Data Integration Strategies

International Delta Roundtable

November 2007

U.S. Department of the Interior
U.S. Geological Survey

USGS Data Integration Vision and Drivers

- **USGS Science Strategy: *Science in the Next Decade***
- **Data Integration and Interoperability Blueprint:**

  "*Provide a robust, well-managed suite of enterprise application development and hosting services for the USGS"*
- **USGS Geospatial Information Office (GIO) Vision:**

  "Provide an efficient and effective integrated information environment that supports science and decision-making."
- **IT Business Strategy call for more efficient delivery of IT services, and improved workforce planning.**
- **Federal IT Governance mandates (EA, CPIC, FISMA, etc.)**

USGS

Data Integration and Interoperability: What is it?

Vision from the Science Strategy:

An individual scientist studying soil geochemistry in the Great Basin connects to a worldwide scientific computing and collaboration platform that includes thousands of USGS and other scientific research databases. She seamlessly visualizes her soil data and compares them to data on paleoclimate, rock geochemistry, dust sources and composition, geomicrobiology of local ecosystems, past and present surface- and ground-water chemistry, current satellite imagery, elevation, slope and other relevant data sets. Scientific inquiry within and outside USGS and external customers all benefit from the enhanced accessibility of decades of observational data and analysis.

Data Integration and Interoperability: What is it?

- A USGS-Wide, Decadal Effort
- A Vision based on the Science Strategy and Existing Efforts in all Disciplines
- A Broad Suite of Projects
- A Large Effort that is Challenging to Convey Requiring an Effective Communications Strategy

Data Integration and Interoperability: What is it?

A Spectrum of Integration:

Access & Discovery:

Portals, Libraries,

Registries, Catalogs

Standards for Exchange:

Protocols, Metadata,

Structure and Semantics

Integrated Data Model:

Ontology or Taxonomy that has been developed and agreed to. Also known as a Data Dictionary.

Data Integration and Interoperability: Why Now?

- Previous efforts have made advances
- Success has been demonstrated
- New and evolving technologies can enable it
- Scientists still need it
- New understanding requires it
- If we don’t, others will

USGS Data and Information Resources

Digital Data

- BASIS+
- Publishing
- Land Cover
- National Map
- ANSS
- NBII
- NWIS

application
database
web service

DATA.USGS.GOV

application
database
web service

application
database
web service

- Visualization Tools
- Models
- Web site
- Downloadable Files
- Science Team Server
- Project Data

Physical Data

- Printed Pubs
- Paper Maps
- Analog Data Files
- Specimens
- Ice Core Lab
- Core Samples

Index

web service

External Data Sources

- NOAA, NASA
- DOI Bureaus
- University
- State/Local
- International
- NEON, GEON

application
database
web service

Science Web Services Bus (connected globally)

Services and Metadata Registries

Data Models

- Hazards
- Ecosystems
- Energy
- Terrestrial Climate

Discovery & Access Portals

Quantitative Scientific Models

- Visualization Tools
- Individual Models
- Modeling Frameworks
- CO-LAB Interfaces

Informatics Applications

- GIScience
- Hydroinformantics
- Bioinformatics
- Geoinformatics

Internal & External User Communities

Data Integration and Interoperability: Key Building Blocks

- Authoritative Data Sources and Discipline-Based Data Stewards
- Data Content Standards (Data Models)
- Comprehensive Science Catalog (Data, Publications, Specimens, Projects, etc.)
- Processes & Tools for Indexing Science Content (by location, topic, time)
- Data Hosting Capability
- Web Services Architecture
- Discovery and Access Portals
- Models and Modeling Tools
- Communities of Practice
- Future Workforce Strategies

Data Integration and Interoperability: Approach

- Actively engage all parts of the USGS in design and implementation
- Use Climate Change as a test bed for data integration approaches that can be used broadly
- Leverage existing best practices from both within and outside USGS
- Divide the effort into high impact segments that can be executed quickly to deliver results and benefits
- Leverage existing funding and programs in the short-term (no new funding) while planning for possible new resources for longer-term goals

Data Integration Blueprint
GIO-Sponsored Building Blocks

- Enterprise Hosting Infrastructure
- Comprehensive Science Catalog
- Geospatial Management Information System
- myUSGS
- The National Map
- Data Modeling
- Data Improvement

USGS Data Integration Strategies Summary

- Leverage existing activities within USGS
- Use data interoperability standards
- Implement GIO-sponsored building blocks
- Educate USGS scientists about efforts
- Partner with others to enable success
- Institutionalize data management as part of science projects

USGS Data Integration Strategies

Contact Information:

Cheryl Morris

US Geological Survey

Central Region Geospatial Information Office

Lakewood, CO

cmorris@usgs.gov

303-202-4774