Global Land Cover Facility

Open House/ Project Initiation

The following are presentations given at the GLCF Open House in AV Williams in October 1998.
The Global Land Cover Facility: First Steps in Implementation

Francis E. Lindsay,
Institute for Advanced Computer Studies
University of Maryland, College Park

James B. Humphries,
Institute for Advanced Computer Studies
University of Maryland, College Park

Catherine Plaisant,
Human Computer Interaction Laboratory
University of Maryland, College Park
What are the Issues of Earth Science Data Access and Use?

1. FINDING THE DATA
   - where do users find data sources?
   - what is the reference context

2. FINDING OUT ABOUT THE DATA
   - metadata
   - temporal and spatial extent

3. GETTING THE DATA
   - data costs
   - on-line versus hard media sources
   - data set size

4. USING THE DATA
   - data formats
   - pre-processing
   - data integration
Our Target Audience(s)

◆ Our main user community will be scientists who need land cover products to further their **Earth Science System** goals.

◆ Science users do not form a homogeneous community but vary across several dimensions including:
  ♦ need for high to low-level products
  ♦ very large to very low volumes of data
  ♦ a broad range of science applications

◆ We anticipate many other types of users as well, including K-12 students, universities, state and local governments, non-profit organizations and commercial vendors.
<table>
<thead>
<tr>
<th>YEAR</th>
<th>Available now!</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>US Coastal Wetlands Health Products&lt;br&gt;Tropical Humid Rain Forest Land Cover&lt;br&gt;Coarse Resolution Global Land Cover&lt;br&gt;Landsat Imagery</td>
</tr>
<tr>
<td>2</td>
<td>User-defined 1 km NOAA products for the land&lt;br&gt;Atmospherically corrected products from Landsat 7&lt;br&gt;Improved global land cover products</td>
</tr>
<tr>
<td>3</td>
<td>Processing of data using the global hierarchical indexing system&lt;br&gt;Land cover products from MODIS &amp; Landsat&lt;br&gt;A Degradation Early Warning System (DEWS)&lt;br&gt;Annual state of global land cover change</td>
</tr>
</tbody>
</table>
Getting Data Products to the Land Cover Research Community

We make GLFC data sets available in a number of differing formats and media types.

- **FTP** for most data sets (depending on size and compression)
- **CDs** for individual and compilation data sets
- **4mm, 8mm, and DAT Tapes** for specialty orders and very large data sets
- **DVDs** for very large data sets (100+ gigabytes)
The Global Land Cover Facility Web Site

- Built in features for users to preview, explore and order needed data.
- Emphasis on rapid data access and useful explanations of data holdings.
- Promote the land cover research community by fostering a forum for information exchange.

http://glcf.umiacs.umd.edu
Improving the User Interface

The End of Zero-Hit Queries: Query Previews for NASA’s Global Change Master Directory

... dynamic query user interfaces, visualization, and overview + preview designs

The Human-Computing Interaction Laboratory (HCIL) group explores how users access data and interact with on-line interfaces.

97-20
Greene, S., Tanin, E., Plaisant, C., Shneiderman, B., Olsen, L., Major, G., Johns, S.
Sample of Web Preview Tools Interface
Existing Earth Science Data Distribution Systems

The Internet has spawned the next generation of data access outlets relating to earth sciences. Currently, the Internet offers numerous sites containing data of use to Earth Science Researchers.
The GLCF Interface
Objectives

• **Assist** the user in finding useful data by…
  – Types of available data
  – Coverage of available data
  – Quality of the available data
  – Other metadata

• **Allow download or ordering of data**
  – Unmodified data (by us)
  – Header and metadata information
Further Objectives

• **Allow visualization and browsing**
  – Overlay multiple data sets at once
    • Independent of data type
    • Independent of map projection
  – Geographic Information System (GIS) functionality
  – Dynamic query previews

• **Experiments in on-demand processing**
  – e.g. 10 year NOAA-AVHRR composite
    • (>3.5 TB of data)
    • Data conversion
Current Version

• Binary geographical coverage previews
  – Show location of available data (using a 5 degree grid)

• Layered data previews
  – Landsat and NOAA satellite images with geographic information system (GIS) data

• Processing
  – On demand processing of NOAA’s AVHRR
    • compositing, atmospheric correction, and more

• Downloadable via http connection
The GLCF End-to-End System

- WWW Interface
- GLCF Web Site
- On-demand Processing
- Kronos Processing
- Database
- IBM DB-2
- Hierarchical Storage
- HPSS System

users
KRONOS: A Java-Based Software for Generating Specialized Data Products Based on the NOAA AVHRR Sensor
EARTH SYSTEM SCIENCE:
NEW CHALLENGES FOR COMPUTER SCIENCE.

John Townshend
Institute for Advanced Computing Studies and Department of Geography
REMOTE SENSING AND EARTH SYSTEM SCIENCE.

• Remote sensing from space is revolutionizing the way we look at the Earth.
  – New sensors, improved frequency, increasing length of records.

• But the data will only be useful if we process and distribute information products effectively.

• The key to success is creative collaboration between Earth System and Computer Scientists.
The Global Land Cover Facility

• UMCP successfully competed for one of the NASA-funded Earth System Information Partnerships (ESIP).

• As a result we have founded the Global Land Cover Facility building on multiple existing activities in Earth System Science and Computer Science.

• Involves collaborations on campus and with many other partners (e.g. with NASA/GSFC and San Diego Supercomputing Center)
The Mission of the Global Land Cover Facility

• To provide the Earth System Science user-community with new products, and services
  – Better
  – Faster
  – Cheaper

• To work within the framework of the new NASA Federation to maximize the overall performance of the Federation
Recent Highlights of Earth System Science Activities

• Humid tropical rain forest monitoring
• Monitoring of the coastal wetlands of the eastern United States
• Producing a base-line for the Framework Convention for Climate Change
Humid Tropical Rain Forest Monitoring

• To reduce uncertainty in the global carbon budget.
• Largest ever use of Landsat data.
• Demonstrated that rates of tropical rainforest destruction, although high, are much less than official international estimates.
• Jointly with Universities of Michigan and Virginia and NASA/GSFC.
Monitoring of coastal marshes in the Eastern United States

- Rapidly declining Marshlands due to Sea Level Rise.
- Use of Landsat data with algorithms developed in Computer Science.
- 1st comprehensive assessment.
- Probable loss of 75% of the marshes in the Chesapeake by the middle of the next century.
- PI Michael Kearney.
Mapping the Earth’s Forests: a contribution to the FCCC.

- Kyoto Protocol added “forest cover” to CO2 emissions for Convention.
- But there is no base-line for the early 1990s available as yet.
- Using results from global assessments of land cover using data from NOAA’s Weather satellites.
- Collaboration with the USGS: PI Ruth DeFries.
Campus involvement with new sensor technologies

- Vegetation Canopy Lidar (PI Ralph Dubayah)
- Landsat 7 Science Team Chair. (Sam Goward)
- EOS MODIS Team Memberships (Eric Vermote and John Townshend)
- Commercial Data Buy Activities
Challenges for Computer Science

• Very large volumes of new Data from new sensors of NASA’s Earth System Enterprise.
  – Multi-terabyte volumes and soon multi-petabyte volumes.

• Improving existing data sets
  – (collaborations between Eric Vermote - UMCP and Jim Tucker - NASA/GSFC).

• More rapid delivery to the Earth Science user-community.
Further challenges

• Integrating data sets from different sensors.

• Creating new products:
  – Running biophysical models with remote sensing data, e.g., the Earth’s biological productivity (Steve Prince and Sam Goward).
  – Atmospherically correcting satellite data.
  – Change detection of the whole of the Earth’s land cover.
What does “better, faster, cheaper” mean to the GLCF?

- Reducing the time from research to product and service delivery.
- Improving the user’s Search Process:
  - improving human-computer interactions.
- Users specifying their own products.
- Users linking data sets to biophysical models
- Exploiting the power of high performance computing.
The GLOBAL LAND COVER FACILITY
SUMMARY

• CURRENTLY AVAILABLE NOW
  • The GLCF Web site offers data sets via FTP
  • Distribution of data on CD (Global Land Cover and Marsh Health Data)
  • On-demand processing proto-type using NOAA satellite data
• COMING SOON (next version by Dec 1, 1998)
  • Integrated web and visualization tools
  • Operational on-demand processing
  • Increased land cover data holdings (<1 terabyte)
  • End-to-end system using high performance computing
• NEAR-TERM GOALS
  • Enhanced data base functionality and ingestion of metadata
  • User-defined processing, using novel algorithms
  • Rapid indexing of very large data sets (>5 terabytes)
Members and Associates of
The Global Land Cover Facility

**PIs:**
John Townshend, Department of Geography and UMIACS
Joseph Ja’Ja’, Director UMIACS, Professor Department of Electrical Engineering

**Earth Science Team:**
John Townshend, Chair
Ruth Defries, Department of Geography
Bruce Douglas, Department of Geography
Ralph Dubayah, Department of Geography
Sam Goward, Department of Geography
Satya Kalluri, Department of Geography
Michael Kearney, Department of Geography
Glen Moglen, Department of Civil Engineering
Steve Prince, Department of Geography
C.J. Tucker, NASA Goddard Space Flight Center
Eric Vermote, Department of Geography and GSFC

**Computer Science Team:**
Joseph Ja’Ja’, Chair
Larry Davis, Department of Computer Science and UMIACS
Jeffrey Masek, Department of Geography
Catherine Plaisant, HCIL and UMIACS
Nick Roussopoulos, Computer Science and UMIACS
Naomi El Saleous, Department of Geography
Ben Shneiderman, Computer Science, HCIL and UMIACS

**Implementation Team:**
Jim Humphries, Implementation Team Leader
Francis E. Lindsay, Science Project Manager

*Programming:*
Bin Gan
Gary Jackson
Jim Kukla
Mike McGann
Allan Tong

*Web Development:*
Frederick J. McCall
W.J. Von Alt

*Customer Service:*
Jenny Hewson-Scardelletti

**Other Collaborators:**
Vivre Bell, Department of Geography
Paul Davis, Department of Geography
Chris Justice, University of Virginia
Rachel Pinker, Department of Meteorology
The GLOBAL LAND COVER FACILITY
Open House

September 25, 1998

Room 2460 A.V. Williams Building
University of Maryland
College Park, Maryland

11:30 - 11:35  Introduction/Welcome (Townshend)
11:35 - 11:40  Remarks from the Provost (Gregory L. Geoffroy, University of Maryland, College Park)
11:40 - 11:45  The ESIP Federation Concept (John Hrastar, GSFC)
11:45 - 11:55  Earth Systems Science: New Challenges for Computer Science (Townshend)
11:55 - 12:05  High Performance Computing Infrastructure for Earth Science (JaJa)
12:05 - 12:15  The Global Land Cover Facility: First Steps in Implementation (Lindsay)
12:15 - 12:25  Advanced Web Tools and Data Processing: Demonstration (Humphries)
12:25 - 12:35  Tour of Super Computing Facilities
12:35 - 1:00+  Lunch and Open Discussion
High Performance Computing
Infrastructure for Earth System Science

Joseph JaJa

Institute for Advanced Computer Studies
University of Maryland at College Park
Main Objectives

- **ESIP:** Provision of novel land cover products and related information services in support of the Earth system science research community, federal agencies, state and local governments, education and the public.

- **Infrastructure:** Development and integration of computer science technologies to support efficient and seamless access, discovery, generation, cross-correlation, and visualization of large scale earth systems science data products.
Contributing Computing Projects

- High Performance Computing for Land Cover Dynamics: NSF Grand Challenge Project
- National Partnership of Advanced Computational Infrastructure led by the San Diego Supercomputer Center.
- Interactive Interfaces for EOSDIS
- AMASE : An Object-Oriented Metadatabase Catalog for Accessing Multi-Mission Astrophysics Data.
- Two Major NSF Infrastructure Awards in High Performance Computing and Networking.
A Possible Scenario

Understanding of Net Primary Production (NPP) variations caused by changes in land use/land cover over the Amazon.

Define a spatio-temporal Query over the Amazon

Get the corresponding land use/land cover/deforestation maps derived from MSS/TM data at different time periods

Generate the appropriate daily/composite AVHRR data

Evaluate causes of changes in productivity in relation to Land cover change

Plug the data in to GLOPEM

(Plug in module available at GLCF)

Estimate changes in productivity
Main Technology Requirements

• Hardware: Distributed, Scaleable, and Open High Performance Processing and Archiving Architecture with High I/O Throughput.

• Software: Development and integration of fundamental computing technologies for:
  – Efficient search across multiple data products through an intuitive Web interface.
  – Generation of user-specified products, integrated with the search and discovery tools.
  – Ability of cross-correlation/visualization of multiple data sets.
  – Ease of incremental evolution and integration of metadata and data sets through a scalable architecture.
  – Efficient access and use by low-end users.
* Each node equipped with 2 4.5 Gigabyte disk drives
Major Computer Science Efforts

- Interface Design Based on Extension of Dynamic Query and Query Preview (B. Shneiderman, C. Plaisant, F. Lindsay)
- Advanced Java Processing and Visualization Interface Based on a Layered Framework (J. Humphries, F. Lindsay, and C. Plaisant)
- Object-Relational Database Infrastructure (N. Roussopoulos and J. Masek)
- High Performance Processing Software (J. JaJa, S. Kalluri, and J. Townshend)
Technology Roadmap

• Multiple Release Plan by the Implementation Team:
  – Each release is an end-to-end system encompassing:
    • Interface
    • Visualization and GIS tools
    • Database Framework
    • Data-types Available
    • Ingestion Process
    • Output Formats
    • Storage Framework
    • Data Processing Modules
    • Testing and Validation
GLCF Interface
Bio-physical products created from satellite data using the Kronos system. These samples were generated in under two minutes.