EGY: PROGRESS IN GLOBAL EARTH AND SPACE SCIENCE INFORMATICS

***D.N. Baker*[1] *and C.E. Barton*[2]**

[1]*Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, 80303-7814*
Email: daniel.baker@lasp.colorado.edu
[2] *Australian National University, Res. School Earth Sciences, Mills Road, Canberra, Australia, ACT 0200*
E-mail: cebarton@gmail.com

ABSTRACT

The past two decades have seen the development of effective ways to acquire, store, and exchange data. This technology requires adoption of community-developed standards for data storage and description to form a basis for sharing data, information, and services. For the 50th anniversary of IGY, scientific societies have promoted the establishment of a system of Virtual Observatories. The Electronic Geophysical Year (eGY) concept embraced all available and upcoming geophysical data and helped organize them into a series of virtual geophysical observatories "deployed" in cyberspace. We describe the essential features of eGY and how it fits into a 21st century vision of informatics.

Keywords: Informatics, Data Sharing, Geophysical Data, Virtual Observatories, Open Access, International Standards

1 INTRODUCTION

In the geosciences, as in other disciplines, providing widespread access to the vast and growing collections of cross-disciplinary digital information is the key to understanding and responding to complex Earth-system phenomena that influence human survival. To that end we have a shared responsibility to create and implement strategies to realize the full potential of digital information for present and future generations (David & Uhlir, 2005).

Two developments have brought us to the threshold of a modern revolution in advancing our understanding of Earth and near-Earth space. First, the ability to collect data has increased dramatically, with pervasive networks of observational stations on the ground and in the oceans, the atmosphere, and space. Second, modern digital communications and methodologies for information management provide us with an unprecedented ability to access, process, and share information (see Figure 1).

Figure 1. More than 20,000 petabytes of digital information are stored in various media in our world every year– and the rate is growing exponentially.

These developments coincide with a heightened awareness by governments of the need to substainably manage the finite natural resources of our planet, the importance of understanding Earth as a complex system, and the central role that ready access to comprehensive information and knowledge plays. That awareness translates into a growing readiness to support so-called e-science and grid infrastructures of computing resources. Broadly, this has led to an initiative to create a global information commons for science (David & Uhlir, 2005).

2 BACKGROUND

Five decades ago, scientists and policy makers embarked on a remarkable experiment in international cooperation: the International Geophysical Year of 1957-58. That successful effort gave birth to the Space Age, launched the World Data Center system, made a commitment to free exchange of geophysical data, and put in place an observatory structure that serves the scientific community to this day. The anniversary years of 2007-2008 were widely known as the "IGY+50" period.

Demographers tell us that there were 2.9 billion people on Earth in 1957. Today there are over 6.8 billion. The world of 50 years ago was dangerously divided into two major political blocs that competed militarily, socially, and ideologically. In many ways, only science provided a common ground for communication and a forum for cooperation. The IGY recognized that all people live in one world. It made clear that we have to study the Earth, the oceans, the atmosphere, and surrounding space as one coupled system. The data gathered about Earth and space systems during the IGY were – at least ideally – to be shared widely, openly, and freely (Korsmo, 2007).

IGY and the launching of the space age gave everyone more than ever before a sense of the smallness, the unity, and the fragility of the world we inhabit. Today we live in a much different world. Not only has the human population more than doubled, but the political and social climate has altered dramatically. But even with all the political and social changes that have occurred in the past five decades, many things have stayed the same. We still have only one world to live in. We must still be good stewards of the land, the oceans, the air we breathe, and outer space. We certainly must still share widely and freely what we learn from science in order to try to make the world a better place. For access to good information is the key to understanding and managing the diverse challenges facing society, and sharing such information openly is the achieving cooperative global action.

Humanity is poised to take the next major steps toward an interdisciplinary, worldwide revolution in the way we store, access, and analyze information (see Uhlir & Schröder, 2007). For the geosciences, our ability to gather data about the Earth and its space environment is unprecedented. We can obtain data and services via the Internet and grid systems from anywhere in the world; we can store and serve data with true interoperability, and we can deal with real-time data applications, assimilate date into models, build virtual observatories, and more (Baker et al., 2008).

The challenges of organizing and using data effectively expand as data volumes, data complexity, the need for interoperability, and our ability to access data and information increase. In particular, there remains great reluctance among research scientists and others to invest time in good data management practices and thereby ensure that publicly funded data are openly available for use and reuse. The reason is simple: research scientists are rewarded only for doing research. The science community lacks a comprehensive system for publishing and citing data sets and for rewarding efforts to make data sets freely available and interoperable.

To help build the foundation for developing such a system, the International Union of Geodesy and Geophysics, led by the International Association of Geomagnetism and Aeronomy, initiated the Electronic Geophysical Year (eGY) of 2007-2008 (see Figure 2). The formal themes of eGY were embedded in the eGY *Declaration for an Earth and Space Science Information Commons* (www.eGY.org). By focusing on eGY, we highlight the growth of informatics and its importance in Earth and space science research. This has led to the vision of an information commons in the 1957-1958 International Geophysical Year (IGY) and its legacy in recent times in eGY. In particular, we highlight the role of virtual observatories and similar cyber-based systems for dealing with the large, diverse data and information requirements of modern research and attitudinal and reward system changes that are required as we adapt to the emerging new and more effective ways of conducting our science.

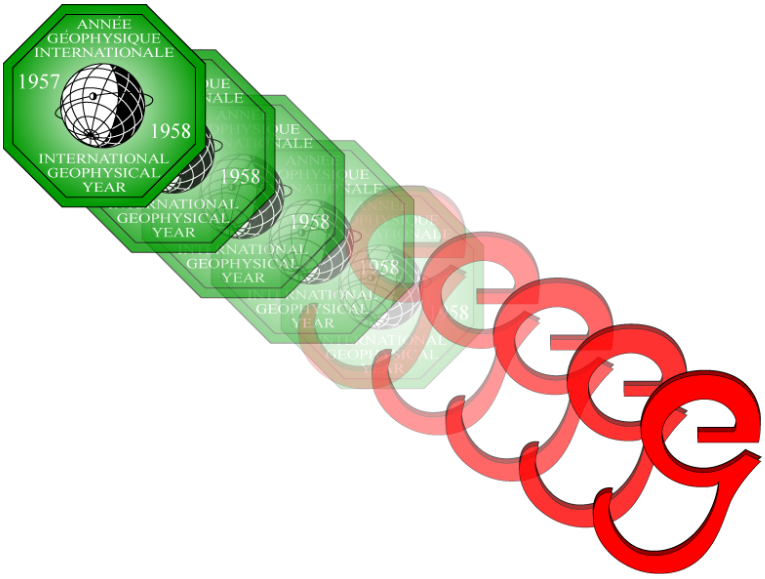

Figure 2. The Electronic Geophysical Year (eGY) is the heir to the IGY principles of free and open exchange of data.

3 THE INFORMATICS REVOLUTION

In simple terms, informatics is the science and engineering that occupies the gap separating information and communications technology (ICT) systems and cyberinfrastructure (computers, grids, Web services, etc.) on one side from the use of digital data, information, and related services for research and knowledge generation on the other (Baker et al., 2008).

Wikipedia defines informatics as "the science of information, the practice of information processing, and the engineering of information systems. Informatics studies the structure, algorithms, behavior, and interactions of natural and artificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields....This has led to the study of informatics that has computational, cognitive and social aspects, including study of the social impact of information technologies."

Three emerging subdisciplines of informatics are illustrated in Figure 3: (1) cyberinformatics, which focuses on the interface with computing infrastructures and has the strongest technical and engineering bias; (2) core informatics, which deals with informatics as a discipline in its own right; and (3) science informatics (also called X-informatics), which delivers relevant and valuable services to a wide range of users such as research workers, decision makers, and the public (Baker et al., 2008).

ICT/computing infrastructure, grids, systems	**INFORMATICS**			**Science/ knowledge generation**
	Cyber informatics	**Core informatics**	**Science Informatics**	

Figure 3. The sub-disciplines of informatics that bridge the gap between information and communications technology (ICT) and computing infrastructure at one end of the spectrum, and end-user activities such as science

research, knowledge generation, and decision-making at the other. In introducing these sub-disciplines, we consider cyberinfrastructure to be discipline/application neutral and thus to fall into science/knowledge generation (from Baker et al., 2008).

The traditional pillars of the scientific method are observation plus experiment, theory, and computation (analysis). Modern information and communication technology capabilities now allow us to address a new class of problems and applications, ones that revolve around the organization of data and information leading to knowledge extraction (see Figure 4). Examples are Google Earth and the real-time tsunami alert system, but others could be drawn from a host of fields ranging from astronomy to consumer marketing. Such new capabilities allow science to deliver benefits to society that were not possible before, particularly in areas such as climate change that involve understanding and modeling of the behavior of complex systems that transcend traditional disciplinary boundaries and are data intensive (Baker et al., 2008).

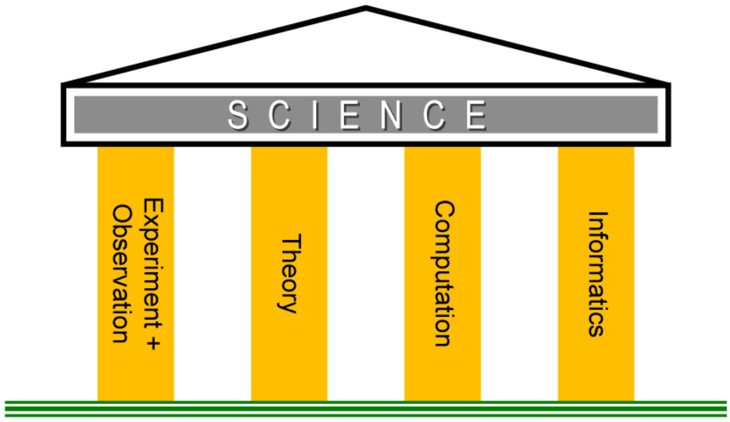

Figure 4. The pillars supporting state-of-the-art science.

It can be argued that the organization of data and information to extract knowledge has always been a core feature of science, so nothing has changed other than the introduction of a more efficient means of accomplishing this. But such an argument devalues the impact of the information revolution. The information revolution is changing radically the way we conduct our science (Baker et al., 2008), as well as affecting all facets of society, and it is little exaggeration to say that informatics has already become the fourth pillar of the scientific method (see Figure 4).

4 EGY: PROVIDING ACCESS TO DATA AND SERVICES THROUGH VIRTUAL OBSERVATORIES

Despite the informatics revolution, the traditional challenges of data preservation, data discovery, data release, and education/outreach remain. For example, as we move toward geographically and functionally distributed information systems, the issue of data permanence takes on new relevance because of the danger of losing data holdings as personnel and funding sources changes over time.

To meet these challenges, eGY called for scientists to adopt practices that optimize access, use, and re-use of scientific data, information and services for the common good (Baker, 2008). The foundations of such a commons were laid during the 1957-1958 IGY.

Traditionally, a "commons" was an area of land for grazing shared by all members of the community. As has been widely discussed in the economic and societal context, the agricultural commons collapsed because the benefit to an individual of increasing the size of his or her herd far exceeded the loss to that individual incurred from the consequent degradation of the land (Hardin, 1968). This principle has clear relevance to our present usage of Earth resources. But the same does not apply to the viability of an "information commons," there being no degradation risk— only the issue of who pays for providing information when the user gets it free.

A central feature of a modern commons approach to sharing data and information is the development of virtual observatories (VOs) - the promotion of which is a key aim of eGY. A VO is a system of interoperable, distributed data repositories made accessible through grid and Web services, allowing large and small groups alike to cope with the complexities of Internet access to data and services regardless of scale. The term (Szalay & Gray, 2001) was coined by the astronomical community (hence the name "observatory") but has since been adopted in other science disciplines (see Figure 5). The architectural features of a VO comprise access through a browser or an application programming interface (API); a registry of Extensible Markup Language (XML) data service schema to construct appropriate queries for each relevant data service; query refinement via the browser or API; and final data transfer direct to the user (no middleman).

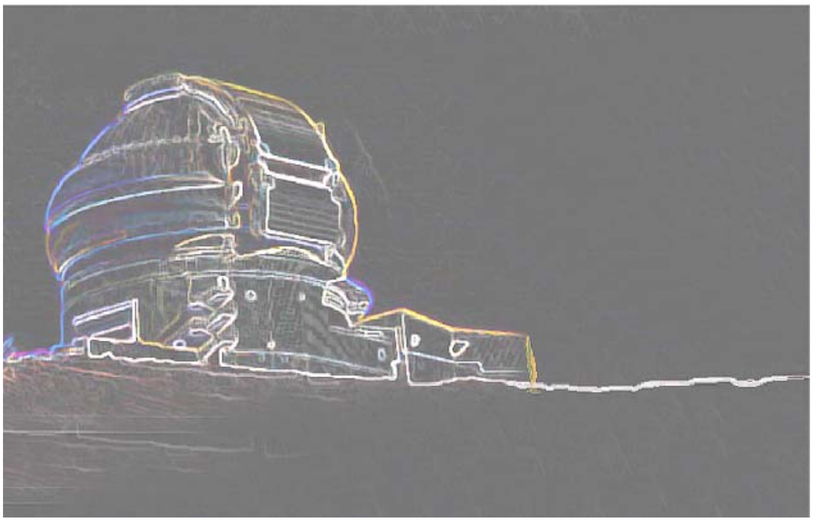

Figure 5. The iconic image of the virtual observatory concept that was used in the Virtual Observatories in Geosciences (VOIG) 2007 Workshop.

Many modern information systems not formally called VOs, share the same architectural elements and offer the same features of data discovery, location, acquisition, format conversion, analysis, and visualization. Large systems based on VO technology are emerging in many discipline areas. For example, the EarthScope program, funded by the U.S. National Science Foundation, (http://www.earthscope.org/) is an ambitious program to foster multidisciplinary research efforts across the Earth sciences utilizing the freely accessible data collected and maintained by EarthScope facilities (for example, Shapiro et al., 2005), Incorporated Research Institutions for Seismology (IRIS), and Stanford University. Data and data products are openly available from thousands of geophysical instruments worldwide that record seismic waves and measure motions and deformation of the Earth's surface (Rundle et al., 2003). The addition of meteorological sensors at Global Positioning System sites extends the use of these data to atmospheric and ionospheric researchers.

Another example is the Space Physics Archive Search and Extract (SPASE) data model, which provides a basic set of terms and values organized in a simple and homogeneous way (Harvey, et al., 2008) to facilitate access to solar and space physics resources (http://www.space-group.org/). The SPASE system is comparable to the approaches developed by the Planetary Data System (PDS) and the International Virtual Observatory Alliance (IVOA) for planetary and astronomical data, respectively. SPASE provides detailed information required for solar and space physics applications at the parameter level, allowing users to more easily access and analyze all available data. The SPASE-based product descriptions are expected to be linked to relevant virtual observatories, data centers, and individual data and model providers. SPASE will continue to evolve in a controlled way as data and service providers and benefiting researchers suggest improvements to extend its framework of common standards.

5 BENEFITS OF A SHARED INFORMATION ENVIRONMENT

The benefits of cooperation and data sharing have led the eGY to help build a framework to begin globally coordinating geoscience (informatics) efforts. This international coordination will help maximize the value of free and open exchange of data and of sharing the benefits equally among nations. Data exchange is evolving so that most of the information transfer is done between machines using state-of-the-art virtual observatory software. A vivid example of the power of virtual observatory technology can be seen on the seismic monitor webpage of IRIS (http://www.iris.edu/seimon/bigmap/index.phtml). Data from worldwide, near real-time seismic sensors and from an archival database are combined to provide a global image that can be used by scientists, teachers, and policymakers. Perhaps the clearest and most compelling example of where the vision of a complex systems integration is needed is in the global climate change program. To understand climate change, and to make policy decisions about it, there will need to be active, integrated research in the atmosphere, the hydrosphere, the cryosphere, and the biosphere (Figure 6). Underpinning all of this is the pressing need to involve policy makers and to engage the public in order to effect change. Given the inherently interdisciplinary requirements of this program, a broad informatics approach will be essential for achieving successful outcomes from climate change efforts.

Figure 6. The global climate change program is an example of the need to understand the complex interplay between many systems. Creating ready access to data and information across diverse disciplines will greatly facilitate such programs.

6 SUMMARY

The mission of the eGY was to foster an international commitment, secure a mandate, and provide an international coordination framework to facilitate, inform, stimulate, encourage, and promote the following:

- Cooperation among scientific agencies, institutions, and programs to reduce duplication and encourage standards.
- Data discovery to determine who holds what, where, and how to promote standards for metadata (information about the data's content, quality, condition, and other characteristics).
- Data release to secure access permission and to ensure active rather than passive release.
- Preservation of data and accessibility to existing data.
- Data integration and knowledge to support information sharing needs and to develop information systems that enable the identification and understanding of relationships between various data seta and measurements.
- Capacity building to boost the scope and output of scientific endeavors and to help develop opportunities for growth of science and reduction of the digital divide in countries in need.
- Outreach to raise awareness and inform students, scientists, decision makers, and the public and to promote environmental literacy.

We hope that eGY, through its organizational and facilitating functions, has produced important benefits to the geoscience, and that the effects will extend well into the future.

7 REFERENCES

Baker, D.N. (2008) A 21st century vision for geophysical data management, *Physics Today 6*(9), pp. 54-55.

Baker, D.N., Barton, C., Peterson, W., & Fox, P. (2008) Informatics and the 2007-2008 Electronic Geophysical Year. *Eos 89* (48), pp. 485-500.

David, P.A. & Uhlir, P. (2005) Creating the Global Information Commons for Science. CODATA. Retrieved from the WWW, August 26, 2009: http://www.codata.org/wsis/GICSI-prospectus.html.

Hardin, G. (1968) The tragedy of the commons. *Science 162* (3859), pp. 1243–1248.

Harvey, C.C., et al. (2008) Virtual observatories for space and solar physics research. *Earth Sci. Inform.* 1:5-13, doi:10.1007/s12145-008-0008-1.

Korsmo, F. (2007) The genesis of the International Geophysical Year, *Physics Today 60* (7), p. 38.

Rundle, J.B., et al. (2003) Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Rev. Geophys. 41*(4), 1019, doi:10.1029/2003RG000135.

Shapiro, N.M., Campillo, M., Stehly, L., & Ritzwoller, M. (2005) High-resolution surface-wave tomography from ambient seismic noise. *Science 307* (5715), pp. 1615-1618, doi:10.1126/science.1108339.

Szalay, A. & Gray, J. (2001) The World-Wide Telescope. *Science 293* (5537), pp. 2037-2040, doi:10.1126/science .293.5537.2037.

Uhlir, P.F. & Schröder, P. (2007) Open data for global science. *Data Science Journal 6*, pp. OD36-OD53.

(Article history: Received 13 January 2009, Accepted 04 August 2009, Available online 30 September 2009)