Georgia Coastal Ecosystems LTER
Information Management

Wade Sheldon, Information Manager
Adam Sapp, Assistant Information Manager
University of Georgia
Athens, Georgia
GCE IM Overview

- Information Management Program
- Current Challenges
- LTER Network IM Participation and Leadership
- GCE IM Outreach and Leveraged Projects
- Future Directions
Goals guiding GCE IM program development

- Develop procedures and technology to facilitate efficient acquisition, standardization, analysis and synthesis of all GCE data and metadata
- Develop an integrated information system to manage all products of GCE research, support site research, and build an archive of long-term ecological observations
- Establish public and private web sites to provide convenient access to project information and research results for GCE members, the LTER Network, broader scientific community and public
- Integrate with the LTER Network Information System to support network-level administration, cross-site comparisons and large-scale synthetic research
Develop Procedures and Technology

Data acquisition and submission

- Focused on aiding the research process, not impeding it
- Provide automatic harvesting, primary processing for sensor data (monitoring data as a service)
- Support submissions in various formats used by each investigator – developed tools to import/standardize
- Support data documentation and augmentation
  - Re-useable metadata templates
  - Automated metadata generation from data introspection, analysis
  - Geographic lookups, taxonomic lookups, coding/decoding, ...
- Provide QA/QC assistance, statistical reports, derived data sets as part of submission process (value added data sets)
Develop Procedures and Technology

- Developed standard data model for all tabular data
  - System metadata
  - Documentation metadata
  - Processing history
  - Attribute metadata
  - Q/C rules
  - Data arrays
  - Q/C flag arrays
Develop Procedures and Technology

- Developed software library for working with standard – GCE Data Toolbox for MATLAB
  - Metadata-driven processing
  - Integrated QA/QC framework
  - Flexible import/export for data & metadata (text files, logger formats, MATLAB arrays/structs, SQL databases)
  - Support for synthesis (re-sampling, unit conversions, data integration)
  - Automated statistical analysis, plotting, mapping

- Supports automated “workflow” processing, data harvesting, server-based applications, and GUI-driven end-user processing

- Provide software to scientific community (~3600 downloads)

- Used at 8 LTER sites

- Used in Marine Science methods class at UGA
Develop Procedures and Technology

- Search tool allows users to discovery and download GCE public data
- Data and metadata can be viewed, and data can be plotted, analyzed, refactored and integrated “on the fly”
**Develop Procedures and Technology**

- Developed support for near-real-time data acquisition (harvesting)
  - USGS NWIS, NOAA NCDC/GHCN/HADS, CSI LoggerNet
Develop Procedures and Technology

- Rapid processing and web posting of data and visualizations from field instruments (CTDs, sondes, weather stations, flux tower)
IM Vision and Accomplishments

- Goals guiding GCE IM program

  - Develop procedures and technology to facilitate efficient acquisition, standardization, analysis and synthesis of all GCE data and metadata
  
  - Develop an integrated information system to manage all products of GCE research, support site research, and build an archive of long-term ecological observations
  
  - Establish public and private web sites to provide convenient access to project information and research results for GCE members, the LTER Network, broader scientific community and public
  
  - Integrate with the LTER Network Information System to support network-level administration, cross-site comparisons and large-scale synthetic research
Develop an Integrated Information System

- Established a strong core IT infrastructure to support IM operations
  - 4 dedicated Dell servers at UGA Marine Sciences
    - Web server, database server, GIS server, software development server
    - All highly fault-tolerant, high availability
    - >6 TB spinning disk, >12 TB LTO tape, offsite backup (CWT)
  - 4 dedicated IM workstations at UGA Marine Sciences
  - 1 dedicated GIS workstations at UGAMI (GIS lab)
  - Desktops and laptops for field staff

- Provide secure storage, backup for GCE field and lab data
- Provide collaboration space for working groups (shared network drives)
- Provide software version control system (Subversion)
- Use institutional and LTER Network resources for basic IT services
  - Email, mailing lists, networking, VTC, desktop support
Develop an Integrated Information System

- GCE Information System
- Database Server
- GIS Server
- Web Server
- Secure Protocols (SSH, SSL, RDP)
- Public URL Filter
- UGA Perimeter
- IPS
- Firewall
- GCE Members
- GCE Partners
- GCE-IM Workstations
- GCE-IM Notebooks
- Development (SVN)

GCE – Information Management
Develop an Integrated Information System

GCE Information Management System

GCE Software
- GCE Data Toolbox
- PyGIS Library
- IIS/ASP + SQLXML

GCE GIS
- ArcSDE
- GeoDatabase

GCE Databases
- Access
  - Access Logs
- Surveys
  - Votes
- Bibliography
  - Citations
  - Reprints

Metadata
- Species Lists

Submissions
- Calendar
- Documents
- Images
- ProjectDB

Taxonomy
- Calendar
- Documents
- Images
- ProjectDB

Products/Services
- Data (.csv, .mat, .txt, .shp, .grd, .xml)
- Metadata (EML/XML)
- Web Pages (XHTML, XML+XSLT)
- Web Services (KML, XML, JSON)
- App Data (ArcGIS, MATLAB, SQL Server, eXist/XML)

Services
- Web Browser Interface (HTTP/HTTPS)
- Direct Connection (SSH/RDP/ODBC)
- Machine Interface (XML/JSON/Files)

User Groups
- LTER/Public
- GCE Members
- GCE IM/Techs
- LTER NIS

GCE – Information Management
Develop an Integrated Information System

- Added major GIS component in GCE2, maintained in GCE3
- Established centralized Geo-database server for GCE (ArcSDE)
  - Reference vector data (roads, shorelines, sampling locations)
  - Reference raster data (aerial photos, satellite imagery, IR)
  - Research data (geo-databases, GPS data, LiDAR, DEMs)
  - Dynamic mapping web services for data sets, sites, projects (Google Earth KML)
  - Extending to support Remote Sensing products
Develop an Integrated Information System

- Also provide research support by assisting with geospatial analysis

Duplin DEM

Vegetation mapping
Develop an Integrated Information System

Changes and challenges for GCE3 (yr 4-6)

- Independent “satellite” information systems for RS, modeling
  - Scalles (Creighton) – NAS Server, workstations for Landsat
  - Meile (UGA) – Mac cluster for biogeochemical modeling
  - Burd (UGA) – Mac cluster for plant model, integrative modeling
  - Castelao & Dilorio (UGA) – Beowulf cluster for hydrodynamic model (moving to HPC nodes at UGA)
  - Alexander (SkIO) – GIS lab and workstations for new UAV
  - Mishra (UGA) – remote sensing lab for MODIS, PHENOCAM

Strategies

- Primary data used for analysis, parameterization managed in GCE-IS
- 3rd party data (imagery, bathymetric grids) managed by PI
- Research products documented, archived in GCE-IS
- Custom code committed to GCE SVN or archived
- Standards for model archiving under discussion in LTER
Methodology

DIGITAL DATA

Level 3, 8 day, (250m and 500m) surface reflectance images for the growing season (March - November) from 2000-'13

Terra

Salt Marsh boundaries

ANALYSIS

Image Pre-processing: Mosaicking, Geo-referencing, and Subset preparation

Calibrations and Model validations (VF, GBM, CHL and GLAI)

Weekly and Monthly composites

IN-SITU DATA

Green Vegetation Factor, Biomass, Leaf Chlorophyll, and Leaf Area Index

END PRODUCTS
Goals guiding GCE IM program

- Develop procedures and technology to facilitate efficient acquisition, standardization, analysis and synthesis of all GCE data and metadata
- Develop an integrated information system to manage all products of GCE research, support site research, and build an archive of long-term ecological observations
- Establish public and private web sites to provide convenient access to project information and research results for GCE members, the LTER Network, broader scientific community and public
- Integrate with the LTER Network Information System to support network-level administration, cross-site comparisons and large-scale synthetic research
Establish Public and Private Web Sites

- During GCE1 we established 3 web sites to meet various end-user needs
  - Public site – general project information, access to research products
  - Private site – password-protected access to raw/provisional data, sensitive files, forms
  - Data Portal – public access to standardized near-real-time and ancillary data

- During GCE2 we integrated all sites into a seamless framework, using database-driven pages and web services to provide dynamic access to the GCE IS

- During GCE3 we are maintaining these websites and services while exploring new mobile-friendly interfaces in collaboration with other LTER sites
Establish Public and Private Web Sites

- General project information
  - GCE news (announcements, calendar, data releases, file uploads)
  - Current field conditions (tide predictions, near-real-time weather and hydrographic data plots, logistics)
  - Research program description (questions, components, products)
  - Personnel contact information
  - IM, GIS, Schoolyard, Outreach information

- GCE research products
  - Data Catalog, Data Portal, Provisional Data
  - Bibliography (GCE, UGAMI, GARLMER), reprint library
  - Species lists (photos, ancillary information)
  - General file/imagery archive (GIS files, photos, documents)

- Links to related programs, institutions, data sources
Establish Public and Private Web Sites

- Web pages are dynamically cross-linked for discovery of related information.
Establish Public and Private Web Sites

- GIS connected to web site through web services and map interfaces
- Access to site info and research data via links in placemark balloons
Establish Public and Private Web Sites

- We developed a generalized file archive for documents/imagery
- Archive cross-referenced to bibliography and taxonomic database photos
- All GCE members can contribute, update their entries

GCE – Information Management
Establish Public and Private Web Sites

- Research projects and reporting database (ProjectDB)
  - Links research objectives, people, products, geography
  - In GCE3 aligned with proposal objectives for accountability
  - Assists with project management, NSF reporting, product discovery
  - XML Schema collaboratively designed with IMs from 11 LTER sites
Establish Public and Private Web Sites

- Research registration system developed for GCE3
  - Applications reviewed by field crew, IM, lead PIs to manage resources, conflicts
  - Ensures compliance with Marsh Protection Act, facilitates CRD/NOAA permitting
  - Compatible system developed for SINERR, UGAMI for info sharing
  - Approved projects registered in ProjectDB, linked to objectives, tracked
  - Field tags generated for display at sites for contact info, ensuring clean up
IM Vision and Accomplishments

- Goals guiding GCE IM program
  - Develop procedures and technology to facilitate efficient acquisition, standardization, analysis and synthesis of all GCE data and metadata
  - Develop an integrated information system to manage all products of GCE research, support site research, and build an archive of long-term ecological observations
  - Establish public and private web sites to provide convenient access to project information and research results for GCE members, the LTER Network, broader scientific community and public
  - Integrate with the LTER Network Information System to support network-level administration, cross-site comparisons and large-scale synthetic research
Integrate with the LTER Network Information System

- All GCE databases designed to support LTER protocols
  - Comprehensive Ecological Metadata Language (EML) support
    - Data set metadata, taxonomic information, research projects, ...
    - Automated EML generation, versioning in GCE Data Catalog
    - EML import/export in GCE Data Toolbox software
    - GCE data published in PASTA for distribution through LTER Data Portal
  - Bibliographic export in EndNote format for all-site bibliography
  - Generalized ClimDB/HydroDB export utility in GCE Data Toolbox
  - Web service schema support (PersonnelDB, ProjectDB, SiteDB)

- Actively synchronize personnel, site information to LNO databases

- Participate on LTER committees, working groups to improve site-network interoperability

- Support LTER X-site science initiatives
GCE IM Overview

- Information Management Program
- Current Challenges
- LTER Network IM Participation and Leadership
- GCE IM Outreach and Leveraged Projects
- Future Directions
Current Challenges

- New GCE3 research initiatives bringing in new data types
  - Genetics/genomics
  - Acoustics (ADCP)
  - Model output, reference data
- Need to develop strategies to archive/share
- LTER lacks standards to document many of these types
- Working with investigators to identify ways to leverage existing repositories and add value from GCE research
  - GenBank accessions linked to environmental data dynamically
  - Online descriptions, request forms for shipping high volume data on portal HD
GCE IM Overview

- Information Management Program
- Current Challenges
- LTER Network IM Participation and Leadership
- GCE IM Outreach and Leveraged Projects
- Future Directions
LTER Network IM Participation and Leadership

- Very active at all levels in LTER Network
  - **W. Sheldon**
    - IM-Exec (2003-2006)
    - NISAC (2007-2012)
    - IM working groups (EML best practices, controlled vocabulary, projects database, web services, personnel database)
    - Led quality control workshop (2007)
    - Co-led 2 PASTA workflows workshops (2012-2013)
    - Co-led Sensor Data Management training workshop (2013)
    - Participated in next generation LNO visioning process (2013-14)
    - Participating in Distributed LTER IM visioning process (2015)
    - Edited 4 editions of DataBits Newsletter
  - **A. Sapp**
    - Assisted with GCE Data Toolbox training at CWT
GCE IM Overview

- Information Management Program
- Current Challenges
- LTER Network IM Participation and Leadership
- GCE IM Outreach and Leveraged Projects
- Future Directions
Using GCE Data Toolbox to harvest USGS streamflow data weekly for 12 LTER sites and 1 USFS site for HydroDB (61 stations)
IM Outreach and Leveraged Projects

- Coweeta LTER:
  - Leveraged GCE-IS in 2013 (Metabase, Biblio, Projects, File archive)
  - Leveraged GCE Data Toolbox for streaming data
  - Provide IIS/ASP hosting (reverse proxy), database support
  - Near-real-time streamflow for 6 stations (http://coweeta.uga.edu/dbpublic/hydrologic_data.asp)
IM Outreach and Leveraged Projects

- AND, HBR, NWT, SEV, NTL
- Leveraged GCE Data Toolbox for streaming sensor data management
IM Outreach and Leveraged Projects

- Bibliography Hosting
  - Hosting UGAMI bibliography (1955-present, ~1000 citations) & GARLMER
  - Developed light-weight search form for embedding on UGAMI web site (including reprint requests skinned to match site)
IM Outreach and Leveraged Projects

- NPS SE Water Quality Monitoring Database
  - Developed database, web portal for NPS (leveraged project)
  - Metadata on >300k parameters at ~17k sites along SE coast
  - Included dynamic Google Maps, web services, multiple output formats
  - http://www.gcrc.uga.edu/wqmeta/
IM Outreach and Leveraged Projects

- Developed database, web portal for NPS (leveraged project)
- Metadata on >300k parameters at ~17k sites along SE coast
- Included dynamic Google Maps, web services, multiple output formats
- http://www.gcrc.uga.edu/wqmeta/
GCE IM Overview

- Information Management Program
- Current Challenges
- LTER Network IM Participation and Leadership
- GCE IM Outreach and Leveraged Projects
- Future Directions
Future Directions

- Upgrading GCE servers in 2015
  - Latest OS and software versions (Win Server 2012, SQL Server 2014)
  - New data server being ordered – consolidate database/GIS
  - Other systems upgraded, refreshed (DB -> web -> software dev)
  - Looking for cost-effective virtual server/storage options

- Collaborating with other LTER sites on shared approaches to IM
  - Visioning for distributed IM services/systems
  - More X-site software/system leveraging

- Website redesigns being considered
  - Mobile-friendly template (or sub-site/redirect)
  - More fluid, up-to-date design

- Much uncertainty – keeping options open