The Electronic Geophysical Year (2007–2008): eScience for the 21st Century

EMILY COBABE-AMMANN, W. K. PETERSON, and DANIEL BAKER, University of Colorado, Boulder, USA
PETER FOX, National Center for Atmospheric Research, Boulder, USA
CHARLES BARTON, Australian National University, Canberra, Australia

Imagine a world in which scientific data can be readily accessed by anyone seeking them. This, in a nutshell, is the premise of the Electronic Geophysical Year (*e*GY), one of the "international science years" commemorating the 50th anniversary of the International Geophysical Year (IGY). In the Earth and space sciences, as in other disciplines, ready and open access to the vast and growing collections of cross-disciplinary digital information is the key to understanding and responding to complex Earth and space system phenomena that influence human survival. We have a shared responsibility to create and implement strategies to realize the full potential of digital information and services for present and future generations.

Two developments have brought us to the threshold of a new revolution in our understanding of Earth and space science. First, our ability to collect data has increased dramatically, with pervasive networks of observational stations on the ground, in the oceans, in the atmosphere, and in space. This "wiring" of the planet is, in part, a legacy of the IGY, a time when a global network of observatories and the World Data Centers were established. As a result, today, petabytes of data about our planet are collected daily—data that are critical for our understanding of our planet and how it is changing.

Second, modern digital communications and information management provide an unprecedented ability to access and share information and processing capability (Figure 1). The Internet has become the "great equalizer" for communities around the world, connecting us all politically, culturally, and scientifically. For scientific data, open access to data and information from around the globe means that everyone can participate in the forward flow of scientific progress in unprecedented numbers. It means that local scientific phenomena can be placed into a global picture readily and with ease.

The *e*GY's focus on data stewardship for the 21st century embraces and extends the principles exemplified in the IGY. The power of universal access to scientific data fulfills an important role and responsibility of developed countries to developing countries, providing a framework that can be leveraged into the free and productive development of societies around the world.

Goals of the *e*GY. The *e*GY is working to develop a 21st century "*e*-science" approach to issues of data stewardship making past, present, and future Earth and space science data openly and readily available. The vast amount of data available to us in the geosciences allows us the opportunity to tackle large-scale questions within the Earth system. To have a further step increase in knowledge and understanding, a cross-disciplinary approach will be essential to address the complex interactions and feedback mechanisms that operate over wide spatial and temporal scales. To achieve this, increased access to data and models from different science areas will be necessary, and this offers new challenges in data management.

The *e*GY provides a cooperative environment and an international mandate to help revolutionize data availability and access worldwide. Through *e*GY, the scientific community, as well as other interested stakeholders, can share experiences and expertise. Coordination of activities reduces replication

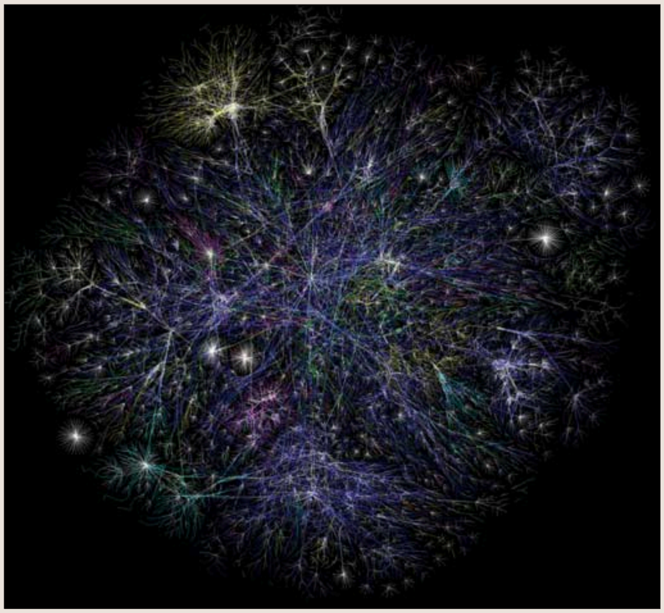

Figure 1. *An artist's conception of the Internet (Courtesy of Charles Barton).*

and promotes standard sets of best practices that will ultimately lead to increases in interoperability. The *e*GY focuses on several areas to meet these goals, including:

- Ready and open electronic access to geoscience data, information, and services; establishment of virtual observatories
- Data discovery (i.e., finding out who holds what, where, and how; promotion of metadata usage and standards)
- Data release (i.e., securing access permission; encouraging active rather than passive release of data)
- Data preservation and archival (i.e., making old data accessible; preservation of existing data)
- Data rescue (i.e., identification of critical data sets at risk; conversion of old data into modern digital form)
- Data integration and knowledge discovery (i.e., support for information integration; development of information systems that enable the identification and understanding of relationships)
- Capacity building (i.e., developing opportunities for growth of science and reduction of the "digital divide")
- Education and public outreach (i.e., raising awareness and informing students, scientists, decision-makers, and the public)

The *e*GY also promotes the development of showcase projects and demonstrations that exemplify the best *e*GY practices. From geohazards to the geospace environment, from programs that support data modeling to education and outreach programs, the demonstrations being developed through the *e*GY year will stand as a lasting legacy.

Virtual observatories: The future of scientific access to data. A main focus of *e*GY is centered on the support, promotion,

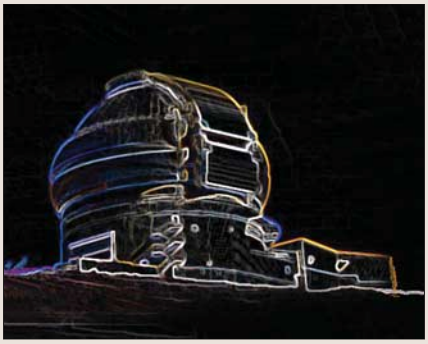

Figure 2. An artist's conception of virtual observatories (Courtesy of Erica Veil).

and facilitation of virtual observatories. A virtual observatory (VO), or any distributed data system, allows users to access multiple databases through a single portal. Data, housed at multiple sites, can range from deep archive data to near-real-time data. The VO portal or interface holds the key for the user, for while every database has its own unique set of characteristics, the user never sees these idiosyncracies. The job of the VO is to mediate the differences between the data bases to provide the user with seamless access to data, regardless of the source (Figure 2).

Increasingly, the field of geoinformatics has been focusing on ways to make that interface more effective, flexible, and responsive to the needs of the user. Using state-of-the-art semantic Web and other information technologies, VOs can be built to identify/locate data from around the world and display them on any Internet-connected machine or integrate the data into many different types programs and displays on Internet-connected workstations.

The Virtual Solar Observatory (http://umbra.nascom.nasa.gov/vso/) is an example of a fully functional, highly developed virtual observatory that exemplifies many of the aspects of distributed data systems that *e*GY is promoting. Data from dozens of solar satellites and ground stations are available through the interface. Searchable by time interval, instrument, source, spectral range, as well as common terms used to describe data products, the Virtual Solar Observatory offers easy access to state-of-the-art solar data.

Another example of VO technology is the IRIS network for seismic activity. Seismic data are not only of interest to active scientists; they are critical input for tsunami early warning systems that are being developed. Another example is the Madrigal Virtual Observatory which provides coordinated access to world-wide incoherent scatter radar data which samples the Earth's middle and upper atmosphere (ionosphere, aurora, etc.).

As VO infrastructure is being accessed on a daily basis by very diverse user communities, it is becoming a critical and often invisible component of the technical infrastructure that must be kept in place—available, robust under load, scalable, and able to evolve as the underlying implementation technologies evolve. This question of sustainability is now one that agencies and institutions are beginning to face.

The importance of connecting the general public from their backyard to a global picture. Scientists aren't the only ones looking for data. From policy makers faced with natural hazards to teachers looking to bring data into the classroom to the "citizen scientists" searching for more information on a topic near and dear to their heart, people without specialized scientific knowledge are coming to the Internet looking for data. What they find isn't always an easy pathway to knowledge. More often than not, virtual observatories, which have been designed to service the scientific community, do not cater to this group of users and haven't developed interfaces that allow nonspecialists to access the data they desire.

Data for the nonscientist needs context and constraint. In order for data to be accessible to the general public, the virtual observatories need to provide supporting information that defines the data for the user, as well as a limited set of choices to keep the user from being overwhelmed. One telling example was brought forward recently by a middle school teacher in Colorado who was interested in showing her class the "real data" of the 11-year solar cycle. She went to a virtual observatory that most logically would have the data she was looking for and began to search. She found that she needed to know either the name of the satellite that collected the data or the specific spectral range (e.g., UV, near-IR) in order to view the data. And though she tried playing with the interface to see if she could stumble on what she was looking for, there was so much data available in so many different forms that she ultimately gave up and turned instead to Google Images to find an image that would suffice. When asked why she didn't simply start with Google, her reply was telling, namely that she saw value in presenting to her students data that "real scientists" use because it helps students connect to the idea that they, too, can be scientists.

It is important to emphasize that the limits the Colorado teacher experienced do not come from faulty virtual observatory design. Virtual observatories can only do what they are designed to do, and most are designed to service the scientific community, not the general public. And yet, it is increasingly the case that the general public wants access to these data. Many virtual observatories have heard this message and have developed solutions that allow access to their data by a broader audience. In the simplest forms, the VO will provide background material and educational material on their subject for people to use.

The USGS has long had some of the best educational resources on earthquakes in its repertoire. While it doesn't directly have anything to do with the data available at the USGS Web site, this material does arm nonspecialists with important background information that allows them to come to the data more informed. At the other end of the spectrum, some virtual observatories are building interfaces for the general user. IRIS (www.iris.edu) and the National Virtual Observatory (www.us-vo.org) each have developed interfaces that both simplifies and constrains the data for the general user.

How to get involved with the *e*GY. The substance and spirit of the Electronic Geophysical Year is something that everyone can participate in. The *e*GY Web site (www.egy.org) offers a wide variety of opportunity to get involved with the effort, including listings of *e*GY meetings and workshops, working groups, listservs, presentations, and supporting documents. You can bring *e*GY into your world by identifying the *e*GY themes that interest you and undertake *e*GY-related activities that build on your own capabilities. You can encourage agencies, programs, national bodies, and committees within your community to become active in issues of modern data and information management. You can access expertise through the *e*GY network and use the *e*GY newsletter and website to promote your activities.

You also can support and promote *e*GY by advertising and promoting *e*GY nationally and internationally, displaying posters and distributing *e*GY literature at meetings, publishing news articles about *e*GY and its messages, and by joining in *e*GY efforts, including working groups and committees. For more information, please contact the *e*GY secretariat (bill.peterson@lasp.colorado.edu). TLE

Corresponding author: EcoBabe@lasp.colorado.edu