EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 89 NUMBER 48
25 NOVEMBER 2008
PAGES 485–500

Informatics and the 2007–2008 Electronic Geophysical Year

PAGES 485–486

Humanity is poised to take the next major steps toward an interdisciplinary, worldwide revolution in the way we store, access, and analyze information. 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 data into models, build virtual observatories, and more.

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 any recognized 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 (IUGG), led by the International Association of Geomagnetism and Aeronomy (IAGA), initiated the Electronic Geophysical Year (eGY) of 2007–2008 (see sidebar describing eGY). By focusing on eGY, we highlight the growth of informatics and its importance in Earth and space science research, as well as elsewhere; the vision of an information commons in the 1957–1958 International Geophysical Year (IGY) and its legacy today in eGY; 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 changes that are required as we adapt to the emerging new and more effective ways of conducting our science. These themes will be explored in greater depth in subsequent *Eos* spotlight articles.

The Informatics Revolution

In simple terms, informatics is the science and engineering that occupies the gap between information and communications technology (ICT) systems and cyberinfrastructure (computers, grids, Web services, etc.), and the use of digital data, information, and related services for research and knowledge generation.

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 1: (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

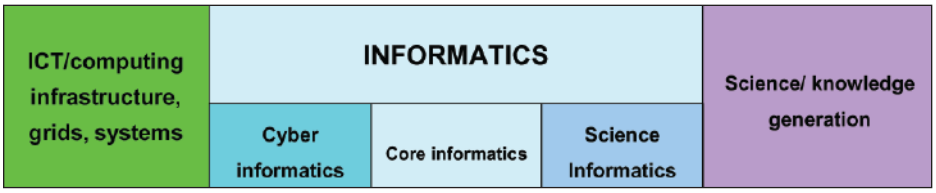

Fig. 1. The subdisciplines of informatics that bridge the gap between information and communications technology (ICT) and computing infrastructure at one end of the spectrum, and end-user activities such as science research, knowledge generation, and decision making at the other end. In introducing these subdisciplines, we consider cyberinfrastructure to be discipline/application neutral and thus to fall into science/knowledge generation.

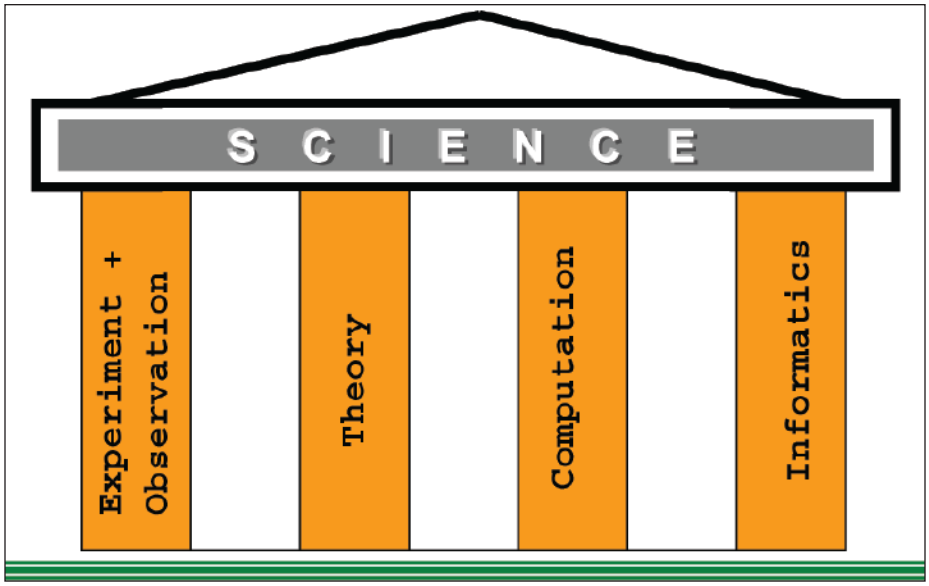

Fig. 2. The pillars supporting state-of-the-art science.

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.

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. Examples are Google Earth and real-time tsunami alert systems, 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.

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. Wheels and laborsaving machines were in use well before the industrial revolution, yet a revolution it certainly was.

The information revolution is changing radically the way we conduct our science, 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 2).

The eGY: Developing an Information Commons 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 change over time.

To encompass all these needs and meet all these challenges, eGY is calling for scientists to work toward developing an "Earth and Space Science Information Commons," which will be a virtual portal to share data. The foundations of such a commons were laid during the 1957–1958 IGY. The formal themes of eGY are embedded in the Declaration for an Earth and Space Science Information Commons (see sidebar on the eGY declaration).

Traditionally, a commons is an area whose resources are shared by all users. As has been widely discussed in the economic and societal context, the agricultural (grazing) commons collapsed because the

What Is eGY?

The eGY is the Electronic Geophysical Year, 2007–2008 (see http://www.egy.org), and is one of the four international science years (together with the International Polar Year, the International Heliophysical Year, and the International Year of Planet Earth) that mark the 50-year anniversary of the highly successful 1957–1958 International Geophysical Year (IGY). The eGY builds on the IGY legacy of providing open and ready access to data and information about the Earth and its space environment, but it does so in the context of modern information and communication technology capabilities.

The vision of eGY is that humanity will achieve a major step forward in Earth and space science capability, knowledge, and usage throughout the world by accelerating the adoption of modern, visionary practices for managing and sharing data, information, and services. The eGY provides an international framework for cooperation and action to achieve this vision.

The official launch of eGY took place on 7 July 2007 at the International Union of Geodesy and Geophysics (IUGG) General Assembly in Perugia, Italy. Activities earlier in the day included an IGY retrospective with films of IGY (courtesy of the U.S. National Academy of Sciences, available at http://lasp.colorado.edu/igy_nas/) and displays showcasing the World Data Centers and state-of-the-art virtual observatories.

The eGY Declaration for an Earth and Space Science Information Commons

"Knowledge is the common wealth of humanity"

The Electronic Geophysical Year (eGY) joins with the International Council for Science, the World Summit on the Information Society, and many other bodies in recognizing that knowledge is the "common wealth" of humanity. We have a shared responsibility to create and implement strategies to realize the full potential of digital information for present and future generations. In the 21st century and beyond, access to digital information and new technologies for information integration and knowledge discovery will influence the free and productive development of societies around the world. Providing 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 system phenomena that influence human survival.

In the Earth and space sciences, as elsewhere, the issues of concern are as follows:

Article 1: Data access

Earth system data and information should be made available electronically with interoperable approaches that facilitate open access.

Article 2: Data release

Owners, custodians, and creators of Earth system data should work together to share their digital information with the world community, though in a manner that respects intellectual property rights and security constraints.

Article 3: Data description

Providers and users of Earth system data and information should share descriptions of structure, content, and contexts to facilitate interoperability and the discovery of relationships within and between information resources.

Article 4: Data persistence

Data and information about the Earth system should be preserved and sustained in forms that are both software and hardware independent so as to be openly accessible today and in the future.

Article 5: Data rescue

Effort should be made to identify and rescue critical Earth system data and ensure persistent access to them.

Article 6: Common standards and cooperation

Standards for interoperability should be identified, created, and implemented through international collaboration.

Article 7: Capability building

Communities with advanced information technology and communications capabilities should contribute to developing such capabilities elsewhere to reduce the digital divide.

Article 8: Education and public outreach

Students, scientists, decision makers, and the public should be informed about and be enabled to contribute to our understanding and management of Earth system phenomena that affect human survival.

benefit to an individual of increasing the size of his or her herd far exceeded the shared loss incurred from the consequent degradation of the common grazing land (see G. Hardin, "The tragedy of the commons," *Science, 162*(3859), 1243–1248, 1968). The 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 creating an informatics commons 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 was coined by the astronomical community (hence the name "observatory") but has since been adopted in other science disciplines.

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). 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, 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. 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 to facilitate access to solar and space physics resources (http://www.spase-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.

Contributions in *Eos* to Spotlight eGY

Eos extends a special invitation to authors to submit contributions that spotlight scientific data and information (informatics) issues in the Earth and space sciences. This invitation is part of the Electronic Geophysical Year (eGY) effort and provides an outlet for news and views as the scientific world adapts to and exploits the opportunities of the information era.

The themes of eGY are open and ready data access, data discovery, data release, data preservation, capacity building (reducing the digital divide), and education and outreach. A central feature of eGY is the promotion of virtual observatories and similar cyber-based systems for providing ready access to data, information, and services. The eGY coverage in *Eos* aims to address all of these topics as well as broad technical and policy issues in data science and informatics.

For further information about eGY and guidelines for contributors to spotlight eGY, visit the eGY Web site (http://www.egy.org).

Tangible Efforts for Virtual Success

Although scientists are currently rewarded for science alone, the future of science and discovery will rely on data interoperability and easy access to the wealth of material that has been generated through the information age. Thus, developing VOs and an information commons requires thoughtful and strong action by scientific societies as well as by individual scientists. The eGY is an important step forward.

To be truly successful in the modern era, we must take greater care to nurture and share data fully and openly. We must also ensure that rewards are provided to community members who undertake these key activities on behalf of all science disciplines. In that vein, *Eos* is inviting authors to submit papers that highlight informatics (see sidebar on eGY contributions to *Eos*).

For further information about eGY, including how to participate and suggestions on activities to pursue, visit http://www.egy.org.

Acknowledgments

The principal sponsors of eGY are the U.S. National Science Foundation, the University of Colorado's Laboratory for Atmospheric and Space Physics, IUGG, and IAGA.

Author Information

Daniel N. Baker, Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder; E-mail: Daniel.Baker@lasp.colorado.edu; C. E. Barton, Australian National University, Canberra, Australia; W. K. Peterson, Laboratory for Atmospheric and Space Physics; and P. Fox, IUGG Union Commission for Data and Information and eGY Virtual Observatories Working Group, National Center for Atmospheric Research, Boulder, Colo.