Sapelo on the world stage: role of the Georgia Coastal Ecosystems LTER in global environmental data networks

Wade Sheldon
Information Manager
Georgia Coastal Ecosystems LTER
Overview

- **Background**
  - LTER Network
  - Georgia Coastal Ecosystems LTER

- **Data sharing networks**
  - Near-real-time data
  - Long-term data

- **Who is using GCE data**
  - Historical data access trends
  - LTER Network Databases (ClimDB/HydroDB)
  - Synthesis projects (EcoTrends)
  - Synthesis tool developers (Kepler)

- **Future trends in environmental data sharing**
  - Where we are now
  - “Metadata-mediated data integration”
  - “Mashups” and other end-user approaches

- **Discussion and food for thought**
Background – LTER Network

- 1977-79: Three NSF workshops to develop LTER framework
  - Philosophy of collaborative research
  - Centralized working hypothesis approach
  - 5 core research areas (1° production, populations, organic matter, inorganic matter, disturbance)

- Series of NSF DEB competitions held to establish sites
  - 1980: NTL, AND, CWT, KNZ, NWT, NIM*
  - 1981: SGS, CDR, JRN, OKE*, ILR*
  - 1987: ARC, BNZ, HBR, KBS, VCR
  - 1988: LUQ, SEV, HFR
  - 1989: Added network office (LNO)

- Expanded LTER by leveraging other NSF directorates/programs
  - Polar Programs (Antarctic sites):
    - 1991: PAL
    - 1993: MCM
  - Social, Behavioral & Economic Sciences (Urban sites):
    - 1997: CAP, BES
  - Ocean Sciences (LMER and oceanic sites):
    - 1998: PIE
    - 2000: GCE, SBC, FCE
    - 2004: MCR, CCE
Background – LTER Network

- Currently 26 sites + network office
- NSF & USFS announcing competition for new urban forest sites (ULTRA)
- Sites funded for 6 years, subject to mid-term review
  - Renewal panel held
  - Probation/withdrawn if poor review
- Research traditionally site-based, thematically tied to core areas
- “Decadal Plan” written to move towards integrated environmental & social science, education, cyber infrastructure
Background – LTER Network

- Major emphasis on Information Management in LTER
  - Standards & practices for data acquisition, quality control, curation
  - Metadata to ensure long-term usability (20 years+)
  - Online access to metadata, data, other research products
  - Promotion of data sharing, synthesis

- LTER Network Resources
  - Web site showcasing network & sites (http://www.lternet.edu)
  - Centralized databases (bibliography, data catalog, personnel, site information, climate/hydrographic data)
  - Collaboration tools (mailing lists, VTC hardware/software)
  - Conference & workshop logistical support
Background – International LTER

- LTER also participates in International LTER (I-LTER)
  - 'Network of networks'
  - Global network of research sites located in a wide array of ecosystems
  - I-LTER has a focus on long-term, site-based research
  - Collaboration encouraged but not funded directly
  - Data sharing through metadata catalog (Metacat servers)
  - WWW: http://www.ilternet.edu/
Background – GCE History

- Originated as follow-on to Georgia Rivers LMER project (Wiegert et al., 1995-2000)
  - GARLMER focused on 5 major coastal rivers (Savannah, Ogeechee, Altamaha, Satilla, St. Marys)
  - Research Goals:
    - Describing hydrodynamic characteristics of the estuaries
    - Understanding biological and biogeochemical aspects of material flux through the estuaries
    - Measuring bioactivity of constituents passing through distinct estuarine communities
    - Simulation modeling of interactions between the terrestrial, intertidal and nearshore systems
  - LMER program folded into LTER, invited to write proposal

- Georgia Coastal Ecosystems LTER time line
  - 1999: Submitted proposal to NSF
  - 2000: GCE project started (May 2000)
  - 2003: Mid-term NSF review (May 2003)
  - 2006: GCE-II project started (Dec 2006)
  - 2009: Mid-term NSF review (Oct 2009)
Background – GCE Research

GCE-I Research Questions:
- How are the effects of spatially and temporally variable patterns of terrestrial, oceanic, and atmospheric inputs propagated through coastal ecosystems?
- How does the resulting variation in nutrients, organic matter, and salinity affect processes in coastal wetlands?

Study Plan:
- Modeling
  - Mechanistic ecological model
  - Modeling and measuring tidal flows and material exchange
  - Effect of land-use changes on Altamaha River water quality
  - Biogeochemical modeling
- Monitoring
  - Altamaha River, groundwater, sounds and tidal creeks, marshes
- Directed Studies of Marsh Structure and Function
  - Angiosperm Production, Community Structure and Population Genetics
  - Prokaryotic-Eukaryotic Decomposer Consortia
Background – GCE Research

- **GCE-II Research Questions:**
  - **Question 1:** What are the long-term patterns of environmental forcing to the coastal zone?
    - Climate signals
    - Watershed inputs
  - **Question 2:** How do the spatial and temporal patterns of biogeochemical processes, primary production, community dynamics, decomposition, and disturbance vary across the estuarine landscape, and how do they relate to environmental gradients?
    - Salinity structure
    - Patterns of dissolved and suspended material
    - Soil processes
    - Plant, animal, microbial dynamics
    - Duplin River interdisciplinary study
  - **Question 3:** What are the underlying mechanisms by which the freshwater-saltwater gradient drives ecosystem change along the longitudinal axis of an estuary?
    - Salt & sulfate interdisciplinary study
    - Associated projects (EPA studies)
Background – GCE Research

GCE-II Research Questions:

- Question 4: What are the underlying mechanisms by which proximity of marshes to upland habitat drives ecosystem change along lateral gradients in the intertidal zone?
  - Hammock interdisciplinary study
  - Groundwater dynamics

- Question 5: What is the relative importance of larval transport versus the conditions of the adult environment in determining community and genetic structure across both the longitudinal and lateral gradients of the estuarine landscape?
  - Population studies
Background – GCE Information Management

- GCE IM Program
  - Developed highly automated information system
    - MATLAB-based data processing tools, automated data harvesting systems (GCE Data Toolbox)
    - Integrated relational databases (all project information)
    - Dynamic web applications, XML-based web services to search/display database content
Background – GCE Information Management

- Archived broad range of data from GCE and related research
  - >300 accessioned “core” data sets (GCE Data Catalog)
  - >250 value-added ancillary data sets (GCE Data Portal)
  - Dynamic USGS data harvesting service for LTER sites, network

- Established GIS infrastructure (GCE-II)
  - ArcGIS software/license servers (UGA & UGAMI)
  - Centralized GeoDatabase Server (raster & vector data)
  - High resolution GPS
Overview

- Background
  - LTER Network
  - Georgia Coastal Ecosystems LTER

- Data sharing networks
  - Near-real-time data
  - Long-term data

- Who is using GCE data
  - Historical data access trends
  - LTER Network Databases (ClimDB/HydroDB)
  - Synthesis projects (EcoTrends)
  - Synthesis tool developers (Kepler)

- Future trends in environmental data sharing
  - Where we are now
  - “Metadata-mediated data integration”
  - “Mashups” and other end-user approaches

- Discussion and food for thought
Data Sharing Networks – Real-time Data

- Marsh Landing climate station (SINERR collaboration)
  - 15-minute met data
  - GOES Satellite uplink
  - Data acquired by NOAA HADS, broadly distributed (NWS, etc)
  - Data archived by GCE, NERR
  - Data transmitted to LTER ClimDB

- Hudson Creek climate/hydro (USGS/SINERR collaboration)
  - 30-minute met/hydro data
  - Microwave uplink
  - Data acquired by USGS
  - Data archived by GCE, USGS
  - Data transmitted to LTER ClimDB, HydroDB
Data Sharing Networks – Long-term Data

- **NADP (GA-33) (SINERR/UGAMI collab.)**
  - Daily precip, weekly conc. of Ca, Mg, K, Na, NH₄, NO₃, Cl, SO₄, pH
  - Analyses performed by NADP
  - Data available from NADP, broadly accessed
  - Data archived by GCE

- **LTER ClimDB/HydroDB**
  - All-site climate/hydrology database
  - Contribute Marsh Landing, Hudson Creek, Altamaha discharge, UGAMI/NWS
  - Daily stats for all parameters
Data Sharing Networks – Long-term Data

- GCE Data Catalog – data sharing through metadata exchange

![LTER Data Catalog](image)

![KNB Metacat](image)

![Geospatial One Stop](image)

![NBII Clearinghouse](image)

![NASA GCMD](image)
Overview

◆ Background
  • LTER Network
  • Georgia Coastal Ecosystems LTER

◆ Data sharing networks
  • Near-real-time data
  • Long-term data

◆ Who is using GCE data
  • Historical data access trends
  • LTER Network Databases (ClimDB/HydroDB)
  • Synthesis projects (EcoTrends)
  • Synthesis tool developers (Kepler)

◆ Future trends in environmental data sharing
  • Where we are now
  • “Metadata-mediated data integration”
  • “Mashups” and other end-user approaches

◆ Discussion and food for thought
Who’s Using GCE Data – Trends

- Public downloads (Data Catalog & Data Portal)
Who’s Using GCE Data – ClimDB/HydroDB

- Access by year (all data access – not just GCE):

<table>
<thead>
<tr>
<th>Year</th>
<th>files</th>
<th>plots</th>
<th>views</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>309</td>
<td>1240</td>
<td>291</td>
</tr>
<tr>
<td>2004</td>
<td>267</td>
<td>566</td>
<td>98</td>
</tr>
<tr>
<td>2005</td>
<td>717</td>
<td>829</td>
<td>191</td>
</tr>
<tr>
<td>2006</td>
<td>1886</td>
<td>978</td>
<td>335</td>
</tr>
<tr>
<td>2007</td>
<td>946</td>
<td>816</td>
<td>210</td>
</tr>
<tr>
<td>2008</td>
<td>1259</td>
<td>888</td>
<td>347</td>
</tr>
<tr>
<td>2009</td>
<td>48</td>
<td>17</td>
<td>1</td>
</tr>
</tbody>
</table>

- Database statistics
  - > 9 million daily values in database
  - 324 measurement stations
  - 26 LTER sites
  - 24 USFS sites
  - 12 sites with USGS stations (data from GCE)
  - 21 total measurement parameters
Who’s Using GCE Data – EcoTrends

- EcoTrends started as LTER synthesis project ~2006
  - Designed to promote and enable the use and synthesis of long-term data to examine these trends in the Earth's ecosystems
  - Acquiring long-term monthly/yearly data from LTER, ARS, USFS, ...
  - Promoting synthesis activities and use of long-term data, not only by scientists, but also more generally by students, teachers, and decision makers.

- GCE contributed climate, nutrients, sea-level, atmospheric deposition data

- Products
  - Printed folio-size book (in editing)
  - Interactive web portal (in beta)
    - Large and diverse collection of standardized long-term ecological datasets
    - Unique data exploration, download, graphing and synthesis tools
    - Information about participating research sites and their parent agencies as contextual information to assist in data and trend interpretation

- Web site: http://www.ecotrends.info/
Who’s Using GCE Data – EcoTrends
Who’s Using GCE Data – Kepler Project

- Kepler – “an analytical system that allows users to construct scientific workflows that model the way data flows among discrete analytical components”

- Java-based graphical programming environment

- Data sources, components (“actors”), workflow directors laid out on graphical canvas
  - Dataset columns configured as “ports”
  - Connector wires used to link inputs, outputs
  - Workflows can be saved, shared, edited and re-run
  - Both local and network data sources can be used in workflows

- GCE public data access through Kepler via Metacat network
  - Metadata retrieved from Metacat used to configure ports
  - Data dynamically retrieved from GCE server when workflow run

- Web site: http://kepler-project.org/
Who’s Using GCE Data – Kepler

- Powerful environment for streaming data processing, manipulation
- Actors for running R, MATLAB – potential for advanced analyses
- Selection of stats actors limited, under-documented, buggy
- Hard to inspect structure of data, choose good analyses
- Work in progress, but promising technology
Overview

◆ Background
  • LTER Network
  • Georgia Coastal Ecosystems LTER

◆ Data sharing networks
  • Near-real-time data
  • Long-term data

◆ Who is using GCE data
  • Historical data access trends
  • LTER Network Databases (ClimDB/HydroDB)
  • Synthesis projects (EcoTrends)
  • Synthesis tool developers (Kepler)

◆ Future trends in environmental data sharing
  • Where we are now
  • “Metadata-mediated data integration”
  • “Mashups” and other end-user approaches

◆ Discussion and food for thought
Future trends

◆ Where are we now
  • Catalogs, portals galore
  • Myriad search interfaces, data formats, metadata formats
  • Data integration largely manual, tedious

◆ Emerging approaches
  • “Metadata-mediated data integration”
    ♦ Computer programs use structured metadata to decipher data formats automatically, ease analysis and transformation
    ♦ GCE Data Toolbox uses this approach
    ♦ LTER “PASTA” framework for EcoTrends
    ♦ “Semantic Mediation” based on “Ontologies” being explored to provide more robust data descriptions for machine operations
  
  • “Mashups” for geo-coded data (Google Earth)
    ♦ Google tools providing novel capabilities for researches
    ♦ Geocoded data sets can be integrated easily, visualized
    ♦ Time sliders can also be used to visualize changes in data over time, trends
Overview

- **Background**
  - LTER Network
  - Georgia Coastal Ecosystems LTER

- **Data sharing networks**
  - Near-real-time data
  - Long-term data

- **Who is using GCE data**
  - Historical data access trends
  - LTER Network Databases (ClimDB/HydroDB)
  - Synthesis projects (EcoTrends)
  - Synthesis tool developers (Kepler)

- **Future trends in environmental data sharing**
  - Where we are now
  - “Metadata-mediated data integration”
  - “Mashups” and other end-user approaches

- **Discussion and food for thought**
Discussion

- NSF is heavily focusing on data archiving, sharing and re-use (top-down pressure on scientists); does this foster new discoveries or take resources from primary research?

- If you have used archived data, how easy was it to understand the research origin (methodology, location, ...)?

- What key info needs to accompany data to prevent misinterpretation and misuse?