EOS
EOS, TRANSACTIONS, AMERICAN GEOPHYSICAL UNION

VOLUME 95 NUMBER 20
20 May 2014
PAGES 165–172

Community-Developed Geoscience Cyberinfrastructure

PAGES 165–166

Discoveries in the geosciences are increasingly taking place across traditional disciplinary boundaries. The EarthCube program, a community-driven project supported by the U.S. National Science Foundation, is developing an information- and tool-sharing framework to bridge between disciplines and unlock the modern geosciences' transformative potential.

The program not only is developing cyberinfrastructure for the geosciences but also is developing and testing new processes for the community that could fundamentally transform its culture. If successful, the project could outline a path for the development, deployment, and adoption of digital capabilities across the wider scientific community. The development of EarthCube and similar projects could result in changes to the practice of science comparable in speed and breadth to the spread of the Internet or other basic infrastructures.

More than 2500 Earth, atmosphere, ocean, computer, information, and social scientists, as well as educators, data managers, and others, have engaged in elucidating the EarthCube concept. This document is a synthesis of the work through August 2013 and a snapshot of an emergent end user vision for EarthCube—a vision that will continue to evolve as EarthCube grows.

The EarthCube Vision

The EarthCube program is a community-driven project focused on building digital infrastructure for managing, sharing, and exploring geoscience data and information with the aim of increasing scientific productivity.

EarthCube aims to foster new, transformational research by enabling scientists to gain access to formerly unconnected software, models, data, and computational resources. Science scenarios developed in conjunction with the community at end user workshops will shape EarthCube's design to ensure that it provides concrete value for geoscientists and addresses the diverse needs of the geoscience community [see *EarthCube Test Governance Project Team,* 2013]. Seventeen such workshops were convened from 2012 to 2013, and the next community review of progress thus far will occur at an EarthCube All-Hands Meeting in June 2014.

The project's goal is to design, build, and maintain an easy-to-use system based on existing resources that embraces open-source culture and methods to align technology development with scientific needs. Perhaps its greatest challenges lie in identifying key capabilities that are widely useful and in focusing efforts on implementing those capabilities without becoming so complex that it discourages use.

Ideally, the EarthCube system will constitute an integral part of everyday research and decision-making workflows, coordinating hardware, software, people, processes, data, and community. A significant part of the EarthCube system will be intangible, including specifications, policies, protocols, and communities of practice, in many ways comparable to the World Wide Web.

EarthCube Workflow

EarthCube will be organized around specific workflow activities.

Data Management. EarthCube will provide a shared archive in which data, tools, and services are documented and curated, enabling reuse of data sets for new analyses. The project will embrace technology that simplifies data curation for archive, publication, and reuse and that can be applied to newly acquired data as well as existing and old data sets.

Resource Discovery. Plug-in components will enable resource discovery and direct data access using scientific software in common use, such as Excel, MATLAB, Python, R, ArcGIS, or ModFlow. Such search plug-ins will be Google-like in simplicity, with text or map-based interfaces, while being efficient and accurate. Alternatively, Web services will provide machine access for searching catalogs and real-time data use.

Data Access, Integration, and Processing Tools. User interfaces tailored for specific communities will simplify data access, visualization, and analysis using software that interoperates with EarthCube data or service providers. Data will be linked to computational capabilities for display, exploration, real-time use, and analysis using data from different sources. Through EarthCube, it will be easy to pipe results between processes to define reproducible workflows. Real-time data sources will be used to improve event-response procedures, optimize sampling, and facilitate scientific analysis.

Data Portal. One or more portals will function as user entry points to support data exploration and access tailored for specific communities. These portals might be organized around various paradigms, but one that has been frequently mentioned is a three-dimensional virtual globe for data discovery and exploration, supporting the ability to spatially integrate and display geoscientific data at varying resolutions.

Technical Challenges

The dominant challenge for EarthCube is to make data management and processing easier and less time-consuming. This challenge is a large one given that the scope of EarthCube includes not only data but also models, workflows, samples, and tools.

To be used by other researchers, data must be understood sufficiently well to be trusted and processed appropriately. This presents a wide variety of technical challenges involved in building workflows for data management, including curation, documentation, access, and integration.

For example, many legacy data issues stem from the difficulty that individual researchers, operating on limited budgets, experience in trying to curate data produced by their research. As a result, data documentation is commonly insufficient to enable cross-domain use or to repurpose data obtained from repositories. In addition, using nonstandard, heterogeneous data requires significant effort. The meaning of data may be unclear because of nonstandard vocabulary usage. Inconsistent practices for data sharing make each new data acquisition a time-consuming learning experience.

By S. M. Richard, G. Pearthree, A. K. Aufdenkampe, J. Cutcher-Gershenfeld, M. Daniels, B. Gomez, D. Kinkade, and G. Percivall

EarthCube will help address these issues by encouraging the development and adoption of community standards for Web interfaces to data, metadata and data formats, and software libraries.

Challenges of Governance and Culture

The goal for EarthCube is to continue innovating while keeping scientific productivity high. To accomplish this goal, EarthCube will need to overcome challenges related to community governance, incentives, development of communities of practice, and education of current and future generations of Earth scientists.

Community representatives involved with EarthCube governance will need to coordinate programs and pilot projects, develop metrics to evaluate system components, identify capability gaps, and promote consensus on requirements. They will also need to develop a portfolio of EarthCube policies and specifications as well as set pathways for collecting, monitoring, and acting on community feedback to refine content, practices, and policies. The community will have to establish priorities for building cyberinfrastructure, making legacy data accessible, and developing standards. Recommendations for funding will need to be based on these priorities, along with usage metrics, ongoing gap analysis, and user requirements.

Social, professional, and financial incentives are necessary to motivate good data management practices, along with the sharing of data, software, or models. Data quality metrics will need to be built into the system, along with practices to minimize data misinterpretation or misuse. Access controls and respect for data ownership are necessary to ensure that data are not shared too soon or inappropriately. Adequate credit must be given for contributing data and models—its absence would be a major deterrent to participation. Cost, effort, and technical barriers, along with concerns about misuse, must not outweigh the tangible benefits of time spent on data management and publication, especially for tenure-track and project-funded researchers.

EarthCube represents a significant shift in culture and will bring new challenges to the geosciences community. To ease the transition, the program has emphasized the involvement of early-career researchers, who will be the vanguard in changing research workflows. Success will be indicated when the geoscience community identifies itself through participation in EarthCube as data providers, data consumers, system developers, maintainers, and managers.

EarthCube Beyond Research

EarthCube can play a major role in training the next generation of cyber-savvy geoscientists by providing intuitive, modular learning objects and self-directed lessons that can be used by teachers from K–12 through to the graduate level. In particular, an EarthCube three-dimensional virtual globe will not only be a data discovery portal for researchers but will also serve as an entry point for students to explore geoscience data.

Success will be achieved when EarthCube helps everyone—including individuals outside of the geosciences—better understand how to use and interpret science data.

References

EarthCube Test Governance Project Team (2013), EarthCube end-user workshops: Executive summaries, report, Ariz. Geol. Surv., Tucson. [Available at http://workspace.earthcube.org/content/earthcube-end-user-workshops-executive-summaries.]

—Stephen M. Richard and Genevieve Pearthree, Arizona Geological Survey, Tucson; email: steve.richard@azgs.az.gov; Anthony K. Aufdenkampe, Stroud Water Research Center, Avondale, Pa.; Joel Cutcher-Gershenfeld, School of Labor and Employment Relation, University of Illinois at Urbana-Champaign; Mike Daniels, Earth Observing Laboratory, National Center for Atmospheric Research, Boulder, Colo.; Basil Gomez, Department of Geography, University of Hawai`i at Manoa, Honolulu; Danie Kinkade, Woods Hole Oceanographic Institution, Woods Hole, Mass.; and George Percivall, The Open Geospatial Consortium, Crofton, Md.