, create models
2. post-processed data
3. altnerate context

data flow graph for creating water budget
process-definition graph: each process is a set of sub-processes

Why use an analytic web?Advantages
- rigorous thinking about scientific analyses
- precise process metadata (incorporate into EML)
- reliability and reporudicibility of data and models
- test for logical and statistical errors
- easily handle loops and exceptions

Challegnes – ease o fuse by ecologists

Bill Michener, LNO
Case Studies in Ecoinformatics education: SEEK, OBFS, and international LTER

<u>SEEK</u>
Seek goals:
Designed to addtress grand challenges:
Expose data through a common architecture
Create a framework for preserving and communications compples analytic processes
Challenges oassciated with integrating heterogeneous data for use in analysis

SEEK components:
Bottom:
Access to data and tools through eco-grid (SDSC): museum data, specimen data, environmental data (SRB), geologrical data (GEON) through Kepler repository analysis libraries
Top:
Overlying this is Kepler workflow analysis, mimics process scientist engages in
Bottom and top are integrated through semantic mediation
Very top is the oommunity of users: (1) biodiversity and ecological analysis and modeling and (2) education, outreach, and training

Kepler scientific workflow system:
Software to design and execute scientific workflows

Direct access to heterogeneous environmental data, including real-time data access to sensor networks
Supports models in many science domains
Kepler components: director, actors, ports

<u>Ecoinformatics training approach</u>
Goal: to inform biologists and ecolgists about ecoinformatics
Presentations, hands on activities, demonstrations

Evaluation: Surveymonkey.com: ask students what they want, assess skill levels before, assess how training met needs

Training facilities: 4 pods of 5 computers plus instructor in center, SOTA audio-visual

Kristin Vanderbilt, Data manager at Sevilleta
Training to post-docs and new faculty
Topics:
Scientific workflows
Kepler analysis and modeling system
Grid technologies
Semantic mediation
Data acquisition: gps, imagery, wireless sensor networks
Database design, web development

Training for PBFS personnel
Topics:
NSF research coordination networks grant
One week ecoinformatics training
One week GIS training
Participants are field station personnel

Metadata: use morpho, to create EML
Database design, using access
Designing dat entry forms
QA/QC
SQL, structured query language
Networking hardware and software
Security/archival/backup
Gps
Dynamic databases, dreamweaver

International LTER ecoinformatics training
Agenda similar to OBFS
3-5 day workshops
LTER IMS are instructors
Hungary, 2000, But no proponent for info management in Hungary

Ulanbatar, Mongolia, 2001
Sosuthern Africa, 2002, but no follow up

Take home messages
Longer workshops are better
Trainees prefer hands-on training
Pre- and post-assessment is valuable
Facilities help
Followup is ideal

Training: for professionals? For academics?
University curricula lag behind demand from grad students

Steve Mitchell, PhD OSU Forest Science
Research using simulation on lt forest dynamics and fire return intervals
International internship at Lancaster Univ, Keith Beven, quantitative Bayesian hierarchical modeler
Steve had little prior knowledge needed to learn languages in a hurry
Catching up to programming capabilities and jargon of other fields
Poster looking at uncertainty in modeling net ecosystem exchange
Advice – ecoinformatics should be part of undergraduate ecology curricula

Zac Kayler,
Many aspects of his research utilizes econiformatics:
Data streams from sensor arrays, need help
Collaboration with staistician and mathematician – Zac would not have been able to find help without his collaborators
Top down approach of asking a question, worked with a mathematican from Vietnam, relatioship started out slow and then communication worked, predicting carbon storage

Chris Graham, Water resource science
BS math, PhD Forest Engineering
Uses math natural fit for him
Long-term datasets from the Andrews, data mining
Project popped up and IGERT funding

Alan Tepley,
Topography effects on fire patterns and forest succession in Oregon
Presented work on fire effects on forests
Then worked with a math student to quantitatively show how diverse successional pathways develop
Collaboration was more than sum of parts

Kristin Vanderbilt
BS and MS degrees in biology
Interested in tools as PhD student

Took a lot of extra classes
Ended up having to train pretty extensively, e.g. in databases
Skills needed for her job, programming skills

Discussion

Alan Covich, Georgia Tech and U. Georgia, stream ecology and engineering
Freshwater ecosystem services

Ecoinformatics post-docs training oversubscribed
Enthusiastic about teaching in this area
Generational change
Bridge programs in the meantime

Judy Cushing
Suggestion for mentoring folks in these programs
Encourage hiring of people with interdisciplinary training
Career awards for young faculty in this area
How can faculty help promote this area?
Top-down piece of defining questions and using ecoinformatics
Aaron: We need to reach out to people outside academia, PhDs may not get academic jobs
Most students in ecology come from other undergrad majors

Top-down questions are not being addressed
Graduate student population is not being served

Aaron Ellison
Constant retooling, why teach tools, instead teach skills for collaboration around the key top-down questions

Bringing new communities of scientists to engage with ecologists: computer scientists, mathematicians, and ecologists, facilitating collaborative interactions among these groups around key research questions

Barbara and Rebecca left biology for CS and EE and learned a way to think

Karen Baker has been creating ocean informatics environment
A learning environment that's significantly different from standard education
Give a profile to the alternative organizational structures that are addressing the needs now
Grad students are expected to know how to deal with all of this
alternative approaches to econiformatics education for various user groups
focus on living with long-term data
how to handle collaborative work

Alan Covich
other motivations are NEON is coming and EI will be needed

challenge of integrating VERY diverse data used by ecosystem scientists
training sessions could work on data standards

Bill Michener
NEON may lead to a centralized training facility
Universities don't have all the expertise, need for centralized training facility within LTER

M. Lopez
Why not create opportunities for IT people to learn ecology?

Judy Cushing
Throughout CS, analysts do not recognize that their training is applicable to other fields, like ecology

Aaron Ellison
LTER sites don't think of outsourcing information managers

Enlarge the community