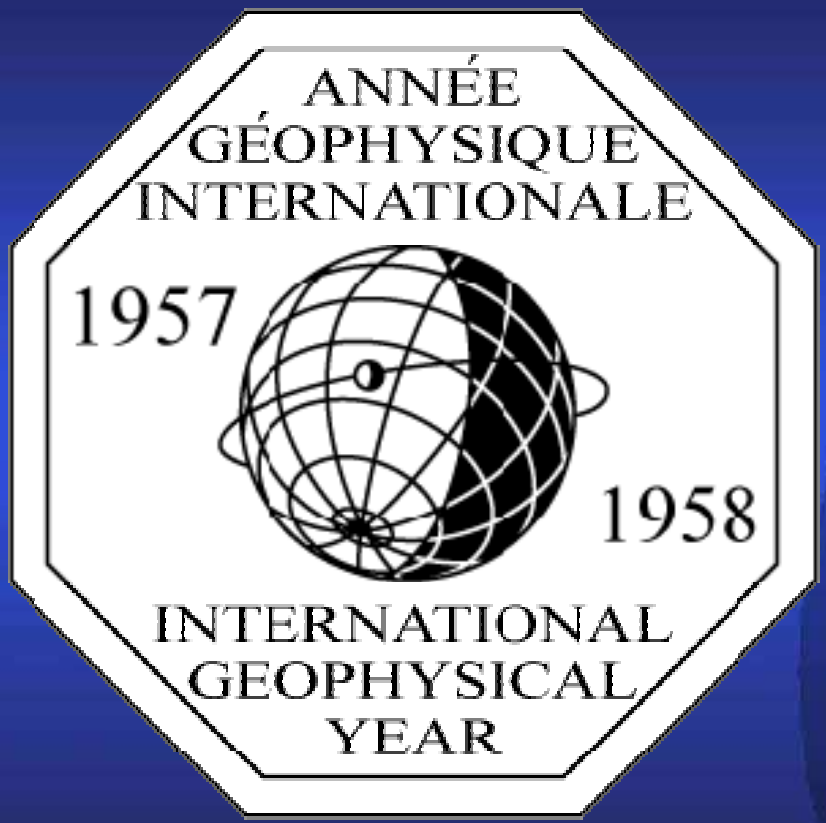
e-Science for Geoscience:

Virtual Observatories in the Framework of "Electronic Geophysical Year"

Vladimir Papitashvili
(SPRL, University of Michigan)
and the *e*GY Team

http://www.egy.org

The *e*GY Motivation: An overwhelming success of the International Geophysical Year, 1957-1958



IGY Legacies:

- Allowed scientists from different countries to participate in global observations of geophysical phenomena using similar instruments and data processing methodologies
- Gathered unprecedented volume of geophysical data from around the World
- Launched first Earth artificial satellites and established the World Data Center System

People Reaction: IGY + 50
New International Initiatives

International Polar Year (IPY, 2007-2009) will expand understanding of a key role of polar regions in the globally-linked environment

International Year of Planet Earth (2007) will be interpreting Earth's history as a basis for forecasting likely future events

International Heliophysical Year (2007) is a grass-root effort with the goals of fostering international cooperation in the study of Universal Physical Processes across the Solar System

Climate and Weather of the Sun-Earth System (2005-2008) continues a series of the SCOSTEP research programmes as IMS (1976-79), SMY (1979-81), STEP (1990-97), and S-RAMP (1998-2002) with a goal to track the solar-terrestrial processes throughout the entire Sun-Earth system

Electronic Geophysical Year, 2007-2008: Sponsored by the International Union of Geodesy and Geophysics and endorsed by ICSU, the *e*GY is to promote a revolution in geoscientific data availability and access worldwide. This coordinated international initiative will make full use of the capabilities offered by the Internet and Web-based information management and digital communications.

Data Collection Paradigm since IGY:
To get scientific data from various, mostly physically distributed sources, a scientist had to:

1. Search through a number of World Data Centers, various research institutions, physical observatories, contact colleagues...

2. Get data via snail-mail and air-mail, but only recently via e-mail and World Wide Web…

3. Then ingest retrieved data into a personal (local) database…

4. Process collected data using mostly proprietary codes, run models…

and then…

5. Finally, do some real science with the collected data!

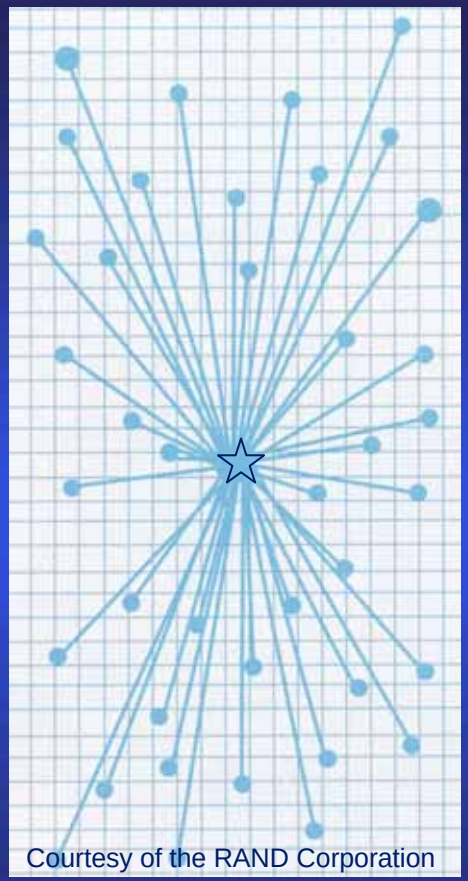
Ever Increasing Requirements:

Geospace and Earth Systems Science

Higher resolution in space and time

Assimilation into models

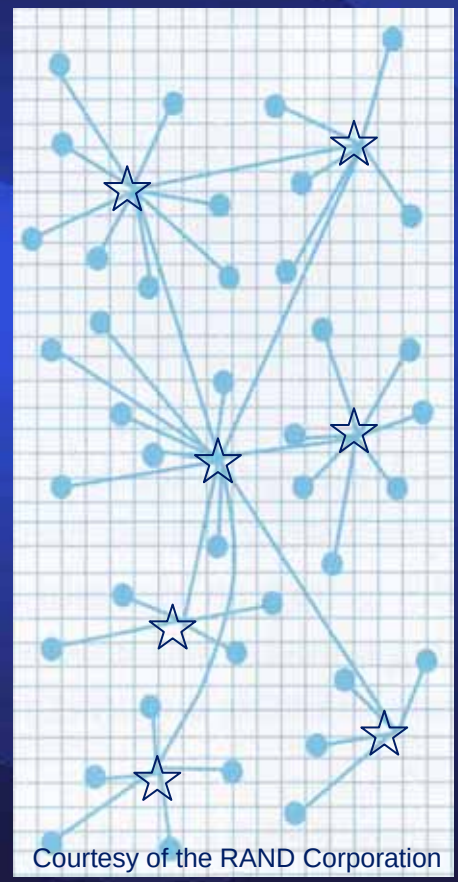
Current Paradigm of Sharing Geoscience Data: Data Must be Submitted to Data Centers



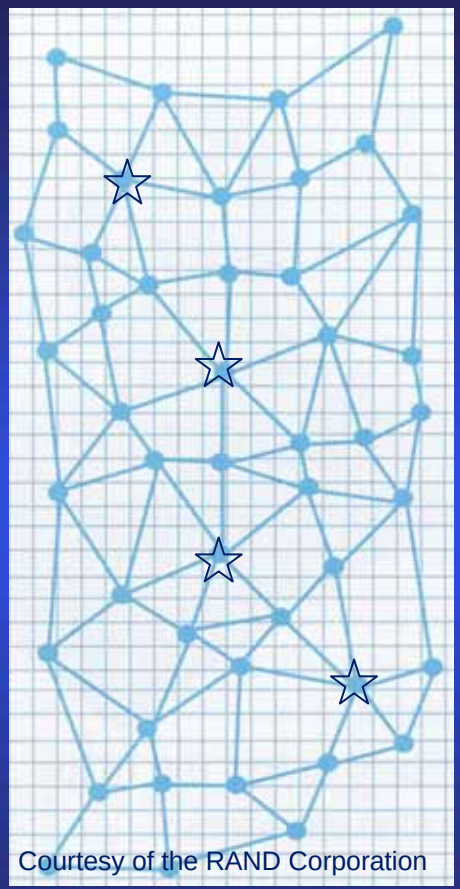
"Push Data" Concept

Centralized distribution schemes – World Data Centers System (WDC, ☆):

- Require continuous support for the data acquisition, storage, and distribution
- Submission of data remains voluntary
- Collected data are often not suitable for submission; e.g., the WDCs only accept absolute geomagnetic measurements



A New Paradigm for *e*GY: Sharing Distributed Geoscience Data via Virtual Observatories Deployed in Cyberspace



"Pull Data" Concept

Publishing and sharing geoscience data through World Wide Web:

- Allows to avoid additional steps in data preparation for submission to WDCs - to where data are now **pulled** from providers
- Data providers achieve greater visibility amongst scientific and user communities
- A **Grid** (or **Fabric**) of interconnected data nodes is **a new vision** of distributed, self-populating data repositories and centers
- **World Data Centers** ☆ become an **integral** and **important** part of the **World Wide Data Fabric**, serving as "clearing houses" for preserving data permanently in the **Fabric**

Main Elements of a Virtual Observatory

Distributed data bases are accessed through the World Wide Web Data Portals and VxO nodes

Data Visualization

Format Conversion

Data Acquisition

Location Discovery

This is a basic concept of the Electronic Geophysical Year initiative we offer to IPY, IHY, IYPE, and World Data Centers

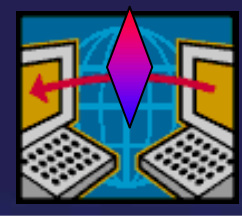
Virtual Observatories Network

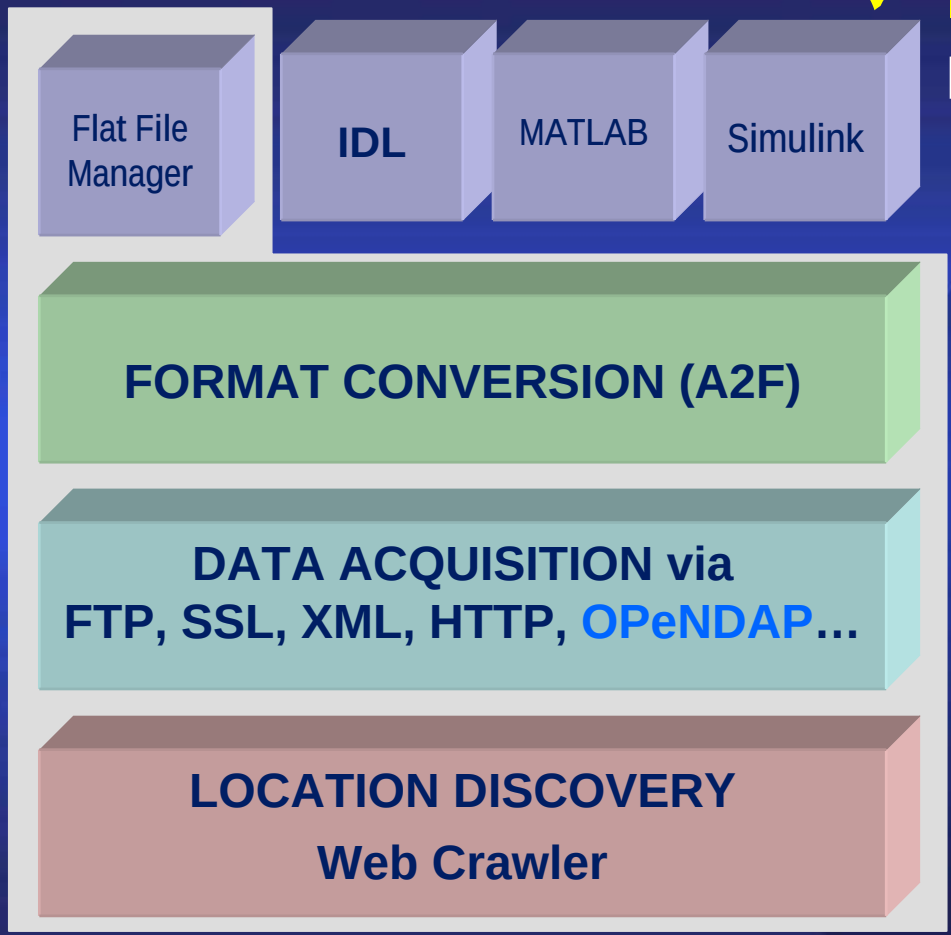
The proposed **VxO.NET** is a middleware that provides a new way for the worldwide geoscience communities to share data and functionality in a platform-independent and location-neutral environment

Design Goals

- Identify prospective geoscience data repositories and provide transparent access to the remote databases through the common interface(s) - data portals
- Perform online acquisition and processing of remote datasets and construct self-populated databases on the portal(s) and individual user nodes
- These self-populating databases can then be made easily available to other users through future requests, thus building Data GRID-type (Data Fabric) access and computing

Virtual Observatories Network (cont'd)





✓ **Integrated Visualization Layer**

Highest Level of Data Analysis

"ASCII to Flat File Format" for ingestion of downloaded data into the Web-based Portal or GRID-node databases

Data Acquisition via World Wide Web and Internet

Lowest layer - Location Discovery Module

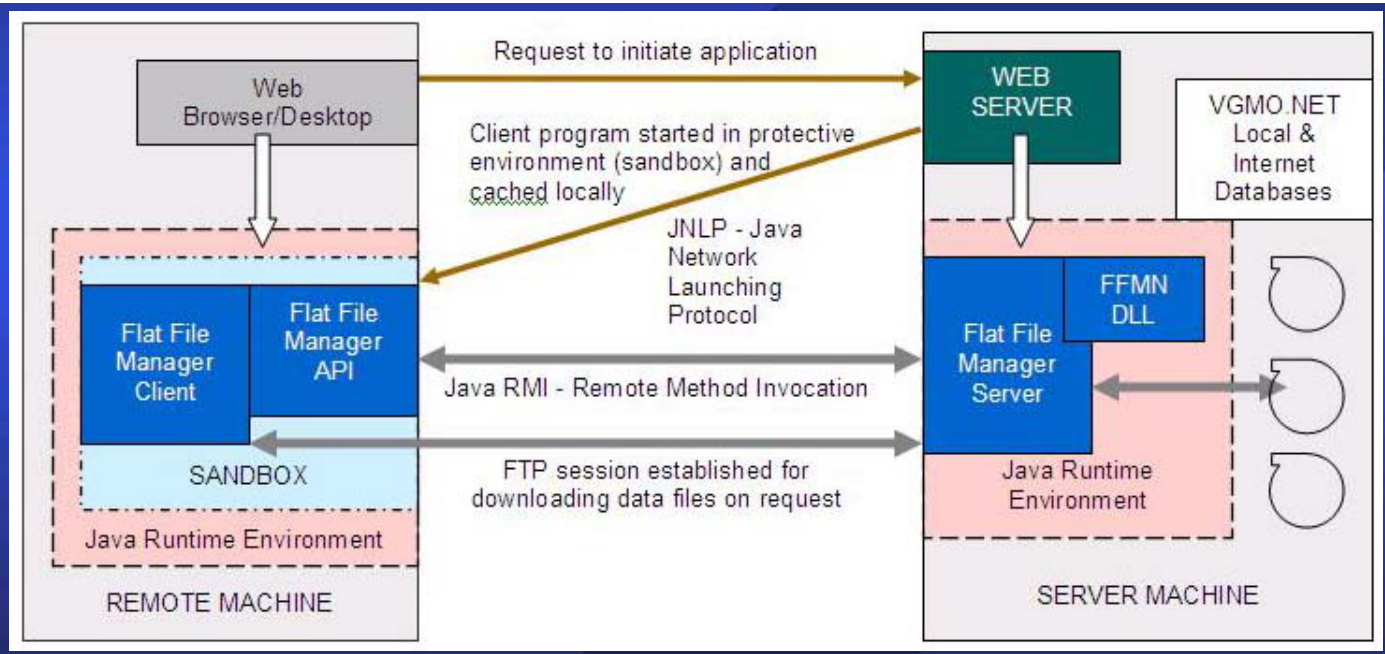
A four-tier architecture of the proposed VxO.NET

VGMO: Virtual Global Magnetic Observatory

A VxO Prototype with Architecture Unleashed

Web-based Portal – runs at the Web site http://mist.engin.umich.edu

- A secure, scalable, platform independent, and user-friendly software framework for remote access to the VGMO Flat File Manager
- The Flat File Manager Client is written to a Java 2 platform that requires a Java Web Start (Java Network Launching Protocol)



Standalone Self-Populating Data Node – get from http://mist.engin.umich.edu

- An alternate version to create, manage, and populate local geomagnetic databases from INTERNET; aims on building geomagnetic GRID access and computing

VGMO.NET: Lookup Tables and Java Interfaces

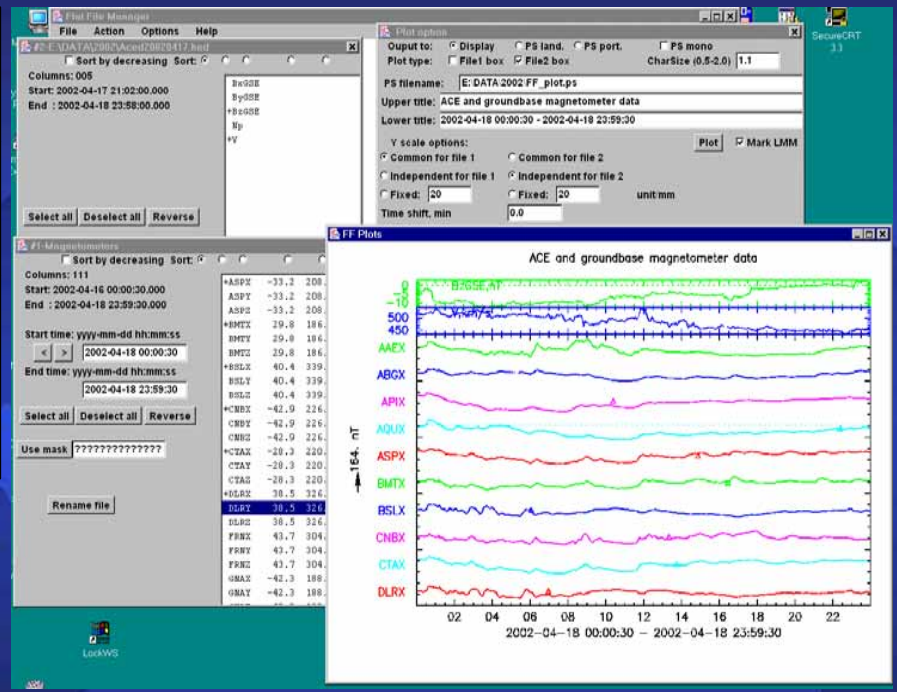
AOSS, University of Michigan

Active Sites

Remote Site	Site Info	Format Information	Conver-sion Pointer
ftp.dmi.dk	1980 - 2002	/pub/wdcc1/obsdata/ 1minval/ YYYY/	dmi.exe
ftp.ngdc. noaa.gov	1970 - 2002	/STP/GEOMAGNETIC_DA TA/ONE_ MINUTE_VALUES/YYYY/	ngdc.exe
.....	.....	.....	.....

A2F - Any Format to Flat File Conversion Module

FFMN
Flat File Manager

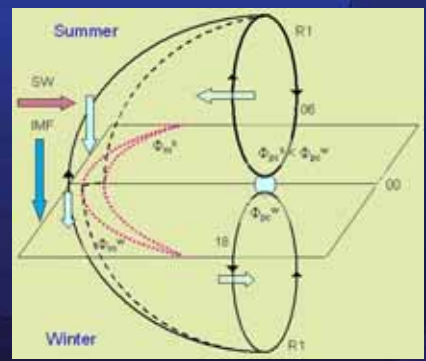


INTERNET

Prospective Sites

Remote Site	Site Info	Format Info	Conversion Pointer
ftp.iki.rssi.ru	-	-	-
ftp.abs.xyz.edu	-	-	-

Geo Magnetic Crawler (GeoMaC)



VRBO: Virtual Radiation Belt Observatory
Another Example of the VxO Architecture

Data system that meets engineering, operational, and scientific needs for:

Near Real Time Data

SAMPEX

NOAA

GOES

POES

Long Term Archival Data

Los Alamos NATIONAL LABORATORY

NOAA

THE AEROSPACE CORPORATION

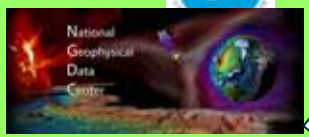
UNIVERSITY OF MARYLAND

NASA NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

LASP
U. Colorado

Gateway to distributed data

NOAA



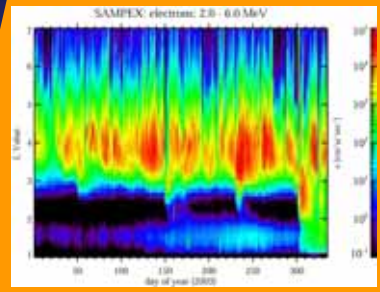
Nowcast/ Forecast Models

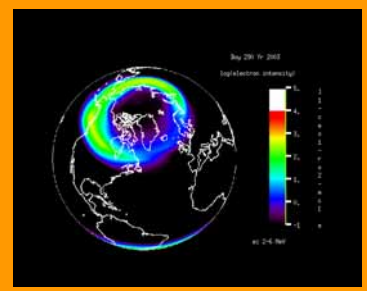
CISM End-to-End Models

Assimilation of Extreme-Event Data

Climatology Models

User Interface and Displays





- Climatology model for designing satellites
- Nowcasts/forecasts that provide situational awareness for satellite operators
- Long term archives of simulated and observed data for testing scientific theory

eGY Participants and Potential Components

International Bodies:

AOGS
CODATA
FAGS
ICSU
IUGG
IAGA
IUGS/IYPE
PAA Data & Information
SCOSTEP / CAWSES
SCAR / IPY
SEG
WDC Panel …

National Organizations

Australia – ANU, UNSW
Canada – SSDP
Russia – IZMIRAN
UK - BAS
USA:
- AGU
- Augsburg College
- NASA / IHY
- NOAA / NGDC
- NOAA / SEC
- NSF / NCAR
- U. Colorado - LASP
- U. Michigan – AOSS …

Virtual Observatories

ASTROGRID Virtual Observatory
NVO – US National Astronomy Virtual Observatory
VCSO – Virtual Carbon Cycle Observatory
VGMO – Virtual Global Magnetic Observatory
VHO – Virtual Heliophysical Observatory
VRBO – Virtual Radiation Belt Observatory
VMSO – Virtual Magnetospheric Observatory
VOO - Virtual Ocean Observatory
VSN – Virtual Seismic Network
VSO – Virtual Solar Observatory

The eGY charter sets out the principles of geoscience data stewardship for accomplishing a 21st Century vision of an "Information Society", in which all of humanity may share the benefits of free and open access to data and information. These principles are enunciated here in a geoscience context, but are applicable across all of the sciences and beyond.

eGY Objectives

- Embrace and extend the IGY principles of data sharing and scientific involvement
- Make existing and newly accrued datasets available – "free access to all"
- Digitize or convert to digital images analog geoscience records making these data available electronically
- Develop a World Wide System of Virtual Geoscience and Geospace Observatories

eGY Characteristics

- **Timeliness:** Virtual Observatories middleware is becoming widely available
- **Interdisciplinary:** Data sharing and data accessibility are common issues in all fields of geosciences
- **Affordable:** Simple networking technology
- **Cost Effective:** More/better science for money
- **Inclusive:** Opportunities for developed and developing countries
- **Capacity Building:** Provides relevant research experience for young scientists
- **Complementary to IPY, IHY, IYPE, and other international initiatives like CAWSES, ILWS, etc.**

*e*GY Timeline

2003 July: IUGG General Assembly, Sapporo, endorsed *e*GY

December: *e*GY planning discussion at AGU

2004 March: Paper describing *e*GY published in Eos

April – May: Presentations at EGU and AGU

July: *e*GY at IAGA Executive Meeting (Paris)

September: *e*GY at IUGG Executive Meeting (Boulder)

Web site: **http://www.egy.org**

2005 February: *e*GY Planning Workshop (Boulder)

April – May: Presentations at EGU and AGU

2005 – 2006 *e*GY Presentations at various IUGG Associations assemblies and other international conferences

2007 – 2008 *e*GY coordinates networks of Virtual Observatories for IPY, IYPE, IHY, CAWSES, ICESTAR, etc.

What does *e*GY mean for me? How do I get involved?

Data Integration and Knowledge Discovery

- With the microprocessors, software, and Internet of today – access to scientific data has become effectively infinite and instantaneous
- The challenges for the future go beyond the data access but towards the objective integration of distributed data based on user-defined criteria to discover new knowledge
- This knowledge facilitates meaningful interpretations and decision-making for the benefit society at global to local scales

Virtual Observatories

- Stimulate, facilitate, and promote the establishment of Virtual Observatories in Geosciences
- Promote sharing of the VO “know-how” and standards, inform the I*Y science community, funding agencies, students, and the public about available Virtual Observatories
- Use http://www.egy.org as a portal to various **VxO**

Summary: How is *e*GY going to help me and my science?

- The *e*GY is an "umbrella" initiative which envisions the massive deployment of Virtual Geoscience Observatories in cyberspace
- The *e*GY major themes:
  - data access and release
  - data discovery
  - data preservation and rescue
  - common data standards and interoperability
  + capacity building and outreach

http://www.egy.org

- The *e*GY promotes the "Pull Data" concept illustrated here by the proposed VxO.NET
- The *e*GY offers a forum to discuss, evaluate, and promote all of the various VxO concepts
- **The *e*GY invites geoscience communities to join in making international and interdisciplinary data exchange easier using the Virtual Observatories approach**