



# **National and International Collaborations for Geoinformatics: Challenges and Lessons Learned from Geoinformatics for Geochemistry**

W. Christopher Lenhardt (CIESIN - Columbia University), Kerstin Lehnert (LDEO – Columbia University), Sri Vinayagamoorthy (CIESIN - Columbia University), and Steve Goldstein (LDEO – Columbia University)

[clenhardt@ciesin.columbia.edu](mailto:clenhardt@ciesin.columbia.edu)



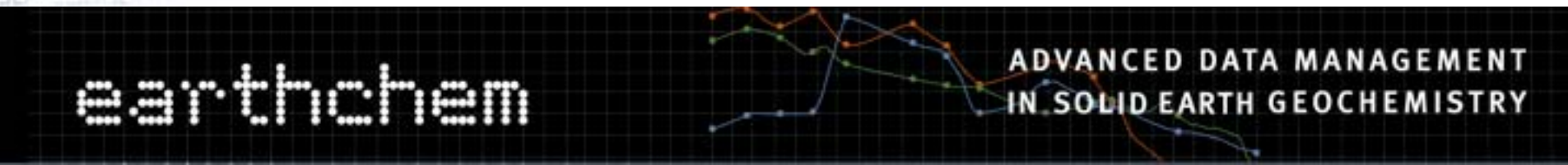


# Outline

- Introduction to geochemical and related projects at the Lamont-Doherty Earth Observatory (LDEO) – Columbia University
  - PetDB (Petrologic Database of the Ocean Floor)
  - SedDB (Sediment Geochemistry Database)
  - Earthchem (Advanced Data Management for Solid Earth Geochemistry)
  - SESAR (System for Earth Sample Registry)
- Similarities
- Challenges
- Data to Information systems and beyond
- Next steps



# LDEO Projects in Geoinformatics for Geochemistry





# Objectives of LDEO Projects

- Maximize the utility of data
  - Build infrastructure that makes data and samples visible and accessible to the broad community
  - Advance the “principle of open access” to data and samples
  - Support the long-term preservation of data (& samples)
    - Provide for persistent archives
    - Ensure comprehensive and accurate documentation
- Support cross-disciplinary approaches in science
  - Facilitate data integration across the Geosciences
    - Technical: interoperability, open access interfaces, better metadata and quality control
    - Cultural:
      - link communities (across related disciplines, nationally, internationally)
      - Facilitate development of relevant expertise



# Collaborative Effort

- LDEO



- Geoscientists
- Information Technology
- Data Managers

- CIESIN



- Information Technology
- Systems Integration
- Database Development
- Data Stewardship
- Operations

## Collaborating Institutions

Harvard (PetDB)  
Boston University (SedDB)  
Oregon State University  
(SedDB)  
Kansas University (EarthChem)  
University of Hawaii (VentDB)  
WHOI (VentDB)





# Reasons for PetDB's Success

- Technical
  - Design guided by scientists
  - Integrative data model
    - Each individual value searchable through flexible query interface
    - Links & integrates disparate data for individual samples
    - Rich metadata
    - Accessible references
  - User interface with flexible data selection
- Organizational
  - Implementation at professional data center
  - Strong ties with the community
    - Users (science)
    - Professional information technology partners
    - National Science Foundation
- Scientific
  - Has enabled new science



# SedDB



<http://www.seddb.org>

## • Integrated Data Management for Sediment Geochemistry

The screenshot shows the SedDB website interface. On the left, a graph for Site 1256 plots Depth (mbsf) from 0 to 250 against Age (Ma) from 0 to 9. It shows data for LO *Discoaster quinqueramus* (5.6 Ma) and LO *Sp. hetero* (13 Ma) with magnetostigraphy and biostratigraphy data points and a linear fit line. On the right, a graph plots Ba/Ti (g/g) from 0 to 40 against Latitude from -15 (South) to 15 (North). It features three data series: Goldberg and Arrhenius (1958) in red, Schroeder et al. (1997) in green, and Murray and Leinen (1993) in black. The website header includes 'SedDB Home - Mozilla Firefox' and the URL 'http://www.seddb.org/SedDBWeb/index.jsp'. The footer contains 'SedDB is supported by the National Science Foundation and managed as part of the GFG Program' and 'Geoinformatics for Geochemistry'.

**Funding Agency:**  
NSF (OCE/EAR)

**Start Date:**  
July 2005

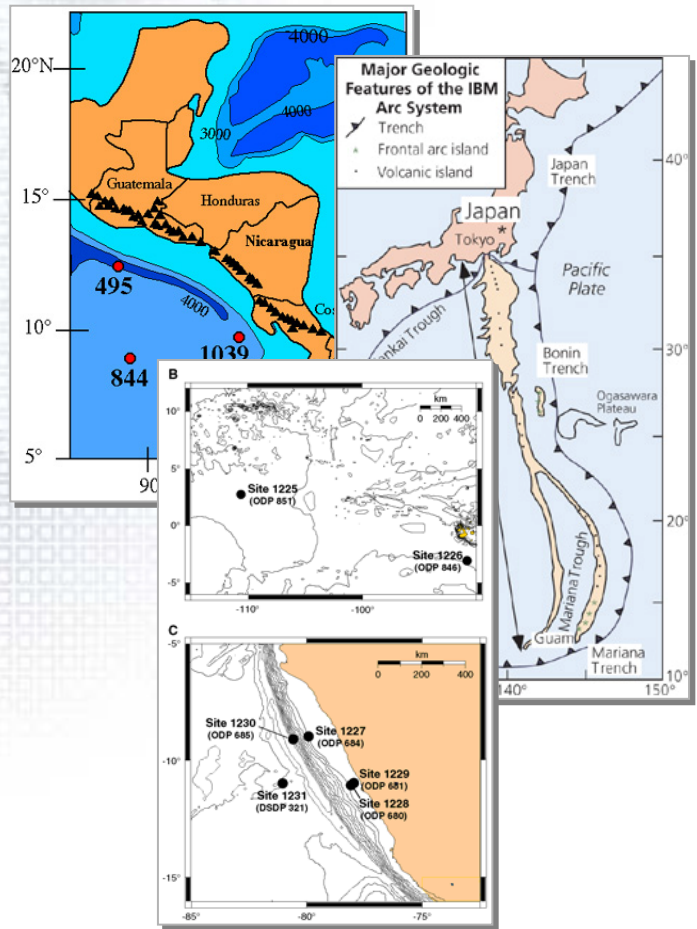
**Duration:**  
3 years

**Investigators:**  
K. Lehnert (LDEO)  
S. Goldstein (LDEO)  
R. Murray (BU)  
N. Pisias (OSU)





# SedDB



- Apply the concept of PetDB to Marine Sediments
  - Design data model based on PetDB schema
  - Compile complete data sets for 3 test bed areas
  - Build interactive query interface
  - Develop data analysis tools for age model conversion & age-depth correlation
  - Ensure integration with other data (interoperability)



# Challenges

- Technical
  - Development of additional aspects of the data model (e.g. age models)
  - Optimize interaction with the data for a broad audience ranging from the casual to the expert user
  - Efficiently populate databases with legacy and new data
  - Data quality control
- Organizational
  - Integration/coordination with other geoinformatics efforts
  - Long-term sustainability
  - Workforce ‘under development’
- Cultural
  - “My data syndrome” and data policies
  - Community education (supporting, not competing with science)
  - Standards for data quality assurance & procedures



# EarthChem

- Consortium founded 2003 by PetDB, NAVDAT, & GEOROC
  - To nurture synergies among projects
  - To minimize duplication of efforts
  - To share tools and approaches
- Collaborative proposal with D. Walker (Kansas University) funded by NSF EAR&OCE (5 years, start 9/2005) to build an integrated data management and information system for solid earth geochemistry.



# The EarthChem Project

- Build the EarthChem portal as a central access point to a system of federated geochemistry databases (“One-Stop Shop for Geochemical Data”)
- Ensure efficient and continuing update and expansion of data holdings



# Project Components

- Data development
  - Data compilation
  - Data quality control
  - Data maintenance
- Data management
  - Data model development
  - Data loading
- Application development
  - User interfaces
  - Interoperability
  - Tools
- User support
- Outreach
  - Community interaction
  - Web site
  - Presentations, publications
  - Advisory committees
  - Workshops
- Project management

# EarthChem Focus: Portal

<http://www.earthchem.org>

EarthChem Home - Mozilla Firefox  
http://www.earthchem.org/EarthChemWeb/index.jsp

earthchem  
ADVANCED DATA MANAGEMENT  
IN SOLID EARTH GEOCHEMISTRY

Home About What's New Links Feedback Contact Us

## WELCOME TO EARTHCHEM

EarthChem is a community-driven effort to facilitate the preservation, discovery, access and visualization of the widest and richest geochemical datasets possible.

### NEWS

#### EARTHCHEM PORTAL NOW OPERATIONAL

The EarthChem Portal (v.0.1) is now available for test searches across the federated databases: PetDB, NAVDAT, and GEOROC. Click [here](#) to get started.

#### GSA SHORT COURSE

Co-sponsored by the GSA Geoscience Education Division

[Using Online Igneous Geochemical Databases for Research and Teaching](#)

October 21, 2006, 1-5:30pm, Philadelphia  
There is still time to register! Click on the link:  
[GSA-Sponsored Short Courses](#)

#### VISIT OUR BOOTH AT GSA AND AGU!

EarthChem staff will be on hand to answer questions and conduct demonstrations at the upcoming GSA and AGU meetings on use of the EarthChem Portal and data retrieval from the recently launched SedDB and [Deep Lithosphere Petrological Dataset](#).

[GSA Meeting](#) in Philadelphia, PA  
October 22-25, 2006  
Booth #215

[AGU Meeting](#) in San Francisco, CA  
December 11-16, 2006  
Booth #328

#### GEOROC

Geochemistry of Rocks of the Oceans and Continents

- Ocean Islands
- Convergent Margins
- Large Igneous Provinces

#### NAVDAT

The Western North American Volcanic and Intrusive Rock Database

- Cenozoic Igneous Rocks from Western North America

#### PETDB

The Petrological Database of the Ocean Floor

- Oceanic Crust generated at Mid-Ocean Ridges (incl. BAB, Seamounts, Old Crust)

Chemical data for more than 230,000 igneous rock samples is accessible through GEOROC, NAVDAT & PetDB.  
[Click on map for an enlarged view.](#)

Done 12:29:36 PM Now: Overcast, 48° F Mon: 51° F Tue: 50° F Wed: 52° F Thu: 49° F



earthchem



CODATA 2006 - Beijing



# EarthChem Focus: Data Holdings

- Create an infrastructure that ensures efficient and community-based growth of data holdings
  - Data entry by dedicated EarthChem personnel
    - New target datasets identified & prioritized via community outreach & the EarthChem Advisory Committee
  - Facilitate Community Contributions
    - Build on-line data submission capability for future data to encourage direct data contributions by investigators
    - Assist investigators with design, implementation, & population of their own databases
    - Serve these databases via the EarthChem portal
  - Expand federation



# EarthChem Focus: Standards

- Promote & implement standards for data management in Geochemistry
  - Ontologies
  - Classification
  - Metadata in publications
    - Analytical information
    - Sample provenance
    - Units
    - Unique sample identifiers (IGSN) ↪ SESAR
  - Data publication & submission
  - (Sample management)





[www.geosamples.org](http://www.geosamples.org)

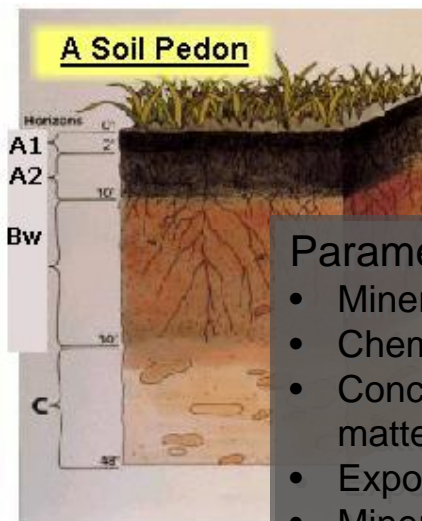
- Providing unique identifiers for Earth samples to allow global sharing, linking, and integration of information and data about these samples.





# SESAR: Rationale

Many data types are generated by the study of Earth samples. Their usefulness is critically dependent on their integration.



- Parameters to be measured:
- Mineralogy
  - Chemistry
  - Concentration of soil organic matter
  - Exposure age
  - Mineral surface area

**Conditions for Joining Individuals from Different Resources**

- Usually we don't make join on individuals cross different resources

We don't know whether 10001 represents the same rock in the two resources. By default, we assume they are not.

- A set of datatype properties can be declared as a key for a class in the ontology. We do join cross multiple resources based on keys.  
 e.g. { hasLatitude, hasLongitude } can be declared as a key of Location  
 Two locations from different resources are same if they have the same latitude and longitude

CSIG GEON

Currently, integration of data derived from the same sample, located in distributed systems is obstructed by ambiguous naming of samples.

Kai Lin (SDSC): "Ontology Based Resource Registration and Integration in GEON", Lecture July 2005





# International Geo Sample Number



IGSN.JDW000001

Unique user code String of (sequential) characters

- Structure
  - String of 9 characters (length limited by use in data publication)
  - First three characters are unique user code (registered with SESAR)
  - Last 5 characters are characters, numbers + letters (one spare character)
  - Allows 2,176,782,336 sample identifiers per registrant
- Managed at a central registry (SESAR)
  - Generated by SESAR or by users.
  - Strict compliance with the IGSN structure required.
- Applied in sample curation, data publication, & digital data management.
- Does not replace personal or institutional names.



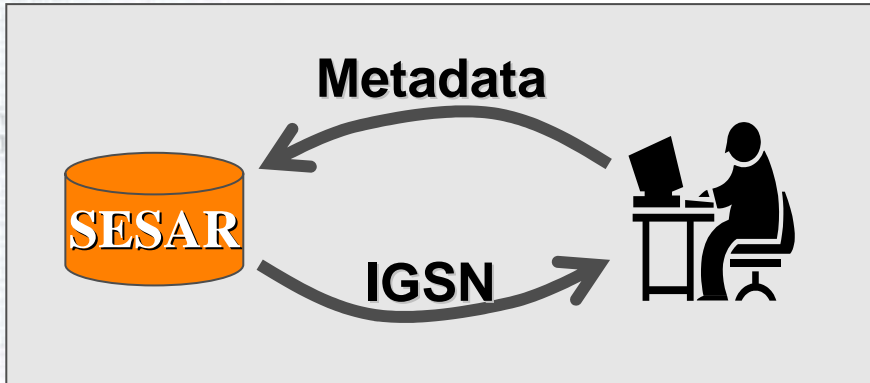
# IGSN: Impact



- Ability to link & integrate data for a single sample will
  - advance interoperability among digital data management systems & the development of Geoinformatics
  - help build more comprehensive data sets for samples
  - foster new cross-disciplinary approaches in science
- Ability to unambiguously identify samples will
  - aid preservation and curation, orphaned samples can be identified
  - ensure proper linking of data from samples and subsamples
  - facilitate sample handling and analysis
- Access to a central sample catalog will
  - allow more efficient planning of field & lab projects
  - facilitate sharing of samples
  - facilitate development of sample profiles



# Sample Registration



- via
- Web site
  - Batch loading
  - Web services

**Review/Edit Individual Specimen Metadata**

Sample Name: All125-D3-1    Collection Method: Dredge

---

**Collection Metadata:** Edit Collection Metadata

Ship: Argo  
 Cruise: ARGAMPH  
 Start Latitude: 44.5 North  
 Start Longitude: 60.7 East  
 Stop Latitude: 44.7 North  
 Stop Longitude: 60.5 East  
 Min Water Depth: 3200  
 Max Water Depth: 3300  
 Location Names: Mid-Atlantic Ridge  
 Additional Information on Location:  
 Feature: Spreading Center  
 Collection Date Start:  
 Collection Date End:

---

**Classification Metadata:** Edit Classification Metadata

Rock Classification: Level 1: Igneous  
 Rock Classification: Level 2: Mafic, volcanic/subvolcanic  
 Field Description: Basalt  
 Additional Information: aphyric, altered

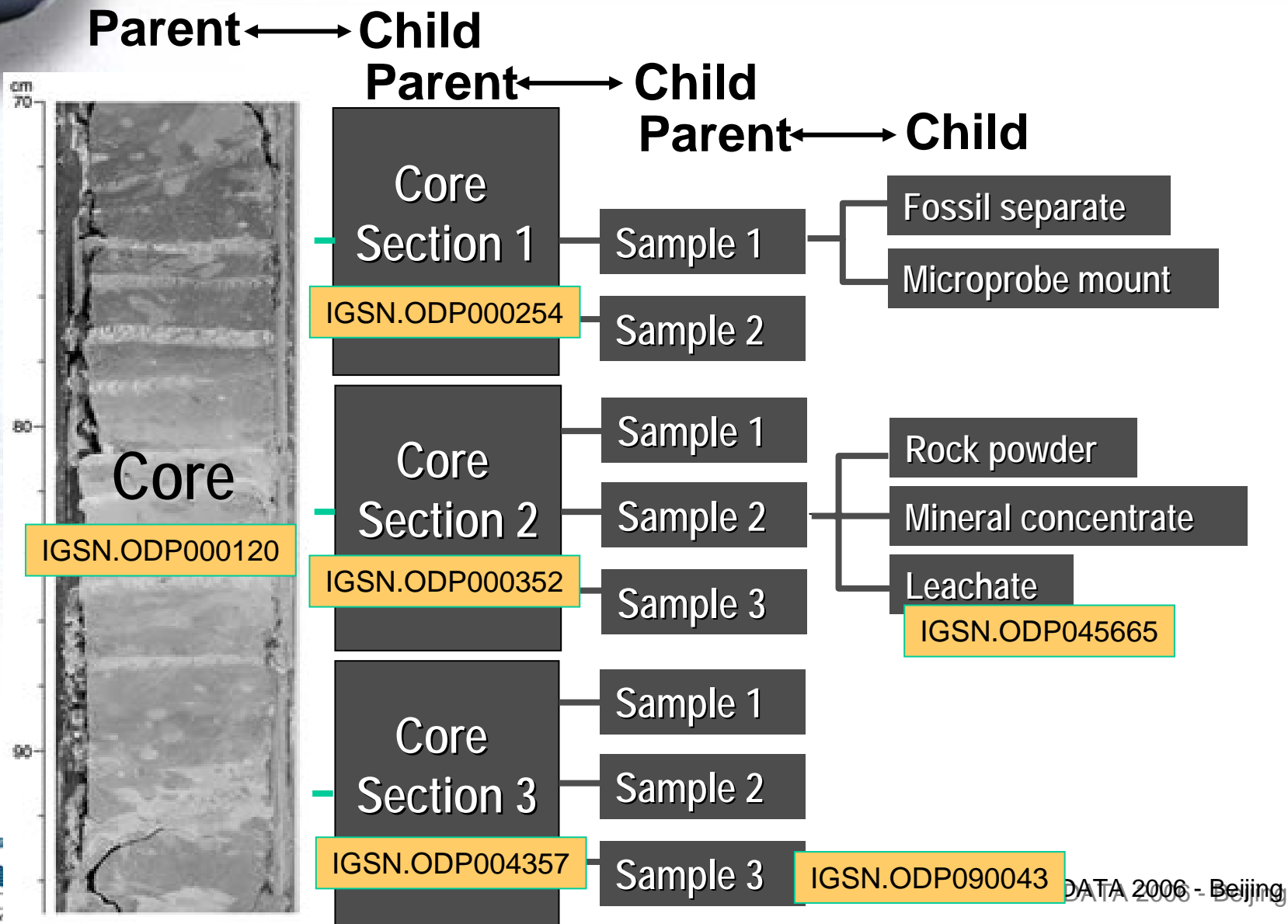
---

**Archival Metadata, etc.:** Edit Archival Metadata

Original Archive  
 Last Name: Lehnert    First Name: Kerstin  
 Institution: Columbia University, Lamont-Doherty Earth Observatory

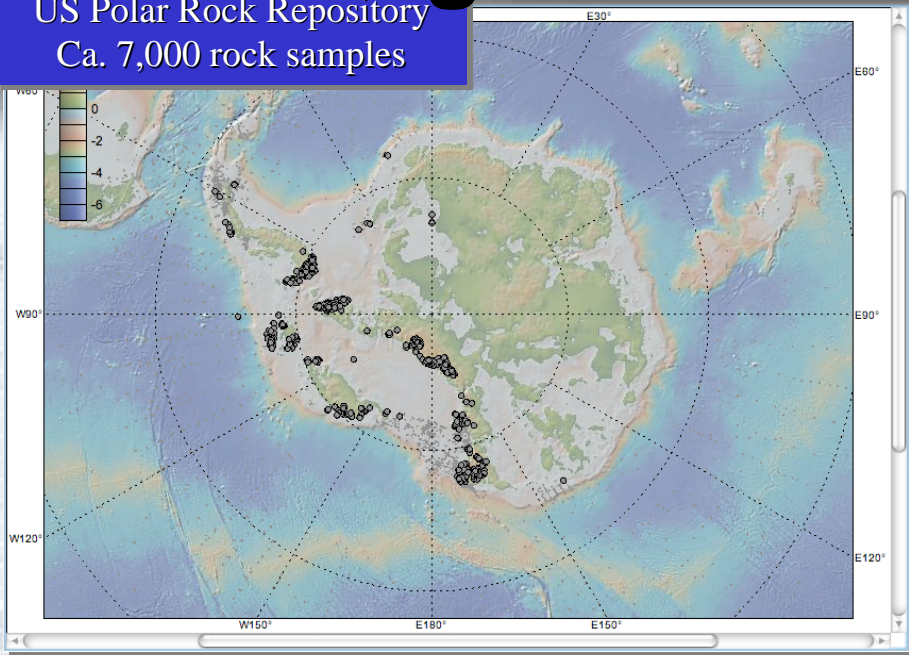
Latest Known Archive  
 Last Name: Lehnert    First Name: Kerstin  
 Institution: Columbia University, Lamont-Doherty Earth Observatory

# Granularity of Registered Samples

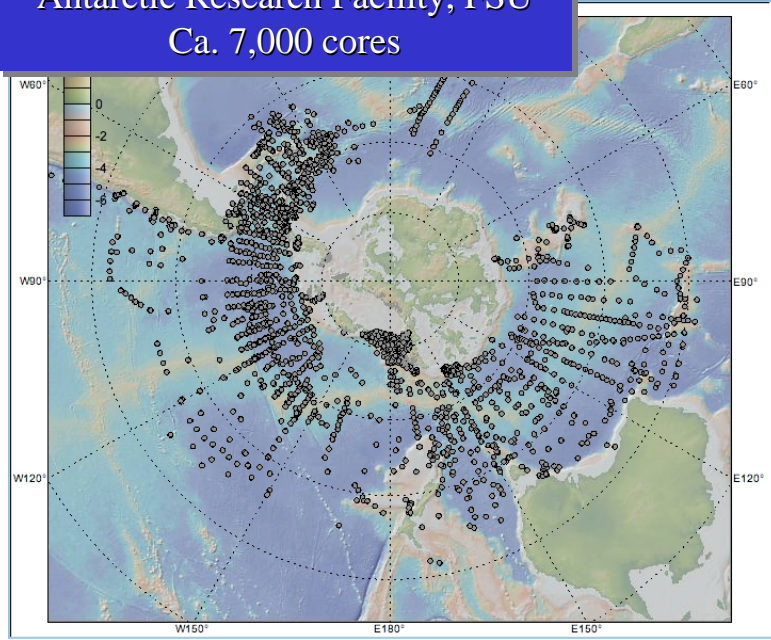


# Building a Global Sample Catalog

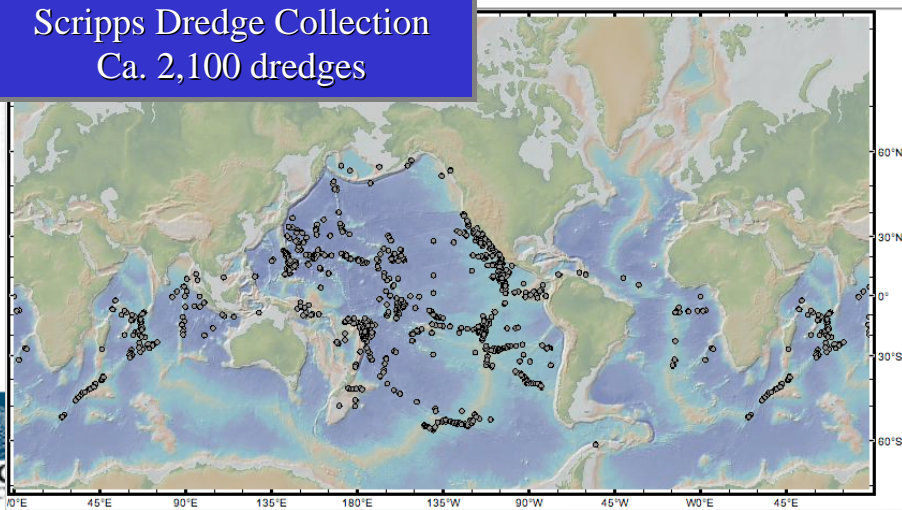
US Polar Rock Repository  
Ca. 7,000 rock samples



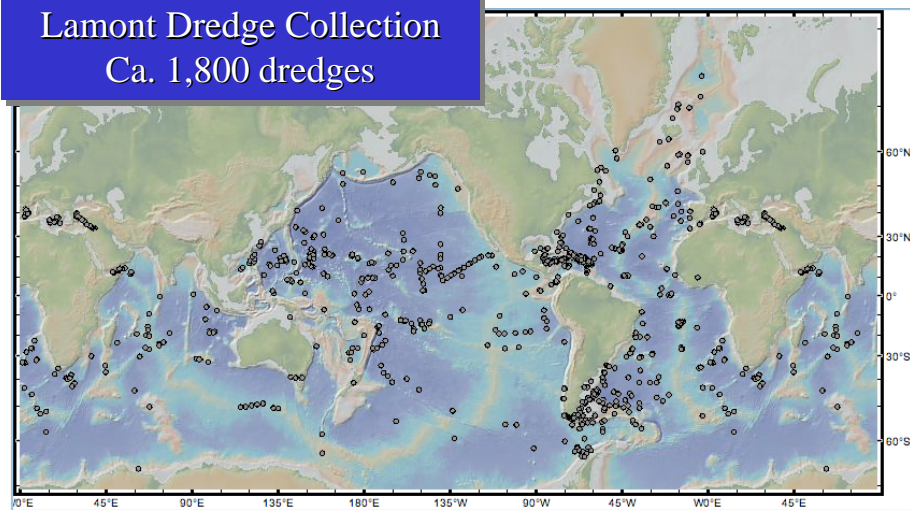
Antarctic Research Facility, FSU  
Ca. 7,000 cores



Scripps Dredge Collection  
Ca. 2,100 dredges



Lamont Dredge Collection  
Ca. 1,800 dredges





# Similarities





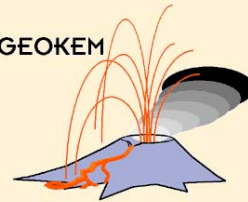
# Many related sources of data and information in the field of geochemistry

## Geochemistry of Igneous Rocks

### Geokem

- **General:**
  - Background to this site (28-Feb-2005)
  - Composition of the Earth's Mantle and Crust (2-Feb-2005)
  - The Average Composition of the Earth's Continental Crust
  - Global Element Distribution (1-Oct-2005)
  - Fingerprints
  - Minerals (25-August-2005)
- **Oceanic Ridge Basalts (ORBs):**
  - Summary (17-Oct-2005)
  - Atlantic Spreading Centres (15-Oct-2005)
  - Pacific/Indian Spreading Centres (3-Jun-2005)
- **Oceanic Island Basalts (OIBs) & Alkaline Series:**
  - Summary (17-Oct-2005)
  - Antarctic (2-Feb-2005)
  - Atlantic Ocean (15-Oct-2005)
  - Iceland (9-Aug-2005)
  - Indian Ocean (25-Aug-2005)
  - Hawaii (22-Aug-2005)
  - Pacific Ocean (24-Oct-2005)
  - Continental alkali basalts-basanites (3-Feb-2005)
- **Continental Flood Basalts:**
  - Gondwana Jurassic Flood Basalts and S Parana, Deccan (3-Feb-2005)
  - Columbia River, Steens Mtn, Snake Rive N, Atlantic-E, Greenland
  - Archaean Sudbury Lopolith, Coppermine

### GEOKEM



An eText of Geochemical Data Interpretation  
This site last updated: Thu, 1 Dec 2005 NZST  
Most recently updated pages:

- Andesite-2.html
- OIB-volcanic-pacific.html
- ORB-oceanic-summary.html
- OIB-summary.html

91500 zircon  
BCR-1 NIST SRM 987  
KL2-G JR-1

**GeoReM**  
Geological Reference  
Materials

News

## Database

GeoReM is a database for geological reference materials and isotopic standards, such as rock powders, synthetic and natural glasses as well as mineral, isotopic, river water and seawater reference materials.

GeoReM contains published analytical data and compilation values (major and trace element concentrations, radiogenic and stable isotope ratios).

GeoReM contains all important metadata about the analytical values such as uncertainty, analytical method and laboratory. Sample information and references are also included.

GeoReM contains more than 750 geological reference materials and 6500 analyses (State: 11/01/2005).

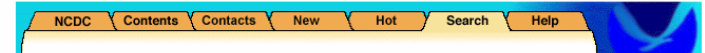
GeoReM complements the three earthchem databases GEOROC, NAVDAT, PETDB.

Sample Names or Material Types

Chemical Criteria

Bibliography

Abbreviations



Search for: sr isotope atlantic  
Results per page: 50 Output format: Long Match: All

Search results: sr : 59, isotope : 199, atlantic : 488

Displaying documents 1-5 of 5.

### 1. Eakin's Coral Literature

- [http://www.ncdc.noaa.gov/paleo/outreach/coral/coral\\_literature.html](http://www.ncdc.noaa.gov/paleo/outreach/coral/coral_literature.html) (text/html) Thu, 20 May 2004 07:49:31 GMT, 193842 bytes

### 2. The Pages/Clivar Intersection

Coordinated research objectives of the IGBP and WCRP Programs. Provides the paleoclimatic perspect...

- <http://www.ncdc.noaa.gov/paleo/reports/interx/refer2.html> (text/html) Wed, 11 Jun 2003 08:17:48 GMT, 14979 bytes

### 3. Coral Data Contribution Form

- [http://www.ncdc.noaa.gov/paleo/coral/coral\\_submit2.html](http://www.ncdc.noaa.gov/paleo/coral/coral_submit2.html) (text/html) Tue, 04 Nov 2003 08:11:36 GMT, 25411 bytes

### 4. NOAA Paleoclimatology Program - What's New for 2000

- <http://www.ncdc.noaa.gov/paleo/news/new2000.html> (text/html) Tue, 04 Nov 2003 08:29:31 GMT, 13389 bytes

### 5. NOAA Paleoclimatology Program - Bibliography of Extramural Research

- <http://www.ncdc.noaa.gov/paleo/reports/pubs.htm> (text/html) Sat, 01 Nov 2003 22:33:00 GMT, 47301 bytes

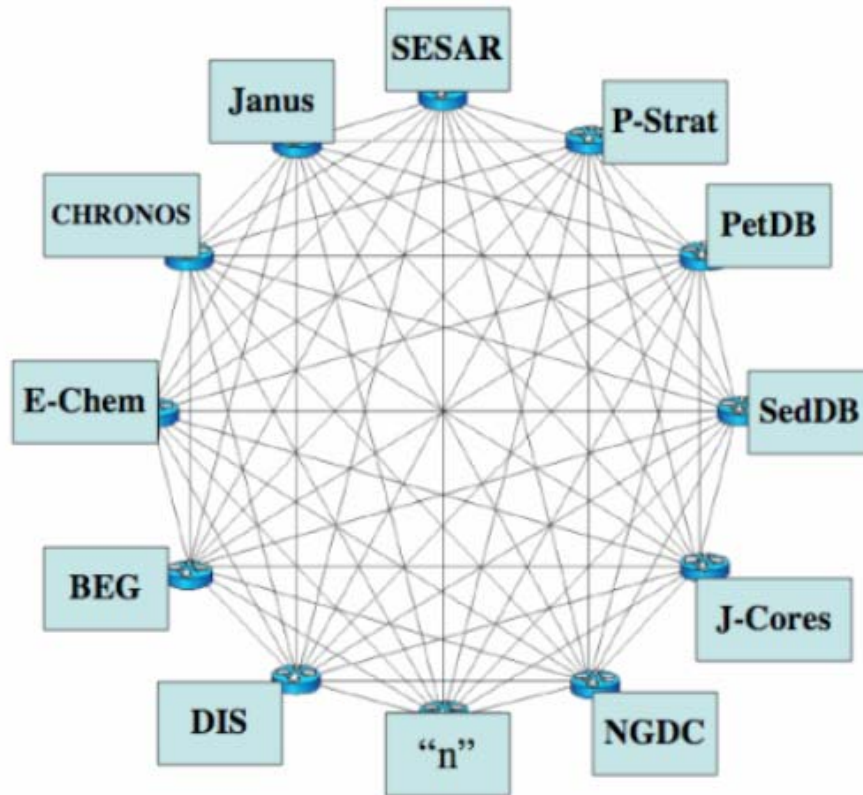
<< Prev 1 Next >>

Powered by UdmSearch



# Many potential interactions

## Interactions among Integrated Geoscience Databases



**F. Rack (JOI):**  
*International Collaboration  
in Data Management for  
Scientific Ocean Drilling,  
AGU 2005*



# Commonalities Across Geochemical Data

- Small volumes
- Complex 'background information' (=metadata)
- Diversity of acquisition methods
- Sample-based
- Producer is 'owner'



# Summary of Lessons Learned

- Data is the foundation
- Science is the driver
- Development of information systems essential
  - Data capture and access
  - Data stewardship
  - Knowledge capture
- Community participation is essential
- Outreach is essential
  - Vertical and horizontal



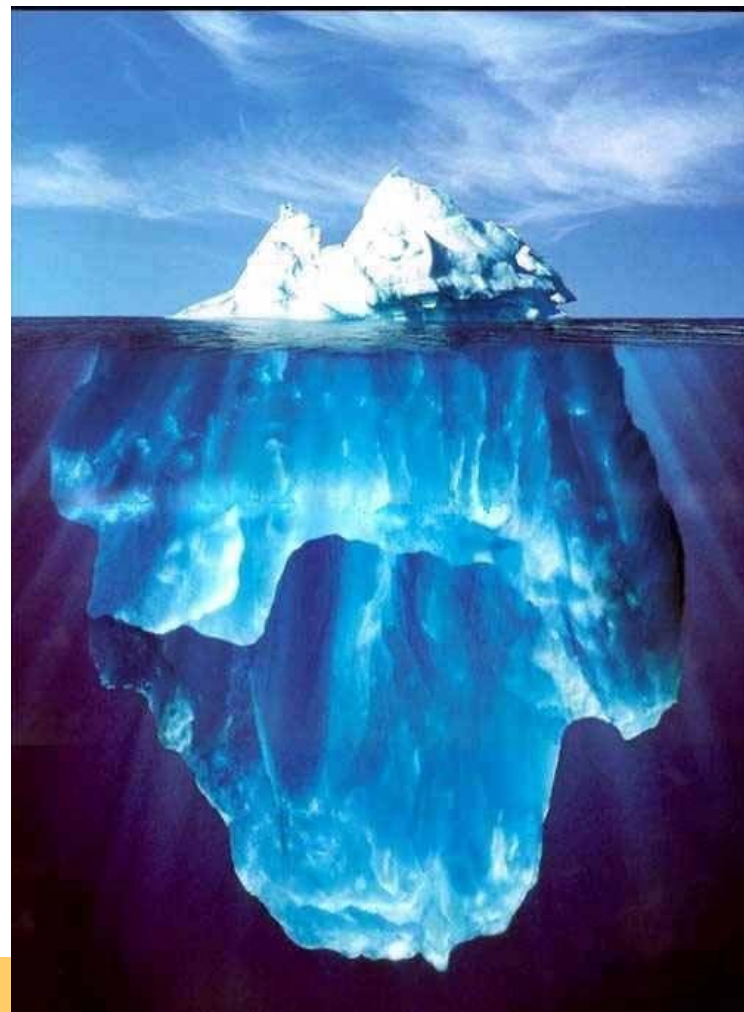
# General Trajectory

- Data to 'information systems'
- Develop and enhance the growing cyberinfrastructure for geoinformatics
- Expand both the data, the systems, interoperability, *AND* participation to move towards a geoinformatics science commons



# Challenges

- How to get the word out
- How to expand participation
- How to promote standardization and interoperability globally



*Most of the technology exists  
Challenges are cultural and organizational*



# Urgency to act

---

- Increasing data volumes
  - Need systems to support data management.
- Large-scale scientific questions
  - Need access to global data compilations.
- New cross-disciplinary approaches
  - Need integration of data with broader Geoscience data set.
- Decreasing funding
  - Need to maximize utility of data (and samples).



# Next steps

- Continue outreach
- Invite participation and collaboration
  - Collaborators
  - Data integration
  - Linkages across systems
- Propose a CODATA task group on geoinformatics?





**Thank you.**

謝謝







# Backup Slides

---





# NSF-OCI



National Science Foundation  
OFFICE OF  
Cyberinfrastructure

SEARCH

NSF Web Site



OCI Home

OCI Funding

OCI Awards

OCI Discoveries

OCI News

About OCI

## Cyberinfrastructure



OCI Home

About OCI

Funding Opportunities

Awards

News

Events

Discoveries

Publications

Advisory Committee

Career Opportunities

View OCI Staff

Search OCI Staff



How to Prepare Your Proposal

Grant Proposal Guide

Frequently Asked Questions

Other Types of Proposals

Regional Grants Conferences

How to Manage Your Award

Grant Policy Manual

Grant General Conditions

Cooperative Agreement  
Conditions

Special Conditions

HOME | FUNDING | AWARDS | DISCOVERIES



National Science Foundation  
OFFICE OF  
Cyberinfrastructure

OCI Home

OCI Funding

Cyberinfrastructure - stimulating  
advances in 21st century  
science and engineering

About OCI

View OCI Staff Directory

HPC Acquisition

## Reports and Workshops Relating to Cyberinfrastructure and Its Impacts

Over the past three years, a number of reports on cyberinfrastructure and its impact on research and education have been compiled. Links to a sample of some of the reports are listed below.

**Building a Cyberinfrastructure for the Biological Sciences**; workshop held July 14-15, 2003

[http://research.calit2.net/cibio/archived/CIBIO\\_FINAL.pdf](http://research.calit2.net/cibio/archived/CIBIO_FINAL.pdf)

<http://research.calit2.net/cibio/report.htm>

**CHE Cyber Chemistry Workshop**; workshop held October 3-5, 2004

[http://bioeng.berkeley.edu/faculty/cyber\\_workshop](http://bioeng.berkeley.edu/faculty/cyber_workshop)

**Commission on Cyberinfrastructure for the Humanities and Social Sciences**;

sponsored by the American Council of Learned Societies; seven public information-gathering events held in 2004; report in preparation

<http://www.acls.org/cyberinfrastructure/cyber.htm>

**Computation as a Tool for Discovery in Physics**; report by the Steering Committee on Computational Physics

<http://www.nsf.gov/pubs/2002/nsf02176/start.htm>

**Cyberinfrastructure for the Atmospheric Sciences in the 21st Century**; workshop held June 2004

[http://netstats.ucar.edu/cyrdas/report/cyrdas\\_report\\_final.pdf](http://netstats.ucar.edu/cyrdas/report/cyrdas_report_final.pdf)

**Cyberinfrastructure for Engineering Design**; workshop held February 28 - March 1, 2005; report in preparation

**CyberInfrastructure and the Next Wave of Collaboration**, D. E. Atkins, Keynote for EDUCAUSE Australasia, Auckland, New Zealand, April 5-8, 2005

**Cyberinfrastructure for Engineering Research and Education**; workshop held June 5 - 6, 2003

<http://www.nsf.gov/eng/general/Workshop/cyberinfrastructure/index.jsp>

**Cyberinfrastructure for Environmental Research and Education (2003)**; workshop held October 30 - November 1, 2002

<http://www.ncar.ucar.edu/cyber/cyberreport.pdf>

**CyberInfrastructure (CI) for the Integrated Solid Earth Sciences (ISES) (June 2003)**; workshop held on March 28-29, 2003; June 2003

[http://tectonics.geo.ku.edu/ises-ci/reports/ISES-CI\\_backup.pdf](http://tectonics.geo.ku.edu/ises-ci/reports/ISES-CI_backup.pdf)

**Cyberinfrastructure and the Social Sciences (2005)**; workshop held March 15-17, 2005

<http://www.sdsc.edu/sbe/>





# Geoinformatics

= Science + Data + Cyberinfrastructure +  
Data Stewardship

*Transforms into...*

“Science Commons” for geochemical data



# Cyberinfrastructure =

- “new research environments in which advanced computational, collaborative, data acquisition, and management services are available to researchers through high-performance networks”

*Report of the NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure (Atkins et al. 2003)*

## Cyberinfrastructure

**is the organized aggregate of technologies that enable us to access and integrate today’s information technology resources –**

*data, computation, communication, visualization, networking, scientific instruments, expertise*

**– to facilitate science, engineering, and societal goals.**



# Cyberinfrastructure =

- "Like the physical infrastructure of roads, bridges, power grids, telephone lines, and water systems that support modern society, "cyberinfrastructure" refers to the distributed computer, information and communication technologies combined with the personnel and integrating components that provide a long-term **platform to empower the modern scientific research endeavor.**"

*Access News Release: "National Science Foundation Releases New Report from Blue-Ribbon Advisory Panel on Cyberinfrastructure,"  
02.03.03 David Hart*



# CI Components



*The cyberinfrastructure should include*

- grids of **computational** centers, some with computing power second to none;
- comprehensive **libraries of digital objects** including programs and literature;
- multidisciplinary, well-curated federated **collections of scientific data**;
- thousands of **online instruments** and vast **sensor arrays**;
- convenient **software toolkits** for resource discovery, modeling, and interactive visualization;
- the **ability to collaborate** with physically distributed teams of people using all of these capabilities.

*Report of the NSF Blue-Ribbon Advisory Panel on Cyberinfrastructure (Atkins et al. 2003)*





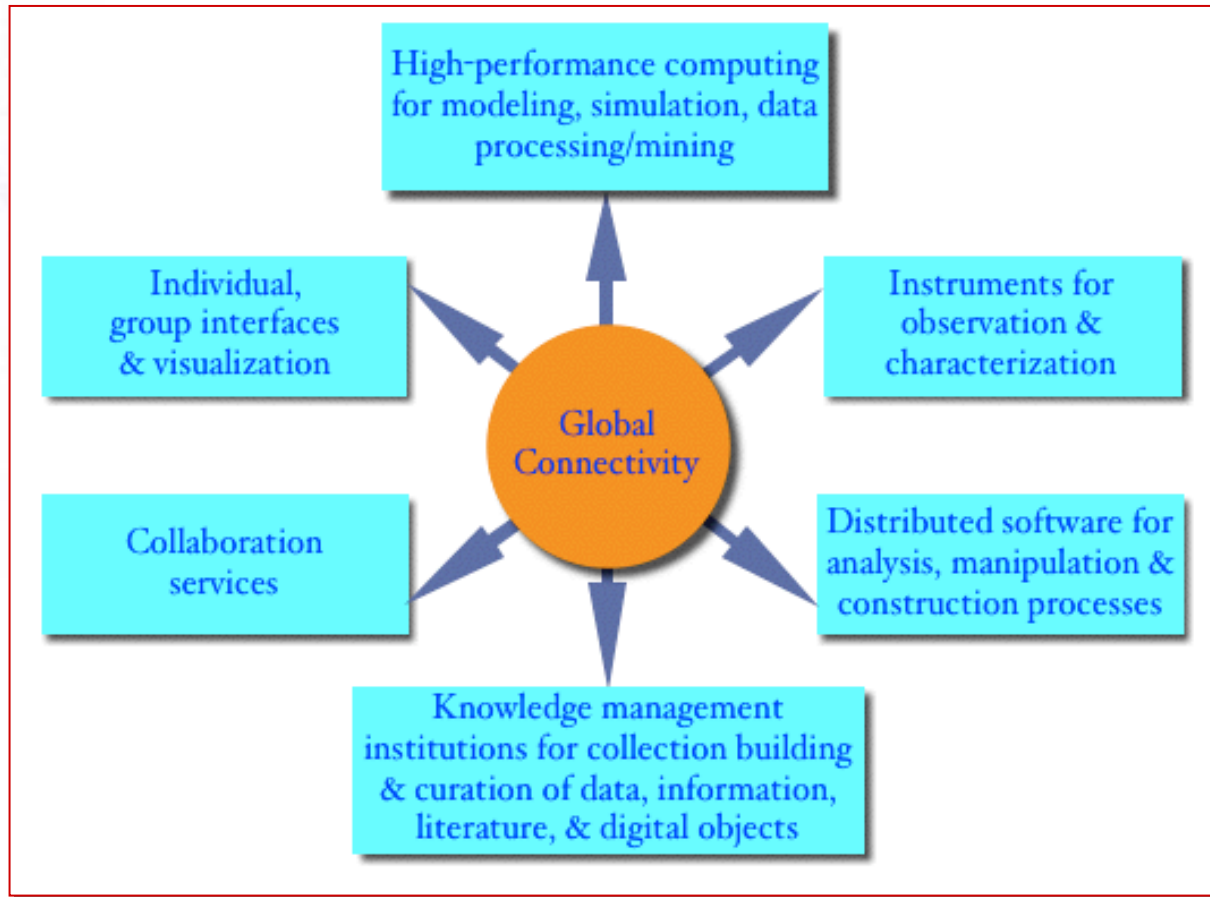
# Geoinformatics

## CYBERINFRASTRUCTURE FOR THE EARTH SCIENCES

- **Geoinformatics is the application of computer technologies and methodologies to scientific results with spatial-temporal coordinates.**
- Geoinformatics encompasses efforts to promote collaboration between computer science and the geosciences to solve complex scientific questions.



# Components of Geoinformatics





# Required Geoinformatics Components

- Interoperability of diverse databases on diverse systems
- Distributed, web-based, web services
- Access to data, tools & Computational resources
- Security

- Large data sets
- Complex data sets
- Data input – ease vs complexity
- Remote sensing & sensor arrays
- Real-time digital field technologies
- Capture analogue “legacy” data
- Data storage and curation

- Band width
- Computational resources: high performance, mid-level, desktop; grids, etc.
- File transfer protocols, etc.

- Real-time collaboration
- Data mining / pattern recognition
- Tools development and maintenance: numerical, statistical, visual
- Online workspace, software, and tutorials
- Community and computational models and Collaboratories
- Model-data fusion
- Skills, career paths and reward structures
- Intellectual property and academic credit
- E-Journals



# CI Challenges

- “The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination.”

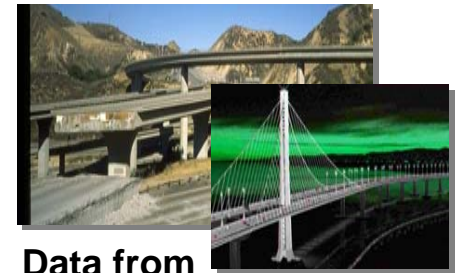
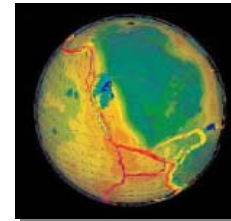
*Fran Berman, Director San Diego Supercomputer Center*

*SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences*

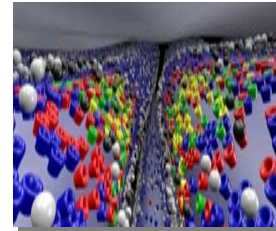


# Data: The Foundation of Geoinformatics

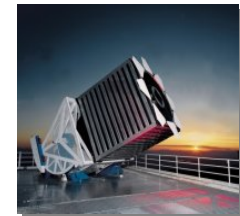
- Data comes from everywhere
  - Scientific instruments
  - Experiments
  - Sensors and sensor-nets
  - New devices (personal digital devices, computer-enabled clothing, cars, ...)
- And is used by everyone
  - Scientists
  - Consumers
  - Educators
  - General public
- Data Cyberinfrastructure environment must support unprecedented diversity, globalization, integration, scale, and use



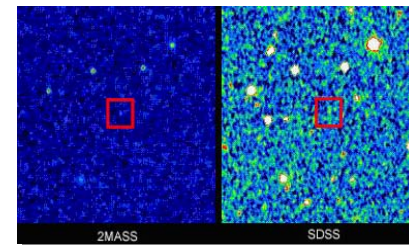
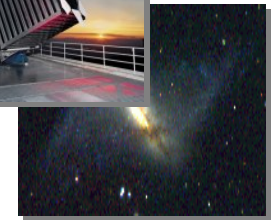
Data from sensors



Data from simulations



Data from instruments



Data from analysis



# Preserving the legacy

- *“The science community has invested vast resources – intellectual and financial - into our present state of knowledge that is bound up in the data it was generated from. **These legacy data are an incredibly valuable resource on which new theories, new discoveries, new knowledge will be based in the future - if they remain available to the community.** Due to limited accessibility, we have under utilized these data in the past, and we are at significant risk of losing them altogether. Capturing legacy data therefore has to be an essential part of Geoinformatics development.”*



# Geoinformatics builds on DATA

"The National Science Board (NSB) recognizes the growing importance of these digital data collections for research and education, their potential for broadening participation in research at all levels, the ever increasing National Science Foundation (NSF) investment in creating and maintaining the collections, and the rapid multiplication of collections with a potential for decades of curation."

*'Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century'*  
*National Science Board Report, September 2005*



# NSB Report

- Recommendations to NSF
  - Develop clear technical and financial strategy
  - Create policy for key issues consistent with the technical and financial strategy
    - Community oversight for data collections
    - Data policies for data generating projects
    - Education & training for using data collections
    - Recognition for ‘data scientists’





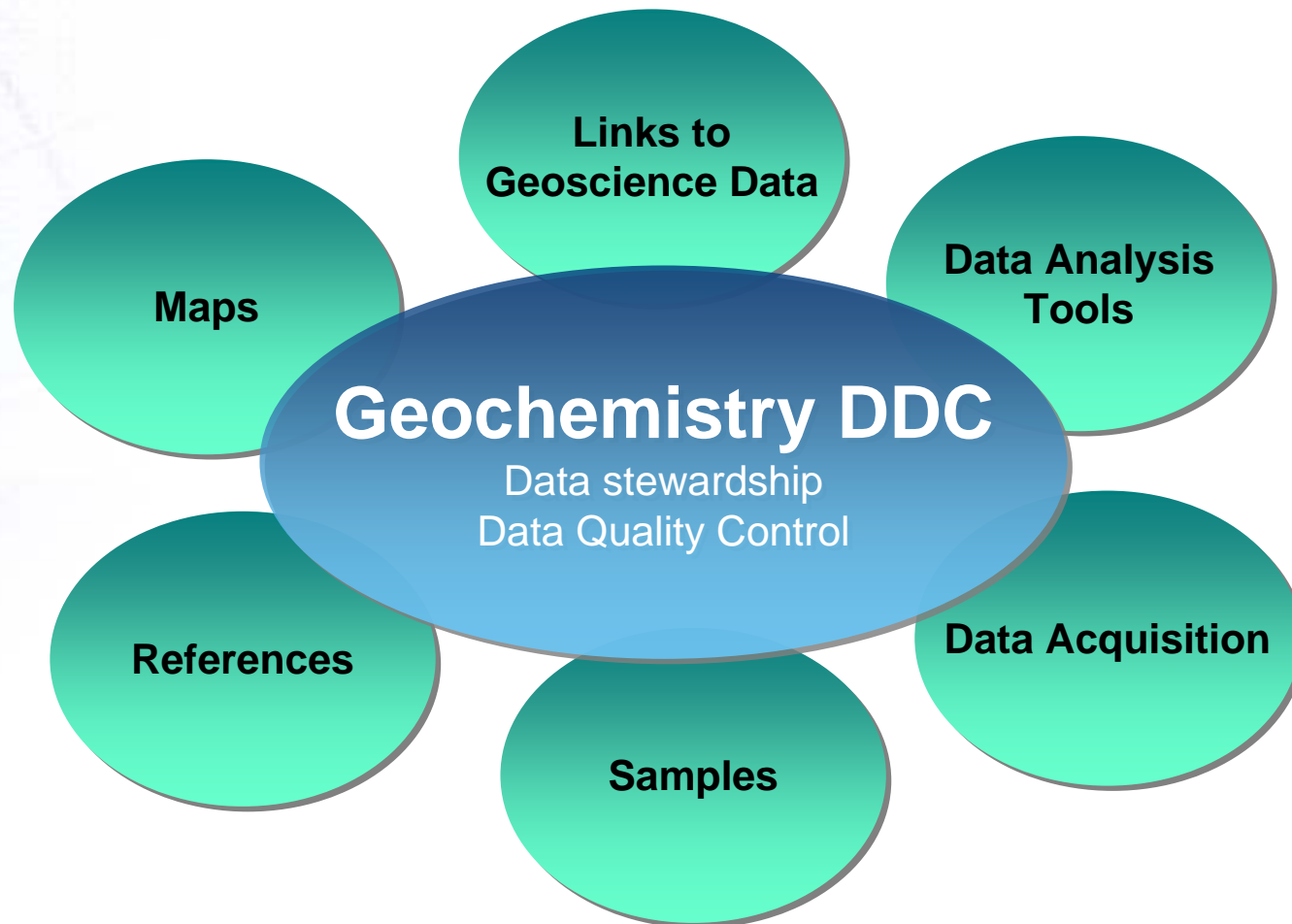
# Digital Data Collections: Benefits

- “Are equally accessible to study at all levels”
- “Serve as an instrument for performing analysis”
  - with an accuracy that was not possible previously
  - from a perspective that was previously inaccessible (by combining information in new ways)

*‘Long-lived Digital Data Collections: Enabling Research and Education in the 21st Century’  
National Science Board Report, September 2005*

**DDC need to be “Information Systems”  
rather than “Data Libraries”.**

# Information Systems in Geochemistry





# Benefits of Information Systems

- Advance scientific discovery
- Maximize utility of the Geochemical data set in science & education
- Allow data integration & visualization across the Geosciences
- Enhance data quality control



# Impact on Science

**National Science Foundation**  
WHERE DISCOVERIES BEGIN

SEARCH  
NSF Web Site

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT | FastLane

**News**

Press Release 05-148  
**Study Reconciles Long-Standing Contradiction of Deep-Earth Dynamics**  
New databases give researchers a look into processes inside the Earth's mantle

News  
For the News Media  
Special Reports  
Research Overviews  
Priority Areas  
Speeches & Lectures  
Multimedia Gallery  
News Archive

News by Research Area  
Arctic & Antarctic  
Astronomy & Space

Lava provides clues that the Earth's deep mantle has been continually moving a  
[Credit and Larger Version](#)

Since 2002, ca.100 articles cite PetDB as the source of data sets used for comparison or synthesis.



# User Survey 2005

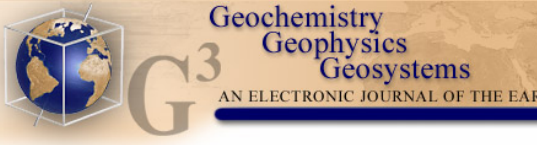
- “More than just a timesaver, these databases make it possible to address both global and regional questions that I would otherwise never bother to attempt. The amount of time saved is such that countless ideas cross from the realm of the totally impractical for a busy working scientist into the realm of easy to squeeze into a spare half hour.” (Paul Asimov, CalTech)
- “I think these online databases are absolutely necessary to ensure some level of access to geochemical data for all. I cannot imagine a more efficient way to compile and distribute this data.” (Garrett Ito, U Hawaii)
- “I use both GEOROC and PETDB regularly and have used them in 2 or 3 publications. I consider them critical for advancing isotope geochemistry.” (Don DePaolo, UC Berkeley).
- “It has been hugely helpful in both my research & teaching activities. One recent paper I have published in Journal of Petrology was on MORB, & I cited PETDB extensively.” (Claude Herzberg, Rutgers Univ)



# A User's Vision

- “... in theory the best thing would be one big Geo-database where all different types of geochemical reservoirs are included and all analytical tools as well and where you can search for either regions or reservoir type or method...

**ok that's a big goal.”**



- Home
- About G<sup>3</sup>
- Publications
- Themes
- Info for Authors
- Info for Reviewers
- Submissions
- Subscriptions
- Search
- Indexes
- News and FAQ
- Help

*Geochemistry, Geophysics, Geosystems* is an electronic journal that publishes research in chemistry, physics, and biology of Earth processes. Interdisciplinary papers are encouraged. G<sup>3</sup> seeks original scientific papers pertaining to understanding the Earth and its processes, including observational, experimental, and theoretical investigations of the solid Earth, hydrosphere, atmosphere, and biosphere at all spatial and temporal scales. G<sup>3</sup> publishes several categories of papers, including Technical Briefs and Data Briefs. G<sup>3</sup> encourages proposals for "theme" issues that encompass the full range of article types at the cutting edge of electronic publishing, encouraging the innovative use of data visualization formats.

- Call for papers for new themes:  
**Formation and Evolution of Oceanic Crust and Eolian Dust...**

**Editors:** Laurent D. Labeyrie, Vincent J. Parnet, John A. Tarduno, and Peter E. van Kesteren



Published by AGU and the Geochemical Society



## Welcome to the CODATA Data Science Journal

The Data Science Journal is a *Journal of the Committee on Data for Science and Technology (CODATA) of the International Council for Science (ICSU)*

ISSN 1683-1470

The Data Science Journal is a peer-reviewed electronic journal publishing papers on the management of data and databases in Science and Technology. Details can be found in the [prospectus](#). The [scope](#) of the journal includes descriptions of data systems, their publication on the internet, applications and legal issues. All of the Sciences are covered, including the Physical Sciences, Engineering, the Geosciences and the Biosciences, along with Agriculture and the Medical Science.

The journal publishes papers about data and data systems; it does not publish data or data compilations. However it may publish papers about methods of data compilation or analysis.

We would like to acknowledge, with thanks, the financial contribution of UNESCO to the funding of the journal.

- [Home](#)
- [Current Issue](#)
- [Previous Volumes](#)
- [Submission Instructions](#)
- [Scope of the Journal](#)
- [Submission Terms & Conditions](#)
- [Sample Submission Paper \(PDF\)](#)
- [For Reviewers](#)
- [Editorial Board](#)
- [Contact](#)

