

Mercury

Distributed Metadata Management, Data Discovery, and Access System

An internet tool to access and distribute data

Tammy Walker Beaty

NASA ORNL DAAC
National Aeronautics and Space Administration
Oak Ridge National Laboratory
Distributed Active Archive Center
Oak Ridge, Tennessee, USA

About the ORNL DAAC

<http://daac.ornl.gov/>

➤ NASA's ORNL DAAC

Distributed Active Archive Center for biogeochemical and ecological data and models useful in environmental research

986 published data sets and counting

4 areas within NASA's Terrestrial Ecology Program

- Field Campaigns, Validation, Regional and Global, Model archive

Web interfaces to ORNL DAAC and NASA's 12 Earth Observing System data centers

Tools for ingesting data and for reporting user activity

Multiple Distribution Mechanisms

- FTP
- Metadata Search Catalog – Mercury (FGDC compliant)
- NASA Earth Observation System-wide clearinghouse (ECHO/Reverb)
- Spatial Data Access Tool (WCS, WMS)
- WebGIS (WMS)
- THREDDS Data Server
- Tools for subsetting and visualization of remote sensing products (MODIS)

Diverse Goals

- ORNL DAAC has a diverse customer base
- Education and Outreach - data contributors, research scientists, educators, legislators, students
- Assembling data bases for major resources
- Developing technologies are as important to provide as the source data
- Security issues related to policy and privacy as well as implementation of governing standards must be addressed.

Data Management Challenges

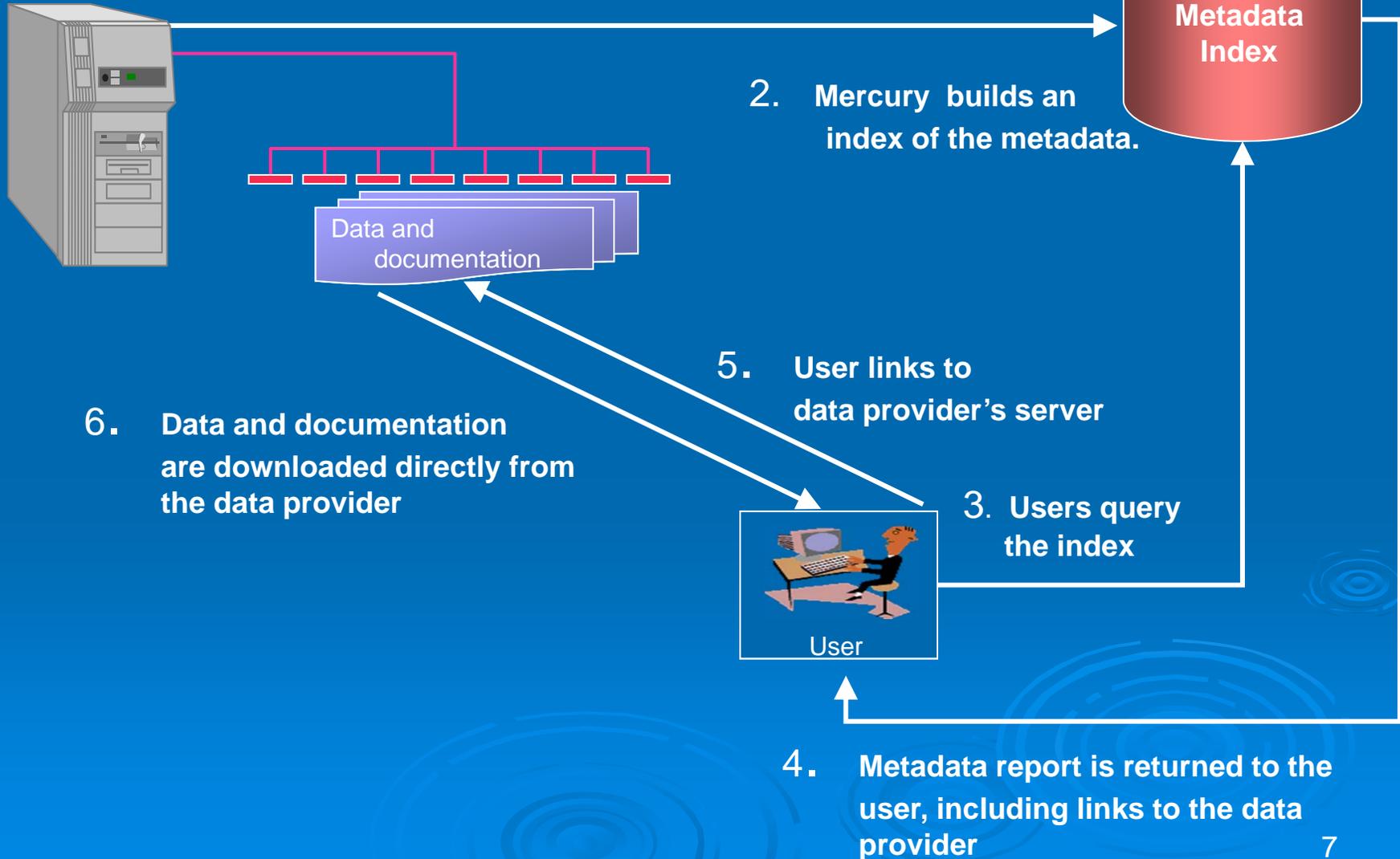
- World full of diverse data, scientists, GIS and DBMS software, custom models, and remote sensing resources.
- Success in science based on publications. Data published in papers, methods generalized, source information cited but usually only results published, not the raw data files.
- You need sophisticated search skills to find exactly what you need through Web search engines like Google.
- Certain data need to be archived with access restrictions in place.
- Similar information in multiple places produced with different tools at different scales.

What is Mercury ?

- Distributed metadata management, data discovery, and data access system
- Metadata Search and Retrieval tool
- Standards based – XML, Z39.50, FGDC, Dublin-Core, Darwin-Core, EML, and ISO 19115 (development pending)
- Based on open source and Service Oriented Architecture

Mercury: *Metadata Search and Data Access System*

1. The data provider uses the ORNL Metadata Editor to create a metadata file containing links to the data, documentation, and ancillary files.



Two Mercury Models

➤ The Virtual Internet Database

- Organizes a new collection of data from informal systems spread across the Internet

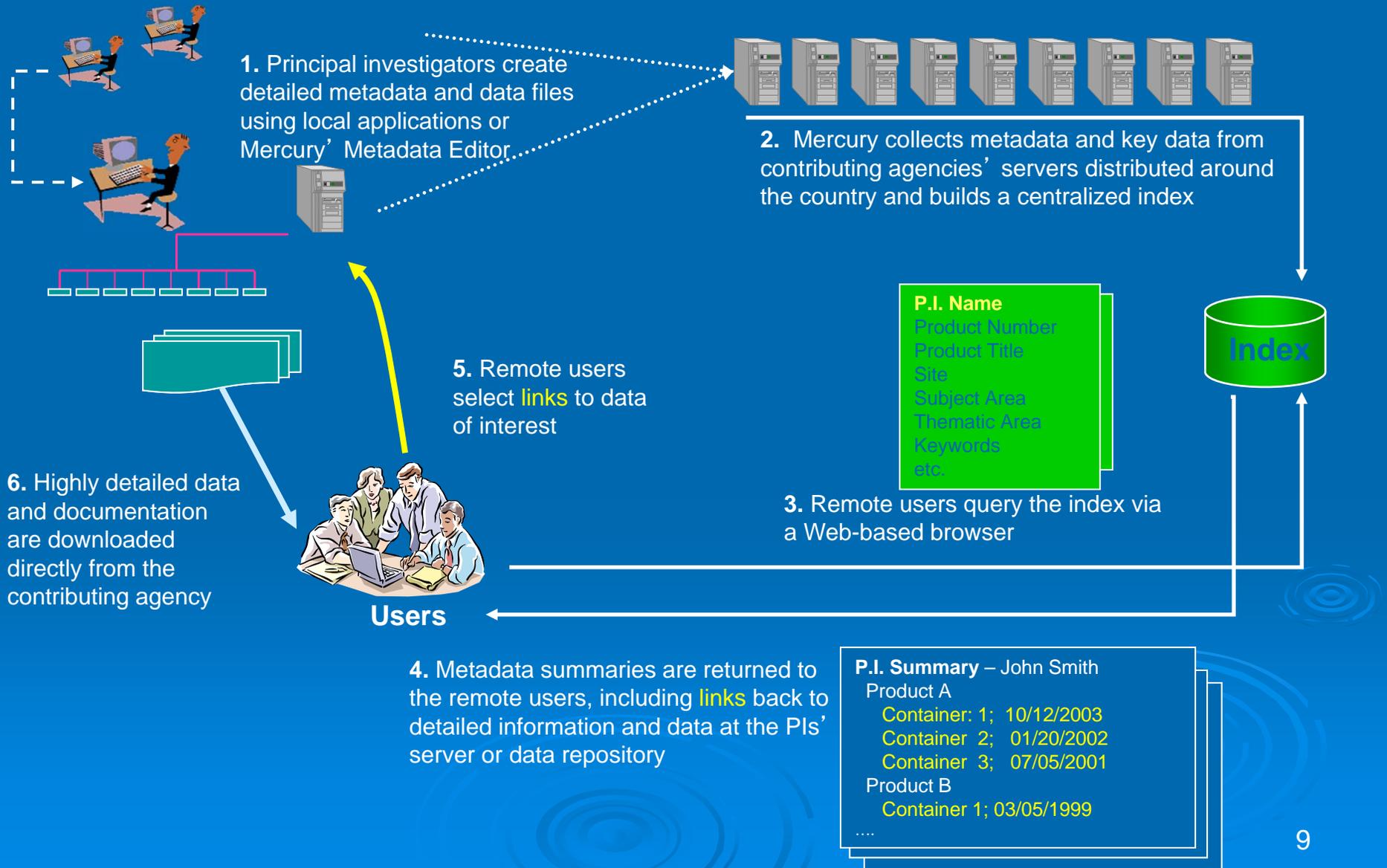
➤ The Virtual Aggregate Database

- Harvests information from existing formal separate DBMS

➤ And Many Hybrids...

Mercury : Virtual Internet Database

Distributed Data Discovery and Access System



A Virtual Aggregate Database via Mercury

Metadata exists in remote legacy databases using any platform, OS or RDBMS



Databases can be of different structures and content

Export programs are easily written and automated



No re-programming of existing systems required

Business as usual for contributing databases

Metadata are extracted into XML files yielding standardized data objects



These files can be remotely harvested via the Internet

Harvested metadata are combined at the central site, transformed (if needed), and indexed



Users work with a single, simple, web-like interface to access all data simultaneously



Frequent, automated harvesting and complete re-building of the index keeps the aggregate database up to date



What Can Mercury Do?

- Provide a **single portal** to information contained in separate data management systems
- Provide **fielded, spatial, and temporal search** capabilities
- Allow individuals and database managers to **distribute their data** while maintaining complete control and ownership
- **Leverage investment** in existing information systems and research

Custom Interfaces

LP DAAC (Land Processes, EROS, USGS)

The screenshot displays the LP DAAC Metadata Clearinghouse website. At the top, there are logos for USGS (science for a changing world) and NASA LP DAAC (LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER). Navigation links for USGS Home, Contact USGS, and Search USGS are in the top right. The main heading is "LP DAAC Metadata Clearinghouse".

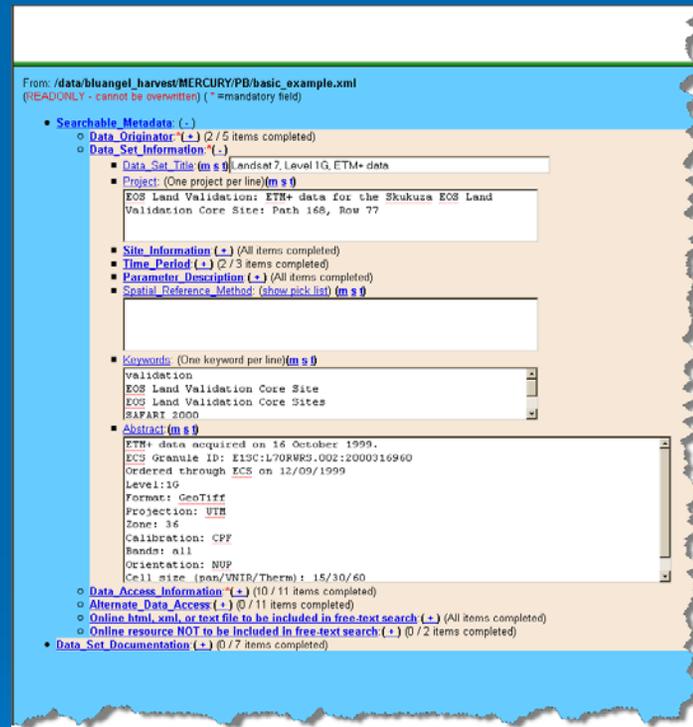
The interface is divided into several sections:

- Simple Search:** A search bar with a "Search For:" label, a "Results/Type" dropdown set to "10", and a "SEARCH" button. A hint below reads: "Hint: boolean operators, wildcards and phrases are allowed. ex: precipitation or (rain* and 'moisture content')".
- Advanced Search:** Includes a "Faceted Search" dropdown set to "FullText" and a "Search by Date Range" section with a "during" dropdown and two date input fields (mm/dd/yyyy).
- Geographic Search:** Features a world map with continent labels (NORTH AMERICA, SOUTH AMERICA, EUROPE, AFRICA, ASIA, AUSTRALIA) and a "List Area in:" dropdown set to "USA". Below the map are "Search Area:" options for "entire" and "entire" (with a globe icon), and a "Click on a [globe icon] to select an area" instruction.
- Query being built:** A section labeled "Not Editable" showing the current search parameters.
- Bottom Search Bar:** A "Results/Type" dropdown set to "10", a "SEARCH" button, a "CLEAR QUERY" button, and a "HELP" button.

At the bottom, the website provides contact information for the LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER: 47914 252nd Street, Sioux Falls, SD 57198-0001. Contact details include Voice: 605-594-6116, Toll Free: 866-572-3222, Fax: 605-594-8962, and Email: lpdaac@usgs.gov. Links for Accessibility, FOIA, Privacy, and Policies and Notices are provided. The footer also includes "U.S. Department of the Interior | U.S. Geological Survey", the URL "https://lpdaac.usgs.gov/lpdaac", "Page Contact Information: lpdaac@usgs.gov", and "Page Last Modified: October 07, 2010". Logos for USA.gov and "Your Planet. America." are in the bottom right corner.

Mercury Demonstration

- Metadata Editor – existing editor created in late 1990s



From: /data/bluangel_harvest/MERCURY/PB/basic_example.xml
(READONLY - cannot be overwritten) (* = mandatory field)

- Searchable_Metadata (-) (All items completed)
 - Data_Originator (+) (2/5 items completed)
 - Data_Set_Information (-) (All items completed)
 - Data_Set_Title (m s g) Landsat 7 Level 1G, ETM+ data
 - Project: (One project per line)(m s g)
EOS Land Validation: ETM+ data for the Skukusa EOS Land Validation Core Site: Path 168, Row 77
 - Site_Information (-) (All items completed)
 - Time_Period (-) (2/3 items completed)
 - Parameter_Description (-) (All items completed)
 - Spatial_Reference_Method: (show pick list) (m s g)
 - Keywords: (One keyword per line)(m s g)
Validation
EOS Land Validation Core Site
EOS Land Validation Core Sites
SAPART 2000
 - Abstract (m s g)
ETM+ data acquired on 16 October 1999.
EOS Granule ID: E15C:L70RWR9.002:2000316960
Ordered through EOS on 12/09/1999
Level:1G
Format: GeoTiff
Projection: UTM
Zone: 36
Calibration: CPF
Bands: all
Orientation: NUP
Cell_size (non/UNIR/Therm): 15/30/60
 - Data_Access_Information (+) (10/11 items completed)
 - Alternate_Data_Access (+) (0/11 items completed)
 - Online_html_xml_or_text_file_to_be_included_in_free-text_search (+) (All items completed)
 - Online_resource_NOT_to_be_included_in_free-text_search (+) (0/2 items completed)
 - Data_Set_Documentation (-) (0/7 items completed)

- New Internal and External metadata editor – in development
 - covered by Viv Hutchinson

Mercury Demonstration

➤ [http://mercury.ornl.gov/ornlDaac/index.jsp?
tab=advanced](http://mercury.ornl.gov/ornlDaac/index.jsp?tab=advanced)

- 8 data sources
 - ORNL DAAC Archives (986)
 - FLUXNET (548)
 - LBA-ECO (614)
 - RGD (452)
 - Land Val (68)
 - LP DAAC MODIS & ASTER (102)
 - Long Term Ecological Research Network , LTER (7048)

Organization of Biological Field Stations, OBFS -
124

Questions?

Additional Slides

Data Processing and Support Challenges in Data Archival

□ Acquisition

- identify how best to serve the audience / scientific community
- establish how to obtain data
- acquire data (local archive) or link to data (mercury, hard link)

□ Ingest – Archival of Data Sets

- perform QA checks
- compile project-provided metadata
- generate additional metadata

□ Enhance (as requested)

- add supplementary variables
- standardize to common units
- aggregate files

□ Documentation

- collect available documentation and write “guide” documents for each data set, as needed

□ Distribution

- advertise data (web, newsletters, customer announcements)
- provide data access to users worldwide

□ Post-Project Data Support

- provide long-term secure archiving of the data
- serve as a buffer between end users and PIs
- provide usage statistics for Sponsors
- survey users about major products and/or projects

□ Stewardship

- security, disaster recovery
- Ensure life of the data (reformat when proprietary software or operating systems reconfigure data formats)

ORNL DAAC: Data Collections

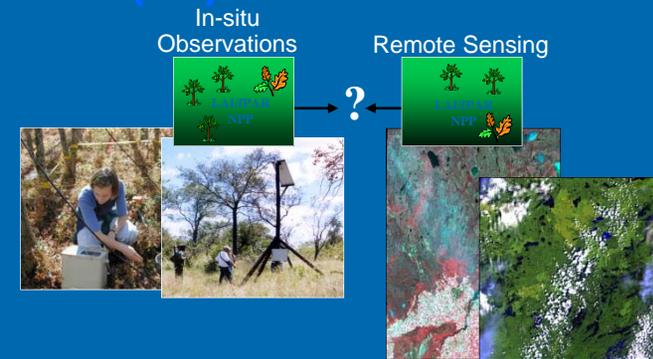
(Number of Data Sets = 986)

1. Field Campaigns (755)

- 6-9 year intensive study of a region:
 - Amazon (LBA)
 - Northern Canada (BOREAS)
 - Southern Africa (SAFARI 2000)

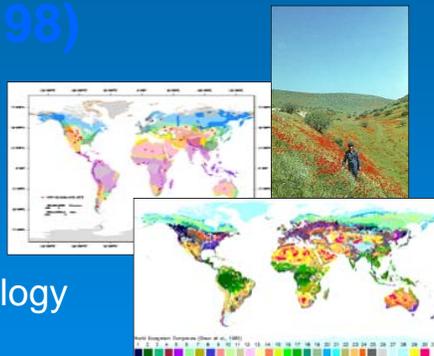


2. Validation of Remote Sensing Products (23)



3. Regional and Global Studies (198)

- Climate
- Soils
- Vegetation
- Hydroclimatology



4. Model Code (10)

- Benchmark Models
 - IBIS, BIOME-BGC, LSM
- Manuscript Models
 - PNeT, Century, Biome-BGC

