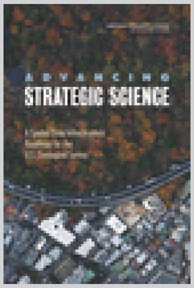

Advancing Strategic Science: A Spatial Data Infrastructure Roadmap for the U.S. Geological Survey

ISBN
978-0-309-26457-0

132 pages
6 x 9
PAPERBACK (2012)

Committee on Spatial Data Enabling USG Strategic Science in the 21st Century; Mapping Science Committee; Board on Earth Sciences and Resources; Division on Earth and Life Studies; National Research Council

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

ADVANCING
STRATEGIC SCIENCE

A Spatial Data Infrastructure Roadmap
for the U.S. Geological Survey

Committee on Spatial Data Enabling USGS
Strategic Science in the 21st Century

Mapping Science Committee

Board on Earth Sciences and Resources

Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Award No. 08HQ4G0145 between the National Academy of Sciences and the U.S. Geological Survey.

International Standard Book Number-13: 978-0-309-26457-0
International Standard Book Number-10: 0-309-26457-X

Library of Congress Catalog Card Number: 2010921862

Additional copies of this report are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu/.

Cover: Cover design by Anne Rogers. Image courtesy of Michael Hitoshi/Stone/Getty Images.

Printed in the United States of America.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON SPATIAL DATA ENABLING USGS STRATEGIC SCIENCE IN THE 21ST CENTURY

ROBERT DENARO (Chair), Nokia Corporation, Chicago, Illinois
GEORGE BRIMHALL, University of California, Berkeley, California
ROBERT CHEN, Columbia University, Palisades, New York
ANDREA DONNELLAN, Jet Propulsion Laboratory, Pasadena, California
MICHAEL EMCH, University of North Carolina at Chapel Hill, North Carolina
IAN JACKSON, British Geological Survey, Keyworth, Nottingham
JOHN KELMELIS, Pennsylvania State University, University Park, Pennsylvania
XAVIER LOPEZ, Oracle Corporation, Nashua, New Hampshire
DENNIS OJIMA, Colorado State University, Fort Collins, Colorado
BRIDGET SCANLON, The University of Texas, Austin, Texas

National Research Council Staff
MARK D. LANGE, Study Director
PEGGY TSAI, Program Officer
NICHOLAS D. ROGERS, Financial and Research Associate
JASON R. ORTEGO, Research Associate (until June 2012)
ERIC J. EDKIN, Senior Program Assistant (from September, 2010)
TONYA FONG YEE, Senior Program Assistant (until September, 2010)
NORMAN GROSSBLATT, Senior Editor

MAPPING SCIENCE COMMITTEE

DAVID R. MAIDMENT (Chair),The University of Texas, Austin, Texas
LUC E. ANSELIN, Arizona State University, Tempe, Arizona
CECILIA R. ARAGON, Lawrence Berkeley National Laboratory, Berkeley, California
BUDHENDRA L. BHADURI, Oak Ridge National Laboratory, Oak Ridge, Tennessee
SUSAN CARSON LAMBERT, EarthWorks, LLC, Frankfort, Kentucky
LEWIS A. LAPINE, South Carolina Geodetic Survey, Columbia, South Carolina
CAROLYN J. MERRY, Ohio State University, Columbus
JAYANT SHARMA, Oracle Spatial, Nashua, New Hampshire
DANIEL Z. SUI, Ohio State University, Columbus, Ohio
MICHAEL F. WORBOYS, University of Maine, Orono, Maine
MAY YUAN, University of Oklahoma, Norman, Oklahoma

National Research Council Staff
ANNE M. LINN, Senior Program Officer
ERIC J. EDKIN, Senior Program Assistant

BOARD ON EARTH SCIENCES AND RESOURCES

CORALE BRIERLEY (Chair), Brierley Consultancy, LLC, Denver, Colorado
WILLIAM E. DIETRICH, University of California, Berkeley
WILLIAM L. GRAF, University of South Carolina, Calcott
RUSSELL J. HEMLEY, Carnegie Institution of Washington, Washington, DC
MURRAY W. HITZMAN, Colorado School of Mines, Golden
EDWARD KAVAZANJIAN, JR., Arizona State University, Tempe
DAVID R. MAIDMENT,The University of Texas, Austin, Texas
ROBERT McMASTER, University of Minnesota, Minneapolis
M. MEGHAN MILLER, UNAVCO, Inc., Boulder, Colorado
ISABEL P. MONTAÑEZ, University of California, Davis
CLAUDIA INÉS MORA, Los Alamos National Laboratory, Los Alamos, New Mexico
BRIJ M. MOUDGIL, University of Florida, Gainesville
CLAYTON R. NICHOLS, Idaho National Engineering and Environmental Laboratory (retired), Sandpoint
HENRY N. POLLACK, University of Michigan, Ann Arbor
DAVID T. SANDWELL, Scripps Institution of Oceanography, La Jolla, California
PETER M. SHEARER, University of California, San Diego
REGINAL SPILLER, Allied Energy, Houston, Texas
TERRY C. WALLACE, JR., Los Alamos National Laboratory, Los Alamos, New Mexico

National Research Council Staff
ANTHONY R. DE SOUZA, Director (until April 2012)
ELIZABETH A. EIDE, Director (from April 2012)
ANNE M. LINN, Senior Program Officer
SAMMANTHA L. MAGSINO, Senior Program Officer
MARK D. LANGE, Program Officer
NICHOLAS D. ROGERS, Financial and Research Associate
COURTNEY GIBBS, Program Associate
ERIC J. EDKIN, Senior Program Assistant
CHANDA IJAMES, Program Assistant

Preface

This report summarizes the findings and recommendations of the Committee on Spatial Data Enabling USGS Strategic Science in the 21st Century. The National Research Council (NRC) has published several reports that have helped to guide the development of the spatial data infrastructure (SDI) both in the U.S. Geological Survey (USGS) and nationally (1993, 1994, 1995, 2001, 2003, 2007). Those reports envisioned an SDI for the USGS and the nation and suggested the research needed to achieve that vision. Over the last decade, the USGS has conducted breakthrough research that has overcome some of the challenges associated with implementing a large SDI. This report is intended to ground those efforts by providing a practical roadmap to full implementation of an SDI to enable the USGS to conduct strategic science.

The committee was charged by the USGS to examine progress made in establishing spatial data infrastructures and the challenges faced by them in the context of the National Spatial Data Infrastructure. The committee examined the role that the USGS can play in continuing to ensure access to high-quality geospatial data and in supporting their use in scientific analyses and decision-making through an SDI construct. The committee was charged with three main tasks: identify existing knowledge and document lessons learned during previous efforts to develop SDIs and their support of scientific endeavors; develop a vision for optimizing an SDI to organize, integrate, access, and use scientific data; and create a roadmap to guide the USGS in accomplishing the vision within the scope of the USGS Science Strategy.

To address its charge, the committee examined SDI development in local, state, national, and international contexts and solicited advice from a variety of sources. Program managers and scientists in federal agencies, state organizations, and academe provided programmatic information and user perspectives on future research directions. The committee also requested written feedback from leaders and data users in the geospatial community who generously provided guidance regarding what has and has not worked in SDI development; the major technical, organizational, cultural, policy, financial challenges still facing SDI development; and their own vision of an effective SDI at the USGS.

The committee was struck by the similarity of challenges faced by other organizations in developing their SDIs; the experiences cited in Chapter 3 on lessons learned are rich with examples of approaches that may be particularly valuable to the USGS. But a recurrent theme in nearly all the case studies was the crucial role of leadership in implementing an SDI. A strong, energetic, and inspirational leader with senior-level authority who stays with the program for the long term is the cornerstone of a successful program. The leader will be instrumental in executing the outside partnerships that are essential to the mission of the USGS and establishment of its SDI.

The names of respondents and other persons consulted by the committee are listed in Appendix B. Many of the conclusions and recommendations reached by the committee reflect ideas articulated in their thoughtful contributions; however, any errors or omissions are the responsibility of the committee, not of the external contributors. Finally, the committee expresses its gratitude to the NRC study director, Mark Lange, for his efforts in managing the committee and editing its report and to NRC staff Peggy Tsai, Jason Ortego, Eric Edkin, and Tonya Fong Yee, who assisted the committee extensively with Web site development, document tracking and assembly, and logistics.

Robert Denaro
Chair

References

NRC (National Research Council). 1993. Toward a Coordinated Spatial Data Infrastructure for the Nation. Washington, DC: National Academy Press.

NRC 1994. Promoting the National Spatial Data Infrastructure Through Partnerships. Washington, DC: National Academy Press.

NRC. 1995. A Data Foundation For The National Spatial Data Infrastructure. Washington, DC: National Academy Press.

NRC. 2001. National Spatial Data Infrastructure Partnership Programs: Rethinking the Focus. Washington, DC: National Academy Press.

NRC. 2003. Weaving a National Map: Review of the U.S. Geological Survey Concept of The National Map. Washington, DC: The National Academies Press.

NRC. 2007. A Research Agenda for Geographic Information Science at the United States Geological Survey Washington, DC: The National Academies Press.

Acknowledgments

This report has been reviewed in draft form by persons chosen for their diverse perspectives and technical expertise in accordance with procedures approved by the National Research Council's Report Review Committee. The purpose of the independent review is to provide candid and critical comments that will assist the institution in making its published report as sound as possible and to ensure that the report meets institutional standards of objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the process. We thank the following for their review of this report:

Paul Birkel, The MITRE Corporation
Virginia H. Dale, Oak Ridge National Laboratory
Ruth Duerr, National Snow and Ice Data Center
John Moeller, JJ Moeller & Associates LLC
Jay B. Parrish, Pennsylvania State University
Cyrus Shahabi, University of Southern California
David G. Tarboton, Utah State University
Bastiaan van Loenen, Delft Technical University

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by William E. Easterling, Pennsylvania State University. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests with the author committee and the institution.

Contents

APPENDIXES

Abbreviations

AIP	architecture implementation pilot
API	application programming interface
BGS	British Geological Survey
BLM	Bureau of Land Management
CBP	Containment Biology Program
CEGIS	Center of Excellence for Geographical Information Science
CEO	chief executive officer
COP	common operating picture
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science
DAAC	Data Analysis and Archiving Center
DAMA	Data Management Association
DEM	digital elevation model
DOE	Department of Energy
DOI	Department of the Interior
DOQ	digital orthophotograph quadrangle
DRG	digital raster graphic
EPA	Environmental Protection Agency
EROS	Earth Resources Observation Systems
ET	evapotranspiration
EU	European Union

FEMA	Federal Emergency Management Agency
FGDC	Federal Geographic Data Committee
FLIR	forward-looking infrared
FWS	U.S. Fish and Wildlife Service
GA	Geoscience Australia
GEO	Group on Earth Observations
GEOINT	geospatial intelligence
GeoSciML	geoscience markup language
GEOSS	Global Earth Observation System of Systems
GIS	geographic information system
GRACE	Gravity Recovery and Climate Experiment
HHS	Department of Health and Human Services
HIS	hydrologic information system
ICT	information and communications technology
IFSAR	interferometric synthetic aperture radar
IFTN	Imagery for the Nation
INSPIRE	Infrastructure for Spatial Information in the European Community
IPCC	Intergovernmental Panel on Climate Change
IT	information technology
IUGSCGI	International Union of Geological Sciences Commission for the Management and Application of Geoscience Information
LIDAR	light detection and ranging
LULC	land use–land cover
NASA	National Aeronautics and Space Administration
NAWQA	National Water Quality Assessment Program
NBII	The National Biological Information Infrastructure
NCAR	National Center for Atmospheric Research
NCGIS	National Center for Geospatial Intelligence Standards
NDCDB	National Digital Cartographic Database
NED	National Elevation Dataset
NEON	National Ecological Observatory Network
NGA	National Geospatial-Intelligence Agency
NHD	National Hydrography Dataset
NHDPlus	National Hydrography Dataset Plus
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NSDI	National Spatial Data Infrastructure

NSF	National Science Foundation
NSG	National Systems for Geospatial Intelligence
ODM	observations data model
OGC	Open Geospatial Consortium
OMB	Office of Management and Budget
OWS	open web services
PAGER	Prompt Assessment of Global Earthquakes for Response
RGE	research graded evaluation
RMSE	root mean square error
SBA	societal benefit areas
SDI	spatial data infrastructure
TM	Thematic Mapper
TNM	The National Map
TNRIS	Texas Natural Resources Information System
UAVSAR	unmanned air vehicle synthetic aperture radar
UN	United Nations
UNSDI	UN Spatial Data Infrastructure
USDA	U.S. Department of Agriculture
USGS	U.S. Geological Survey

Summary

Science is increasingly driven by data, and spatial data underpin the science directions laid out in the 2007 U.S. Geological Survey (USGS) Science Strategy. A robust framework of spatial data, metadata, tools, and a user community that is interactively connected to use spatial data in an efficient and flexible way—known as a spatial data infrastructure (SDI)—must be available for scientists and managers to find, use, and share spatial data both within and beyond the USGS. In the opinion of the Committee on Spatial Data Enabling USGS Strategic Science in the 21st Century, an SDI is so important for supporting the six Science Strategy directions that it could have had its own chapter in the Science Strategy report as an underpinning of the six directions. The committee hopes that this report can serve as the "missing chapter" of that important document.

STUDY SCOPE

The charge for this study is to describe a vision for an SDI for the USGS and create a roadmap for executing that vision starting from the current state of the SDI at the agency (see Box S.1). It is not within the scope of this study to design an SDI or to present an exhaustive list of recommended datasets for the SDI. Those activities will be the work of the USGS if it chooses to move forward with the plan outlined in this report, and some of this work is already in progress through efforts such as the USGS Council on Data Integration and other agency initiatives.

It is important to note the distinction between an SDI at the USGS and the broader and more ambitious goal of a National Spatial Data Infrastructure[1] (NSDI). The NSDI is the work of the Federal Geographic Data Committee (FGDC, 2011), and the USGS is an important contributor to this multi-partner effort. The purpose of this study was specifically to provide an SDI roadmap to support the USGS Science Strategy (USGS, 2007), therefore this report focuses on how an SDI can support science within the agency. By extension, a functional SDI at the USGS will be a key component of the NSDI to support science, analysis, and decision requirements in other federal agencies, state, local, and tribal governments, academe, and the private sector.

The USGS recently dissolved the four core disciplines of water, geology, biology, and geography and reorganized around the missions outlined in the landmark 2007 Science Strategy (USGS, 2007). The reorganization is important for SDI development because it establishes an Associate Directorship for Core Science Systems, which includes the National Geospatial Program. Because the Science Strategy outlines future science directions for the USGS, the present committee adopted the six directions in the Science Strategy—ecosystems, energy and minerals, climate and land-use change, environmental health, water, and natural hazards—as the focus of its report for optimizing an SDI. The members selected for this committee were identified to address each of the directions in the Science Strategy.

There are not likely to be any surprises in our definition of an optimal vision for an SDI at the USGS. Much has been written and debated publicly on the subject, and the agency has recently held workshops to review the concepts. A focus on vision and execution—to define a roadmap as called for in the third item of the Statement of Task (Box S.1)—is the USGS's primary need with regard to an SDI. Although it is neither appropriate nor feasible for the committee to recommend changes in the organizational structure of the USGS, there are critical elements of successful SDI implementation that pertain to the entire organization that are appropriate to highlight, and these are described in this report.

LESSONS LEARNED

There is no established, validated process for developing an SDI, and past efforts have produced mixed results. However, past efforts yielded lessons that can provide valuable guidance for the USGS. The committee chose to look at lessons learned in several types of organizations to gain the broadest perspective possible. The committee examined 14 entities in the following five categories: USGS analogues in other countries, multinational organizations, U.S. public and

---

[1] Executive Order 12906, published in 1994 and amended in 2003, initiated the development of a coordinated

BOX S.1
Statement of Task

This study will examine progress made in establishing spatial data infrastructures and the challenges faced by those infrastructures, within the context of the National Spatial Data Infrastructure. The study will examine the role that the USGS can play in continuing to ensure access to high quality geospatial data and support its use in scientific analyses and decision-making through a spatial data infrastructure (SDI) construct.

The committee will undertake three main tasks:

1. Identify existing knowledge and document lessons learned during previous efforts to develop SDIs and their support of scientific endeavors;

2. Develop a vision for optimizing an SDI to organize, integrate, access, and use scientific data; and

3. Create a roadmap to guide the USGS in accomplishing the vision within the scope of the USGS Science Strategy.

private institutions, large discipline-specific organizations, and spatial data at the USGS.

Successful implementation of an SDI depends on an agency's roadmap and strategy, organizational leadership and culture, standardization, technical competence, funding and contracting, workforce competence, and cooperation and partnerships. SDI roadmaps that were well developed and consistently reviewed and updated were the ones that were most successful. Roadmaps that were essentially well-written business plans clearly articulated the merit of an SDI and the community that it would serve. Organizational leadership and culture influence how roadmaps and strategic goals are carried out on a daily basis and probably determine the success or failure of SDI implementation. Establishing standards for the data community is critical for SDI success. Implementation was more seamless and effective in organizations that incorporated the needs of the user community to develop and improve standards, and ones that also accepted the need for data and information products to conform with consensus standards developed by domestic and international standards bodies. Standards should serve the widest range of user types possible.

Technology and tools of the underlying database structure will need to adapt constantly in anticipation of data types beyond the current set, such as multispectral data and an expansion of data layers. Funding and contracting mechanisms affect SDI development. One key factor was adequate funding—not overfunding or underfunding—for carrying out critical activities. Workforce competence is another contributing factor: successful SDI implementation requires the presence throughout the organization of well-trained and respected professionals who understand the technology. Finally, partnerships with state and federal agencies are essential for SDI implementation and for the long-term sustainability of an SDI.

A VISION FOR OPTIMIZING A U.S. GEOLOGICAL SURVEY SPATIAL DATA INFRASTRUCTURE

The USGS has the unique role and responsibility of acquiring, preserving, and archiving geospatial data on a national scale. As envisioned, an optimal SDI at the USGS would need to include data acquisition, data standards, modern data management services, and a set of key application services essential for supporting USGS in addressing scientific questions and questions of societal impact. The SDI would also need to consider the importance of data sharing and data discovery and would need flexible methods of preserving geospatial data across extended time frames and through numerous changes and updates because the ability to document and analyze temporal changes on a national scale is of immense scientific and societal value. Thus, data acquisition, data discovery, data sharing, and data archiving form the core vision for an effective USGS SDI. The committee believes that the effort to implement an SDI can be best framed by the phrase "discover and share for the long term" and hopes that this phrase can become the mantra for spatial data handling throughout the USGS.

First, data discovery is an important task of the USGS, and it will need to ensure the discoverability of prime datasets that were acquired in each division. Once prime datasets have been identified and indexed, there is a need to make them searchable and accessible in a corporate data management system. That will require the development of new institutional policies and series of standards on metadata and data discovery and will require compliance with new policies and standards.

Second, data sharing is a critical task that requires data to be structurally and semantically interoperable so that they can be shared and integrated with other datasets in the USGS, across the nation, and with international partners. As a multidisciplinary organization, the USGS will need to be able to readily combine and synthesize data from various disciplines to contribute to its cross-domain missions. The USGS is also a major international player and will need to collaborate with international partners to address data standardization and to comply with international protocols.

Third, the USGS has the responsibility for maintaining data for the long term. The third fundamental component of a USGS SDI is an effective institutional strategy for data archiving.

Carrying out the vision for an SDI at the USGS requires synergistic partnerships with agencies and organizations that have already contributed substantially to the SDI. A judicious selection of partners will enable the USGS to leverage its limited resources while adopting best practices and furthering interagency standardization.

Finally, the success of implementing a vision for such a large program as the USGS SDI will depend in large part to its leadership. There is a need for empowered leadership and for USGS ownership of a comprehensive and reliable national dataset. Supportive leadership will be critical for developing carefully planned, staffed, budgeted, and executed governance and policies.

A SPATIAL DATA INFRASTRUCTURE ROADMAP

A well-designed SDI program that is based on best practices and focused on the agency's mission will have a high probability of success provided there is adequate planning and execution. To that end, the committee proposes a roadmap for SDI implementation that divides it into three broad phases: (1) preparation and planning; (2) design, development, and testing; and (3) rollout and refinement. The committee proposes some general steps in each phase to assist the USGS in carrying out its task in implementing an effective SDI.

Programmatic preparations and plans are critical in the first phase of SDI implementation. A first critical step is the appointment of key leaders and personnel for envisioning, establishing, and carrying out the vision for an effective SDI. The leadership team will need to determine and define SDI system requirements (based on the six directions in the USGS Science Strategy and with consideration of user needs in other agencies, local governments, academe, and the public), determine the organizational structure of the SDI, identify goals, establish timeframes and milestones, and develop performance metrics. Once the initial planning is complete, it will be important to announce a general outline for implementing the SDI program; communication and outreach will play a decisive role in its success.

The second phase would entail designing, developing, and testing the SDI program. Once standards are determined, the next steps are process identification and development and software development. The former identifies common and documented processes that can enable the SDI to function smoothly across the USGS, and the latter establishes tools for discovery, management, recording, archiving, and sharing of data. With standards, processes, and software in place for an SDI prototype, a training development program would be needed to allow staff to become acquainted with the prototype. The training program would be crucial for providing technical training and support and for building organizational support and buy-in at all levels. After the prototype is introduced, it would be beneficial to unveil a pilot program on a small scale within the USGS to test how well the prototype works and to identify and rectify glitches.

The third and final phase in implementing an SDI will need to include a process for rolling out the SDI throughout the USGS and a process for fine-tuning the program. To ensure that people are properly informed and trained, an institution-wide training program will need to be in place before the SDI is unveiled, and retraining will need to be offered periodically for users to understand the system as it develops. The SDI program will need to implement follow-up metrics to determine how well it is being executed to meet its strategic goals. On the basis of findings gathered with those metrics, there will need to be a process for making adjustments to serve users and fulfill USGS priorities.

A series of organizational and technical considerations are necessary for following the roadmap. It is important that SDI implementation have high priority for USGS leadership. A designated senior SDI staff officer will need support from all levels of leadership—from senior managers to the USGS director—and would need to be given commensurate authority to develop and deploy standards. Implementation of an SDI is a major program for establishing a geospatial base for USGS professional staff and outside users, and it would need to be viewed as such by the Survey. The incentive structure for scientists may need to be modified to reward sharing of spatial data. The USGS will need to consider expanding partnerships of five kinds: strategic partnerships with agencies such as the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration, and the National Science Foundation; data partnerships with agencies such as the Census Bureau; standards partnerships; academic partnerships; and technology partnerships with the commercial sector.

Among the technical considerations, supporting the diverse science workflows will require the Survey to evaluate its current information-technology infrastructure to ensure that it is aligned with the USGS Science Strategy. In light of that assessment, the USGS can implement robust enterprise data management, begin the transition to using the Web as a computing platform, and ultimately implement a comprehensive, long-term knowledge-management infrastructure that supports end-to-end spatial data management, including the collection, integration, maintenance, and delivery of multidisciplinary scientific data.

1

Introduction

Science is increasingly driven by data. The U.S. Geological Survey (USGS) has developed several discipline-specific spatial data infrastructure (SDI)[1] programs over the years and has begun developing a comprehensive SDI through The National Map program. The 2007 USGS Science Strategy outlines the immediate and future science directions at the Survey, and any future SDI will need to be designed to serve these strategies. There are several technical challenges to developing a coherent SDI for any institution, but some of the largest challenges may be organizational. Establishment of a coherent SDI in the USGS to connect spatial data, metadata, tools, and a user community offers a potential for great advances in how science is conducted at USGS and elsewhere.

STUDY SCOPE

The charge for the present study is to describe a vision for a USGS-wide SDI and to create a roadmap for executing that vision (see Box S.1). It is not within the scope of this study to design an SDI, create an exhaustive list of recommended datasets or recommend specific funding for SDI development. Those activities will be the work of the USGS if it chooses to move forward with the plan outlined in this report, and some of this work is already in progress through the USGS Council on Data Integration and other agency initiatives.

It is important to note the distinction between an SDI at the USGS and

---

[1] A spatial data infrastructure (SDI) is a framework of spatial data, metadata, tools, and a user community that are interactively connected so that spatial data can be used in an efficient and flexible way (Nebert, 2004).

the broader and more ambitious goal of a National Spatial Data Infrastructure (NSDI).[2] The NSDI is the work of the Federal Geographic Data Committee (FGDC, 2011), and the USGS is an important contributor to this multi-partner effort. The key design focus of the USGS SDI, and therefore the focus of the present study, is to support science in the agency and to address key disciplines of water, geology, biology, and geography. An important secondary goal is to support science in other federal agencies, state, local, and tribal governments, academe, and the private sector.

The USGS recently dissolved the organizational structure around the four core disciplines of water, geology, biology, and geography and reorganized around the strategic directions outlined in the landmark 2007 Science Strategy (USGS, 2007). The reorganization is significant with respect to SDI development because it establishes an Associate Directorship for Core Science Systems, which includes the National Geospatial Program. Because the Science Strategy outlines the future science directions for the agency, the present committee adopted the six science directions in it—ecosystems, climate, energy and minerals, hazards, environmental health, and water—as the focus of this report for optimizing an SDI. Indeed, the members selected for the committee were identified to address each of those directions.

The Science Strategy clearly defines a need for geospatial data to support each of the science directions. In the opinion of this committee, an SDI is so important for supporting the six directions that it probably deserved its own chapter in the Science Strategy report as an underpinning to those six directions. The committee hopes that this report can serve as the "missing chapter" of that important document.

This report incorporates state-of-the-art SDI concepts for consideration by the USGS. Our review of contemporary SDIs in use today in government, academe, and private industry provided the basis for adapting these concepts to the needs of the USGS. Clearly, keeping an SDI relevant over time will require the USGS to regularly review developments in SDI components.

There are not likely to be any surprises in the committee's definition of an optimal vision for an SDI for the USGS. Much has been written and debated publicly on the subject (e.g., NRC, 1993, 1995, 2001; Onsrund, 2007), and the Survey has held recent workshops to review the concepts. A focus on execution and defining a roadmap as called for in the third item of the Statement of Task (Box S.1) is the USGS's primary need with regard to an SDI. Although it is neither appropriate nor feasible for the committee to recommend changes in the

---

[2] Executive Order 12906, published in 1994 and amended in 2003, initiated the development of a coordinated National Spatial Data Infrastructure and National Geospatial Data Clearinghouse and called for the establishment of spatial data standards, partnerships for data acquisition, and a National Digital Geospatial Data Framework.

organizational structure of the USGS, some critical elements of a successful SDI implementation pertain to the entire organization and are described in this report.

ORGANIZATION OF THE REPORT

This report is organized according to the Statement of Task. Chapter 2 provides background on the Science Strategy and its dependence on geospatial data and identifies the challenges for a successful SDI. Chapter 3 addresses Task 1, lessons learned from SDI implementation in other organizations that the committee felt were pertinent to the USGS situation. Most of the information for the chapter was drawn from briefings to the committee and a survey of key people and organizations involved in SDI development and implementation. Chapter 4 addresses Task 2, outlining the committee's vision for optimizing an SDI for the USGS and discussing key goals for desirable constituent elements of an SDI. Finally, Chapter 5 addresses Task 3 with key recommendations on organizational considerations and a general roadmap for SDI implementation. The committee understands that execution is USGS's key need with respect to its SDI; Chapter 5, although brief, might be the most valuable contribution of this study.

REFERENCES

Buxton, H.T., Griffin, D.W., and Pierce, B.S., eds., 2007, Earth Science and Public Health: Proceedings of the Second National Conference on USGS Health-Related Research, Reston, Virginia, February 27-March 1, 2007: U.S. Geological Survey Scientific Investigations Report 2008-5022, 47p. Accessible at http://pubs.usgs.gov/sir/2008/5022/.

Cramer, B. and DeMulder, M., 2009. The National Map Science Workshop. Lakewood, CO Sept. 2-3, U.S. Geological Survey.

Federal Geographic Data Committee (FGDC). 2011. National Spatial Data Infrastructure. Available online at http://www.fgdc.gov/nsdi/nsdi.html (accessed May 2, 2011).

Ginsberg, H.S., K. E. Hyland, R. Hu, T.J. Daniels, & R.C. Falco. 1998. Tick population trends and forest type. Science 281:349-350.

Ginsberg, H.S., P.A. Buckley, M.G. Balmforth, E. Zhioua, S. Mitra, and F.G. Buckley. 2005. Reservoir competence of native North American birds for the Lyme disease spirochete, Borrelia burgdorferi. Journal of Medical Entomology 42:445-449.

Goodchild, M.D. 2008. Geographic Information Science: The Grand Challenges, in The Handbook of Geographical Information Science, J. P. Wilson and A. S. Fotheringham eds., Blackwell: Oxford. p1-22.

Lubick, N. 2009. Ocean mercury on the increase: Rise may affect neurotoxin levels in fish. Nature doi:10.1038/news.2009.218.

Nebert, D. (ed) 2004. Developing Spatial Data Infrastructures: The SDI Cookbook. Global Spatial Data Infrastructure, 171p.

NRC (National Research Council). 1993. Toward a coordinated spatial data infrastructure for the nation. Washington, D.C.: National Academy Press, 192p.

NRC. 1995. A data foundation for the national spatial data infrastructure. Washington, D.C.: National Academy Press, 45p.

NRC, 2001. National Spatial Data Infrastructure partnership programs: rethinking the focus, Washington, D.C.: National Academy Press, 94p.

NRC, 2003. Weaving a National Map: Review of the U.S. Geological Survey Concept of The National Map, Washington, DC: National Academy Press, 140p.

NRC, 2004. Future Challenges for the U.S. Geological Survey's Minerals Resources Program, Washington, DC: National Academy Press, 154p.

NRC, 2007. A Research Agenda for Geographic Information Science at the United States Geological Survey, Washington, DC: National Academy Press, 156p.

NRC, 2009. Landscapes on the Edge: New Horizons for research in Earth Surface Processes, Washington, DC: National Academy Press, 163p.

Onsrud, H. (ed) 2007. Research and Theory in Advancing Spatial Data Infrastructure Concepts, Redlands, CA: ESRI Press, 306p.

Sunderland, E.M., Krabbenhoft, D.P., Moreau, J.W., Strode, S.A., and Landing, W.M. 2009. Mercury sources, distribution and bioavailability in the North Pacific Ocean--Insights from data and models: Global Biogeochemical Cycles, doi:10.1029/2008GB003425.

Sunderland, E. M. 2007. Mercury exposure from domestic and imported estuarine and marine fish in the United States seafood market: Environmental Health Perspectives, v. 115, no. 2, p. 235-242, doi:10.1289/ehp.9377.

USGS (U.S. Geological Survey). 2007. Facing tomorrow's challenges—U.S. Geological Survey science in the decade 2007–2017/ U.S. Geological Survey Circular 1309, 70p.

United Nations. 2008. United Nations Spatial Data Infrastructure Framework: Interim Report January 2009 to December 2010, United Nations: New York, 36p.

2

Background

DEFINING A SPATIAL DATA INFRASTRUCTURE

A spatial data infrastructure (SDI) is a framework of spatial data, metadata, tools, and a user community that are interactively connected so that spatial data can be used in an efficient and flexible way (Nebert, 2004). Spatial data and metadata are distributed, accessed, and exploited with software tools and services over computer networks. To achieve a well functioning SDI, it is necessary to define standards and to have good coordination between all actors; because an SDI is large (in size, cost, and number of interactors) and is usually government-related. An example of an existing SDI is the U.S. National Spatial Data Infrastructure (NSDI). Infrastructure for Spatial Information in the European Community (INSPIRE) is a European Commission initiative to build a European SDI beyond national boundaries, and the UN Spatial Data Infrastructure plans to create the same type of SDI for over 30 UN funds, programs, specialized agencies, and member countries (United Nations, 2008). INSPIRE sets out a framework and timetable that obliges public-sector organizations to publish key spatial datasets in ways that not only support the discovery of the data but provide access to them through visualization and downloading services. This chapter examines how an SDI can facilitate the collaboration of engineers, scientists, policy-makers, and community groups working across disciplines, on temporal and spatial scales, and in different types of geologic surfaces and subsurfaces.

Science agencies such as the U.S. Geological Survey (USGS) are tasked with providing reliable scientific information to support scientists, researchers, policy-makers, and the general public in making informed decisions on Earth science and natural resource issues facing the nation. As society demands action and responsiveness to growing environmental issues, new process-based solutions grounded in scientific research are needed to generate the knowledge necessary to inform practical decisions. Addressing those challenges requires the synthesis of data and model projections that may routinely span length scales (from micro to global) and time scales (from a few tens of milliseconds to millennia). An SDI for Earth system science makes use of tools for data creation, curation, analysis, and archiving and leverages the Web as a platform for collection, analysis, reporting, and publication.

Understanding large-scale human-stressed environments over long time periods requires observing multiple variables on regional scales. This includes monitoring and measuring how the characteristics and functioning of environmental systems change and determining the cause and extent of such change. The resulting knowledge can lead to improved predictive models that inform decisions about more effective adaptive management policies and practices. Developing these predictive models requires not only establishing new environmental sensor networks, but also integrating data from existing sources and available sensors to provide high-resolution and integrated data. It requires a cyberinfrastructure capable of collecting, managing, and using integrated geospatial datasets. Having a robust data infrastructure would facilitate research investigations aimed at improving understanding of interacting environmental system processes.

The tools needed to conduct science and inform policy-making are changing. Meeting many of the challenges faced by the USGS Science Strategy (USGS, 2007) requires information from the basic sciences, but they also require new scientific approaches that focus on integrating physical, biogeochemical, engineering, and human processes. This cultural and organizational shift becomes more challenging as science becomes more computational and data-intensive. In this new research environment, scientific data are captured by instruments or generated by simulations and then processed by software into models that can be examined by scientists, policy-makers, and the public using the Web. As the Earth becomes increasingly instrumented with interconnected, low-cost, high-bandwidth sensors that are linked through the Web, scientists will be in a better position to sense the environment and predict possible environmental outcomes.

CURRENT STATUS OF THE USGS SDI: THE NATIONAL MAP

The USGS has multiple disciplinary data infrastructures, but recent efforts to develop a science SDI at the USGS have been conducted largely under the umbrella of The National Map (TNM). TNM is a collaborative effort with the USGS and federal, state, tribal, and local partners to create "a database of continuously maintained base geographic information for the United States and its territories that will serve as the Nation's topographic map for the 21st century" (USGS, 2001). TNM provides a common set of base information for use by public and private stakeholders. The 2007 National Research Council report

A Research Agenda for Geographic Information Science at the United States Geological Survey recommended that the focus of the USGS Center of Excellence for Geographic Information Science be on TNM, with key extensions into information access and dissemination, integration of data from multiple sources, and data models and knowledge organization systems (NRC, 2007).

In creating and implementing TNM, the USGS is targeting improvements in data characteristics, such as currency, seamlessness, consistent classification and formatting, variable resolution, completeness, variable positional accuracy, spatial reference systems, standardized content, metadata, and temporal dimensions (USGS, 2001; Cramer and DeMulder, 2009). The data themes in TNM are orthoimagery, elevation, hydrography, geographic names, land cover, transportation, structures, boundaries of government units (such as states and counties), and publicly owned lands (such as national forests and state parks). These data themes were chosen to fulfill a gap and for use for the USGS topographic maps, therefore there are plans to retain data characteristics that are more useful to users of the USGS topographic maps, such as consistent feature identification and classification (NRC, 2007).

The USGS offers several methods for accessing data in TNM. For users seeking to view a data map, a map viewer is available (see http://viewer.nationalmap.gov/viewer) and provides basic Geographic Information System-type (GIS) query and analysis tools. Users can also retrieve national coverage data through interactive and preprocessed methods that can be accessed online, via the map viewer, or physical media. For users that create their own map viewer and that need access to tools and inventoried services, the USGS offers program applications and an online catalog of metadata entries that can be discovered and "harvested" into the Geospatial One-Stop portal. The USGS also has service-level agreements with agencies to provide more advanced Web-based access to national databases.

Perhaps the most fundamental change in TNM approach is the transition from reliance on internal USGS resources for collecting new data to reliance on partners for providing new data (NRC, 2007). These USGS partnerships involve a value-based exchange: In exchange for partners' data, the USGS provides funding, expertise, data, data models, data-collection software tools, information technology, Web and other data-management services, access to contracts, and access to related management and quality-assurance processes.

TNM resides in a large environment that includes electronic mapping products and services provided by the government, academe, and private industry. In the USGS, there are multiple datasets (see Box 2.1) that could eventually be fed into a larger USGS SDI. In the private sector, the emergence of commercial products, such as Google Earth and Microsoft Bing Maps, has captured the interest of the public and professional users. To remain relevant, the 2007 NRC report A Research Agenda for Geographic Information Science at the United States Geological Survey states that the TNM "must be a trusted [emphasis added] geospatial information source for all of these constituencies," and that "the mea-

Box 2.1

Examples of USGS Spatial Datasets

The following are examples of the various types of spatial datasets maintained by the USGS. It should be noted that this is not a comprehensive list of datasets required to support the USGS Science Strategy, although many of the datasets listed are useful to several USGS Science Strategies.

National Land-Cover Dataset — A 21-class land-cover classification scheme that includes urban, agricultural, rangeland, forest, surface-water, wetlands, barren-lands, tundra, and perennial ice and snow classes.

National Orthoimagery Dataset — Data that combine the visual attributes of an aerial photograph with the spatial accuracy and reliability of a planimetric map.

National Elevation Dataset — 10-meter and 30-meter digital elevation models and some higher of resolution derived from light detection and ranging (LIDAR) and Interferometric synthetic aperture radar (IFSAR).

National Transportation Dataset — Data on roads, ports, railroads, and other features associated with the transport of people or commerce.

National Boundaries Dataset — Data on major civil areas, including states, counties, federal, and Native American lands, and incorporated places, such as cities and towns.

National Structures Dataset — Data on selected structures, including locations and characteristics (such as physical form, function, name, location) of man-made structures.

Geographic Names Information System — Federally recognized names of physical and cultural geographic features (excluding roads and highways) in the United States and their locations by state, county, USGS topographic map, and geographic coordinates.

The National Hydrography Dataset — Data on surface waters of the United States, such as lakes, ponds, streams, rivers, canals, and oceans.

Watershed Boundary Dataset — Data on hydrologic units that establish a baseline drainage boundary framework, accounting for all land and surface areas in the United States.

sure of success for TNM will be the extent to which the diverse users embrace and depend on the product" (p. 36, NRC, 2007). However, as previously stated in another previous NRC report Weaving a National Map: A Review of the U.S. Geological Survey Concept of the National Map, it is impossible to be all things to all users at the outset (NRC, 2003). TNM can best serve its users by first focus-

Box 2.2

Additional Attributes for The National Map

The 2007 National Research Council report A Research Agenda for Geographic Information Science at the United States Geological Survey provides a vision for the next generation of The National Map (TNM). In addition to including the existing features, TNM would consist of the following additional attributes:

- an authoritative geographic knowledge base of topographic features based on a geographic feature ontology,
- a comprehensive database of official geographic feature names, and local, regional, and historic variants in TNM gazetteer,
- an enhanced spatial-temporal integration framework for organizing and synchronizing with other USGS data collections,
- a geographic semantic reference system,
- multiple levels of spatial detail,
- feature histories (for spatial locations, attributes, and names),
- user-supported local validation,
- flexible product generation (for example, responding to fact queries, process-model data packages, maps on demand, and traditional topographic maps), and
- smart adjustment of maps or other visual display settings for different devices.

ing on high-impact research areas and by identifying what differentiates TNM in the crowded geospatial-product field.

As indicated by the 2007 National Research Council study, a successful TNM requires a high level of quality, accuracy, national coverage, standardization, and continuous updates (see Box 2.2 for additional attributes of TNM).

As the largest ongoing effort to develop a science SDI, TNM will likely play a key role in underpinning the future USGS SDI and the committee in no way envisions the TNM and the SDI to be mutually exclusive. TNM has the potential to be the go-to resource for the USGS and other federal, state, and local agencies across the nation. This report provides a roadmap to guide the USGS as it decides whether to expand, adapt, or subsume TNM in service of the agency-wide SDI (see Chapter 5).

THE USGS SCIENCE STRATEGY

The 2007 report Facing Tomorrow's Challenges—U.S. Geological Survey Science in the Decade 2007–2017 (USGS, 2007) outlines a strategic scientific approach for the USGS SDI. It is a watershed report by the USGS that guides its direction in its Science Strategy. The USGS Science Strategy breaks away from the conventional approach of organizing strategically by discipline and instead focuses on key future science-based challenges. A private-sector analogy would be a company organizing around markets and customers rather than according to products and technologies. The focus on the users of an organization's products and services can be highly motivational and empowering for its employees. The USGS science-based challenges are categorized into six strategic areas: ecosystems, energy and minerals, climate and land-use change, environmental health, water, and natural hazards. The USGS Science Strategy report identifies the science needs that the USGS SDI will have to address, and the committee concurs with the recommendations provided in that report for addressing the needs.

Throughout the USGS Science Strategy report, references are made to the need for geospatial data and, by extension, the need for an effective SDI. An effective SDI will be essential to the success in each of the six science directions. Many of the characteristics of an SDI are already outlined in the USGS Science Strategy report, such as providing a framework for interactively connecting data users, open standards, and the ability to integrate data from environmental sensor networks and land imaging with spatial modeling capabilities (USGS, 2007). This committee's report attempts to further outline an SDI roadmap that cuts across the six strategic science directions to show the need for and value of spatial data, and to discuss how a well-designed SDI can benefit each strategy.

Ecosystems

A key element of the USGS Science Strategy is "Understanding Ecosystems and Predicting Ecosystem Change" (USGS, 2007). The plan recognizes that ecosystems are multi-scalar in space and time, and that information-analysis tools are needed that can accommodate analysis of ecosystems on multiple scales and incorporate modeling tools. Analysis of ecosystems is thematically considered in various ways, from a geographic point of view as a unit of study and in the context of ecosystem services that are related to the structure and function of ecosystems on diverse space and time scales. A fundamental framework is needed for the Survey to measure, map, understand, monitor, predict, and engage in relevant issues, and a geospatial framework for the analysis of ecosystems will be important to advance the Science Strategy. The geospatial framework can be used to integrate various sets of information for informing ecosystem models and for analysis of interactions between biophysical and societal effects on ecosystems.

The Science Strategy calls for a coordinated effort to produce a scientifically rigorous map of national ecosystems on scales that have meaning for land managers and for understanding interactions between, biophysical, anthropogenic, and biological processes that can be used in the prediction of ecosystem change. The USGS and organizations such as the Environmental Protection Agency (EPA) and the National Ecological Observatory Network (NEON) have already developed versions of ecosystem maps and these can provide a foundation for developing a suite of ecosystem layers within the SDI. A geospatial platform will directly aid in the study of effects of land-use changes on ecosystem dynamics and in the prediction of ecosystem change that may affect ecosystem services.

Energy and Minerals

Science and information on energy and mineral resources underpin private and public decisions that determine resource availability, costs, and conditions for producers and consumers. Scientific research and information collection and dissemination benefit society through four federal roles in energy and mineral resources: (1) unbiased national source of science and information, (2) basic research, (3) advisory functions, and (4) international compilation. The emergence of a global economy affects the demand for all resources, and the use of natural resources is occurring on a scale that may modify terrestrial, marine, and atmospheric environments. The use of and competition for natural resources on a global scale and the natural threats to these resources have the potential to harm the nation's ability to sustain its economy, national security, quality of life, and natural environment.

Under the USGS Science Strategy, the energy and minerals direction is sufficiently broad to deal with resource availability and related land, water, and environmental concerns. Linking the USGS energy and minerals mission area to the USGS ecosystems, water, and climate and land-use change mission areas provides a scientific foundation for resource security, environmental health, economic vitality, and resource management.

The major challenge with respect to energy and minerals is to provide a scientific foundation for resource security, environmental health, economic vitality, and stewardship of the nation's lands. This challenge is complicated by the fact that the appropriate response will need to address evolving U.S. domestic and global priorities and a broad, diverse user base. Responding to national priorities and global trends requires a strategy that builds on existing USGS strengths and partnerships and that advances unprecedented innovations that will be made possible only if they are effective in integrating USGS digital data resources, data viewing, and hypothesis testing. An important challenge is to create user interfaces that do more than view single datasets, instead providing overlaid comparative datasets and map features that support analysis that can lead to new scientific interpretations.

The National Geologic Map Database (NGMDB)[1] effectively locates USGS publications that pertain to energy and minerals, including those on geology (bedrock and surficial), geophysics (magnetic, gravity, and radiometric), and resources (metals, nonmetals, petroleum, and coal). However, the value of this search capability could be enhanced with search capabilities that span a broad sequence of steps involved in scientific investigation and data integration. Inte-

---

[1] http://ngmdb.usgs.gov/ngmdb/ngm_compsearch_map.html.

grating the NGMDB with an SDI offers much potential, but much work remains to develop and implement the new infrastructure before it can take on additional capabilities to serve as a much needed multidisciplinary 3-D GIS database.

Climate and Land-Use Change

With regard to climate variability and change, the USGS brings to bear an impressive array of capabilities for research, monitoring, modeling, and assessment on various relevant topics, including land and ecosystem dynamics, hydroclimatology, coastal processes, and biogeochemical cycles. The USGS climate and land-use change strategy underscores the key role that the USGS can play in helping the nation to "understand and prepare for climate change and its effects" (USGS, 2007). It identifies three critical subjects for focused efforts: monitoring, research, and assessment.

Monitoring activities build on the USGS's long and distinguished record of environmental observations in key arenas, such as land-use and land-cover change, species distribution, hydrology, glaciology, and geochemistry. Proposed strategic initiatives include development of a national phenology (plant and animal cycles) network, a baseline soils database, and a collection of altitudinal transects to study population dynamics and adaptation of life cycles in vulnerable environments across the United States. In the research arena, the USGS brings to bear strong capabilities in analysis and modeling of the terrestrial carbon cycle, climate–land use–ecosystem interactions, hydrologic effects of climate variability and change, and climatic effects on ecosystem health. Monitoring and research together form a key basis for assessment, which encompasses the development of simulation models and predictive tools that can support understanding and decision-making about climate changes and their effects on spatial and temporal scales. The USGS climate strategy recognizes the imperative to integrate multidisciplinary research in developing complex models and assessing feedbacks and linkages between land, water, biological, ecological, and human systems.

The strategic actions related to climate variability and change proposed by the USGS depend on the development and use of geospatial data and models, which would benefit from an enhanced SDI. For example, modernizing existing USGS observing networks and developing new capabilities integrated with those operated by other federal agencies will require the implementation of state-of-the-art geospatial technologies, interoperability standards, and data-management processes and procedures. The ability to identify, evaluate, assess, and predict environmental and natural resource responses to climate change would be enhanced by ready access to harmonized geospatial data (i.e. the combination of common data components) from different disciplines and observational networks that combine historical and baseline data with current observations and by model simulations and forecasts. Analysis and reporting of climate trends and effects—and effective support of decision-making—require the ability to aggregate and disaggregate data on different spatial scales and to provide geospatial visualizations geared to the needs of decision-makers and other users.

In the face of climate variability and change, the USGS Science Strategy appropriately recognizes the important role of the USGS in working with the Bureau of Reclamation, the National Park Service, the Bureau of Land Management, and other agencies of the Department of the Interior to directly manage about one-fifth of the U.S. land resources. Land-related data and models are needed to demonstrate the impact of climate change on land areas. Integration and interoperability of land-related data and models would directly benefit land and resource managers and contribute to the overall adaptation and resilience of the United States to future climate change.

Similarly, close cooperation with other federal agencies, such as the Department of Energy and the U.S. Department of Agriculture, is needed in areas such as carbon capture and storage and soil management. The USGS also needs to coordinate with the international scientific community, as in the case of its research and observational efforts related to biogeochemical cycles, paleoclimate, and cryospheric processes and the USGS's contribution to the Intergovernmental Panel on Climate Change process. All those efforts would benefit from the ability to build on and connect the USGS SDI to the broader NSDI and the emerging global SDI.

Environmental Health

The USGS Science Strategy includes the goal of understanding the role of environment and wildlife in human health. Most USGS efforts in this arena have been on health problems caused by environmental contamination and emerging infectious diseases, and research has focused on public health threats that result from the relationship between people and the physical, chemical, and natural environments. The USGS specializes in research in vector-borne and zoonotic diseases, water contamination, airborne contaminants, bioaccumulative contaminants in the food chain, and environmental threats to public health. It does not, however, focus on health effects that result from the built environment, although the aforementioned research certainly involves these parts of the environment in addition to the natural environment.

The USGS efforts in environmental health can best be described as investigator-driven science, so research is dispersed throughout the agency. For the most part, this research is excellent and is published in top scientific journals, such as work on Lyme disease (Ginsberg et al., 1998, 2005) and methylmercury in ocean fish (Sunderland et al., 2007, 2009). One of the benefits of USGS's research on human health is project longevity which often allows longitudinal health studies, compared with university-based health research studies that are typically 2-5 years for the project duration.

Even though high-quality research on the role of environment and wildlife in human health is being conducted at the USGS, the decentralized structure creates substantial challenges. Funding for individual studies does not come through sources directly targeted to human health, so consistent funding has been a challenge for USGS researchers. A well-functioning SDI could assist USGS health researchers in this struggle by centralizing data access and pairing data with other relevant datasets. The lack of a centrally organized health research agenda makes it difficult to establish systems for efficiently supporting the sharing of spatial data within the USGS, and it is even more difficult to share with other federal agencies or researchers outside the federal government. Investigators do not have much incentive to make their datasets available, let alone discoverable (see discussion later in this chapter on "Incentives for Scientists").

Organizing USGS research efforts around the Science Strategy on the role of environment and wildlife in human health will help to alleviate some of those challenges. In the last few years, the USGS has organized meetings to bring the research community together on human health issues as they are related to the environment and wildlife. The USGS convened meetings titled "Natural Science and Public Health: Prescription for a Better Environment" (USGS, 2003) and "The Second National Conference on USGS Health-Related Research" (Buxton et al., 2007); each had more than 60 paper presentations on the aforementioned USGS health-research topics that underscored the importance of spatial data. The purposes of the conferences were to enhance communication among federal agencies; to identify common science interests for leveraging science research, results, and resources; and to establish joint science investigations and cooperative partnerships to increase the use of USGS information by the public health community. The goals of expanding access to existing data and strengthening partnerships and collaborations are consistent with the principles underlying the USGS Science Strategy.

The USGS Science Strategy also includes a partial listing of USGS environmental health–related databases that could be incorporated into a science SDI. The USGS Science Strategy report specifically calls for developing an "online data atlas of potential environmental health threats" to improve the ability of the United States to respond quickly to health threats and for developing a national "environmental health information system" that integrates basic USGS scientific data with GIS decision-support tools (USGS, 2007). Administrative structures, incentives, and funding are some challenges that make it difficult to make the aforementioned goals a reality.

Water

A goal of the USGS Science Strategy is to develop a National Water Census that includes quantifying, forecasting, and securing freshwater for America's future. Equal emphasis is given to water quantity and water quality as related to water availability and to meeting both human and ecological needs in the context of land-use change, climate change, and changing demands for water resources. Previous USGS studies and monitoring programs focused on human needs; the added dimension of quantifying ecological needs is new. The USGS is uniquely qualified to develop a National Water Census because of its unique capabilities with its diverse scientific staff and technicians (diverse geographically and in expertise) and its state-of-the-art groundwater models. The goal of the National Water Census is to inform the public on: (1) the status of and trends in freshwater resources, (2) water use for human, environmental, and wildlife needs, (3) relationship of freshwater availability to storage and transport in natural and engineered systems, water use, and related transfer, (4) identification of uncommon water sources, and (5) forecasts of effects of changes in land use and land cover, natural and engineered infrastructure, water use, and climate on water availability, including water quality, and aquatic ecosystem health (USGS, 2007).

The National Water Census will need to pull together several disparate data sources at multiple scales. The StreamStats application will provide water availability information on precipitation, runoff, baseflow, and trends. The USGS National Water Quality Assessment (NAWQA) Program will be used to assess the role of water quality in water availability. The detailed field studies throughout the nation conducted as part of the NAWQA Program will provide spatial data on natural and anthropogenic contaminants in aquifers. Relationships to geology and land use will be used to assess sources and mobilization mechanisms of contaminants. The intensification of the hydrologic cycle will likely result in longer-term droughts and more intense floods (USGCRP, 2009), which could further burden already stressed water systems. Quantifying the role of groundwater in water availability will require spatial and temporal information on recharge, groundwater yields, changes in storage, saltwater intrusion, trends in groundwater indexes, groundwater demand, and groundwater–surface water interactions. The National Water Census will rely primarily on the Regional Groundwater Availability Studies for much of that information. Data on water withdrawals will be obtained from the National Water Use Information program. Satellite data on evapotranspiration could also be used to estimate water demand in irrigated regions. Information on the geohydrologic framework of aquifers will be obtained from the National Cooperative Geologic Mapping Program.

Because water is inherently multidisciplinary, a well-designed SDI is essential for the success of the water census, connecting climate forcing with detailed measurements and monitoring of surface and subsurface hydrology. Linking satellite data, including evapotranspiration maps and water storage changes from the Gravity Recovery and Climate Experiment with ground-based measurements of water storage and demand, will be important for ground referencing satellite-based estimates. Because the water census will rely heavily on datasets produced by many state programs and will also rely upon water infrastructure data from the U.S. Environmental Protection Agency, an open framework SDI, one that allows individual components to be easily added or replaced, will be helpful for data-sharing. An SDI that allows discovery and provides detailed metadata information will be valuable because water is of general interest to the public in addition to providing essential research data for various studies. Because of the potentially large number of groups providing data to the census, common standards and tools will be essential for the success of the National Water Census.

Natural Hazards

The USGS is involved in every part of the disaster cycle from scientific study to disaster response. It is charged with identifying, understanding, monitoring, and mapping hazards. It is an active participant in preparation for, response to, and recovery from disasters that occur when humans and cultural artifacts are exposed to a hazardous event. USGS scientists study the geophysical phenomena that lead to volcanic eruptions, earthquakes, landslides, tsunamis, floods, coastal inundation, droughts, and other natural disasters. The USGS is also responsible for the International Charter on Space and Major Disasters for the United States and in the event of a disaster can activate the International Charter to direct assets to the disaster.

The USGS operates and participates in various monitoring networks, such as seismic and stream monitoring networks, which are fundamental to hazards-related work. Data produced by those networks are necessary for both early warnings and pre- and post-disaster assessments, and they are often the primary data used in fundamental and applied research. The data are used operationally for mapping hazard zones, which are used by (1) communities for planning, (2) responders for identifying emergency routes and the potential severity of an event, (3) industry for siting facilities and infrastructure, (4) insurance companies for determining risk, and (5) private individuals for housing and other decisions. Information generated from the monitoring networks are used directly in the science and disaster management communities, where they are combined with additional data, information, and models to advance knowledge and provide valuable, usually critical information to those who need it.

The USGS provides geospatial information in response to domestic disasters and provides assistance whenever possible in response to international disasters (see Box 2.3). For example, in anticipation of Hurricane Katrina, the Survey was on site with hard-copy maps, digital data, analytical tools, and computer hardware 3 days before the hurricane made landfall. In its continuing effort to ensure maximum access to needed data and information, USGS scientists and technicians work with the National Geospatial-Intelligence Agency (NGA) to ensure that software is prepared to facilitate the integration of NGA data with data from USGS and other civilian agencies. In the international arena, USGS is the lead agency for the International Charter to provide remote sensing data to nations that request them during times of disaster. It was a major asset in response to the May 12, 2008, earthquake in Sichuan, China, when USGS worked on a team

BOX 2.3

Prompt Assessment of Global Earthquakes for Response

Prompt Assessment of Global Earthquakes for Response (PAGER) is an automated system that quickly estimates the number of cities and people worldwide exposed to severe shaking after significant earthquakes. PAGER began operation on March 28, 2005.

While PAGER was still in the design and testing phase, a major aftershock of the 2004 Sumatra–Andaman earthquake occurred about 300 kilometers to the southeast of the original quake. Massive devastation took place on the island of Nias, major damage on the island of Simuelue, and damage on other islands. A 4-meter tsunami hit parts of the coast of Sumatra. Information from the then experimental PAGER program was used to identify where the most affected populations would be. The earthquake occurred at 11:09 p.m., and rescue workers worked through the night to prepare flight plans for rescue helicopters. The PAGER data were transmitted to the flight planners so that pilots could reach the most at-risk populations. The rapid and effective transmission of the PAGER data allowed the launch of an efficient rescue mission, a much more successful one than would otherwise have been possible. PAGER has since become an important operational component in the global disaster-response toolbox.

Source: Kelmelis et al., 2006.

Table 2-1 Major USGS Geophysical Monitoring Networks Central to Data Collection, Hazard Monitoring, and Event Warning

Monitoring Network	Hazard Type
Advanced National Seismic System	Earthquake
National Volcano Early Warning System	Volcano
Stream gage (stable-core network)	Flood and drought
Marsh Surface Elevation Table Network	Sea-level rise
Light detection and ranging (LIDAR)	Land-cover change and coastal inundation

to broker access to government and commercial satellite data to assist with the response and recovery effort and worked with Chinese scientists to develop a new understanding of the fault zones and other geologic conditions in the earthquake area. Those types of situations require information that combines geographic data with a variety of geophysical data.

As one of its major future investments, the USGS Science Strategy describes the modernizing of major geophysical monitoring networks (see Table 2-1). It also discusses the importance of USGS's taking advantage of new and emerging technologies for network communication, characterizing and assessing hazards, providing forecasts based on understanding of physical processes, and developing partnerships that will advance the state of the art of decision support systems and intelligent access to data. An important responsibility for the emerging SDI will be to support those advances while ensuring the continuation of traditional natural-hazard and disaster-management functions until the advances are online.

THE CHALLENGE OF AN SDI FOR SCIENCE AND DECISION-MAKING

Creating a functioning SDI presents numerous technical challenges. As previously stated, SDIs are inherently complex because they are large and have multiple administrative and technical components—such as software and hardware, standards definition and adoption, and institutional agreements—on a wide variety of scales. The task of integrating and visualizing data generated on different spatial scales remains one of the grand technical challenges in spatial science (Goodchild, 2008). Adding the dimension of time to spatial datasets to enable spatial–temporal correlations remains a painstaking process because spatial data collected decades apart use different standards and equipment. Organizational challenges include training and developing partnerships across different disciplines in and outside the Survey. This section discusses the technical and organizational challenges in the context of a USGS SDI.

Scale

The challenges faced by the nation and identified in the USGS Science Strategy call for support of geospatial information on multiple spatial scales in a timely manner to address widely varied issues: national hazards and risks, human health, climate variability, ecosystem dynamics, water quality and quantity, and energy and mineral resources. Advances in geospatial computing and information technology are beginning to provide tools that can integrate information across multiple scales of space and time (such as www.geo.data.gov). The framework to work across multiple scales must provide consistency from one level to another and provide the ability to "telescope" down or up in scale depending on the need of the analysis. That is far from trivial and requires planning and coordination to work across two or more spatial datasets.

The ability to provide geospatial support across multiple spatial, elevational, and temporal scales (4-D analysis) will give researchers, decision-makers, and managers the ability to assess hazards and risks as they occur, to evaluate ecosystem and hydrologic changes across scales useful to wildlife managers and resource managers dealing with regional drought conditions, to track health risks due to mobile animal populations or more stationary water quality changes that result from mining effects, and to respond to storm surges. A robust and integrated SDI will provide a basis for those kinds of analyses to be conducted in a timely manner that is appropriate for researchers, decision-makers, and managers.

Spatial–Temporal Correlation

A major challenge for the USGS will be to garner spatial–temporal correlations for understanding and predicting ecosystem changes over space and time and to use such interdisciplinary information to inform decision-making. As documented in the USGS Science Strategy (USGS, 2007), existing databases and related maps in USGS programs are not functionally integrated, and this severely limits the use of science and spatial–temporal correlations for making informed decisions. Until scientific geospatial analysis protocols are established in the USGS and requisite application software tools are widely implemented, decision-making based on spatial–temporal correlations will remain tentative or will at best be successful on a limited scale.

Gaining scientific understanding, mapping, monitoring, modeling, and advising the Department of the Interior by using a systems approach to temporal and spatial change are central issues. A geospatial framework requires the integration of various sets of information for informing ecosystem models and for analyzing interactions between biophysical, anthropogenic, and biological processes to projecting ecosystem change. A scientific foundation is critical for effectively managing the use of energy and mineral resources, the environment, and lands. The geospatial platform will be an essential tool for gauging the effects of land-use changes on ecosystem dynamics, for monitoring the effects of climate change, and for predicting ecosystem change over time. Nation-wide datasets with consistent definitions are required to meet the challenges in spatial–temporal correlation. Topographic map coverage in raster, elevation, and land cover constitutes a starting point. Key datasets to add include multi-temporal (time-series) data for hydrography and data on invasive species for ecosystems analysis. Until geospatial science is fully integrated into all the USGS strategic directions, it will be difficult to address broader questions related to sustainability and how Earth's surface will evolve in the Anthropocene—a new geologic period in which human influence dominates the Earth system (NRC, 2009).

High-resolution systematic data need to be made available in a large, seamless archive to support activities that detect changes over space and time. Emergency responders rely on a variety of spatial–temporal correlations for information, including hazards information based on detailed time-series analysis (such as coastline changes with severe storms). Also needed is rapid access to real-time and archived geospatially referenced data, including access to satellite imagery over time, airborne and ground-based light detection and ranging (LIDAR) land-surface images, airborne imagery photos, and forward-looking infrared (FLIR) thermal imaging camera data.

Historical and Baseline Data

Another function of a well-designed SDI will be to provide a coherent framework for collecting, organizing, accessing, and analyzing vast amounts of historical USGS data that are vital for understanding baseline conditions and for modeling and projecting future environmental behavior. Long-time series data are essential inputs into efforts to improve prediction of earthquakes, volcanoes, floods, and other hazards; to understand and predict climate variability and change; and to manage land, water, and ecological resources. Unique data collections, such as the Landsat and Corona data archives at the USGS EROS Data Center, provide invaluable information about land-cover and land-use changes over the last 50 years. Historical data are often essential for calibrating new instruments and monitoring activities and for developing reference or baseline datasets used to assess the effects of future trends or policy changes.

Historical data have often been stored in separate, disconnected databases and datasets and have often been collected and stored with different methods, standards, and tools. That has led to spatial and temporal discontinuities in data, unnecessary barriers to data integration, reduced data quality, and duplication of effort. Another important challenge is the large amount of historical data and documentation that has not yet been digitized and may be at risk of deterioration because it remains on physical media (such as paper, film, and microfiche).

An SDI would serve as a common foundation for historical and baseline data that would enable more flexible integration of different data types and facilitate discovery, access, and use of historical data in conjunction with new observational and monitoring efforts. It could also facilitate priority-setting among efforts to rescue datasets at risk by providing a synoptic view of the most important data gaps and needs across the USGS and partner agencies.

Multidisciplinary and Collaborative Activities

Environmental challenges today arise from complex interactions between multiple disciplines including human activities, ecosystems, and biophysical features, such as climate, land-use change, hydrologic dynamics, and disturbance regimes. Integration of information and analysis across those activities is necessary to ascertain emergent properties of environmental changes related to challenges being met in the USGS Science Strategy. For instance, evaluation of avian influenza needs to account not only for instances of outbreaks but for migration of bird populations, wetland locations, and settlement patterns if it is to accurately predict the emergence and spread of avian influenza. Similarly, issues related to hazards and risks rely not only on environmental factors but on patterns of transportation, settlements, waterways, topography, weather, and other factors that have particular spatial patterns that need to be integrated in assessment of risks posed by environmental conditions.

A well-known important challenge in moving an SDI forward is the cross-disciplinary collaborative ethos necessary in the USGS to develop and maintain an SDI that was previously lacking. In evidence and testimony provided to the committee, stovepipes and silos were terms used to describe USGS structure and operation although the recent USGS reorganization is a substantial step toward removing many of the silos. The USGS has an amalgam of scientific disciplines and areas of expertise. The data, processes, and cultures in the USGS are also diverse, and it is governed by agency-specific and domain-specific goals. Part of that discreteness could be explained by the presence of several factors: legitimate discipline differences, reluctance of scientists to engage outside their specialties, the competitive nature of science funding, and a natural tendency toward precision in one's own field. Implementing an SDI will require effective collaboration among multiple disciplines and will require agreement on issues such as data standards, sharing, and maintenance. In addition, a well-implemented TNM would replace the focus from one-off base data acquisition with a comprehensive, nationally-consistent dataset, just as desired in an SDI. The USGS will also need to ensure that key datasets are discoverable and accessible over the long term.

Externally, the USGS has an extensive array of stakeholders to engage and collaborate with, including federal, state, tribal, local, international, academic, and public entities (see Box 2.4). It will be difficult for the USGS to achieve its strategic goals alone; it will need to partner with others to achieve them. Surveys of external users show that they use the spatial data that the USGS makes available (e.g. orthophotography, elevation, transportation) and would like data that is currently more difficult to find and access, such as the location of wells and springs (Sugarbaker et al, 2009).

Program vs. Project

System studies are often anchored by place-based studies in which greater detail can be extracted on a wide array of processes and components of a system, such as tower-flux studies, acid-deposition studies, and watershed studies. Those placed-based studies are often associated in a network of sites so that various environmental factors can be evaluated to understand how controlling variables affect, for instance, water quality, plant productivity, or carbon emissions. These typically one-off studies also can provide process understanding of system interactions when linked to other studies in a region or linked to similar studies in other locations or conducted at different times. However, the information collected in and shared between USGS programs and individual projects can vary considerably. For instance, satellites to assess land cover in the United States provide consistent, almost seamless, coverage of land cover across the nation. But a number of specialized land-cover databases have been developed by various groups or projects for special purposes to elucidate processes that feature specific characteristic or spatial patterns that are not captured by the national efforts.

Box 2.4

External Users of USGS Spatial Data

The USGS has a wide variety of stakeholders that use spatial data. The examples below illustrate the needs of representative user groups.

Climate Modelers — The effort to downscale climate projections to the regional level requires models of regional land change that, in turn, require authoritative data on land use and cover, much of this derived from satellite imagery. These data are often combined with socio-economic data to project the impoacvt of policy decisions on land cover changes.

Coastal County Managers — Coastal managers are currently using many of the core data layers provided by the USGS, such as elevation and land use and cover. One of the greatest challenges to coastal managers, particularly in small and medium sized counties, is the lack of knowledge about what spatial data are available to them.

Natural Hazards — The combination of orthoimagery, transportation network, and elevation data provide a powerful tool for mitigating and responding to natural hazard events. However, these data are only useful together if they are truly integrated by sharing a common georeference framework and are updated regularly.

Sources: NACo, 2008; Sugarbaker et al, 2009.

For example, boundaries of seasonally inundated areas or wetland areas may be established in detail for a specific project, whereas national datasets on the same phenomena are more generalized.

To resolve the program–project discontinuity, criteria for developing harmonization of program-level data in a national geospatial database would be useful. The scheme would need to provide a basis for regional cross-comparison analysis and for future data-product improvements. The information for the harmonization effort might also include enough information to reconcile data from the program-level analysis and the national database. This effort in cross-linkage between various geospatial products will greatly enhance the transparency of derived products relative to the base heritage geospatial information sets.

For program- and project-level datasets, the spatial data infrastructure will need to allow for easy tagging of information in a common data format. Tools to facilitate the tagging of geospatial and temporal information from program and project analyses will enhance the integration and synthesis of studies across programs and agencies. The ability to conduct synthetic analysis across project and program data will greatly improve USGS's ability to further understand how critical ecosystem and natural resources dynamics are affected by human and environmental changes.

Incentives for Scientists

Staff incentives for sharing data are an important component of successful implementation and management of an SDI. The USGS workforce consists primarily of research scientists—there are nearly 2,000 full-time research-grade scientists. Like their academic counterparts, USGS scientists are evaluated through an annual research graded evaluation (RGE) process and rewarded on the basis of their scholarly productivity. Productivity is typically measured by scientific outputs such as publication in peer-reviewed journal articles and original research. Salary increases for and career advancement of scientists typically require prolific publication and high-impact research. Because research scientists are rewarded through publication and research, there are few compelling reasons for scientists to invest time and effort in sharing data unless these include scientific collaborations that directly involve co-authorship of scholarly output. As Nobel laureate and current US Secretary of Energy Steven Chu has remarked, "We seek solutions. We don't seek—dare I say this—just scientific papers any more" (Del Vecchio, 2007).

The RGE process does not consider data-sharing as an output category, and USGS budgets do not sufficiently allocate for resources that are needed to share data efficiently. As a result, scientists are often reluctant to make their data available until they have been interpreted and published. A potential solution is to modify the RGE process in such a way that USGS scientists are evaluated on the data they shared. Another possible solution for this "publish or perish" problem could be a requirement that USGS-funded projects have part of their budget devoted to data-sharing and that proposals include plans for data-sharing. One example is the funding model in the National Aeronautics and Space Administration, which requires free and open access to data in an SDI; such an arrangement would probably result in greater utility of USGS data and improved relevance of the USGS itself. The USGS will also need to devise career incentives for scientists to provide open access to data and to partner with non-USGS researchers and organizations. That would require the support of skilled employees whose responsibility would be to facilitate data-sharing among USGS divisions. An SDI could be the integration tool for data-sharing inasmuch as most Survey data are inherently spatial. The diversity of data and of technological and scientific approaches makes it necessary for scientists to partner and to leverage expertise and knowledge.

REFERENCES

Buxton, H.T., D.W. Griffin, B.S. and Pierce (Eds.). 2007. Earth Science and Public Health: Proceedings of the Second National Conference on USGS Health-Related Research, Reston, Virginia, February 27-March 1, 2007: U.S. Geological Survey Scientific Investigations Report 2008-5022, 47p. Available online at http://pubs.usgs.gov/sir/2008/5022/ (last accessed May 17, 2011).

Cramer, B., and M. DeMulder. 2009. The National Map Science Workshop. Lakewood, CO Sept. 2-3, U.S. Geological Survey.

Del Vecchio, R. 2007. "UC Berkeley: Panel looks at control of emissions," San Francisco Chronicle, March 22, 2007.

Ginsberg, H.S., K.E. Hyland, R. Hu, T.J. Daniels, and R.C. Falco. 1998. Tick population trends and forest type. Science 281:349-350.

Ginsberg, H.S., P.A. Buckley, M.G. Balmforth, E. Zhioua, S. Mitra, and F.G. Buckley. 2005. Reservoir competence of native North American birds for the Lyme disease spirochete, Borrelia burgdorferi. Journal of Medical Entomology 42:445-449.

Goodchild, M.D. 2008. Geographic Information Science: The Grand Challenges, in The Handbook of Geographical Information Science, J. P. Wilson and A. S. Fotheringham eds., Blackwell: Oxford. p1-22.

Kelmelis, J.A., L. Schwartz, C. Christian, M. Crawford, and D. King. 2006. Geospatial information response to the Sumatra-Andaman earthquake and tsunami of December 26, 2004. Photogrammetric Engineering and Remote Sensing 72(8):862-876.

Nebert, D. (ed) 2004. Developing Spatial Data Infrastructures: The SDI Cookbook. Global Spatial Data Infrastructure, 171p.

NACo (National Association of Counties). 2008. NACo/NOAA-CSC Coastal County Geospatial Needs Questionnaire Analysis Report. Washington, DC, 18p.

NRC (National Research Council). 2003. Weaving a National Map: Review of the U.S. Geological Survey Concept of The National Map. Washington, DC: The National Academies Press, 140p.

NRC. 2004. Future Challenges for the U.S. Geological Survey's Minerals Resources Program. Washington, DC: The National Academies Press, 154p.

NRC. 2007. A Research Agenda for Geographic Information Science at the United States Geological Survey. Washington, DC: The National Academies Press, 156p.

NRC. 2009. Landscapes on the Edge: New Horizons for research in Earth Surface Processes. Washington, DC: The National Academies Press, 163p.

Sugarbaker, L., K.E. Coray, and B. Poore 2009. The National Map Customer Requirements: Findings from Interviews and Surveys. U.S. Geological Survey Open File Report 2009-1222, 34p.

Sunderland, E.M., D.P. Krabbenhoft, J.W. Moreau, S.A. Strode, and W.M. Landing. 2009. Mercury sources, distribution and bioavailability in the North Pacific Ocean--Insights from data and models: Global Biogeochemical Cycles 23(2):1-14.

Sunderland, E. M., 2007. Mercury exposure from domestic and imported estuarine and marine fish in the United States seafood market: Environmental Health Perspectives 115(2):235-242.

United Nations. 2008. United Nations Spatial Data Infrastructure Framework: Interim Report January 2009 to December 2010, United Nations: New York, 36p.

U.S. Global Change Research Program (USGCRP). 2009. Global Climate Change Impacts in the United States. Available online at http://downloads.globalchange.gov/usimpacts/pdfs/water.pdf (accessed March 12, 2012).

U.S. Geological Survey (USGS). 2001. The National Map: Topographic Mapping for the 21st Century. Available online at http://nationalmap.gov/report/national_map_report_final.pdf (last accessed May 11, 2011).

USGS. 2003. Natural Science and Public Health: Prescription for a Better Environment. Available online at http://pubs.usgs.gov/of/2003/of03-097/of03-097.pdf (accessed May 17, 2011).

USGS. 2007. Facing tomorrow's challenges—U.S. Geological Survey science in the decade 2007–2017: U.S. Geological Survey Circular 1309, 70p. Available online at http://pubs.usgs.gov/circ/2007/1309/pdf/C1309.pdf (accessed May 16, 2011).

3

Key Challenges and Lessons Learned

ORGANIZATIONS AND TYPES OF SDIs EXAMINED

At all levels of domestic and foreign government, within academe, and in the private sector, organizations have struggled with the development of spatial data infrastructures (SDIs). There is no established, validated process for developing an SDI, and past efforts have produced mixed results. However, there are lessons to glean from past effortss that could be applied in developing and refining the SDI for the U.S. Geological Survey (USGS). In reviewing the past efforts, the committee noted several relevant experiences that can provide valuable guidance for the USGS. The missions of various organizations may differ from that of the USGS, and those organizations may have unique requirements, but there are common lessons from each that can serve as a roadmap for successful SDI development for the USGS. The examples selected for this chapter have particular relevance to some aspect of the USGS requirements, and some have been successful.

The committee chose to look at lessons learned from efforts of several types of organizations to gain the broadest perspective possible. Fourteen organizations were examined in the following five categories: USGS analogues in other countries, multinational organizations, U.S. public and private institutions, large discipline-specific organizations, and spatial data at the USGS (see Box 3.1). For the USGS, planning a unique SDI that serves a variety of scientific domains means that no single SDI example can be translated directly to the USGS. However, the committee's examination revealed several themes that recurred in differ-

Box 3.1

Examining Spatial Data Infrastructures

USGS Analogues in Other Countries — The British Geological Survey has made cultural adjustments and committed an impressive budget commitment to managing spatial data. Geoscience Australia is beginning to recognize the high value of scientific collaboration through data-sharing enabled by an SDI. These cases provide lessons at the organizational level and are the closest organizational analogues to the USGS.

Multinational Organizations — The Infrastructure for Spatial Information in the European Community and the Global Earth Observation System of Systems are ambitious multinational efforts at standardization and collaboration with direct relevance to USGS's role in the National Spatial Data Infrastructure.

U.S. Public and Private Institutions — In the United States, the National Geospatial-Intelligence Agency, the National Aeronautics and Space Administration, the Texas National Resources Information System, and NAVTEQ each take different approaches to integrating datasets from multiple sources. Standardization plays a particularly large role and varies among these institutions, and it provides a valuable comparison for the USGS.

Large Discipline-specific Organizations — The National Ecological Observatory Network and the Consortium of Universities for the Advancement of Hydrologic Science, Inc. provide lessons from large-scale data integration and access efforts.

Spatial Data at the USGS — The USGS Topographic Mapping Program is the seminal agency-wide commitment to an ambitious spatial data program that established the core value of spatial data at the USGS. Research at the Center of Excellence for Geospatial Information Science is providing much of the technology needed to implement an agency-wide SDI through its work on The National Map. The National Biological Information Infrastructure and National Hydrography Dataset are successful integrations of multiple, dissimilar datasets with direct relevance to spatial dataset integration for the USGS SDI. These programs provide examples of how SDI development has occurred at the USGS.

ent organizations. Other lessons are drawn from single incidents that are directly relevant to some aspect of a USGS SDI.

Geoscience Australia

Geoscience Australia (GA) is the national geoscience research and information agency for Australia. GA was formed in 2001 as a result of a merger of the Australian Geological Survey Organization with the government bodies for topographic-mapping and remote-sensing functions. Like the USGS, GA operates in a federal system, in partnership with the states and territories of Australia. Spatial data are a prime responsibility, and activities focus on providing key information for Australia with an emphasis on onshore and offshore environmental hazards and natural resources. GA is also responsible for coordinating the implementation of the Australian government's policy on spatial data access. The following information is synthesized from a questionnaire provided by the GA information-management team supplemented by information drawn from a recent report by the Australian National Audit Office (2010).

Key Challenges

In designing and implementing SDIs in GA and in other state and science organizations in Australia, there have been a number of common and important challenges that range from organizational and cultural concerns to policy and financial issues. SDI development had been difficult for highly competitive, inwardly focused organizations and ones that focused on the final deliverable. Self-taught experts dominated discussions about SDI development, rather than the necessary highly trained technical informatics experts who fully understood an SDI and who were committed to its successful implementation. Science funding has been increasingly competitive in the last 3 decades. Although collaboration on issues such as data-sharing and agreement of standards is critical for the development of an SDI, competition for shrinking funding has made it difficult for scientists to collaborate. In other cases, scientists did not share data, because they believed the data were unfit for release and had no timeframe for completing the data-improvement processes. Agreed policies were imperative at the organizational level in that properly implemented and articulated policies can be an enabler for SDIs. Spending large amounts of funds in a short period became unsustainable for the financial health of those efforts.

Lessons learned

GA personnel reported that the most important factors for successfully building SDIs were the ones that focused on collaboration to develop and improve data standards (in accordance with international standards) and the ones that focused on making data accessible to the broader community. In developing data standards, once the standards are defined and agreed on, they must be applied consistently.

Another factor that led to the Australian government's successful SDI design and implementation was a well-developed roadmap that was based on sound scientific and business practices; that encompassed technological, computational, and engineering viewpoints; and that was consistently reviewed and updated as required. A well-written business case articulated the value proposition of an SDI, and the efforts were championed by a leader who was knowledgeable and respected in the community and could clearly articulate the value of an SDI in the organization. College educated and respected professionals who understood the technology were needed. Incremental SDI implementation was also important; it was more effective to establish progressive goals than a final deadline.

The culture of the organization played a role in successful SDI implementation. The introduction of an SDI was initially disruptive. Realistic expectations

Box 3.2

Excerpts of Key Findings about Geoscience Australia from the Australian National Audit Office

"Feedback from government agencies and key industry stakeholders confirmed that Geoscience Australia's work is valued and often essential to their outcomes. Notwithstanding this positive feedback, Geoscience Australia's website, its key interface with customers, is complex to use and more data and information could be made publicly available. In addition, the management of many product and service projects lacked project plans, risk assessments and key performance indicators."

"In addition, there is no inventory that documents the purpose, extent and nature of Geoscience Australia's data and information holdings and physical collections. It is therefore not well positioned to appropriately maintain and store its data holdings or make informed decisions about the accessibility of that data."

SOURCE: Australian National Audit Office, 2010

were needed as inevitable improvements were made after the introduction of the SDI. A policy of under-promising but over-delivering is useful in such situations. Support by the executive level can foster commitment and enthusiasm in senior, middle, and junior members of staff. Adequate funding was also important.

A recent Australian National Audit Office report (2010) provides additional lessons for SDI development (see Box 3.2). The findings are pertinent for public-sector organizations responsible for custodianship and delivery of public-sector data and information, such as the USGS. The key points of the audit report echo findings stated by GA employees. GA's value is in its spatial data but has yet to be fully appreciated. GA has not yet developed a clear spatial data plan, cataloged and shared data, improved communication with partners, or implemented standards. In many ways, the Survey is further along than GA in SDI implementation, but many of the missed opportunities outlined in the GA audit report can also apply to the USGS.

British Geological Survey

The British Geological Survey (BGS) is the national geological survey of the United Kingdom (UK). Unlike the USGS, which covers many disciplines, BGS examines only geoscience. However, both the BGS and USGS have national responsibility for the acquisition, analysis, management, and delivery of geoscience data in their countries. The BGS budget is roughly £48 million, and approximately half of it is funded by their national government (British Geological Survey, 2011).

The UK treasury agency, HM Treasury, conducted an audit of BGS in 1992 and found its data fragmented, questioned its accuracy, and concluded that the existing information systems could not support BGS's mission for providing geologic data. It also expressed reservations about the value of the unique data holdings as a major competitive strength of the BGS (Griew, 1990). In 1996, an additional external review of BGS found little improvement in data management. In 2000, after 2 years of pilot studies and development of a new strategy that recognized BGS as an information organization, BGS was restructured from a hierarchically managed organization to a matrix-managed organization. An Information Directorate was created, assigned one-third of the BGS budget, and given corporate priority to work on metadataa, data standards, data-product development, and delivery.

The investment of one-third of the BGS budget in data and the priority given to data activity have resulted in clear benefits for its partners and data users. The result has been up-to-date, quality-assured, and interoperable national versions of all the primary geoscience datasets and internal and external access to an extensive variety of core and value-added Internet information services.

Key Challenges

The organizational and cultural challenges that BGS faced in the 1990s in improving the poor condition of its data and information policy and practice are probably similar to those faced by the USGS. A systematic approach was lacking for setting priorities among research projects according to national needs, and the focus was instead on localized independent research projects. Fieldwork and research were accorded high priority, whereas data management was seen as an inherently tedious and unproductive task and received lower priority. Scientists claimed ownership of data, were protective of their data, and were afraid that others might misuse them. Furthermore, individual approaches to data management meant that data standards, either technical or semantic, were not complied with or developed. There was also a cultural divide between scientists who gather and use data and the information system and technology experts who develop and understand how to manage data. Another difficulty was that scientists lack a proper understanding of the needs of society and their stakeholders and often were unable to engage and communicate with them effectively to establish and realistically meet their needs.

Lessons Learned

In the decades before 2000, data had not received high priority and had not been highly valued. The involvement of scientists in information management is crucial, but data management typically does not have high priority, and placing that responsibility on scientists in the absence of strong and prescriptive directions has proved unsuccessful. Over the last decade, BGS has come to recognize that its Unique Selling Proposition is "national, long-term, and strategic" and that its core competence consists of both expertise and data. BGS adopted a corporate-

Box 3.3

British Geological Survey

Stakeholder Benefits from a Corporate- and Asset-based Approach to Data Infrastructure

- Reduced staff effort in finding data.
- Reduced duplication of effort (building databases and applications).
- Improved quality of data available to staff and customers.
- Allowed corporate implementation of standards and best practices.
- Facilitated collaboration within BGS.
- Provided the opportunity to integrate data of diverse types and to create innovative products and services.
- Enabled BGS to be more responsive to customer needs.

based and asset-based approach in developing its data infrastructure, and this benefited both its staff and stakeholders (see Box 3.3). External stakeholders are generally appreciative of the benefits of a professional and corporate approach to information, and their encouraging responses have has led to improvements in engagement, services, and internal processes. The cost-recovery model that funds national mapping in the UK has provided a powerful incentive for public-sector organizations and their employees to focus on customer requirements and to produce datasets that are complete and up to date. Organizations that lack this cost-recovery model, such as the USGS, will need to establish another way of incentivizing scientists to communicate. The cost-recovery model also limits the free availability of data as required for U.S. federal agencies. Finally, implementing an effective information strategy is not a one-time action but requires enduring responsibility.

Infrastructure for Spatial Information In Europe

The Infrastructure for Spatial Information in Europe (INSPIRE) is a directive of the European Commission that establishes an infrastructure for spatial information throughout the European Union (EU). The directive went into effect on May 2007, signaling that the EU decided that a coherent SDI was essential for environmental policy-making across its national boundaries. INSPIRE is a distributed infrastructure and will be based on existing SDIs operating in the 27 member states of the EU. In June 2010, the Krakow Declaration was approved which recommended that participating governments and organizations (1) maintain efforts and investments needed to establish INSPIRE; (2) increase international collaboration; and (3) support implementation of SDIs in non-EU countries.

INSPIRE is being implemented in stages; full compliance is required by 2019 (see Table 3.1 for major milestones). It addresses 34 spatial data themes

Table 3.1 Major Milestones for Implementing INSPIRE

Milestone date	Description
May 2007	Entry of INSPIRE directive into force
December 2008	Entry of INSPIRE metadata regulation into force
May 2009	Entry of provisions of directive into force in all European member states
December 2009	Adoption of INSPIRE regulation on discovery and view services
December 2009	Adoption of rules governing access rights of use to spatial datasets and services for Community institutions and bodies
November 2010	Establishment and running of geoportal at community level by the European Commission
December 2010	Metadata available for spatial data corresponding to Annexes I and II
November 2011	Discovery and viewing of services operational
December 2012	Transformation and downloading of services operational
November 2012	Newly collected and extensively restructured Annex I spatial datasets available
December 2013	etadata available for spatial data corresponding to Annex III
November 2017	Newly collected and extensively restructured Annex II and III spatial datasets available
February 2018	Other Annex I spatial datasets available in accordance with Implementation Rules for Annex I
October 2020	Other Annex II and III spatial datasets available in accordance with rules

SOURCE: http://inspire.jrc.ec.europa.eu/index.cfm/pageid/44.

needed for environmental applications (see Box 3.4). Some of the themes are within the purview of the USGS, but many extend well beyond the mission of the USGS. The directive is specific and provides detailed technical implementing rules, which cover metadata, data specifications, network services, data-sharing, service-sharing, and monitoring and reporting. The intent of INSPIRE is to enable the sharing of environmental spatial information and to facilitate better access to data held by public-sector organizations throughout Europe.

Key Challenges

INSPIRE is being implemented in 27 countries that have different languages and cultures, different levels of geographic-information maturity, varied legal systems, and varied approaches to public-sector data access. There are many challenges in introducing an effective SDI, and INSPIRE has defined a number of technical challenges that it has addressed as a part of its basic principles, including

Box 3.4

INSPIRE Spatial Data Themes

Annex I
1. Coordinate reference systems
2. Geographic grid systems
3. Geographic names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

Annex II
1. Elevation
2. Land cover
3. Orthoimagery
4. Geology

Annex III
1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and government services
7. Environmental-monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution and demography
11. Area management, restriction, and regulation zones
12. Natural-risk zones
13. Atmospheric conditions
14. Meteorological geographic features
15. Oceanographic geographic features
16. Sea regions
17. Biogeographic regions
18. Habitats and biotopes
19. Species distribution
20. Energy resources
21. Mineral resources

SOURCE: http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2/list/7.

- Collection of data only once and their being kept where they can be maintained most effectively.
- Ability to combine seamless spatial information from different sources throughout Europe and share it with many users and in many applications.
- Possibility for information collected at one level or scale to be shared at all levels and scales and to be detailed for thorough investigations and generalized for strategic purposes.
- Availability of geographic information for good governance at all levels that would be transparent and readily available.
- Discoverability of the available geographic information and awareness of how the data can be used to meet particular needs.

Reaching an agreement on the scope and design of INSPIRE was a major challenge, and implementing the directive is an even greater one. As of June 2010, Cyprus, Finland, France, Greece, and Luxembourg had failed to enact key INSPIRE components in their national law (EU, 2010). Although INSPIRE is coordinated by the European Commission, it is dependent on the consent and close involvement of stakeholders and experts in all member states. This international group develops, scrutinizes, and reviews rules and specifications before they enter into law.

Lessons Learned

INSPIRE is a multinational undertaking, and there are many lessons to glean from its implementation. Three overarching lessons are especially relevant for the USGS to consider. First is the importance of stakeholder collaborations for developing appropriate parameters for an SDI. The INSPIRE directive is a legal instrument that has been transposed into national law in 27 EU member states. To implement an SDI in a complex system in a reasonable timeframe, the EU decided that a legislative approach was necessary. However, INSPIRE still depends on open and transparent stakeholder involvement; it would not be viable without multinational collaboration to define and review specifications and processes.

Second is the importance of having relatively straightforward goals. INSPIRE has been able to distill the purpose of the SDI to making spatial data throughout Europe discoverable, viewable, interoperable, and downloadable and therefore removing barriers to the access and use of data.

Third is the importance of reasonable expectations and timelines given limited resources. The committee believes that despite the simple goals of INSPIRE, the expectations and timeline were too ambitious, and the resources necessary to carry out the goals were underestimated. With diverse stakeholders and data domains, a lesson for the Survey in implementing an SDI is the necessity of simplifying the vision and creating pragmatic objectives.

The Global Earth Observing System of Systems

In 2005, the Group on Earth Observations (GEO) launched efforts to create a Global Earth Observing System of Systems (GEOSS) that would link many different Earth observation systems into a common framework. The framework would not only support science but support decision-making and have applications in a wide array of "societal benefit areas" (SBAs). The nine defined SBAs include disasters, public health, energy, water management, weather, climate, agriculture, ecosystems, and biodiversity (GEO, 2011). With a 10-year implementation timeframe, GEOSS is still in the process of being implemented: it is building on a diverse set of contributed components, and a GEOSS common infrastructure is under development. There remain many challenges, and preliminary lessons can be derived from the experience to date. GEO has made noteworthy progress in SDI development in various ways.

Key Challenges

Technical interoperability is a key concern, in that it is difficult to interconnect diverse systems that were developed by different organizations and countries for different purposes. For a system of systems to function effectively, clear and open interfaces need to be defined between systems regardless of their specific structures and implementation of the component systems. Thus, a major thrust of GEO's technical efforts was on developing and agreeing to an open architectural approach that could be implemented through widely accepted and transparent interoperability standards. Several groups that are responsible for standards development and implementation were involved at the outset. Prototypes and testing activities were attempted in developing consensus on the most appropriate standards and specifications for exchanging data, metadata, and interlinking tools and services. With GEOSS addressing diverse applications and data types, a key challenge continues to be to develop and implement standards and specifications that can be interoperable among applications and disciplines while providing flexibility to allow tailoring of outputs and interfaces to specific user needs.

The voluntary nature of GEO meant that organizational and institutional cooperation and participation would be a key challenge for GEOSS implementation (GEO, 2011). The implementation plan includes an explicit expression of GEOSS data-sharing principles, which call for full and open exchange of data, metadata, and products within GEOSS and recognition of relevant international instruments and national policies and legislation. It will have to be determined how to enable more open and flexible use of data by GEOSS users while respecting the rights and concerns of data providers, all in the context of a voluntary international initiative.

Lessons Learned

Perhaps the most useful lesson learned from GEOSS to date is that a voluntary, intergovernmental framework has the potential to create a functional, global-scale SDI. The voluntary nature of the initiative has encouraged a focus on both short- and long-term incentives for participation, cooperation, and collaboration. In the case of GEOSS, such incentives include

- The expected benefits of shared data and services.
- The need to reduce unnecessary duplication in data collection, processing, analysis, and dissemination.
- The need for cooperative decision-making on regional and global levels on pressing environmental and resource problems.
- The desire to make progress on shared international goals for poverty reduction and sustainable development through better access to vital data.
- The importance of expanding the use of Earth observation and related geospatial data in a variety of SBAs.

Incentives like those are likely to be just as important for the long-term success and sustainability of an SDI as a legal or government mandate.

National Geospatial-Intelligence Agency

The National Geospatial-Intelligence Agency (NGA) relies on imagery and geospatial information to "provide timely, relevant, and accurate geospatial intelligence (GEOINT) in support of national security" (NGA, 2011). NGA was formed in 2003 and is both a combat-support and national intelligence agency, so it is staffed, funded, and guided by the Department of Defense and the U.S. intelligence community. By using imagery intelligence, mapping, charting, and geodesy, NGA uses GEOINT to form a common operating picture (COP) for military and senior decision-makers. A COP consists of a model of Earth (such as a chart, map, or composite image taken from a variety of sources) and then layers on locations of friendly forces, enemy positions, roads, power lines, buildings, or geologic features. This multi-layered complementary information is used to build detailed pictures and enables decision-makers to work with the best available data.

Key Challenges

The NGA has two continuing challenges relevant to the USGS: (1) developing and maintaining data-sharing partnerships with diverse national and international stakeholders, and (2) working with standards development organizations for the development and adoption of standards. The National System of Geospatial Intelligence (NSG) is a broadly-defined SDI that supports GEOINT across the many defense, military, and private-sector organizations involved in it. Those partners have geographic information services specialists, imagery intelligence officers, geographers, meteorologists, and others with a GEOINT perspective. The NSG collectively harnesses the skills and energy of those agencies to tackle the highest-priority challenges of the U.S. government. An NSG Senior Management Council meets twice a year to review unified operations, improve information-sharing, re-evaluate methods, and define the most difficult challenges ahead. NGA has an international program that provides a unified direction in building and maintaining international partnerships. There is a continuing need to provide information and address the concerns associated with releasing information to allies and coalition partners, and forging relationships with these partners is increasingly important because of the growth of coalition activities, evolving international threats, and the expanding globalization of GEOINT.

The second challenge facing the NGA is the critical role of universally accepted and agreed-on standards. Standardization ensures that NSG system components perform as they should and are integrated in a way that allows GEOINT to be exchanged between them. The National Center for Geospatial Intelligence Standards (NCGIS) is the coordinating organization in the NGA that is responsible for setting and implementing GEOINT standards-management policies for NGA and the NSG community. The NCGIS was established to ensure a coordinated standards-based approach to achieving data and system interoperability, implement collaborative business practices, and act as an advocate for the needs of the NGA and the NSG community. Through strategic planning and enterprise architecture-based analysis, the NCGIS strives to optimize NGA resources as it implements a comprehensive NSG-wide standards-management policy. The NCGIS sponsors the Geospatial-Intelligence Standards Working Group, an NSG community forum that addresses issues on the latest standards that are critical for achieving the systems interoperability necessary for mission success. An NSG-wide plan for standards and continued involvement of the NSG community are crucial for developing and implementing standards that enable sharing of timely, relevant, and accurate GEOINT.

Lessons Learned

Several lessons can be taken from the NGA approach. First, it is important to establish a plan that is based on a vision and shared values among all partners involved in spatial data. The USGS could use NGA's mission-driven approach to providing geospatial information and products to relevant users of USGS information, in the same way that NGA provides GEOINT to its stakeholder communities. Second, it is important to identify common standards for participation. Technology and data standards are key to enabling interoperability of information resources and services throughout a broad community. Third, it is critical to address the business requirements of the community. The purpose of an SDI is to provide an information infrastructure and a service to a community. The NGA explicitly recognized the service role as central to its mission, so priorities had to be set among the information needs of the users.

National Aeronautics and Space Administration

The National Aeronautics and Space Administration (NASA) is the U.S. operations and research organization for space and aeronautics. NASA provides numerous types of Earth-observing data, primarily from space-borne and airborne platforms through its Earth Observing System (EOS). The EOS is a coordinated series of satellites that produce long-term observations of land surfaces, biosphere, solid Earth, atmosphere, and oceans. The National Research Council report Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond (NRC, 2007a) recommended that NASA launch a set of 15 missions in the next decade to continue expanding its space-borne missions.

Key Challenges

Data typically flow systematically from space-borne missions, but some mission events result in other data acquisitions in addition to or at the expense of the routine observations. That is analogous the program-vs-project tensions described in Chapter 2. For example, an active airborne science program results in intermittent, campaign-driven observations that are less continuous or consistent in time and location. However, one advantage of the airborne program is its flexibility to tailor data collections to observed phenomena. Numerous instruments collect the observations, and this results in a wide variety of data of different measurement types, coverage, and resolution. The diversity of data streams presents a challenge in relation to workflow, coordination, and standardization.

Lessons Learned

In collecting and analyzing data from different types of observations systems, it is necessary to have an industry data standard for ensuring that data are properly integrated and georeferenced. Standardizing data formats increases the user base and makes it easier to integrate and access different types of data. NASA has learned that it is imperative to avoid frequently changing formats (Friedl and Donnellan, 2010), inasmuch as a change in detection or classification of data can suffer large errors from small co-registration or geo-registration errors (Townshend et al., 1992). Resampling or reprojection of data can also cause geospatial errors. Similarly, metadata need to be in a standard format and be easily interpretable. In the absence of consistent metadata, researchers run the risk of using the data and derivative products improperly; a single convention for variable names would help to avoid improper use of data.

It is important to archive existing data, so NASA uses several Data Analysis and Archiving Centers (DAACs). However, the distributed and locally controlled DAACs can make it difficult for users to access and use NASA data. Simple changes, such as an agency-wide single sign-on Web portal, could greatly improve data access. NASA has also learned that a consistent core set of capabilities would be beneficial, such as powerful, flexible, and consistent visualization tools that enable researchers to more fully explore and use data from NASA and other organizations. One such approach that NASA used is the Open Geospatial Consortium visualization and data-delivery services.

Texas National Resources Information System

The Texas National Resources Information System (TNRIS) was established (as the Texas Water-Oriented Data Bank) by the Texas legislature in 1968; its mission was to provide a “centralized information system incorporating all Texas natural resource data, socioeconomic data related to natural resources, and indexes related to that data that are collected by state agencies or other entities” (TNRIS, 2011). TNRIS evolved into one of the first state-wide clearinghouses for Geographic Information System (GIS) data with staff trained in the natural, computer, and library sciences to supply data to government, academe, the private sector, and the public. The TNRIS data catalog includes about 1 million frames of aerial photography and more than 50 unique datasets equal to about 50 terabytes that it distributes through an estimated 10,000 data downloads per month (TNRIS, 2011). Datasets include data on elevation, land cover, geology, soil survey, meteorology, hydrography, mineral resources, energy resources, orthoimagery, Landsat, light detection and ranging (LIDAR), and census. TNRIS has cooperative agreements with many data-collection agencies; for example, it regularly combines funding with the U.S. Department of Agriculture (USDA) Natural Resources Conservation Service and makes specific recommendations for the soil-survey data.

Making data discoverable and viewable and archiving data are important functions of TNRIS. TNRIS ensures that its data conform with open geospatial data standards, which allow it to provide imagery to Google and Microsoft for their Web mapping systems. TNRIS publishes metadata about its holdings according to standard federal metadata practices, but it also associates additional "tags" for search-engine optimization and discoverability via Web searches and its own Web mapping viewer. TNRIS has periodically received grants to digitize over 1 million air photos from 1920 to the 1980s, and the scanned images are available to the public. TNRIS also maintains a limited historical map collection and has worked with other agencies to scan historical maps into accessible archives. The success of the TNRIS SDI program lies in the economic impact of making data available at no cost and the fact that TNRIS uses various metrics to track the use of the data provided, including number of downloads and frequently downloaded files. About 1 terabyte of data is downloaded monthly at an original equivalent cost of roughly $1 million. TNRIS is working to increase the granularity of the statistics to track its performance.

Key Challenges

One of the primary challenges for TNRIS has been to maintain constant base funding for acquiring new data. To maintain funding of about $1 million per year, TNRIS has had to demonstrate the use and value of the data archives. Therefore, it is constantly seeking ways to improve how it measures data access and use. Data-storage requirements have become an issue as demand for increased frequency, accuracy, and quality of data has continued to drive the need to make more data available. Infrastructure costs are rapidly declining, and new cloud platforms present potentially efficient services for hosting and dissemination of data. Keeping pace with lowered infrastructure costs is important to offset increases in data-storage requirements. A review by the state Council on Competitive Government found that programs typically prohibited sharing of data for 2 years and presented barriers to free access to data. As published data lose currency and as changes occur, there are opportunities for incorporating community stewardship of key data, such as road and hydrography networks. Maintaining authoritative datasets with "crowdsourced" data will require more sophisticated technology and processes to strengthen the quality and accuracy of an SDI.

Lessons Learned

Some lessons can be learned from TNRIS that range from strategic programmatic and management concerns to technical ones. The alignment of large-scale data repositories with clear priority issues contributes to long-term sustainability of data-acquisition programs. In the case of Texas, current and accurate data have been associated with water-resources management in the state for over 50 years. It has been important for TNRIS to have active partnerships that serve all levels of government and the public: Texas has a long history of working with federal agencies, such as the USGS and USDA. Adaptive technology strategies are key to keeping current on migration to Web-based data services and to lowering long-term costs. Open standards for data and Web platforms are essential, and open access to public-domain data has been important for withstanding cycles in funding and priorities for public geodata. Texas's statutory authority to designate members of a Texas Geographic Information Council has reinforced a long-standing culture of data-sharing and coordination. Adopting a state-wide data-acquisition contract has allowed greater response by state, tribal, and local governments to identify mutual projects and initiate procurement within weeks rather than months. Dedicated capital funding makes it possible to allocate funds for data purchases and supports clear and meaningful metrics for tracking data priorities and results. Organizational culture is important for developing a strong culture of sharing and value creation, so it is essential to have executive support in many agencies to foster that development. Also, a commitment to a long-term vision of open access to data is also necessary for success.

NAVTEQ

NAVTEQ, formerly a subsidiary of Nokia Corporation and now fully integrated into it, provides highly attributed digital roadmaps with extensive coverage throughout 85 countries. The U.S. database consists of more than 5.5 million miles of roads organized into five function classes with up to 260 attributes per segment. Data are gathered principally by driving the roads with GPS-equipped vehicles to record and verify visual attributes, such as address ranges, median types, and lane markings. When possible, the company also sources data from local authorities, government records, and, increasingly, individuals who submit "map reports" online, indicating a change or error in the database. Increasingly, the company processes data from contracted "probe" vehicles—position data gathered from vehicles equipped with positioning systems that can substitute for dedicated drives and help to identify changes in the road network. More recently, the company has fielded vehicles with scanning LIDAR and high-resolution video, providing dense point clouds and imagery of all features along the road. The latter generates a massive amount of data with each vehicle drive; each map feature is versioned, and a history of prior versions is maintained. Because of the complexity and volume of the data, a well-defined SDI is essential to NAVTEQ. NAVTEQ licenses data in a wide variety of formats to a wide variety of customers and applications; thus, data-sharing is already the nature of the business.

Key Challenges

Worldwide, over 1 million changes or additions are made in the NAVTEQ database daily, presenting a global data quality and consistency challenge that requires numerous coding tools and in-line checks. Field teams around the world upload their data to the main database, with coding of the data collected during a drive. Maintaining data quality over such a widely distributed data-collection system is a substantial challenge. NAVTEQ's approach is to embed a series of redundant checks into each step of its data production and validation process. An independent quality department maintains all quality processes and conducts a large number of tests during processing and after the database is updated, just before its public release. Another challenge is integrating data from sources outside NAVTEQ. Because data quality and thoroughness vary widely beyond the core 85 countries, NAVTEQ found it necessary to introduce two lower classes of digital roadmaps in addition to the primary, fully navigable map. The two new classes of map include third-party data that require NAVTEQ to translate submitted data into the required format and process them in a multistage pipeline that ensures basic soundness and validity before inclusion in the production map.

Lessons Learned

During the evolution of its SDI, NAVTEQ learned some important lessons related to the processes instituted that would achieve maximum levels of data quality. Key aspects of the quality program included attacking problems at the source, such as when the data are initially collected by the field team, because they are most familiar with their local areas, and when third-party data are acquired where some sources had provided poor-quality data that required extensive rework by NAVTEQ. It was also important to qualify all sources extensively and ensure that the same standards were applied to third-party data as to internally generated data. The analogous solution for the USGS would be to invoke standardized quality requirements at its sources, including state- and county-level data inputs.

A second key lesson learned is that the technology of the underlying database structure and tools needs to adapt constantly. NAVTEQ went through a massive reformulation of its database over several years and continues to modify its data structures with the inclusion of 3-dimensional data and imagery. The technical tools and structures used for the database need to be designed for current and known future expansion of requirements. For the USGS, that would mean that database structures will need to be planned beyond the current set to handle anticipated data types, such as multispectral data and an expansion of data layers.

NEON

The National Ecological Observatory Network (NEON) is funded by the National Science Foundation (NSF) to build on research findings of the last century and create a set of continental observations over the next 30 years. These observations would allow greater understanding of ecological changes by observing biological, biophysical, and geochemical interactions of various ecosystems across the nation's landscapes. The ability to understand ecological dynamics calls for a suite of observations to be conducted in an integrated manner, which few studies have been able to do. NEON was established as a structure to manage the environmental data cycle. That would include the use of sensors in the field and in space, processing and visualizing data, and sharing data and providing tools for collaboration from local to global scales. NEON is designed to answer grand questions in environmental science in a way that has not been possible with the individual investigator-driven efforts of the past. To accomplish that goal, NEON is building a cyberinfrastructure to manage large volumes of spatial data, such as physical geography, human geography, and satellite data.

Key Challenges

The complexity of data integration and synthesis is enormous. Much thought has been given to incorporating cyberinfrastructure and geospatial analysis tools needed to analyze ecological and environmental observations and the network of sites. The information framework provides synthetic capacity and provisions for forecasting ecological dynamics at specific sites and across regions. Other agencies have data that are critical for NEON's mission, thus NEON is partnering early with these organizations to share and process data. The USGS is negotiating a memorandum of understanding to establish a relationship with NEON so that cyberinfrastructure framework development will have an interface with the USGS SDI.

Lesson Learned

Because NEON is still in the planning phase, the lessons learned are limited. However, in constructing a purpose-built cyberinfrastructure from scratch, NEON has had the advantage of customizing an SDI that would meet its specific needs. Scientific data in the environmental sciences have a spatial component, and an SDI can form the backbone for handling scientific data. NEON has recognized the importance of formatting and checking the quality of data at every step as they move through the cyberinfrastructure—an approach to spatial data quality that is similar to that taken by some private companies. It includes a multistage pipeline approach that feeds data through a series of quality checks, including an examination by a scientist before data are published on the Web portal. To accommodate data that do not meet high NEON standards, NEON also categorizes data by level of quality. Another lesson from NEON is the importance of early partnerships to ensure that SDI constructs are integrated with existing SDI infrastructures of partner agencies.

CUAHSI

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is an organization that was established in 2001 to further hydrologic sciences. CUAHSI is supported by NSF and represents 122 U.S.

universities in advancing hydrologic science through programs such as a network of hydrologic observatories, the Community Hydrologic Modeling Platform, a Synthesis Center, and a Hydrologic Information System. The latter develops infrastructure and services to improve access to hydrologic data and is most relevant to a USGS SDI. Hydrologic data can be classified in several categories, including water-observations data, geographic data, climate and weather data, and remotely sensed data. The data are reported as time series measured at fixed geographic locations indexed by latitude and longitude. Major national holdings of time-series information include the USGS National Water Information System, the Environmental Protection Agency (EPA) STORET Data Warehouse, Climate Data Online from the National Climatic Data Center, and snow and soil water observations from the USDA National Resources Conservation Service. Additional water-observations data are collected by state and local water agencies and by academic investigators. A complete inventory of water-observations data on a particular subject and geographic region requires accessing and synthesizing the various data sources into a consistent form.

The CUAHSI Hydrologic Information System (HIS) project has designed a new language for transmitting time series of water-observations data through the Internet called WaterML, and the USGS now publishes its time series of real-time data and daily data by using WaterML (Zaslavsky et al., 2009). CUAHSI HIS has also designed an Observations Data Model (ODM) for storing time-series data and a Hydro Server for publishing such data in WaterML (Tarboton et al., 2008). A number of universities and NSF-supported research centers use this approach to publish their water-observations data. The Texas Water Development Board is also publishing a database of coastal observations in Texas collected by various state agencies as a series of ODM databases and corresponding WaterML time-series services. The Open Geospatial Consortium and the World Meteorological Organization have together formed a Hydrology Domain Working Group, which is evolving WaterML towards becoming an international standard for transmission of time series of water-observations data through the Internet. CUAHSI HIS has a service oriented architecture that is organized into multiple data servers for publishing data, a centralized catalog for collecting and publishing metadata to support data discovery, and a desktop client for data retrieval and analysis. What is emerging is a service-oriented architecture for water-observations information.

Key Challenges

It has become apparent that a functional hydrologic information system in the United States will require data sources to be labeled with standard terms so that searches of multiple information sources will be consistent. At first, CUAHSI attempted to label the time series indexed at HIS Central with a standard set of concepts, but it found it to be a nearly overwhelming task. Instead, the community agreed that standard hydrologic ontology is necessary so that this task can be performed by the data providers, which demonstrates the importance of semantics and ontologies as a necessary component of a dynamic SDI.

It is also apparent that a fairly formidable "digital divide" exists between GIS data and time-series water-observations data and between GIS data and continuous spatial arrays of weather, climate, and remote sensing data. With Hydro Desktop, it is not yet possible to ingest continuous arrays of information; smoothly combine them with time series of water data defined on discrete spatial objects, such as points, lines, areas, and volumes; and then link the resulting datasets with simulation models and analysis routines. That would create a true hydrologic information system, but this goal is still some distance away in research and technology development.

Lessons Learned

A key lesson learned from CUAHSI's experience is the importance of standards. There is a vast array of time-series data on water observations, and it is possible to access them in a common language. However, a hydrologic scientist will need to find and access the specific data in a format that is user-friendly. The emerging service-oriented architecture is being developed for water-observations information in which a client application, such as Hydro Desktop, is built specifically to search, view, and download water-observations data in WaterML. The generality of this construct is a way of organizing the functionality present in many spatial data information systems.

U.S. GEOLOGICAL SURVEY SPATIAL DATA INFRASTRUCTURE INITIATIVES

The USGS has successfully provided surveys and maps in support of the nation's science and economy for the last 125 years. In fulfilling its mission to map the nation, many lessons in conducting successful mapping programs have been learned. There have also been relevant lessons in conducting successful partnerships, ensuring the continuity of data and information, and keeping pace with changing needs and technologies. And there have been lessons in the importance of adequate and enforced specifications and standards, the benefits and difficulties of integrating data of disparate types, and the importance of conducting research on data needs, sources, production, and applications. This section describes several SDI-related initiatives at the USGS that are in addition to The National Map discussed in the previous chapter and outlines their key challenges and the lessons learned from them.

National Biological Information Infrastructure

The National Biological Information Infrastructure (NBII) was developed by the USGS as a platform to enable federal, state, and public partners to coordinate ecological and biological data holdings through the use of protocols that enhance data-sharing, data transfer, and geographic investigations (Rugg, 2004; NBII, 2011). The partner community provides work on standards, tools, and technologies that make it easier to find, integrate, and apply biological resource information in a geographic framework. A key goal is to make these data available to land managers, other scientists, and the public.

Key Challenges

The system includes the use of the Global Ecosystems Data Viewer to perform customized viewing and data selection and to download ecosystems data layers. Ecological and biological data are available as a continuous raster in which each pixel value represents class codes that are described in the metadata for each dataset. The effort to map standardized, meso-scale ecosystems for the contiguous United States provides a biophysical stratification system for the United States. The data used to develop the mapping process were not all of the same quality or spatial resolution, but each dataset obtained and used for the mapping was considered to contain "best available data" for a given theme at a national extent.

The NBII has developed various geographic data tools, but these have not been consistently applied among the diverse holdings of the NBII. An example of a GIS spatial modeling tool that links various databases can be viewed on the National Institute of Invasive Species Science Web site. This predictive spatial modeling tool is an online statistical tool used to help to develop predictive models by using various user-defined regression techniques, and it generates a predictive surface based on the selected model. The results can be overlaid on Google Maps, allowing the spatial distribution of a given species to be visualized with the original species occurrence data that were used to create the predictive species. The output results can be saved as a map or in a pdf file, depending on a user's needs. Although this is one success, a truly integrated data search and analysis portal is still not available.

Lessons Learned

One critical lesson from the NBII experience has been that establishing and distributing standards for the biological data community has been critical for the NBII's success. Another lesson applicable to a USGS SDI is that the best integrated data are of little value if they are not easily discoverable. The NBII Web portal will need to make some progress in this regard, and it is a formidable challenge when the data are as diverse as those maintained by the NBII.

National Hydrography Dataset Plus

The National Hydrography Dataset (NHD) was developed by the USGS as the surface-water map layer for The National Map. The National Hydrography Dataset Plus (NHDPlus) was first released in late 2006 and is a suite of geospatial products that builds on and extends the capabilities of the NHD. The NHDPlus integrates the NHD (1:100,000 scale) with the National Elevation Dataset (30 m) and the Watershed Boundary Dataset (WBD). Interest in estimating NHD stream flow volume and velocity to support pollutant fate-and-transport modeling was the driver behind the joint EPA–USGS efforts to develop the NHDPlus. The NHDPlus includes improved NHD names and networking; value-added attributes (such as stream order) that enable advanced query, analysis, and display; and elevation-derived catchments that integrate the land surface with the stream network, catchment attributes (such as temperature, precipitation, and land cover), stream discharge and velocity estimates for pollutant dilution modeling, and associated flow direction and accumulation grids. The NHDPlus represents the initial implementation of the national surface-water geospatial framework envisioned by the Subcommittee on Spatial Water Data, a group cosponsored by the Federal Geographic Data Committee and the Advisory Committee on Water Information.

Key Challenges

A major production-related challenge was integrating the vector-based NHD and raster-based National Elevation Dataset to produce the NHDPlus catchment (local drainage area) for each NHD stream segment. The catchments were used to associate temperature and precipitation attributes with each stream segment in estimating stream flow volumes. The underlying method used to produce the NHDPlus catchments is described in USGS SIR 2009-5233 (Johnston et al., 2009). One step in addressing this challenge is to align vector streams with hydrologically conditioning elevation data better during catchment production. Improved integration of data across international boundaries with both Canada and Mexico is also needed. For both countries, coarse representations of the portions of drainage areas that fell outside the United States border were used in the initial NHDPlus.

Good stewardship of the underlying data used to produce the NHDPlus is crucial. The federated data model with stewardship has been working well for the NHD, but it is threatened by limited resource support in the USGS. There is also concern that private efforts based on the NHD could eventually supplant the dataset rather than build on it. A potential solution is to encourage private-sector entities to compete vigorously to provide useful services based on the data but not allow them to own the data. The public could maintain ownership of the data themselves and keep them free and in the public domain. The NHD and WBD stewards have made major commitments of resources to support their side of stewardship, but it remains a challenge for the USGS to continue finding the necessary resources to support its obligation to data stewardship.

Lessons Learned

From an organizational perspective, there is much to be gained from multiagency cooperation in spatial data development. The NHDPlus team was able to leverage the collective interest and resources of EPA and USGS to complete what has since become a highly valued data source for the water-resources community. One of the biggest technical lessons learned is the need for the production process to be as automated as possible so that it can be updated regularly as underlying ingredient datasets are improved through the stewardship process. That is the impetus behind the current EPA–USGS effort to develop new NHDPlus production tools on the basis of the latest GIS technology.

There is tremendous demand for consistently produced nationwide and continental datasets, and the user community has been very supportive of the NHDPlus in particular, because it is easily digested by existing computer applications. Although it is challenging to find the necessary resources to produce such nationwide datasets, the long-term benefits will probably exceed the costs.

The federated data model appears to work and has been beneficial to all involved, but no single entity can afford to provide all the resources required to improve the data. The community currently supports the data through a stewardship process, and each partner benefits from improved data and reduced duplication of effort.

The Topographic Mapping Program

The USGS Topographic Mapping Program (TMP) began in 1884, and its topographic maps have become the signature product recognized by the public and industry as a versatile tool for viewing the nation’s landscape. It has served as an essential instrument in integrating and analyzing place-based information and is a seminal model of a federal agency that has successfully created and supported a comprehensive SDI for the United States. Almost from its beginning, topographic mapping was a cooperative effort of federal and state governments: Massachusetts and New Jersey cooperated with USGS in topographic mapping as early as 1885–1887 (Kelmelis et al., 2003). Since then, all states and many federal agencies have worked with USGS to make topographic mapping a cooperative effort.

Technological advancements have transformed topographic mapping science from printed products to digital data and online-based applications for accessing digital topographic maps. The USGS began developing the National Digital Cartographic Database (NDCDB) by converting existing maps to both raster and vector forms and developing new data to update them or create new maps. The USGS began to develop those data to be used in geographic information systems as well. A timeline of recent USGS developments in topographic mapping and GIS is provided in Box 3.5.

Key Challenges

The continual and necessary co-evolution of topographic mapping, technology, and emerging applications has presented a series of challenges for the USGS. For example, integration of existing data layers has included transitioning from analogue to digital maps for the NDCDB, separating topographic maps into data layers for the National Spatial Data Infrastructure, and recombining data layers to form The National Map (TNM) (Kelmelis, 2003). An additional challenge, one that the USGS continues to make progress on, is ensuring consistency with each new product release. For example, the National Land Cover Dataset (NLCD) released in 2001 incorporated many improvements learned in developing the previous release in 1992 (Homer et al., 2004). These improvements resulted in slight incompatibilities between these two releases that have since been rectified, but improving datasets while maintaining consistency among releases remains an ongoing challenge.

As previously mentioned, the USGS in 2001 released its vision for TNM, the topographic map of the 21st century. TNM is a seamless, continuously maintained, nationally consistent set of base geographic data that is available on the Internet. As a source of revised topographic maps, TNM serves as a national spatial data foundation for a broad array of issues such as land and resource management and homeland security, and the USGS recognizes the importance for which TNM can serve as the nation's trusted resource for current, consistent, and integrated topographic information. There are eight layers of topographic information provided in TNM: boundaries, elevation, geographic names, hydrography, land cover, orthographic images, structures, and transportation (Usery et al., 2010). TNM uses data from seamless databases developed in the 1990s and early 2000s, and has added data from federal, state, local, and tribal sources. The USGS Center for Excellence in Geographic Information Science (CEGIS) has spent much of the last decade in finding ways to integrate these layers across the various spatial scales.

A challenge for TNM is the eventual integration of mapping and scientific data beyond the current data layers. Developing an SDI to organize, integrate, access, and use scientific data within the scope of the USGS Science Strategy requires the technical advancement of present capabilities of TNM, which was developed to meet a different set of objectives and is less focused on complex multifaceted geoscience domain databases. Adding geoscience domain data involves more than simply adding layers to available GIS records. The TMP has propelled the creation of 1:24,000-scale and 1:100,000-scale topographic maps for most states, but there is not yet a consistent standard data format for geologic map legends across state boundaries. National bedrock geology with a resolution sufficient to satisfy USGS scientific staff is a large challenge.

Research in energy and mineral resource geology requires much more complicated datasets that convey rock-forming processes, aerial distribution, age relationships, geochemical and geophysical data, and resource attributes. One potential solution is a raster-format geoTIFF geologic coverage of the United States on 1:24,000 and 1:100,000 scales. Later, vector data formats could be developed that could support functionalities beyond viewing, including searchable formats. Standardized formatting and metadata could replace the less pro-

Box 3.5

Timeline of Recent USGS Efforts in Topographic Mapping and GIS

1987 — Introduction of digital orthophotograph quadrangle (DOQ). The USGS generated digital images with correct map geometry, which were created using photographic stereo pairs. The USGS partnered with USDA to generate digital orthophotographs at 1-meter resolution, and the USGS's DOQ became the standard base image for GISs in the 1990s.

1991 — Completion of analogue map coverage of the contiguous United States. The USGS generated a map of the contiguous United States on a 1:24,000 scale, which included more than 55,000 7.5-minute quandrangles for the National Mapping Program. The most recent versions of the 1:24,000-scale topographic maps were converted to digital raster graphics (DRGs), which are geocoded and are a critical layer in GIS and used for applications such as feature extraction and image rectification.

1991-1992 — Transition to seamless nationwide layer-based datasets. Using seamless nationwide layer-based datasets, the USGS was able to first complete the National Elevation Dataset (NED) and then the National Land Cover Dataset (NLCD). The NED was created by using existing USGS databases to provide a seamless, nationwide, multi-resolution mosaic of elevations, with improvements now available to show 10-meter horizontal spacing and even 3-meter horizontal spacing with LIDAR generated elevations. The NLCD was created by using Landsat Thematic Mapper (TM) images to provide a seamless mosaic of land cover for the United States. The USGS cites the NLCD as one of its most frequently downloaded datasets, with the most recently released version in 2011 based on 2006 TM satellite data.

SOURCE: Usery et al., 2010.

ductive efforts involved in reformatting large datasets. However, there is still a discernible cultural divide among USGS scientists in how they perceive, share, and use interdisciplinary data. There is a need for incentives that serve the process of integrating science; save time in discovering, visualizing, and handling data; and propel the effective use of information.

Lessons Learned

There are many lessons to learn from the evolution of the TMP. First, partnerships with state and federal agencies are essential and allowed the USGS to share costs and to access data that would not otherwise be available. Second, compiling and managing spatial data require a long-term investment in evolving technology; it took over 100 years to complete the first coverage of the 48 contiguous states. An SDI would be best viewed as an ongoing initiative that would adapt with changing user needs and technical advances. Third, an SDI would need to be designed with the future in mind. The rate of spatial data collection is increasing exponentially: more data have been collected in the last decade than in the entire previous history of the TMP. Establishing standards for data formatting that anticipate needs many years down the road can enable useful data integration in the future.

Center of Excellence for Geospatial Information Science

CEGIS was created in 2006 to "identify, conduct, and collaborate on geospatial information science research issues of national importance; assess, influence, and recommend for implementation technological innovations for geospatial data and applications; and maintain world-class expertise, leadership, and a body of knowledge in support of the National Spatial Data Infrastructure" (USGS, 2011). The role of CEGIS is not to collect or process data but to develop technology that aids data-processing, specifically to develop tools and data formulation for TNM. CEGIS plays a major scientific role in defining the standards and structure for the USGS SDI, so it has the role of implementing the SDI for the agency. Following the recommendations of the National Research Council report A Research Agenda for Geographic Information Science at the United States Geological Survey (NRC, 2007b), CEGIS now includes strong emphasis on three high-priority research areas: (1) investigating new methods for information access and dissemination; (2) supporting integration of data from multiple sources; and (3) developing data models and knowledge organization systems.

Key Challenges

The USGS underwent transitions in recent years that led to a declining ability to coordinate national-level geospatial research. As a result of waning leadership, the USGS lacked a dynamic, nimble, cutting-edge research unit that could lead national efforts and harness capabilities in academia, government, and industry (NRC, 2007b). Furthermore, the 2007 National Research Council report recognized that challenges inherent in geographic information science would need to be addressed before TNM could be successfully implemented. TNM requires data to be generalized and fused with different scales, resolutions, and quality, and the standardization and integration of such disparate spatial data sources for TNM has been a serious challenge for CEGIS (NRC, 2007b).

Lessons Learned

In recommending how CEGIS could realize its potential, the 2007 National Research Council report emphasized the importance of collaboration with other agencies and organizations that carry out geographic information science research and emphasized the critical need for CEGIS and the USGS to establish effective leadership in geographic information science (NRC, 2007b). It would be difficult to coordinate an effective research agenda without external networks of partners and without cohesive leadership to drive the agenda.

FINDINGS

A science organization's core competence consists of expertise and data, and it follows that implementing an SDI would require both for success. Through its examination of SDI implementation in various agencies and countries and their key challenges and lessons learned, the committee found similar themes that are relevant as the USGS moves forward in implementing its own SDI. The committee found that successful implementation of an SDI depends on an agency's roadmap and strategy, organizational leadership and culture, standardization, technical competence, funding and contracting, workforce competence, and cooperation and partnerships. Individuals who provided testimony to the committee expressed great hope that large benefits can come from a fully functioning SDI at the USGS (Box 3.6).

Roadmap and Strategy

The committee found that developing a roadmap and strategic goals are integral to implementing an SDI. In the case of INSPIRE, legislation was necessary for implementing an SDI in a complex federated system in a reasonable timeframe. In the absence of a legislative mandate, the committee found that SDI roadmaps that were well developed and consistently reviewed and updated were the most successful ones. The BGS and Geoscience Australia are the closest analogues to the USGS and the BGS in particular has a well-written business plan

Box 3.6

Sample Testimony Provided to the Committee

(See Appendix D for additional responses.)

"Correctly organized, an SDI will give the USGS the flexibility and agility to increase its capability in the rapidly emerging field of computational geosciences and enable it to unlock the breadth and depth of its scientific data to a far wider group of clients and stakeholders."—State-level respondent

"Carefully structured an SDI will give the USGS the flexibility and adaptability to meet not only its current 6 key strategic science directions: it will also enable the USGS to rapidly change directions to meet new Geo-scientific challenges in the decades beyond 2017."—Federal-level respondent

"Properly managed an SDI will enable the USGS to conduct multidisciplinary, collaborative science projects that are focused on delivering influential scientific solutions to the current six key strategic science directions identified in the document US Geological Science in the Decade 2007-2017."—International respondent that clearly articulates the merit of an SDI and the community that it would serve. As demonstrated by the assortment of SDI roadmaps the committee examined, it was important for the accompanying strategic goals to be straightforward and for these goals to undergo periodic evaluation because compiling and managing spatial data require a long-term investment in evolving technology. In addition, incentives (such as reduced discovery costs and reduced duplication) are likely to be just as important for the long-term success and sustainability of an SDI as a legal or government mandate.

Organizational Leadership and Culture

The committee also found that organizational leadership and culture influence how roadmaps and strategic goals are carried out on a daily basis and probably shape the success of SDI implementation. Incremental SDI implementation was key to success: leadership that established progressive goals rather than a final deadline found greater adherence to those incremental goals and thus greater success. Executive support was essential for developing an institutional commitment to a long-term vision of open access to data and value creation. Executive-level support drove the commitment and enthusiasm of senior, middle, and junior members of staff. Agencies that found success were the ones that included knowledgeable and respected leaders in the community that could champion and articulate a strong case for an SDI in the organization. In examining several other agencies and their SDIs, the committee found that the organizations that were most successful in building SDIs were the ones that had a long history of collaborating with others and a culture that focused on making data and information accessible to the broader community. Also important was that these organizations developed a mantra of under-promising but over-delivering on deadlines and products.

Standardization

Standardization was another key theme that echoed through the various case studies. Establishing standards for the data community and distributing them are critical for SDI success because technology and data standards enable information resources and services to be interoperable. Implementation was more seamless and effective for SDIs that incorporated the needs of the user community to develop and improve standards and for the ones that also accepted the need for data products to conform to international standards. For example, metadata standards included standardized variable and parameter definitions. Labeling data with standard terms allows searches of multiple information sources to proceed consistently. It was also essential to have open standards for data and Web platforms. Finally, it was important for data to be properly formatted and quality-checked as they entered the system and throughout each step as they moved through the cyberinfrastructure pipeline. Once standards have been determined, the committee found, it was essential to move forward by consistently applying the standards and to avoid indecision over which standards to follow.

Technical Concerns

In examining the various SDIs and how they were implemented across different agencies, the committee found that agencies had to overcome some technical concerns. Data quality was an issue, and it was best addressed at the time of collection before data were propagated through the SDI. On a technical point, the primary requirement for fusing data is accurate georeferencing of data products; changing the detection or classification of data can result in large errors that arise because of small co-registration or geo-registration errors. The committee found that with evolving technology, the technology and tools of the underlying database structure would need to adapt constantly in anticipation of data types beyond the current set, such as multispectral data and an expansion of data layers. In this case, the automation of a stewardship process is valuable so that updates can occur regularly. It is imperative to avoid frequently changing formats. Large-scale data repositories with clear priority issues depended on the long-term sustainability of data-acquisition programs. In addition to data collection and analysis, it is important to archive data: the best integrated data in the world are of little value if they are not easily discoverable.

Funding and Contracting

The committee found that funding and contracting mechanisms affected how well implementation could be carried out. One key factor was adequate funding for carrying out activities—not overfunding or underfunding. Overfunding can lead to waste, whereas underfunding can lead to frustration and the inability to reach goals in a reasonable time, and the exact level of adequate funding for the USGS SDI will vary with each phase of the roadmap suggested in Chapter 5. With dedicated capital funds, resources can be properly allocated for data purchases and for developing clear metrics to track data priorities and results. An organization-wide purchasing contract allowed an organization to acquire technology and data in weeks instead of months. An open-data policy is fundamental for long-term support by stakeholders, and this long-term approach was necessary to withstand cycles in funding and priorities for public geodata.

Workforce Competence

The committee observed that workforce competence contributed to successful implementation of SDIs. Training and retaining a skilled workforce will be critical for introducing and maintaining an SDI. An SDI introduction will initially be disruptive, and there will need to be a full understanding of that fact and a need to develop realistic expectations. As data are available and new areas emerge that are relevant to an SDI (for example, data science), it will be important to recruit talented experts in those areas who will continue to make the SDI useful and relevant. The committee also found that it was important to have highly trained and respected professionals who understood the technology.

Cooperation and Partnerships

Partnerships with state and federal agencies are essential for SDI implementation and for the long-term sustainability of an SDI. An SDI partnering plan can be successful if it is based on a common vision among its partners; there is much to be gained from multiagency cooperation on spatial data development. In the case of GEOSS, a voluntary intergovernmental framework has the potential to create a working global-scale spatial data infrastructure.

REFERENCES

Australian National Audit Office. 2010. The Auditor-General Annual Report 2009-2010. Report No. 22 Available online at http://www.anao.gov.au/uploads/documents/2009-10_Audit_Report_No.22.pdf (Accessed June 8, 2011).

British Geological Survey, 2011. Annual Report of the British Geological Survey 2009-2010. Nottingham: UK. 80 pgs.

Friedl, L., and A. Donnellan. 2010. Earth Observations Missions Applications Workshop: Charge to Workshop and Overview of Needed Products and Outcomes. February 1-3. Available online at http://appliedsciences.larc.nasa.gov/pdf/2010EOMAW/Overview.pdf (Accessed June 9, 2011).

GEO (Group of Earth Observations). 2011. About GEO. Available online at http://www.earthobservations.org/about_geo.shtml (Accessed June 9, 2011).

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. Photogrammetric Engineering and Remote Sensing 70(7):829-840.

Johnston, C.M., T.G. Dewald, T.R. Bondelid, B.B. Worstell, L.D. McKay, A. Rea, R.B. Moore, and J.L. Goodall. 2009. Evaluation of catchment delineation methods for the medium-resolution National Hydrography Dataset: U.S. Geological Survey Scientific Investigations Report 2009–5233, 88 p. Available online at http://pubs.usgs.gov/sir/2009/5233/pdf/sir2009-5233.pdf (Accessed June 13, 2011).

Kelmelis, J.A. 2003. To The National Map and Beyond. Cartography and Geographic Information Science 30(2):185-198.

Kelmelis, J.A., M.L. DeMulder, C.E. Ogrosky, N.J. Van Driel, and B.J. Ryan. 2003. The National Map From Geography to Mapping and Back. Photogrammetric Engineering & Remote Sensing 69(10):1109-1118.

EU (European Union). 2010. Environment: Six member states face Court for failing to apply EU laws on their statute books. 3 June. Available online at http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/686 (Accessed June 9, 2011).

Griew, P. V. 1990. IS Strategy Scoping Study for the British Geological Survey. Consultancy Report.

McMahon, Gerard; Susan P. Benjamin; Keith Clarke; John E. Findley; Robert N. Fisher; William L. Graf; Linda C. Gundersen; John W. Jones; Thomas R. Loveland; Keven S. Roth; E. Lynn Usery; and Nathan J. Wood, 2005. Geography for a Changing World: A Science Strategy for the Geographic Research of the U.S. Geological Survey, 2005-2015. Circular 1281 U.S. Geological Survey: Sioux Falls, SD, 76 pg.

NBII (National Biological Information Infrastructure). 2011. National Biological Information Infrastructure at a Glance. Available online at http://www.nbii.gov/portal/server.pt/community/nbii_home/236 (Accessed June 13, 2011).

NGA (National Geospatial-Intelligence Agency). 2011. National Geospatial-Intelligence Agency: Vision, Mission, and Goals. Available online at https://www1.nga.mil/About/WhoWeAre/VisionMissionGoals/Pages/default.aspx (Accessed June 9, 2011).

NRC (National Research Council). 2007a. Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond. Washington, DC: The National Academies Press.

NRC. 2007b. A Research Agenda for Geographic Information Science at the United States Geological Survey. Washington, DC: The National Academies Press.

Rugg, D.J. 2004. Creating FGDC and NBII Metadata with Metavist 2005. General Technical Report NC-255. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. Available online at http://www.ncrs.fs.fed.us/pubs/gtr/gtr_nc255.pdf (Accessed June 13, 2011).

TNRIS (Texas National Resource Information System). 2011. About TNRIS. Available online at http://www.tnris.state.tx.us/About/Index.aspx (Accessed June 9, 2011).

Tarboton, D. G., J. S. Horsburgh, and D. R. Maidment 2008. CUAHSI Community Observations Data Model (ODM), Version 1.1, Design Specifications . Available online at http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf (accessed May 18, 2012).

Townshend, J.R.G., C.O. Justice, C. Gurney, and J. McManus. 1992. The Impact of Misregistration on Change Detection. Geoscience and Remote Sensing 30(5):1054-1060.

Usery, E.L., D. Varanka, and M.P. Finn. 2010. A 125 Year History of Topographic Mapping and GIS in the U.S. Geological Survey 1884-2009, Part 2. 1980-2009. Available online at http://nationalmap.gov/ustopo/125history_part_2.html (Accessed March 10, 2012).

USGS (U.S. Geological Survey). 2011. Center for Excellence for Geospatial Information Science. Available online at http://cegis.usgs.gov/about_us.html (Accessed June 14, 2011).

Zaslavsky, I., D. Valentine, D. Maidment, D.G. Tarboton, T. Whiteaker, R. Hooper, D. Kirschtel, and M. Rodriguez, 2009. The Evolution of the CUAHSI Water Markup Language (WaterML)", EGU General Assembly 2009, Geophysical Research Abstracts, Vol. 11, 21 April, EGU2009-6824-3.

4

A Vision for Optimizing the USGS Spatial Data Infrastructure

The U.S. Geological Survey (USGS) has the unique role of and responsibility for preserving and archiving geospatial data on a national scale. An optimal spatial data infrastructure (SDI) for the USGS would include data standards, modern data-management services, and a set of key application services that are essential for addressing scientific questions. The SDI would also need to consider the importance of data-sharing and data discovery and to have flexible methods for preserving geospatial data for long periods through numerous changes and updates because the ability to document and analyze changes in temporal values on a national scale is of immense scientific and societal value.

Carrying out the committee's vision for an SDI at the USGS requires synergistic partnerships with agencies and organizations that have already contributed to the SDI. A judicious selection of partners will enable the USGS to leverage its resources while adopting best practices and furthering interagency standardization. Finally, the success of a vision for a large program like the USGS SDI depends on supportive leadership and carefully planned, staffed, budgeted, and executed governance and policies.

DISCOVER AND SHARE FOR THE LONG TERM

Data are major tangible assets of the USGS. Although these assets are unique national and international resources, they are resources that have yet to realize their full potential, because many components remain inaccessible or are not interoperable. The lack of a single information space where data can be discovered and accessed is also an issue. The existence of data may be known by USGS programs and individuals but is largely unknown and much less understood outside the organization. An effective USGS SDI would need to enable and facilitate broad data discovery, sharing, and archiving across the research community.

Data discovery is the first important task of the USGS, and it will need to ensure the discoverability of prime datasets in each division. Data discovery simply means that basic information about the existence of a spatial dataset and how it can be obtained is widely available. Once prime datasets (the datasets most critical to the pursuit of each science mission) have been identified and indexed, they must be searchable and accessible in a corporate data-management system. That will require the development of new institutional policies and series of standards on metadata and data discovery. It will also require compliance with these policies and standards.

Data sharing is the second critical task for a functional USGS SDI. Data must be structurally and semantically interoperable so that they can be shared and integrated with other datasets in the USGS, around the nation, and with international partners. As a multidisciplinary organization, the USGS will need to be able to combine and synthesize data from various disciplines to contribute to its cross-domain missions.

USGS has the responsibility for maintaining data for the long term. Thus, an effective institutional strategy for data-archiving is needed as the third fundamental component of a USGS SDI to support temporal analysis. The USGS has a long history of creating of authoritative spatial datasets and, therefore, data creation is not included in the ‘discover and share for the long term’ mantra that was developed to help the USGS focus on the remaining steps beyond data creation.

STANDARDS

Standards and interoperability are essential elements of an SDI, whether implemented in a region (such as the Infrastructure for Spatial Information in Europe) or in an organization (such as the British Geological Survey). Standards apply not just to data but to the array of processes that operate in an SDI. Standards require consistency of operation, which will be somewhat challenging for a scientific organization that needs to function within defined parameters but at the same time to innovate for future needs. Not all standards meet every user community’s needs and, in some cases, non-standard derivative products may be necessary. However, the committee believes that existing standards that have been developed with the input from across the user community are the best way of providing the widest possible access to outside users. The USGS also needs systems that are interoperable and that follow internationally agreed-upon consensus standards, such as the Open Geospatial Consortium (OGC) and Geoscience Markup Language (GeoSciML), if they are to advance national and international multidisciplinary science. That will require the USGS to design and build an information-management system within the SDI so that information can be effectively managed, analyzed, and delivered to the appropriate stakeholders within and outside USGS. As a major international player, the Survey will need to collaborate with international partners to address data standardization and to comply with international protocols. Expanding ongoing efforts to make spatial data available in Keyhole Markup Language (KML) and other formats compatible with popular web-based map viewers such as Google Earth and Bing Maps will provide great value to USGS spatial data.

Open Geospatial Consortium

The OGC is an international organization consisting of 420 government, industry, and academic entities that participate in a consensus process to develop open spatial data interface standards. Its core mission is "to develop standards that enable interoperability and seamless integration of spatial information, processing software, and spatial services" (OGC, 2004). It allows users of geospatial technology to work with technology providers. The OGC has defined some key interoperability standards for geospatial data that are supported by U.S. federal agencies, international data providers, national SDI organizations, and commercial software providers. Many OGC standards have been incorporated into International Organization for Standards (ISO) standards and, conversely, many OGC standards incorporate ISO standards. OGC standards provide an essential infrastructure for SDIs that are designed to integrate fully onto the Web, and the OGC specification process and products have been adopted by nearly all SDI programs worldwide.

Geoscience Markup Language

GeoSciML is a major geoscience interoperability standards initiative that is being developed and supported by geological organizations worldwide. It is a geography markup language application schema that transfers and shares geologic information typically in the form of geologic maps. GeoSciML standards are based on the World Wide Web Consortium (W3C), the OGC, and standards and specifications of the International Organization for Standardization (ISO). The OneGeology project is an initiative that uses GeoSciML to increase the accessibility of geologic map data on Earth that are delivered in real time by merging data from several national geological surveys. To establish a common suite of features, GeoSciML draws from geoscience-data model efforts, geologic criteria (such as units, structures, and fossils), and artifacts of geologic investigations (such as specimens, sections, and measurements). Supporting objects are also considered (such as timescale and lexicons) so that they can be used as classifiers for the primary objects. GeoSciML meets the short-term goal of providing geoscience information associated with geologic maps and observations, and it could be extended in the long term to other geoscience data.

GeoSciML is governed by a working group of the International Union of

Geological Sciences Commission for the Management and Application of Geoscience Information. It would benefit from substantially increased USGS involvement because of USGS's major role as a global geoscience-data provider, and the USGS would benefit greatly from making its data more interoperable, both within USGS and externally.

ENTERPRISE DATA MANAGEMENT FOR AN SDI

Enterprise data management (EDM) refers to the ability of an organization to define data precisely, integrate them easily, and retrieve them effectively for both internal applications and external communication (DAMA, 2011). A common objective of EDM services is creating and maintaining data content that is accurate, precise, granular, consistent, transparent, and meaningful. There is an emphasis on integrating data content into business applications and facilitating the transfer of data from business process to another. EDM applies to the management of spatial data resources and other types of scientific and business data, and it commonly tries to address circumstances in which users in organizations or in collaborative environments independently source, model, manage, and store data. Although EDM is not dependent on a specific data-type or technology strategy, it still requires a strategic approach when selecting appropriate technologies, processes, and governance structure. That can often be a challenge for organizations because EDM requires an aligning of activities (such as data-content management) with their multiple user groups (such as finance, information technology, and operations). Moreover, in scientific organizations in which data have typically been managed by individual researchers or small teams, responsibilities for data management have fallen on individual researchers with uneven results. Uncoordinated data-management approaches can result in data conflicts and inconsistencies in quality, which makes it difficult for users to rely on such data for generating models, providing estimates, and informing decision-making.

The USGS SDI efforts can benefit from EDM techniques that have been adopted by others (see lessons learned and case studies in Chapter 3). For example, consolidation of data-management resources (such as database licensing, performance, backup and recovery, and archiving) helps to improve economies of scale. Those benefits apply whether data are centralized in an organization, distributed over multiples sites, or hosted in the cloud.

Data-Centric Research Challenges

The SDI concept is an outcome of a data-centric approach that is changing the management of information resources that are needed to support science and transforming scientific research and environmental policy-making.. As data collections used to support specific research projects increase to petabytes, desktop Geographic Information Systems and statistical tools alone will be insufficient to support the complex workflows required by scientists. Goble and De Roure (2009) have noted that "we are in an era of data-centric scientific research, in which hypotheses are not only tested through directed data collection and analysis but also generated by combining and mining the pool of data already available." The data-driven landscape is expanding in scale and diversity. Spatial and scientific datasets grow in size and number, but they are often poorly coordinated and incompatible, so discovery and integration tasks present serious challenges. The data mismatch may result from incompatibilities in scale, resolution, feature class, or temporal resolution or from semantic mismatch. In this environment, it is important to go beyond relegating file management to individual scientists or research teams and to begin providing them with a robust data-management infrastructure.

Integrating Diverse Data

The types of digital spatial data created, maintained, and produced by USGS scientific teams are diverse. Some data resources are reference-data collections generated by sensors; others, like the Topographic Map Series, are maintained as national collections. Other pertinent types of USGS information include traditional scientific collections, such as reports, publications, drawings, and videos. Although a standard coordinate reference system would be invaluable in integrating various spatial datasets, it may impractical. Nevertheless, there is a need to generate linkages among various spatial and nonspatial data collections (such as spreadsheets, published reports, documents, photographs, and engineering drawings) to support the USGS Science Strategy objectives. Integration of multiple data sources to support analysis (for example, to generate computational mashups) will be critical for an SDI and warrants high priority. The infrastructure will need the fundamental ability to cross-reference and cross-correlate information, facts, assumptions, and methods from different research domains on a global scale.

Analysis and Modeling

Sophisticated data-mining techniques allow scientists to explore spatio-temporal patterns in data. The use of the data to model phenomena such as floods, landslides, and influenza pandemics can provide information helpful in projecting the course of events, which can be useful for prevention and intervention efforts with even just a few hours of lead time. A USGS SDI would enable spatial and nonspatial data from multiple sources and disciplines to be integrated and linked to a unified model of a given system (such as a watershed, ecosystem, or regional climate). The SDI should serve as a powerful Web platform for supporting search and analysis capabilities on a rich corpus of interlinked spatial data and conventional research publications. Moreover, to avoid moving massive amounts of data around, an SDI enables computations to be pushed as close to the data sources as possible.

Security and Confidentiality

Deploying an SDI that is available to a wide user community requires attention to maintaining appropriate privacy, security, and provenance. Documenting provenance is particularly important in the publication and reward incentives to encourage data and model-output contributions. Security and privacy are critical in handling data related to human populations, endangered species, and sensitive environments. A security and confidentiality policy that addresses every stage of the geospatial data stream, from collection to end user, must be in place.

APPLICATION SERVICES TO ENGAGE AND SUPPORT SCIENTIFIC QUESTIONS

The data required to understand and model many of the environmental interactions highlighted in the USGS Science Strategy are massive (petabytes) and growing rapidly. The transition to wired and wireless distributed sensor networks that monitor the solid Earth and oceans is a major driving force for the increase in data volume. Others are the spread of data collection and archiving capabilities around the world, which are interlinked through the Internet, and the increasing use of new Web-enabled capabilities, such as cloud computing and social networking. However, the software tools and applications needed to fully manage and exploit this vast and distributed set of information resources are still emerging. Scientists and computer experts are developing a new generation of software applications to access, visualize, analyze, and interpret large amounts of diverse data for use in research and models that are designed to improve predictions and inform policy and decision-making.

Ideally, an SDI would not only enable the effective management of and easy access to distributed spatial data and information but support the tools and applications needed for research and decision-making, especially in support of the USGS Science Strategy. In the past, spatial data were often incorporated directly into scientific models and decision-support systems to meet specific research and policy needs. However, that led to “stovepiping” of information with poor transparency of data and methods and to numerous inconsistencies between approaches, which then inhibited data integration across multiple problems and discouraged reuse and repurposing of data and derived products. The common data layers, such as those that are hosted by The National Map, are an exception to this stovepiping.

Spatial Data Infrastructure as an Applications Platform

An essential function of an SDI is to provide an overall framework and architecture within which new applications can be developed and integrated. That does not necessarily mean that the SDI would itself need to encompass all those applications. Rather, the SDI could serve as a platform to support a large community (both in and outside USGS) to develop and operate a rich set of application services. In some cases, specific USGS elements or teams may need to build new applications to address specific science questions or meet specific mission needs. Using the SDI as a community applications platform would allow users to take advantage of existing applications that perform functions that they need rather than having to develop their own applications. In turn, when users develop new or improved services, they could more easily make them available to others through the SDI. By opening up the SDI to application developers in other federal agencies or in the much larger geospatial data community, this could yield substantial benefits in shared resources, reduction in duplication of effort, greater innovation, and expanded capabilities.

Spatial Data Infrastructure as a Workflow Platform

Recent work in other fields of electronic science, or e-science, demonstrates the valuable concept of high-throughput workflow processes as a means of processing and analyzing large and complex data resources (Taylor et al., 2007). Some workflow methods are designed to perform routine jobs and utilize the necessary computational protocols to undertake data-centric science. They enable scientists to focus on scientific discovery rather than having to spend effort and resources in routine data-processing. They also permit the development of more sophisticated tools for monitoring and detection (such as alert services for unusual or extreme events). Workflow approaches offer the opportunity to facilitate cross-disciplinary transfer and application of data-processing and analytic techniques in support of interdisciplinary research and problem-solving. Workflows that have been developed to address a specific problem in one field of science often are directly applicable to other, seemingly unrelated fields (Goble et al., 2008). Such cross-disciplinary uses of workflows can help to generate new analytic approaches, improve data quality and timeliness, reduce analysis costs, and speed the transfer of scientific knowledge to applications. Another important benefit of workflow approaches is the detailed record of data transformation and data-processing that is generated, which is a vital part of tracking the provenance of data and is critical for the long-term curation and reuse of data. Documenting and curating the workflows themselves may be an important role for an SDI in capturing geospatial expertise and ensuring the long-term reusability of the spatial data supported by the SDI.

INSTITUTIONAL LEADERSHIP AND CULTURE

It is important for USGS management to appreciate the significance of USGS ownership of a comprehensive, reliable, national dataset. Developing and implementing the crucial elements of an SDI will require sponsorship and support at the highest levels in the USGS. Those in leadership that are specifically responsible for delivering the implementation will need to have the authority and resources to carry the task through. However, the biggest challenge in successfully implementing an SDI will be in facilitating a cultural shift in the approach to data: that data should be viewed as a corporate asset and not held as individual or divisional resources or as liabilities. Instituting the cultural change will be difficult given the natural tension between data-management activities and science research. As previously mentioned, incentivizing researchers for sharing data would be an example of a change in USGS management practices that could affect the necessary cultural shift.

EXTERNAL SERVICE INTEGRATION, PARTNERSHIPS, AND GOVERNANCE

The development and maintenance of an SDI requires collaborative partnerships with cooperating agencies, research organizations, nonprofit organizations, private organizations, and the public. The USGS Science Strategy acknowledges the need for its scientists and policy-makers to partner with other organizations in the sharing of scientific resources. With the creation of an SDI, the USGS is in a unique position to catalyze linkages between national science organizations and government environmental-assessment efforts, the broad research community, and the public. If properly coordinated and managed, the USGS SDI could provide considerable value to government efforts through data management, application of models, and other analysis tools. Through proper coordination and linkages with other, more science-based observatories, the USGS can focus its resources on high-priority science and environmental policy issues that are of national importance.

The public is an increasingly important partner in localized Earth observations through its use of devices such as GPS-enabled mobile telephones and cameras. The use of citizen-scientist field observations—such as observations of plant species and growth and in fish and bird counts—will need to be integrated, and such diverse volunteered information will be challenging but necessary to include for science analysis. A clear policy will need to be established regarding if and how to incorporate a given spatial dataset, whether from citizen-scientists or private industry.

Effective partnerships will require workable and fair policies for the full and open exchange of data in compliance with U.S. federal government policies. In the United States, the White House Office of Management and Budget (OMB) establishes a framework policy for data access and reuse among federal agencies. In recognizing that government-generated information can be a valuable public resource, OMB Circular A-16 Revised (OMB, 2002) states that federal agencies have a responsibility to "collect, maintain, disseminate, and preserve spatial information such that the resulting data, information, or products can be readily shared with other federal agencies and non-federal users, and promote data integration between all sources."

International partnering will be especially challenging. International science collaborators typically express a commitment to data access and data-sharing. For example, Europe has a policy framework with a specific directive that establishes an Infrastructure for Spatial Information in the European Commission (INSPIRE) and a directive on public access to environmental information (Europa, 2003) that requires environmental information to be provided to the public in a timely manner. Furthermore, the European Commission has a policy that enables reuse of public sector information (Directive 2003/98/EC) and more recently in December 2011 issued Directive 2011/833/EU so that documents could be provided on an open basis (Europa, 2011). In the absence of a supporting national policy, legal framework, and good data-management practices, however, such objectives are at risk of not being implemented. Many governance issues can arise when cooperating organizations span different legal jurisdictions or must comply with different organizational data-dissemination requirements. National policies and organizational practices that support these data-access systems will be necessary to ensure that research data flows as intended.

REFERENCES

DAMA, 2011. The DAMA Dictionary of Data Management, 2nd Ed. Data Management International, 260 pgs.

Europa. 2003. Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003. Available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:041:0026:0032:EN:PDF (Accessed June 29, 2011).

Europa. 2011. Commission Decision of 12 December 2011 on the reuse of Commission documents. Available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:330:0039:0042:EN:PDF (Accessed March 17, 2012).

Goble, C., and D. De Roure. 2009. The Impact of Workflow Tools on Data-centric Research. In The Fourth Paradigm: Data-Intensive Scientific Discovery, edited by T. Hey, S. Tansley, and K. Tolle, Microsoft Research. Pp. 137-145.

Goble, C.A., R. Stevens, D. Hull, K. Wolstencroft, and R. Lopez. 2008. Data curation + process curation=data integration + science. Briefings in Bioinformatics 9(6):506-517.

OGC (Open Geospatial Consortium, Inc.). 2004. The OGC – A Unique Organization Offering Unique Benefits. Available online at http://portal.opengeospatial.org/files/?artifact_id=7376 (Accessed June 24, 2011).

OMB (Office of Management and Budget). 2002. Coordination of Geographic Information and Related Spatial Data Activities. Circular No. A-16 Revised, August 19. Available online at http://www.whitehouse.gov/omb/circulars_a016_rev (Accessed June 29, 2011).

Taylor, I.J., E. Deelman, D.B. Gannon, and M. Shields (Eds.). 2007. Workflows for e-Science: Scientific Workflows for Grids. London: Springer.

5

A Roadmap for Spatial Data Infrastructure Implementation

A well-designed spatial data infrastructure (SDI) for the U.S. Geological Survey (USGS) that is based on best practices and focused on the agency's mission will have a high probability of success provided there is adequate planning and execution. Planning and execution will need to entail both a broad vision and steps that address specific needs, resources, and organizational structure in the USGS. On the basis of the experience of other agencies that have implemented SDI programs, the Survey may encounter some challenges as it prepares to implement an SDI, with challenges ranging from general organizational concerns to technical considerations. A roadmap tailored to address issues specific to the USGS can provide the help to ensure that its implementation is successful. This chapter provides a roadmap of both general guidance and specific actions needed for SDI implementation that is based on the committee's findings and its vision for the USGS SDI.

A ROADMAP

A roadmap can serve as a starting point for planning, developing, and implementing a comprehensive SDI program at the USGS. The committee proposes a roadmap for SDI implementation in Figure 5.1 and divides SDI implementation into three cyclical phases: (1) preparation and planning; (2) design, development, and testing; and (3) rollout and refinement. It proposes some general steps in each phase to assist the USGS in implementing an effective SDI. The general steps are described below.

Although the committee proposes a roadmap with phases and steps for the Survey to consider, it was beyond its task and expertise to provide details of

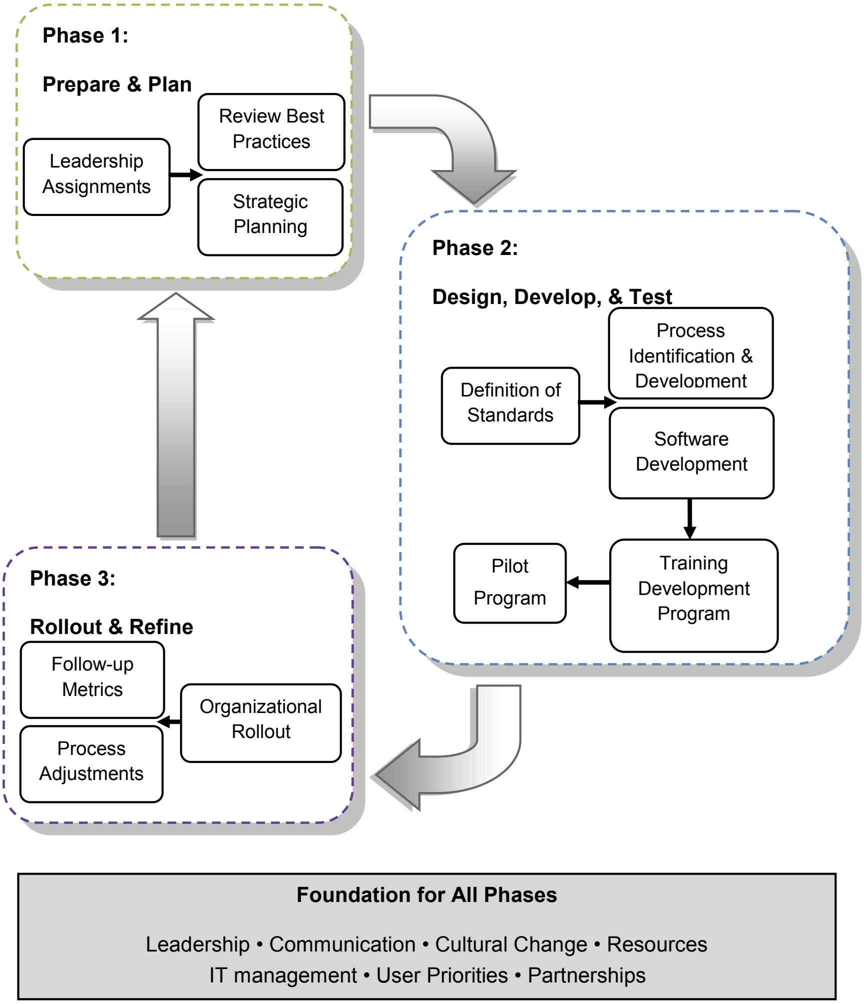

Figure 5.1 Roadmap for implementing the USGS spatial data infrastructure.

specific tasks for carrying out its vision for the SDI. Such a detailed roadmap will need to be developed within the agency and tailored to the specific needs and resource limitations of the USGS, and decisions will have to be made by the organization's leaders who understand its strengths and limitations.

Phase 1: Preparation and Planning

Programmatic preparations and plans are critical in the first phase of SDI implementation. A first step is the appointment of key leaders for envisioning, establishing, and carrying out the vision for an effective SDI. One notable position is that of the SDI program director, who will need to report to the USGS Directorate level for such an agency-wide effort. The core team taking on the challenge of initially scoping and envisioning an SDI will need to consist of talented and skilled professionals that understand the value and purpose of an SDI and are capable of carrying out the vision through effective project planning and management. It may be necessary to attract outside professionals through financial incentives to complement the SDI talent already at the USGS. The work of the core team will include reviewing similar SDIs—such as those of the National Aeronautics and Space Administration (NASA), the National Geospatial-Intelligence Agency (NGA), and the British Geological Survey (BGS)—for best practices and lessons learned at a higher level of detail than those already outlined in this report; determining and defining SDI system requirements on the basis of the six science directions of the USGS Science Strategy and user needs of other agencies, local governments, academe, and the public; determining the organizational structure for the SDI; and identifying goals, establishing timeframes and milestones, and developing performance metrics. Once the initial planning is complete, it will be important to announce a general outline for implementing the SDI program because communication and outreach will play a decisive role in the success of the program.

Phase 2: Design, Development, and Testing

The second phase would entail designing, developing, and testing the SDI program. One of the first steps to undertake in designing the SDI is to identify and define standards that are specific to the USGS mission. Once the standards are determined, the next steps to consider are process identification and development and software development. The former identifies common and documented processes that can enable the SDI to function smoothly across the USGS, and the latter establishes tools for the acquisition, discovery, management, recording, archiving, and sharing of data. With standards, processes, and software in place for an SDI prototype, a training development program would be needed to allow staff to become acquainted with the SDI prototype. The training program will be crucial for providing technical training and support and for building organizational support and buy-in at all levels. After the prototype has been introduced, it will be beneficial to deploy a pilot program on a small scale in the USGS to test how well the pilot program works and to identify and rectify glitches and incorporate these improvements into the prototype and education efforts.

Phase 3: Rollout and Refinement

The third phase in implementing an SDI will need to include a process for rolling out the SDI throughout the USGS and a process for fine-tuning the program. To ensure that people are properly informed and trained, an institution-wide training program will need to be in place before the SDI is unveiled, and retraining will need to be offered periodically for users to understand the system better. The SDI program will need to implement follow-up metrics that feed back into review and planning (Phase 1) to determine how well it is being executed to meet its strategic goals. On the basis of findings gathered with those metrics, there will need to be a process whereby adjustments can be made to serve users and fulfill USGS priorities.

ORGANIZATIONAL CONSIDERATIONS FOR SUCCESSFUL SDI IMPLEMENTATION

Leadership

A designated project official would need to be identified early in the process to oversee SDI implementation. The SDI implementation program will require the efforts of a full-time staff officer, and the USGS could explore whether such an official should be placed at the executive level so that information would be clearly articulated both in and outside the Survey.

An oversight body would also be essential in advising the SDI staff official on strategic goals for SDI development and implementation. Oversight could take the form of a board of experts and could include key external stakeholders inasmuch as it would be important to determine how an SDI would best serve the needs of its users and to solicit input from external stakeholders.

SDI implementation would need to have high priority with USGS leadership. The SDI staff official would need support from all levels of leadership—from senior managers to the USGS director—and would need to be given commensurate authority to develop and deploy standards. Doing that will require a sense of ownership on the part of all levels of staff and a fundamental shift in corporate culture that will be partly driven by bringing employee rewards and incentives in-line with SDI goals.

The relationship between the considerations above and ongoing efforts at data integration and SDI development at the USGS will need to be determined early in the process. The National Map currently forms a central part of the SDI at the USGS, but the USGS also has many other data infrastructures, such as the National Biological Information Infrastructure, and these multiple SDI's would need to be incorporated into the coherent SDI envisioned in this report. The newly created USGS Directorate for Core System Science appears to be a positive step towards this goal and the organizational considerations above may be best carried out under that new structure.

Management

Implementing a complex, organization-wide system will require effective methods in program management. Communicating the goals and progress of SDI development is a critical component of leading and implementing an SDI that allows for transparency and accountability by internal and external parties. It is also a useful way to leverage favorable publicity—publicly committing to developing an improved and comprehensive SDI and publicizing progress during implementation.

It will be important to manage expectations at all levels because implementing an SDI will be disruptive and the process will take time to complete. Thus, it would be helpful to set priorities among deliverables to identify low-hanging fruit in order to build on early successes. To militate against back-sliding, it would be useful to consolidate steps that help with incremental progress.

A corporate information policy will need to be developed to address all aspects in and feeding into the SDI. If there is already an existing policy, it will need to be reviewed comprehensively and revised accordingly to accommodate SDI priorities. Incentives and protocols could be developed and deployed to encourage adherence to policy.

SDI implementation is a multi-tiered, multi-year process, therefore the USGS will need to consider a project-management structure that is flexible in allowing for personnel changes and reassignments. An adaptive-management approach would allow the USGS to respond to an evolving technological, data, and application environment. Furthermore, managers involved in implementing the SDI will need to clearly define current and future job responsibilities for their staff so that written job responsibilities reflect and comply with the corporate information policy. A paradigm shift in management practices will need to occur that embraces integrated team achievement and includes incentives and resources to encourage USGS scientists to share data. As mentioned in Chapter 2, an extra layer of accountability would be to include SDI implementation as a component of annual staff-performance evaluations.

In measuring the progress of SDI development, it would be helpful to develop realistic benchmarks based on the performance of analogous national and international organizations. The USGS could consider an international peer-review process to gauge the progress of SDI implementation.

Cultural Change

Success will take clear vision, clear direction, and, most important, buy-in from all USGS staff ranging from the responsible officer to the individual contributing scientist. Implementation of an SDI is a major program for establishing a geospatial base for USGS professional staff and outside users, and it would need to be viewed as such by the Survey. Individual scientists are typically committed to their research; thus in implementing a new SDI, a similar level of commitment and shared responsibility is needed from researchers to format and share data, adhere to standards, and generate metadata. An example of cultural change that resulted in effective institutional leadership and staff buy-in is the Six Sigma quality program developed by Motorola (see Box 5.1). Such staff participation, adherence to new established processes, and instilling commitment constitute a necessary paradigm shift in the USGS.

Resources

Adequate resources are needed to design, develop, establish, operate, maintain, manage, and lead an SDI. These resources include personnel who have the proper training, knowledge, skills, and motivation including sufficient understanding of hardware, software, data, and support infrastructure (such as space and electric power) to keep up with expectations and technological changes. Financial resources will also need to be sufficient for the SDI program, but it should not be overfunded; overfunding can lead to waste, whereas underfunding can lead to frustration and the inability to reach goals in a reasonable time. For example, in implementing its SDI, BGS budgeted one-third of its expenses to the SDI program (Ian Jackson, personal communication, 2012). A similar percentage of the USGS budget may not be practical or required, but the BGS experience indicates that commitment and a substantial budget are necessary. Finally, SDI implementation will require substantial time and effort.

Partnerships

Partnerships are vital to a successful SDI and can be used to support the USGS Science Strategy. Partnering can help the USGS to align itself with federal and international priorities and with mandates to share scientific data. Benefits of partnerships include resource-sharing, access to high-value information resources that are not available otherwise, and the ability to network shared assets and address problems with information that a single agency usually cannot address (for example, in the case of climate change). In addition to leveraging expertise and experiences in different fields, such collaborations can also enable partners' use of each other's existing data and knowledge assets to be repurposed, creating greater efficiency through sharing of data resources. There is also the ability to mobilize quickly in response to critical studies and the ability to connect to domain users via application programming interfaces. Furthermore, partnering

BOX 5.1

Driving Cultural Change: Implementation of Six Sigma Quality at Motorola

In the 1970s, Motorola embarked on a process to improve product quality by implementing a statistical process control called Six Sigma; it represented the goal of minimizing defects in all aspects of a product's development (Montgomery and Woodall, 2008). Achieving a level of quality of about three to four defects per million opportunities required a substantial culture change in the corporation.

To invoke culture change, Motorola mounted a campaign with the goal of training and garnering buy-in from every employee (Luther Siebert, personal communication, 2011). An essential tenet of institutional culture change is commitment by top management, and the CEO of Motorola embraced the Six Sigma program. The support of senior managers enabled company officers to deploy the initiative, and its implementation was reinforced by tying the success of the initiative with individual performance, ensuring that each employee had a personal stake in quality. For example, scorecards weighed quality-performance achievement, and dashboards were developed for reporting up the chain of command. In addition, a bonus program, in which a high percentage of employees participated, was partially tied to quality.

The Six Sigma process was a multi-tiered campaign. The first step was training the trainer. Departments designated candidates to become experts in statistical process control and designated them as the Six Sigma "black belts" who would also be "change agents". Staffing was determined as a percentage of organization population to ensure consistent saturation of the organization with program experts. Certification of black belts and entry level green belts was determined by length of training, examination completion, and practical project completion. It was conducted through Motorola University, the organization's employee-training program that established a thorough curriculum in quality and quality management.

Senior management was then trained in the Six Sigma process so that they would understand the basic concepts and the benefits of such a change to the corporation. In addition to accountability for company oversight, the role of senior management included removing barriers and making sure that success was recognized. Lower-level executives and managers were also trained in courses that included reference textbooks and other materials. Simultaneously, training teams traveled to all corporate offices and provided extensive staff training courses with emphasis on engineering and manufacturing. Six Sigma training was required for employees, and each department had metrics and goals for Six Sigma performance. For example, managers were periodically evaluated on their performance in achieving Six Sigma goals, and training resources remained available in case a manager required assistance.

As a result of the Six Sigma process at Motorola, there was a gradual but dramatic elimination of defects and an increase in product quality, and Motorola eventually became known for its product quality. In 1988, Motorola was the recipient of the Malcolm Baldridge National Quality Award (given by the president of the United States) for its processes and documentation to determine a high-threshold measure of organizational quality.

and making data publicly available to partners present an opportunity to garner goodwill and political support.

Partnerships require time and effort to develop and maintain and each partner will need to reassess full control and ownership; this may cause participants to shift from their agencies' missions or visions. It will also be difficult initially to establish and agree on shared domains and standards.

There are five categories of partnerships to consider: (1) strategic partnerships (with agencies such as NASA, the National Oceanic and Atmospheric Administration (NOAA), and NSF); (2) data partnerships (with agencies such as the Census Bureau);(3) standards partnerships;(4) academic partnerships; and (5) technology partnerships (with the commercial sector). The USGS will need to develop and expand strategic partnerships with key federal players—NASA, NOAA, NSF, NGA, the U.S. Department of Agriculture (USDA), Department of Energy (DOE), and the Department of Health and Human Services (HHS)—on science data to address common SDI issues; these are strategic in that they provide authority, expertise, and critical information resources for addressing the USGS Science Strategy. The Survey will also need to establish strategic partnerships with selected state, tribal, and local agencies to implement SDI functions that support effective exchange of spatial data among the local, state, and federal levels, perhaps starting with hazards or water data. It will also be important to engage key domestic and international scientific organizations in domain standards development (such as the Consortium of Universities for the Advancement of Hydrologic Science, Inc.) and interoperability standards (for example, OGC and ISO). Many of these partnerships will also increasingly include volunteered geographic information from the public and that is provided to the USGS through partners and from the USGS's own volunteered science data programs. The USGS will need to develop partnerships with the science community, the private sector, and other federal agencies to address technical issues of SDI data quality, value-added data, and online services, perhaps through centers of excellence and working groups.

Federal

Once SDI implementation has begun in the USGS, the SDI will need to be accessible to external agencies and expand to encompass data contributed by external partners. USGS partnerships will be important for creating or augmenting USGS data assets through a collaborative process. Conversely, USGS scientists provide research support for an extensive array of external agencies while maintaining interfaces with state agencies that create statewide data content.

Ongoing USGS research support to external agencies is an example of successful partnerships already under way. The unique combination of biological, geologic, hydrologic, and mapping programs of the Survey provides independent high-quality data, research support, and assessments needed by federal, state, local, and tribal policy-makers, resource and emergency managers, engineers and planners, researchers and educators, and the public. The USGS provides applied research and data to agencies such as the Federal Emergency Management Agency, the Army Corps of Engineers, NASA, NOAA, the U.S. Forest Service, the National Park Service (NPS), and the Bureau of Reclamation. The USGS Contaminant Biology Program collaborates with other agencies and organizations—including the Fish and Wildlife Service, the Bureau of Land Management, NPS, the National Park Service, U.S. Environmental Protection Agency (EPA), the Bureau of Reclamation, the Office of Surface Mining, and other federal, state, and local agencies—in the conduct of its research to improve understanding of the environmental effects of current and emerging contaminants.

Given the multidisciplinary activities of the USGS in support of research and of external agencies, digital data will need to be shared within the USGS and with other agencies. Digital data on various scales will need to be archived so that they are discoverable by researchers and so that they exist in standard formats to keep reformatting to a minimum. Both the USGS and states provide data, and the USGS fills the roles of coordination and dissemination.

States

Cooperation with states will be essential for ensuring that information is available and usable. States are users and providers of USGS data, so it is necessary for data to flow in both directions. Small, densely-populated states, such as Delaware, seldom use USGS datasets, because most of their data needs are on a fine spatial scale. In those small states, local and state governments often create datasets, such as high-resolution elevation databases and air photographs, whereas large states with low population densities, such as Montana, might use USGS datasets. However, state cooperation will be essential for updating the national-level USGS datasets whether a state is small or large.

Other Researcher Communities

The USGS is in the unique position of being a national provider of spatial data and the home for a large cadre of scientists and therefore has the potential to be a more effective national asset of the larger research community. Many research collaborations already exist with USGS scientists; these should continue, and new collaborations should be encouraged. It is important that all USGS data be discoverable, so at a minimum there should be data catalogs that describe the details of the data that currently exist. The government-wide effort to catalog spatial data on-line through Data.gov[1] includes a large number of USGS spatial datasets and this effort provides a basis for a future catalog that is fully integrated into the USGS SDI. Online catalogs of USGS data can foster research collaborations with other research communities and ensure that data are used by as many researchers as possible to answer research questions that are important

---

[1] http://www.data.gov.

to the nation. Therefore, datasets need to be discoverable, and resources will need to be allotted specifically for data-sharing.

Public

The general public is a user of USGS products, and public interest in USGS products has increased as new ways of producing, displaying, and distributing the results of USGS data have evolved. The USGS will need to consider the public's needs as it develops methods to ensure that the SDI provides public access to its vast array of high-quality products. Access can be provided through portals or by making USGS data accessible to others for distribution through other popular access capabilities. The ability to ingest volunteered geographic information from the public will become increasingly important.

Implementation Partnering

To ensure the broadest utility of the SDI, the USGS will need to consider implementing it by using open standards and making use of existing infrastructure. Previous SDIs typically made use of Geographic Information System (GIS) tools to manage and display data, and users of GIS tools typically need training and need to license or buy GIS software. Open standards for description and display of geologic and other Earth science data exist. If open standards are embraced, USGS data can be used and fused with other data for both traditional and new applications; this would result in a much wider use of the data than if they were represented in proprietary or internally developed formats and systems.

Adopting open standards for implementing its SDI will enable the USGS to leverage and partner with other agencies and organizations effectively. For example, NASA, NSF, EPA, and USDA collect or disseminate geospatial data. In addition, adoption of open standards will allow informal or implicit partnerships to be established without the need for formal and often expensive partnerships. For instance, ground-based field data that are collected by the USGS, USDA, or NSF can easily be used to validate or calibrate data collected from the air or from space, such as data on forest height or speciation observed with radar, light detection and ranging, or spectral imagers. In another example, crustal deformation data collected by NSF's Plate Boundary Observatory can be fused with data collected by NASA through uninhabited aerial vehicle synthetic aperture radar and paleoseismic and seismic data collected by the USGS to improve earthquake understanding and forecasting.

TECHNICAL CONSIDERATIONS

Information Technology Management

Information technology (IT) is a fundamental aspect of the research environment, and the concept of an SDI is merging with similar concepts, such as "eScience" and "eResearch", used in the scientific community. Those all point to the important role of defining a “cyberinfrastructure” for undertaking research, collaborating, sharing data and documents, collecting sensor data, and archiving for provenance and long-term preservation. For a science organization like the USGS to support its diverse science workflows would require an evaluation of current IT infrastructure to ensure that it is aligned with the USGS Science Strategy.

The USGS will need to implement a comprehensive, long-term knowledge-management infrastructure that supports end-to-end spatial data management, including the collection, integration, maintenance, and delivery of multidisciplinary scientific data. To carry that out, the USGS would need to

- Identify data assets most critical for supporting the Science Strategy and give high priority to making them discoverable and interoperable.
- Implement structural and syntactic interoperability of USGS knowledge resources, starting with the highest-priority data assets, followed by semantic interoperability.
- Develop an effective data-archiving strategy to ensure a persistent and cumulative knowledge resource.
- Initiate a corporate process to create a comprehensive and consistent inventory and ensure the quality of the data and create standard metadata for it.
- Make “discovered” data structurally (and eventually semantically) interoperable so that they may be shared and integrated with other datasets in the USGS, elsewhere in the United States, and internationally.
- Implement internationally agreed-on data standards.

The Web as a Computing Platform

The USGS lacks a robust data-management structure that can manage and process knowledge on a continental or global scale. Such a system will need to serve as a foundation for the next generation of knowledge-driven services and applications. As mentioned in Chapter 3, several other governments and scientific organizations have recognized the necessity of layering infrastructure and have begun to develop systematic and general approaches to their SDIs with architecture that can scale into the future. The USGS will need to be supported by an adequate cyberinfrastructure that comprises both hardware (computing resources, data centers, and high-speed networks) and software tools and middleware. It will also require an SDI that can help to align the germane scientific research literature with research data. Additional IT developments in cloud-based computing infrastructures, semantic technologies, database-centric computing, and data integration on the Web are platform components that warrant further evaluation as the USGS embarks on its SDI development (See Box 5.2).

It is beyond the scope of this committee to draft a system design for the USGS SDI, but the agency can embark on the process by learning from others. It could establish an IT management team that solicits requirements from the new

Box 5.2

Information Technology Platforms to Consider for the Spatial Data Infrastructure

Cloud-Based Computing Infrastructures — Cloud-based computing infrastructures enable groups to host, process, and analyze large volumes of multidisciplinary data. Such consolidation and hosting help to minimize many organizational and technical barriers that impede multidisciplinary data-sharing and coordination. Cloud hosting can also facilitate long-term data preservation, a task that is challenging for universities and government agencies and is critical for conducting longitudinal experiments. Provisioning key core geospatial datasets on the cloud would increase the value of the data and facilitate exploitation by a large scientific community. Scientific organizations that currently manage their own scientific cyberinfrastructure will probably turn to cloud-based hosting as a more efficient alternative. Cloud-based computing infrastructures (such as that of Amazon.com) are being complemented by a new generation of data-intensive computing services, such as Google's MapReduce, Apache's™ Hadoop™, and Microsoft's Dryad. However, a "smart" SDI still does not exist for automatic data discovery, acquisition, organization, analysis, and interpretation for information that is available on the Internet or on individual researchers' hard drives.

Semantic Technologies — Semantic technologies will need to be considered as an SDI creates necessary links between various disciplines that use different jargon if information is be interlinked as part of a global network of facts and processes. The use of semantic technologies is gaining traction in scientific and engineering communities, including the life sciences, health care, ecology, marine science, and computer science. With increasing volume, complexity, and hetero-

generation of researchers and from external partners and research organizations. The IT management team could also solicit expertise and best practices from existing commercial software, hardware, and service vendors.

Enterprise Data Management

The USGS will need to implement enterprise data-management techniques in order to

- Create a robust data-management structure that enables individual scientist to make their data interoperable.
- Generate linkages between spatial and nonspatial data collections (such as spreadsheets, published reports, documents, photographs, and engineering drawings).
- Link spatial and nonspatial data to a unified model of a given environmental system (such as watershed, ecosystem, and regional climate).
- Direct substantial attention to maintaining appropriate privacy, security, and provenance.

geneity of data resources, scientists are turning to semantic approaches (in the form of ontologies—machine encoding of terms, concepts, and relations among them) to interconnect the different sources of data. The interoperable exchange of scientific data requires scientific communities to explore common vocabularies for capturing facts and information specific to their domains of expertise. In that way, the world's data become joined together on the Web as a common interconnected database.

Database-Centric Computing — In database-centric computing, computations are brought to the data rather than the data being brought to the computations. This will become increasingly important as datasets increase in size beyond terabytes. The volume of data that Earth scientists will work with can be daunting, and new approaches to scientific discovery, collection, and analysis will increase the volume of data. A new generation of parallel database systems from Oracle, Teradata, and others leverage "MapReduce" as an effective data-analysis and computing paradigm that performs computations as close to the data as possible (Dean and Ghemawat, 2004).

Data Integration on the Web — SDI architects will need to determine the best way to aggregate huge amounts of semantically rich information and consider how the resulting information is generated and analyzed. It will be important to consider such requirements now to support higher-level knowledge-generation processes of the future. Cloud computing services currently focus on offering a scalable platform for computing, and future services will need to be built around the management of knowledge and reasoning over it. Services such as OpenCyc (www.opencyc.org), Freebase (www.freebase.com), and Wolfram/Alpha (www.wolframalpha.com) demonstrate how facts can be recorded in such a way that they can be combined and made available as answers to users' questions.

Priorities for Users

Internal Use and Needs

Internal use–only databases that can be discovered and shared effectively within the USGS will continue to be needed and the proprietary nature of the information means that such databases will not be discoverable by external parties until there is official agreement about the broader accessibility of data to external agencies and the public. Given the multidisciplinary research activities of the USGS, digital data will need to be made available on various scales, archived so that they are discoverable by USGS researchers, and exist in standard formats so that reformatting is kept to a minimum when they are made accessible to outside users.

Application Services to Engage and Support Scientific Efforts

An SDI will need to create a foundation for a more open approach to application services that would promote greater transparency, easier integration,

Box 5.3

Key Application Services

Data Transformation — Data transformation services convert data between different formats and coding systems, between different spatial coordinate systems and projections, or between different levels of aggregation or disaggregation. Such services include certain types of image processing services that manipulate images from remote sensing instruments or aerial photography, geoparsing and geocoding services that convert location-based information (such as place names, addresses, and administrative and postal codes) to and from spatial coordinates, gridding algorithms that convert vector-based to raster-based data formats, and semantic translation services that facilitate interpretation of data between languages, disciplines, and applications.

Data Integration — Data integration services typically build on data transformations to support the assembly of data from different sources (for example, instruments, models, and disciplines) into a combined dataset or database or other derived product, such as a map. They can include linking services to identify data that have overlapping geographic coordinates or similar geographic features or characteristics, alignment services that adjust geometric models to improve spatial matching between different spatial datasets or images, and filtering services that select data for inclusion or exclusion on the basis of specified constraints and data characteristics.

Spatial and Statistical Analysis Spatial and statistical analysis services facilitate assessment of possible relationships within and between geographic distribu-

and greater reuse and repurposing of distributed data and services. Current approaches to SDI development, such as the Open Geospatial Consortium (OGC) Web Services Architecture (Whiteside, 2005), support an array of interoperable services focused on data transformation, integration, analysis and statistics, modeling, and visualization (see Box 5.3).

The SDI will need to adopt and implement an open Web services architecture and take a leadership role in establishing a framework for collaborative applications development and operations. For that to become a reality, the USGS will need to play a leadership role in establishing key elements of a community applications platform, including

- Specific open standards needed for data and service interoperability.
- Key reference datasets and parameters that would facilitate the interoperability and integration of data and services.
- Coordination mechanisms for ensuring development and implementation of new standards and reference datasets when and where needed.
- Mechanisms for quality control, measurement, and improvement.

tions, often on the basis of integrated data. A major development over the last several decades has been that of image classification methods that focus on characterizing and interpreting imagery data, for example, from satellites and aerial platforms. More recently, a new generation of statistical analysis methods tailored to spatial data has emerged that can provide new insights about spatio-temporal phenomena and processes and new methods for transformation, visualization, and prediction.

Modeling Services — Modeling services usually apply algorithms based on a combination of theory and statistics to generate estimates of past or current environmental conditions and changes and to project future events, trends, or risks. Models that have significant spatial dimensions or elements are increasingly prevalent in virtually all environmental and social science disciplines and many engineering and public health fields. The USGS has pioneered the development of diverse spatially enabled models, such as those related to earthquake prediction, hydrological resource management, and land-cover change.

Visualization — Visualization services not only support hypothesis generation, analysis, and modeling but provide a mechanism for scientists to use in communicating their findings to other scientists, applied users, and the public. An example of an online visualization service that is available at the USGS is the Disease Maps Web site (http://diseasemaps.usgs.gov). It is a simple tool that allows users to see the spatial distribution of wildlife and zoonotic diseases (such as West Nile virus) in different years at national or state levels. Another USGS Web site, WaterWatch (http://waterwatch.usgs.gov/), facilitates spatial data visualization and has daily local-level stream flow data, which can also be accessed through Google Maps. It displays real-time stream flow and is able to compare with historical stream flow by station.

The Spatial Data Infrastructure as a Workflow Platform

Over the last decade, various scientific programs have begun to incorporate workflow methods into their best practices. Most of the programs highlighted in Chapter 3 use workflows to collect raw observation data and make them available to thousands of researchers worldwide through specialized analytical and visualization tools (such as National Center for Atmospheric Research and National Science Foundation collaboratories). Workflow approaches are being developed by the OGC through its Open Web Services testbeds and through the Architecture Implementation Pilot of the Global Earth Observing System of Systems.

The USGS will need to consider implementing high-throughput workflow processes as a means of advancing the overall scalability of the SDI in support of the USGS Science Strategy and the SDI's long-term role in geospatial knowledge capture, preservation, and reuse. Workflow techniques will need to become an essential technology layer in an SDI—one that enables research on a large scale by automating complex data preparation and analysis pipelines and by facilitating cross-disciplinary analysis, visualization, and predictive modeling. Providing computing, analytics, visualization, and other application processes as an SDI layer moves them closer to the data and makes it possible to leverage unprecedented power and bandwidth to understand and predict the interaction of natural and human processes better.

CLOSING REMARKS

The intent of this roadmap is not to be prescriptive regarding SDI development and implementation but to identify keys to success on the basis of lessons learned from other SDI implementation efforts. The committee recognizes that there are many feasible ways of implementing an SDI program successfully, but all will require that an SDI program be properly defined, led, and supported. The committee believes that such an effort can be best framed by the phrase "discover and share for the long term", and it hopes that this phrase can become the mantra for spatial data handling throughout the USGS.

REFERENCES

Dean, J. and S. Ghemawat, 2004. MapReduce: Simplified Data Processing on Large Clusters, OSDI, doi: 10.1145/1327452.1327492.

Montgomery, D.C. and W. H. Woodall 2008. An Overview of Six Sigma, International Statistical Review, 76:329-346.

Whiteside, 2005. OpenGIS Web Services Architecture Description, OGC 05-042r2, 22pgs.

Appendixes

A

Committee and Staff Biographies

COMMITTEE MEMBERS

ROBERT P. DENARO (Chair) is vice president of Advanced Driver Assistance Product Marketing at Nokia Location & Commerce. Mr. Denaro came to Nokia through its purchase of NAVTEQ, a company that specialized in digital roadmap data and services for navigation, and he lead NAVTEQ's Advanced Driver Assistance Systems business. Mr. Denaro joined NAVTEQ from Rand McNally & Company, where he was senior vice president and general manager of Global Business Solutions, responsible for business-to-business applications and consumer technology products and services in mapping and routing. Before joining Rand McNally, Mr. Denaro was vice president and director of Motorola's Consumer Telematics Products, a division that he launched after heading the company's global positioning system (GPS) business for 5 years. Earlier in his career, Mr. Denaro launched Trimble Navigation's Fleet Management and Vehicle Tracking Division and was co-founder of TAU Corporation, producer of the first commercial differential GPS systems. He started his career in the U.S. Air Force, where he served for 9 years, initially working on research, development, and flight testing of the first cockpit digital map displays and ultimately carrying out research, development, and field testing as an Air Force captain at the Navstar GPS Joint Program Office. Mr. Denaro is chair of the Department of Transportation Intelligent Transportation Systems (ITS) Program Advisory Committee and was previously a member of the National Research Council Mapping Science Committee. He is a member of the Transportation Research Board ITS Committee and a member of the International Cartographic Confernce 2017 U.S. Organizing Committee. Formerly, he was a member of the Board of Directors of the Intelligent Transportation Society of America, a Policy Board director of the 511 National Traveler Information Number Deployment Coalition, vice president of the Institute of Navigation, vice chairman of the U.S. GPS Industry Council, and lecturer for the NATO Advisory Group for Aerospace Research and Development. He holds an M.S. in electrical engineering from the U.S. Air Force Institute of Technology, an M.S. in systems management from the University of Southern California, and a B.S. in engineering sciences (astronautics) from the U.S. Air Force Academy.

GEORGE H. BRIMHALL (NAE) is a professor of geology in the Department of Earth and Planetary Sciences of the University of California, Berkeley, where he has taught and conducted research for nearly 33 years. Previously, he taught in the Department of Earth and Planetary Sciences at Johns Hopkins University and worked as a project and underground mine geologist for the Anaconda Company. Dr. Brimhall's research interests include digital field mapping, exploration and mining geology, ore deposit genesis and geochemistry, surface process geochemistry, and mineral resources issues. He has been active in the Society of Economic Geologists and the Geological Society of America; he was associate editor of the Geological Society of America Bulletin from 1992 to 1995. Dr. Brimhall was elected to the National Academy of Engineering in 2001 and received the University of California, Berkeley Noyce Distinguished Teaching Award in 1999. He holds a Ph.D. in geology from the University of California, Berkeley.

ROBERT S. CHEN is director and senior research scientist at the Center for International Earth Science Information Network (CIESIN) of Columbia University. He served as CIESIN's deputy director from 1998 to 2006 and as CIESIN's interim director from 2006 to 2007. Dr. Chen is also the manager and co-principal investigator of the Socioeconomic Data and Applications Center, a data center in the National Aeronautics and Space Administration Earth Observing System Data and Information System. He is secretary-general of the Committee on Data for Science and Technology of the International Council for Science and an ex officio member of the National Research Council Board on Research Data and Information. He has contributed to activities of the Intergovernmental Panel on Climate Change (IPCC) for more than a decade and serves as an ex officio member of the IPCC Task Group on Data and Scenario Support for Impacts and Climate Analysis and co-manager of the IPCC Data Distribution Center. Dr. Chen received his Ph.D. in geography from the University of North Carolina at Chapel Hill and holds master's degrees in technology and policy and in meteorology and physical oceanography from the Massachusetts Institute of Technology (MIT). His undergraduate degree was in earth and planetary sciences from MIT.

ANDREA DONNELLAN is a geophysicist at the National Aeronautical and Space Administration (NASA) Jet Propulsion Laboratory (JPL) and a research professor at the University of Southern California. Dr. Donnellan integrates satellite technology with high-performance computer models to study earthquakes, plate tectonics, and the corresponding movements of Earth's crust. She is NASA's Applied Sciences Program area lead for natural disasters and principal investigator of NASA's QuakeSim and other projects. Dr. Donnellan has also been the project scientist of a mission to study natural hazards, ice sheets, and ecosystems and deputy manager of the JPL Science Division. She has conducted field studies in California in the region of the Northridge earthquake, in the Ventura basin, and on the San Andreas fault. She has also carried out field work on the West Antarctic Ice Streams, in the Dry Valleys, and in Marie Byrd Land of Antarctica; on the Altiplano of Bolivia, in Mongolia; and on Variegated Glacier in Alaska. Dr. Donnellan received the Presidential Early Career Award for Scientists and Engineers in 1996, the Women in Aerospace Award for Outstanding Achievement in 2003, the Women At Work Medal of Excellence in 2004 and was the MUSES of the California Science Center Foundation Woman of the Year in 2006. She has held a National Research Council Postdoctoral Fellowship at NASA Goddard Space Flight Center and has been a visiting associate at the Seismological Laboratory at the California Institute of Technology. Dr. Donnellan has a B.S. in geology from the Ohio State University, an M.S. in computer science from the University of Southern California, and M.S. and Ph.D. in geophysics from the California Institute of Technology.

MICHAEL EMCH is associate professor of geography, a Fellow at the Carolina Population Center, and an adjunct associate professor of epidemiology at the University of North Carolina at Chapel Hill. His expertise is in infectious disease ecology, neighborhood determinants of health, and geographic information science applications in public health. He leads the Spatial Health Research Group, which conducts research that explores spatio-temporal patterns of disease, primarily infectious diseases of the developing world (www.unc.edu/depts/geog/spatialhealthgroup/). Disease patterns are studied with a holistic approach by investigating the role of natural, social, and built environments in disease occurrence in different places and populations. Diverse statistical and spatial analytical methods are informed by theory from the fields of medical geography, epidemiology, and ecology. Those theories and methods are used to examine diverse topics, such as the role of population–environment drivers in viral evolution, how social connectivity and spatial connectivity simultaneously contribute to disease incidence, and the use of environmental indicators to predict disease outbreaks. Dr. Emch holds a Ph.D. in medical geography from Michigan State University, M.A. in geography from Miami University, and B.A. in biology from Alfred University.

IAN JACKSON is the chief of operations at the British Geological Survey (BGS). In 2000–2007, Mr. Jackson was the director of information at BGS. He is a member of a European Commission (EC) team drafting regulations for the new EC spatial data infrastructure directive, Infrastructure for Spatial Information in the European Community (INSPIRE), and serves on the council of an International Union of Geological Sciences commission. From 1997 to 1999, he was the project manager of a major European Union-funded project to create a European geoscience metadata service. Mr. Jackson has worked for BGS for over 35 years, initially on mineral assessment programs in the United Kingdom and overseas and later as a field geologist undertaking applied geologic mapping in the North-East England coalfield. Use of relational database and computer-aided design systems to handle the large borehole and mine plan datasets associated with these projects led to his appointment as the manager of the BGS Digital Map Implementation project in 1990. That was followed by responsibility for BGS Information Systems. During his career with BGS, he has also undertaken geoscience information systems consultancy in Canada, Australia, South America, and Europe. He was responsible for the development of the UK digital geologic map database and closely involved in designing the BGS program for 3D modeling. Mr. Jackson is a graduate of the University of Newcastle, UK.

JOHN A. KELMELIS is a professor in the School of International Affairs of Pennsylvania State University and an affiliate professor in the Department of Geography. Previously, Dr. Kelmelis served as senior counselor for Earth science in the office of the science and technology adviser to the secretary of state, where he provided policy advice to the White House, the Department of State, and other high-level government entities on geology, hydrology, biology, geography, and related sciences and technologies in establishing and executing U.S. foreign policy. He concurrently served as senior science adviser for international policy in the office of the director of the U.S. Geological Survey (USGS), where he served as principal staff adviser on incorporating science into international policy. Dr. Kelmelis has coordinated the USGS Global Change Research Program, directed the White House Scientific Assessment and Strategy Team, managed the U.S. Antarctic Mapping Program, and conducted research on many geographic scientific topics. From 1997 through 1999, he served as the chief scientist for geographic research at the USGS, where he provided research and guidance on infrastructure resources in the United States (such as drinking water, abandoned mine lands, urban hazards, and ecosystem restoration in South Florida, the Chesapeake Bay, and San Francisco) and international issues and research. From 1999 to 2004, he served as chief scientist for geography, providing scientific leadership for The National Map, land remote sensing, and geographic analysis and monitoring programs. He is active in professional societies, including the American Geographical Society and the Association of American Geographers. Dr. Kelmelis has provided scientific and technical leadership to various national and international committees, including the Planning Committee of the Global Dialogue on Emerging Science and Technology 2008 (in Africa), the U.S. African Command Transition Team, and the U.S. Department of State Working Group on Populations at

Risk. He has led official U.S. delegations to several countries and has worked on and participated in many UN events. His current research addresses the linkage of scientific findings to the policy process. Dr. Kelmelis received a B.A. (magna cum laude) from Central Connecticut State College, an M.S. from the University of Missouri–Rolla, and a Ph.D. from Pennsylvania State University.

XAVIER R. LOPEZ is director of Oracle Corporation's Spatial Technologies. He is responsible for incorporating location and semantic technologies across Oracle's database, application server, and business applications. He has 18 years of experience in geospatial technologies. He has been active in numerous academic and government research initiatives on geographic information; has served on the National Academies Transportation Research Board and on the boards of directors of the Geographic Information Technology Association and the International Geographic Information Foundation; and was editor of the Journal of the Urban and Regional Information Systems Association. He is author of a book on spatial information policy and of over 100 scientific and industry publications pertaining to spatial information technology. He is the recipient of Fulbright, Ford, and University of California, Berkeley postdoctoral fellowships. He holds a Ph.D. in spatial information engineering from the University of Maine and an M.A. in urban and regional planning from the Massachusetts Institute of Technology and was an independent major in geography at the University of California, Davis.

DENNIS OJIMA is senior research scientist at the Natural Resource Ecology Laboratory of Colorado State University (CSU), where he was interim director in 2005–2006. He is also a senior scholar and codirector of mitigation programs at the H. John Heinz III Center for Science, Economics, and the Environment and assistant professor in the CSU Rangeland Ecosystem Science Department. His current U.S. research contributes to the North American Carbon Project. His research is in global change effects on ecosystem dynamics and regional climate change assessment for the Central Great Plains and in international efforts in Central Asia, Mongolia, and China. His research with the Chinese Academy of Sciences includes development of regional carbon management. Dr. Ojima is also member of the U.S. National Committee for the Scientific Committee on Problems of the Environment and was a member-at-large of the Governing Board of the Ecological Society of America (2005–2007). He received a B.A. in botany from Pomona College, an M.S. in botany from the University of Florida, and a Ph.D. in rangeland ecosystem science from Colorado State University.

BRIDGET R. SCANLON is a senior research scientist in the Bureau of Economic Geology of the Jackson School of Geosciences of the University of Texas at Austin. The primary objective of her research group is to assess sustainability issues with respect to water resources in the context of climate variability and land-use change in semiarid regions. Her group is working in the southwestern United States, India, and China and collaborating with groups in West Africa and Australia. Her research focuses on evaluation of the effects of land-use change on groundwater resources; quantification of groundwater recharge on the basis of soil physics, environmental tracers, and numerical simulations; assessment of paleoclimate effects on groundwater recharge in semiarid and arid regions; and evaluation of groundwater contamination related to geogenic and anthropogenic sources. Dr. Scanlon has participated in focus groups on global recharge in the International Atomic Energy Agency and has served on National Research Council committees on radioactive-waste disposal and integrated observations in the hydrologic sciences. Dr. Scanlon received a B.S. in geology from Trinity College, Dublin (Ireland), an M.S. from the University of Alabama, and a Ph.D. from the University of Kentucky.

NATIONAL RESEARCH COUNCIL STAFF

MARK D. LANGE (Study Director) is a program officer with the National Research Council's Board on Earth Sciences and Resources and director of the Geographical Sciences Committee. He is a geomorphologist and has expertise in river and coastal processes, Geographic Information Systems, and science policy. He was a Tyler Environmental Fellow and a U.S. Congressional Fellow where he managed federal environment and natural resources policy for a member of Congress. He is a member of the Association of American Geographers and the American Geophysical Union and holds a Ph.D. and graduate certificate in geographic information sciences from the University of Southern California.

PEGGY TSAI is a program officer with the National Research Council's Board on Agriculture and Natural Resources. She joined the National Research Council in 2004 and has worked on various studies ranging from agricultural biotechnology to animal health to international agriculture. She began her work with the National Academies as a Christine Mirzayan Science and Technology Policy Fellow. Ms. Tsai received an M.A. in science, technology, and public policy from George Washington University and a B.S. in microbiology and molecular genetics with a double major in political science from the University of California, Los Angeles.

NICHOLAS D. ROGERS is a financial and research associate with the Board on Earth Sciences and Resources, National Research Council. He received a B.A. in history, with a focus on the history of science and early American history, from Western Connecticut State University in 2004. He began working for the National

Academies in 2006 and supports the Board on Earth Sciences and Resources on a wide range of areas from earth resources to mapping science.

ERIC J. EDKIN is a senior program assistant with the National Research Council's Board on Earth Sciences and Resources. He began working for the National Academies in 2009 and has supported the board on a broad array of earth resource, geographic science, and mapping science projects.

B

Presentations to the Committee

The following individuals provided information and advice to the committee:

M. Lee Allison
Arizona Geological Survey

Phyllis Altheide
U.S. Geological Survey

Gerald Bawden
US Geological Survey

Gregory Black
National Geospatial-Intelligence Agency

Steve Blanchard
U.S. Geological Survey

David Briar
U.S. Geological Survey

Patrick Comer
Natureserve

Daniel Cole
Smithsonian Institution

Bryant Cramer
U.S. Geological Survey

Kari Craun
U.S. Geological Survey

Mark DeMulder
U.S. Geological Survey

Kevin Dennehy
U.S. Geological Survey

Mike Dettinger
U.S. Geological Survey

Robin Fegeas
U.S. Geological Survey

Mike Frame
U.S. Geological Survey

John Gardenier
Centers for Disease Control and Prevention, Retired

Marty Goldhaber
U.S. Geological Survey

Linda Gunderson
U.S. Geological Survey

John Haines
U.S. Geological Survey

Kevin Hope
U.S. Geological Survey

Bruce Jones
U.S. Geological Survey

Susan Carson Lambert
Earthworks, LLC

Todd LaPorte
University of California, Berkley

Katherine Lins
U.S. Geological Survey

Mike Mahaffie
National States Geographic Information Council

David Maidment
University of Texas at Austin

Gerald McMahon
U.S. Geological Survey

Sara McLafferty
University of Illinois at Urbana-Champaign

Douglas Nebert
Federal Geographic Data Committee

Bruce Quirk
U.S. Geological Survey

Roger Sayre
U.S. Geological Survey

David Schimel
National Ecological Observatory Network

David Soller
U.S. Geological Survey

Larry Sugarbaker
U.S. Geological Survey

Ione Taylor
U.S. Geological Survey

E. Lynn Usery
U.S. Geological Survey

Leslie Wyborn
Geoscience Australia

C

On-Line Questionnaire

The questionnaire below was developed and used by the committee to gather information from key individuals and organizations involved in spatial data infrastructure development and implementation. Respondents submitted their answers to these questions using an online form.

On-line Questionnaire: Spatial Data Enabling USGS Strategic Science in the 21st Century

At the request of the U.S. Geological Survey, the National Research Council is conducting a study that will: (1) identify existing knowledge and document lessons learned during previous efforts to develop Spatial Data Infrastructures (SDI) and their support of scientific endeavors; (2) develop a vision for optimizing an SDI to organize, integrate, access, and use scientific data; and (3) create a roadmap to guide the USGS in accomplishing the vision within the scope of the USGS Science Strategy. For the committee's full statement of task, [click here].

Because the committee cannot hear from all the individuals and organizations that have valuable experience and ideas on this topic during its few scheduled meetings, the committee seeks your help in the form of written contributions on the following set of questions.

Based on the last five years working with spatial data infrastructures:

1. What has worked well?
2. What has not worked?
3. What are the major challenges (technical, organizational, cultural, policy, financial)?

4. What would you do differently?

5. In what domain(s) are your data (e.g. biological, hydrologic, cultural, etc.)?

6. What is your vision for an SDI to meet the needs of the USGS Science Strategy?

Comments received by December 6, 2009 will be considered at the committee's next meeting (December 10-11, 2009). However, the committee welcomes all input through February 2010. Please note that any written comments submitted to the committee (whether by mail, e-mail, fax, or this comment form) will be included in the study's public access file.

D

Questionnaire Responses

In formulating the lessons learned described in Chapter 3, the committee interviewed experts, heard presentations, and created an online questionnaire that was distributed to the broader community. The online questionnaire received responses from national and foreign individuals who had an interest in spatial data infrastructure and proved to be a rich source of opinions from users, planners, and policy-makers. The responses are grouped below by issue.

Lessons Learned Questionnaire Responses

Issue	Lesson
Standardization	Full acceptance of the organization of the need for rigid standardization of its data and information products to agreed international standards
	Organizational commitment to internationally agreed metadata standards
	[Successful organizations] work within the community to make/improve standards…
	[Challenges] Let a thousand flowers bloom – In the past year we are witnessing a global convergence in thinking on how spatial data should be integrated. This is occurring in all technical fields as well as in the library sciences. There are scores (hundreds??) of projects moving forward and we risk duplicating effort or worse, creating divergences in standards, protocols, processes and methods that will make later data integration much more difficult or effectively impossible.
	[Challenges] The amount of data. Data from countless agencies and data formats, gathered with varying standards, documented with varying accuracy and amount. It can be difficult to get many people and agencies to agree on standards and work together unless some plan and mutual benefit is in place.
	[Challenges] Technical - appropriate standards (metadata, vocabulary)

Issue	Lesson
	[Does not work] Implementation of OGC standards suffers from performance issues. There is lack of leadership. [Worked well] The development and adoption, though limited, of open standards for geospatial Web Services is a key capability that promotes interoperability. This, in turn, allows for neighboring or overlapping SDIs to work well together without special agreements or translators.
Data	Scientists must make available data that underpin knowledge products Federal data are created to some minimum achievable standard Census using local roads data Landsat 7...we've done the best we could. We need the continuity mission now. Everything online [what has worked well] - We have seen a paradigm shift in thinking about data and especially spatial data, in the last four years or so. Prior to this, data owners were generally unwilling to share their data for fear of them being misused, losing control over the data, of them being used to scoop the originators of the data, or others getting credit for the data. In the past few years however, there is widespread recognition of the value of making ones data more widely available for others to use. This coincides generally with the release of Google Earth and the rapidly growing expectation that everything, including scientific data, should be readily available online at no cost. [what has not worked well] Mandated uniformity – Everyone has invested vast resources in their databases and spatial data infrastructure. So, when discussion of data integration came up, the fear was that we would all be forced to convert what works for us, into some format (and operating system, and server configuration) that would be imposed. Many data providers have custom systems and applications that will be prohibitively expensive to re-do. Plus, with cyberinfrastructure in a constant state of change, how could we adopt a system that would not be obsolete before it was implemented. [Do differently] Put some good people into cataloging existing reports and data sets. Build better metadata tools. Make management accountable to publish geospatial data from all projects. Make all projects identify spatial data results, plan for, and publish them before project is considered complete. [Challenges] Geography has spent the last decade trying to justify their existence, rather than meeting customer needs. Many of Geography programs have become largely irrelevant, with some very notable exceptions, all of which are long-term commitments of resources focused on data content, such as NED, NHD, NLCD.

Issue	Lesson
	[Challenges] The major challenge is to establish a NSDI organization that is viewed by authors as a robust clearinghouse of their spatial datasets. Authors should be glad to submit their data and should be delighted that others will have easy access, instead of having to handle 'data requests' every time someone wants it. The current cultural views the NSDI as an unfunded mandate, with a lot of hassles to submit data, with very little benefits in return. Trying to establish a new organization - or revamping the existing one - is always extremely difficult, and the establishment of this one is even more difficult because of the lack of the basic understanding of the true value.
	[What has not worked?] US Topo is a solution looking for a problem. The focus should be more on content, rather than packaging. GeoPDF may satisfy a certain niche, but without excellent content, it serves little purpose.
	[has worked] Several national seamless datasets have been very successful, including NED, NHD, NLCD, and NHDPlus. These are providing very useful data that is nationally consistent, well organized, and easy to access.
	[not worked] The WRD NSDI node is just a tabular list of 646 datasets - some datasets are listed by theme, such as, ag, aquifers, etc; a lot are listed with obscure names, such as , darea, diffus, etc.; some are listed by OFR #, by SIR # and by WRIR#. How in the world can anyone find what their [sic] after? We need a better way of assigning searchable 'key words' to the datasets and tools that can search and retrieve datasets that meet a specified query.
	[not worked] Main sticking points are access to updated, high-quality satellite-derived imagery, and access to sufficient field-based observations of vegetation (i.e., we need 500,000 current georeferenced samples – with sufficient vegetation composition and structure documented - maintained and accessible) to support map development and accuracy assessment.
	[Challenges] technical challenges come mainly from a lack of certain critical data sets required to develop robust spatial models. We work across local/regional/national/ continental scales, so access to data that are standardized across these scales presents the greatest challenge.
	[Domain] Our work is centered on biological data, including the characterization and assessment of ecological systems and habitats for species of concern. But in order to address this domain successfully, we rely on a wide range of non-biological data inputs, such as imagery (of varying types and resolutions), digital elevation, synthesized climate data (past, current, future), surficial geology, soils, surface drainages, wetland location, hydrography, land use, land ownership, and land use policy.
	[worked well] The understanding that data needs to be consistent and of known quality so that decisions can be more easily made on how the data can/should be used.

Issue	Lesson
	[didn't work] We are severely lacking in field-based observation data for about 10,000 plant and animal species in the U.S. that are of conservation concern. With these observations and a foundation, habitat models can be developed to apply to many forms of environmental decision making processes. [not worked well] Data content standards and schemas have not been widely adopted for key base data sets. This makes the exchange and co-development of data, at least in the US, more difficult.
Metadata	[what works] State and local grants for data creation and metadata training. [what did not work well] Difficult metadata standards – It is such a challenge to generate FGDC-compliant data that many individuals, programs, and agencies, do not even try, and instead have their own internal systems. The head of a geoscience division in a large federal agency told us that, yes, the USGS standards are nice, but they could not invest the time and resources to meet them, so they developed their own in-house way of doing things, because they had to get things done. The chief scientist for one of the world's largest multinational oil company described to me how they were scraping their fourth internal attempt to create a company-wide data base, after spending millions on it. Before they could make significant progress, various offices and branches had gone off in other directions because they could not wait, and they had their own needs to address. [What has not worked?] It has been too hard to develop metadata [for water data], and the process is usually left to the end of a project and then not done. Management bears much of the blame for this, as they have not enforced the requirement to publish metadata, even though it has been required by executive order since 1994. Although much of our work is supported by geospatial data, much of the data supporting that work is never published because of this. [What has worked?] Publishing metadata through the NSDI node, when it is done, does work fairly well. I can find datasets I published in the past easier with Google than I can find them on my hard drive or backup media. [not worked well] Outside the Federal government: metadata collection and dissemination are often non-existent.
Distribution, Serving	[what works] The National Map accepting state data and using it on … servers Having the National Map portal at EDC is very useful

Issue	Lesson
Distribution, Serving Continued	Work with Google [private industry in general] to make it even more useful and friendly. Market expansion for geospatial data could be exponential if the tool people sue to access the data is easy enough to use and does enough analysis. (all we need is overlay analysis and we are in the GIS business online) [What has worked well] Market driven solutions – the pervasive adoption of free online visualization tools for spatial data (e.g., Google Earth, Google Maps, Bing, ArcGIS Explorer, etc) by even the smallest retailer and organization has made it the norm to share (and promote) your data online. Also, data providers find these free tools of tremendous value, so that when more data are made public, more tools and applications become available to use them. [Do differently] Make the USGS SDI more powerful by giving it better search and data discovery mechanisms. Requires support from the top, a budget, and a dedicated team - not just USGS but from all agencies. Standards need to be defined. Robust software tools need to be developed to create standard metadata, and to provide the ability to search all NSDI nodes. This really requires a lot of coordination from all agencies to enforce the standards so 'searches' have the potential to retrieve all data that meets a query. [What has not worked?] The National Map is [ineffective]. Viewer 2 is better, but it still lacks compelling content. Much of the current content is not much better than we had in the early 1990's from TIGER/Line and 100K DLGs. [worked well] New map services are providing data access in new ways. These include NWIS web services, NWIS Mapper, real-time earthquake maps, and StreamStats. StreamStats provides analytical services rather than just raw data, and is a good example of how far the web service model can be taken. We were able to work successfully with USGS to access global climate, digital elevation, lithology, and other data sets for critical new advancements in classification and mapping of terrestrial ecosystems. Much new work has been advanced in the U.S., Latin America, Canada, and Africa, due to the accessibility of key data sets from USGS. [Worked well] The idea of centralizing data so that it can be easily accessed. [Did not work well] Up until the National Map, we had access to localized lists of available spatial data from both the USGS and other agencies. You had to read through title upon title to find things and getting a grip on what was available for a given geographic region was difficult. You would have to visit each agency where you think data might be available and then it may not be documented very well. One-stop portals have not materialized and efforts such as the National Map have failed to reach their potential.

Issue	Lesson
	It has worked OK for me. I use several sources for data and those sources are reliable and well documented. However, certain USGS efforts have not reached a level of usefulness (e.g., the National Map). [Worked well] The increased use of and migration to database technology for storing spatial data.
Tools	[Challenges] Technical - appropriate standards (metadata, vocabulary) and tools (gazeteers, spatial and keyword search) are still lacking. Existing systems need to lessen reliance on proprietary software. Tools to integrate diverse data types need development. [worked well] Cutting edge development of tiled map services by Google and others
Public Relations	Under promise and over deliver The more successful organizations in building SDIs are the ones that have a long history of collaborating with other organizations and ha[ve] a culture which is focused on making data and information available to the broader community. State liaisons. Relationships in the field cannot be beat…should be some of the most intelligent and motivated folks…part of their performance evaluation (not sure if this is possible) should come from the people they serve in the states. Partnership maintenance, state and local venues…local professional organizations…representation on state and local geospatial decision making bodies [Challenges] High expectations - Increasingly, scientists as well as decision makers, business and the public not only want, but expect all data will be instantly available online at no cost, and fully interoperable. Such systems are standard on a number of popular network television crime shows where all data of any kind sought are brought to the desk top instantly and fully integrated with no need to convert, process, or interpret them. [Do differently] Demonstrate the results and benefits earlier – Until recently, we have been talking among ourselves primarily and not to the users of the infrastructure. Audiences glaze over instantly with the mention of data exchange standards and semantic ontologies. So, we are starting to showcase what the system will look like and deliver to the average user. A demonstration of the Geoscience Information Network (GIN – http://usgin.org) to the Arizona Legislature in November 2009 was hugely successful, not only in showing decision-makers the potential but to many of our stakeholders and participants who are still somewhat fuzzy about how this will all work and what it will do.

Issue	Lesson
Public Relations Continuted	[do Differently] Make the USGS SDI more powerful by giving it better search and data discovery mechanisms. - Requires support from the top, a budget, and a dedicated team - not just USGS but from all agencies. Standards need to be defined. Robust software tools need to be developed to create standard metadata, and to provide the ability to search all NSDI nodes. This really requires a lot of coordination from all agencies to enforce the standards so ‘searches’ have the potential to retrieve all data that meets a query.
Planning	Develop a roadmap that encompasses the business , information, technology, computation and engineering viewpoints, and consistently review and update as required Well developed business case that articulates to the organization what the value of the proposition of SDIs are Successful projects are done incrementally…low hanging fruit Successful projects initially focus only on those projects that are staffed by fully committed people. [do differently] Appropriate data management starts at the planning phase and proceeds through data collection, processing and use. Tools must be provided that reduce the burden to individual projects/users throughout this process - and that ultimately provide them access to more data than would be otherwise available (or easily discovered/accessed). [Do differently] 1) Promote data lifecycle management objectives and outcomes as performance indicators for federal agencies, 2) create government centers of excellence for highest priority data sets and require cross agency funding mechanisms for collection and maintenance, 3) promote standards-based, optimized, geospatial data service hosting for federal agencies to increase capacity and uptake.
Organization	FGDC or some other entity has not been given adequate authority to carry out the mission they were put in place to do. If they are meant to be successful, they need to be put in some place other than USGS...like OMP.

Issue	Lesson
	[What has not worked] Central or concentrated control (e.g. Data Czars) – in the early days of the Web, researchers starting creating centralized databases for each domain or sub-domain. These required scouring the archives and literature for analog ("legacy") data, digitizing them, and building an ongoing capability to update and maintain the repository. Very soon, data providers could be barraged by multiple data base owners for copies of their data and constant demands for the latest updates. No one had the time or resources to be repeatedly feed the demands of external bodies for their data. As the number, size, and diversity of data bases grew rapidly, the communities wrestled with how to share and integrate data from disparate sources. Proposals to 'coordinate' data integration or oversee standards were viewed skeptically or hostilely by many as creating the potential for 'data czars' to impose their will on the rest of the community. This concern was one of the biggest stumbling blocks to getting community consensus in building cyberinfrastructure for the earth sciences in the past decade.
	[Challenges] This new organization is not just [about the] USGS, but all stakeholders from all agencies. Since the current organization is disjointed, it almost appears the past approach was to allow agencies to do whatever they wanted, and the 'best practice' would float to the top becoming the de facto standard. But the reality is, nothing floated to the top and it is still disjointed.
	Security needs and concerns also challenge most government programs
	[worked well] Recognition of benefits of web services
	[Challenges] The major challenges are primarily organizational, confounded by financial challenges. USGS has not had consistent leadership with the goals of leveraging our geospatial data and the enterprise licenses The majority of geospatial issues in the Department of the Interior (DOI) and USGS is the result of too little attention to the fundamentals of data standards and data applications across the spectrum of spatial data services in USGS. There is a partitioning of data collection among themes and funding of these themes, as well as partitioning of support services for Geospatial Data collections and the research scientists requiring GIS support to use our enterprise license. A very small part of the GIO is able to see the big picture and the result is that GIS application support has fallen through the cracks of constant reorganization.
	[Challenges] The most significant challenges to success of SDI are: 1) Clarity of responsibility and government-wide recognition of the stewardship responsibilities, 2) clear governance with regard to collaborative development and stewardship within and beyond the federal government, 3) greater leverage of public and private data resources and value-add capabilities, 4) lack of wide adoption of Web Services infrastructure.

Issue	Lesson
Organizational Commitment	Executive level support as well as commitment from senior, middle, and junior levels of staff
	Champion who is knowledgeable and respected by the community
	Full understanding of the impacts of the introduction of what is essentially a disruptive technology
	Collaboration in sharing of data, agreement of standards etc is critical to the development of an SDI
	Scientists must make available data that underpin knowledge products
	… properly articulated polices can be an enabler for SDIs….needed at the organizational …whole government level.
	[Challenges] Sustainability – Hundreds of millions of dollars have been spent on myriads of projects that, while individually successful, have not led to the creation of an integrated or sustainable spatial data infrastructure. Hundreds of stove-piped projects have been funded, but too many disappeared when they could not get funding renewals. Or the technology has changed and the results are in obsolete formats or buried on a hard drive somewhere. NSF is requiring new informatics projects to address the question of sustainability but having recently reviewed a large stack of proposals on an external panel, the community practitioners are not even close to dealing with this problem realistically or satisfactorily.
	[Do differently] Once the new infrastructure is in place, all projects should be required to budget time and money to prepare and submit all spatial data - as intended.
	[What has not worked?] The NSDI was initial established in 1994 and was intended to be a repository of all spatial data referenced in reports/publications. I'm not sure how many spatial datasets have been referenced in Water Resources Division (WRD) publications since 1994 to the present, but I would estimate well over 10,000. Keep in mind, GIS started to become main-stream in WRD in the mid 80s.WRD currently only has 646 datasets in the WRD NSDI (http://water.usgs.gov/cgi-bin/lookup/getgislist), so as you see, there is a huge problem getting authors to participate and I'm glad to see this finally getting addressed. To the authors defense, the reasons listed below are why they did not participate.
	[worked well] Not much has worked well; no support; standards not well defined; very little guidance; very little incentive; software tools to create consistent metadata lacking; datasets are almost considered a burden, especially large ones; search mechanisms of data in NSDIs lacking.
	[Challenges] Cultural - Incentives, if not mandates, need to be provided and a culture needs to be developed that recognizes data management and provision as part of the public trust responsibility of federal and state agencies. This culture will not arise because of theoretical benefits, it will develop when real benefits accrue to users through a) facilitation of data access and use and b) when systems provide relief from burdens of data and metadata development and management.

Issue	Lesson
	[do differently] Make a real commitment to Enterprise GIS and geospatial data management, development, and integration. Current support is nominal and based on the minimum support required to fulfill requirements of enterprise GIS licensing agreements. [do differently]The USGS Geospatial programs are primarily outward looking, and driven what they feel is public demand. This does very little to support USGS science. USGS Management needs to define a geospatial science commitment and plan
Personnel	Tertiary trained professionals who understand the technology and are respected Important to accept the high level of technical skills required to develop an SDI. …people become overnight experts…can annihilate a project very quickly. [do differently] Requirements for, and funding for, comprehensive data management within a shared infrastructure should be explicitly required in funding requests and performance evaluation. USGS lacks staff that are as skilled as in the private sector. The USGS is very salary burdened and as such has limited funds to go to outside vendors who could develop infrastructure. I think it is important to make sure that USGS researchers have a clear stake in the development and maintenance of world class data bases. In line with one of the recommendations of the NRC report (Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data Committee for a Pilot Study on Database Interfaces, National Research Council 1995) I think that USGS has to find a way to enable researchers to RGE "credit" for ongoing involvement in the development and maintenance of databases. Leaving database development/management to IT people or masters-level scientists will inhibit the researcher-driven experimentation, brainstorming, and interdisciplinary mindset needed for the creation and ongoing development a database that serves [an] ambitious science agenda.
Resources	Adequate funding but not over funding Avoid big projects with big funding that promise to deliver everything to everyone SDIs work that have provided economic revenue…easy to get additional funding. Is economic revenue the only benefit that will work? It works when funding is applied from the fed – state level supplement long term partnerships between fed and state. And it doesn't take many $$... Work with Google [private industry in general] to make it even more useful and friendly

Issue	Lesson
Resources Continued	Uncoordinated federal/state/local geospatial budgets and expenditures do not work
	Funding geospatial data programmatically rather than strategically does not work
	The USGS has not been adequately funded to carry out their mission of civil domestic mapping over the U.S.
	What I would do first and immediately is figure out what the SDI is worth in the U.S. and to whom it is worth what? Once you know what everyone does with it, where the gaps are and put a $$ value on closing those gaps you could begin creating the necessary partnerships both programmatically and fiscally to complete a sturdy and useful SDI. We worry so much about the sexy technologies that we forget people just need this stuff to get their jobs done. Those who have worked beside me for years have heard this before. We need to understand the econometrics of our SDI to be able to spread the cost and responsibility in a useful and meaningful way. Maybe we need to get economists and intergovernmental programmatic folks together to monetize the SDI.
	[what has not worked well] Non-sustainable business models - Early on, NSF and other agencies funded the creation and population of databases but after a few years it became clear that NSF did not have the mission or the resources to maintain this infrastructure permanently. Data bases shut down for lack of funding. Resources disappeared and people moved on to other projects. Even today, many funding proposals describe their sustainability plans as simply returning to the original funding agency and asking for more money.
	[Do differently] Integrate with other domains – To say we would do things differently may be misleading. The problem has been finding resources to do all the things we know need to be done, including integrating our work with that being done in other domains.
	[Do differently] Once the new infrastructure is in place, all projects should be required to budget time and money to prepare and submit all spatial data - as intended.
	[Worked well] Spatial data infrastructures (SDI) have worked well at the federal level, and have mostly worked well at the state level. With funding problems, SDI has faltered somewhat at the state level, and for the same reason, many counties and other local jurisdictions have had mixed results varying from robust SDIs to non-existent SDIs.
	[Not worked well] Outside of the federal government: un-funded mandates for SDI tend to be ignored;
	[Challenges] The major challenge is financial: support for SDI requires additional personnel, with changes technology and cultural behaviors. Many academic and non-governmental organizations (as well as a number of governmental entities below the federal level) will not undertake participating in SDI unless the financial support is available since it would take time away from existing activities.

Issue	Lesson
	[did not work] Broad and generic mandates or reliance on "good will" to drive participation in development of community information resources. [Challenges] Financial - Data management, provision, and integration are the infrastructure for both science and management applications. The resources to build this infrastructure are lacking.
Coordination	It doesn't work when there is competition within the state to be the single point of contact. i. e. a state GIS coordinating council. Helping the states get coordinated is a very useful activity for USGS. Through their liaisons and field offices. (suggest NSGIC for these activities - they live and die by coordination and cooperation). [does not work] States who are not coordinated and have a state level geospatial coordinating body. There must be an entity who can speak with authority on funding issues for geospatial data at the state level, otherwise fed state partnerships are very difficult to put together. The state entity must be recognized by state agencies, and the executive and legislative branches of govt. along with the local governments. [Challenges] Agency cultural, data, fiscal you name it....silos. I was a fed and a state person for a long time. I know first hand how difficult it is to do intra and inter-agency coordination of anything, let alone intergovernmental cross coordination. But it is critical to the success of an SDI. If geospatial funds and programs were (pipe dream here) coordinated (not consolidated) across the fed level - by OMB – the only people with a big stick in the fed govt. - just the slosh factor of $$ being expended on geospatial activities at the fed level could fund coordination activities at the state level. I always did think that if we took the lines of business (or whatever the current lingo is at the fed level) not just across the bureaus and down through the fed agencies but on down to the state and local level there would be a logical pathway of responsibilities. In those pathways there is a common need for the same kind of data, geospatial data and practices. How hard would it then be to monetize the value of the necessary data and applications to get the job done at every level it needs doing?? Something like the old a-16 process.[OMB Circular A-16 Coordination of Geographic Information and Related Spatial Data Activities Revised 2002]

Issue	Lesson
Coordination Continued	[Challenges] Community adoption and buy-in – The geoscience community has been wary of cyberinfrastructure (including spatial data infrastructure) due to concerns over control of and access to data, recognition of data ownership, costs of converting data and systems, mandates, and how decisions are made. Every domain is dealing with similar issues, and coming up with generally similar approaches. Yet, we are all still mostly working within our community stovepipes. We have much to learn from each other and much we can share so we don't have to duplicate or relearn what others have done. The library sciences in particular are making dramatic strides in aggregation, archiving, and disseminating digital data in a multitude of formats. We have not made the connections yet with them. Even within the geosciences community, we are only part way there. Our network is based on geological surveys with only a few example external partners. The NSF-funded National Geoinformatics System (NGS) project to evaluate community needs and wishes has been dormant for more than a year. Could they be watching to see how GIN (and NGDS) develop and serve as core elements of an NGS? We also need to nurture preliminary linkages with the biological, oceanographic, atmospheric, and geographic communities as well as computer sciences
	[don't do well] We also need much greater coordination and dialogue across this community to minimize wasted effort and maximize accomplishment of shared goals.
	[Challenges] ...the most critical challenge stems for the inadequate dialogue and coordination among developers and users of these critical data. This is a combination of policy (e.g., stovepiped federal agencies), cultural (basically, a 'stovepiped' mindset), and financial issues (we're all scrambling for resources).
	[Do differently] The FGDC, USGS, and other bodies need to be better supported, more open in membership (i.e., to science NGOs), and empowered to support more robust dialogue, clarify shared goals, and facilitate sharing of financial resources.
	[what worked] Development of systems/processes that engage the "user community" in defining requirements and reflect the technical capabilities available to the users. And, in response, focusing on provision of tools that facilitate use of existing systems (FGDC/GOS) by reflecting the particular search, discovery, and access needs of the users. Working with a specific but broad user community (coastal and marine researchers and managers) to develop tools that facilitate integration of data and model output using open source standards in response to identified needs.

Issue	Lesson
Vision	[Do differently?] Could we have gotten here earlier? – The debates at workshops, forums, professional meeting sessions, and in the corridors over the past decade were part of a process of exploring and testing ideas in a fast-changing technical and social environment. It is only in hindsight that we see where we were heading. But I doubt that if we presented our current model to ourselves 10 or even 5 years ago, that we would be ready to embrace it. There has been an evolution in thinking that was crucial to developing current models. Based on conversations with colleagues in other fields, and in tracking the literature superficially, it appears that the solid earth geosciences are just a bit ahead of other communities in coming to our present realization and acting effectively on it.
	a. Interoperable – data should be seamlessly delivered to desktops regardless of the originating database software, version, operating system, or server. b. Open-source – data and services need to be compliant with open-source standards such as OGC and ISO. This will help avoid the problem of data that cannot be accessed in obsolete or priority software c. Distributed – data providers should provide their latest available data directly into the network. They decide what is made public and when. They do not have to continually pass along their revisions to a growing number of data aggregators or central databases. The data network then looks more like the Web – each provider is responsible for what they want to share. There will be a continuing need for archive and orphan data repositories for data that do not have permanent homes, and for data scavenged from historical and analog sources. But even these central databases will be another layer of distributed nodes in the network. d. Web-based (SOA) – services and applications are increasingly being served on the Web rather than being on the desktop. This allows for large resources beyond the standard desktop to handle and greatly diminishes bandwidth requirements. Referencing an online resource also means you are using the latest version as are others. e. Flexible, dynamic, organic, modular – the system has to open to users to choose what tools and applications they want to use and to allow them to develop and implement their own applications. Just as there is not only one Web browser, there should not be components beyond the most fundamental standards and protocols that are mandated to users. Technology is moving too fast to be locked into restrictions that will limit and ultimately make the system obsolete. A modular approach allows anyone with a better idea to link into the network and make their service available. It also means the network developers don't have to build everything. They can choose among the best work done by others in order to quickly assemble a functioning system, while leaving open the potential for alternatives to be networked

Issue	Lesson
Vision Continued	f. User-friendly – The first Web sites had to be tediously programmed in html, but now user-friendly commercial software and ubiquitous free applications, allow everyone to easily and quickly build Web sites, including specialized sites like blogs. The early stages of the spatial data infrastructure will require fairly sophisticated developers but emphasis should be on off-the-shelf cookbooks and guides, and eventually smart applications that almost anyone can use to provide data or services to the network. g. Community of practice – The changes being brought about by the widespread use of digital data delivered via the Web requires that we develop new communities of practice in how we qualify the vast amounts of data that we might otherwise use indiscriminately and how we recognize and reward those who provide data and services in data networks.
	Improved search engines should make it easier to find everything. A web service should index everything we have, allowing users to subscribe to any content desired. The system should be distributed, and should aggregate datasets from Science Centers. The Science Centers would go through a streamlined process to document and publish their data sets, and to set access, e.g. local use only, USGS only, or public dissemination. From that point on, it should be automatically harvested and pushed out to the appropriate user groups. The content could be live services, or could be extracted to a local geodatabase, and this could be maintained and updated automatically. Most of the pieces of such a system exist and could be implemented today.
	[Do differently] Put some good people into cataloging existing reports and data sets. Build better metadata tools. Make management accountable to publish geospatial data from all projects Make all projects identify spatial data results, plan for, and publish them before project is considered complete.
	[Vision] Very simply, my vision for SDI is that it should enable the scientific community to freely access and exchange spatial data with sufficient metadata to allow an interchange.
	To look at an image of the U.S./globe, zoom in on an area, and get a listing of ALL the available data for that patch of land. Then be able to view detailed documentation on what the data are and how it should be used and then be able to download a single geodatabase of that information for the patch of land I am interested in.
	USGS has a critical role to play in facilitating dialogue among the federal agency, academic, non-government science, and state agency sectors to clarify shared goals, data standards, and data sharing technology. Success in this area will allow us to collectively maximize utility in our investments in spatial data.

Issue	Lesson
	SDI should benefit data collectors and users from planning (evaluation of existing data), collection (standards and requirements), metadata development, archiving, search & discovery, and integration. The system will not be seen as an "overhead" on research activities - but rather as a way to facilitate research, ensure data preservation, and will enhance and expand the application and integration of information resources. Performance will be evaluated not simply on "availability" of data - but on success in enhancing data application to meet diverse research and application needs.
	A system that is integrated that provides readily available information from local to national scales. A on-stop integrated portal would be a nice start. Also, the SDI should have a set of tools and interfaces that permit the integration of data ... e.g., downscaled climate data and models.
	Promotion and development of fast, reliable, web services that provide discovery and access to geospatial data. The users will figure the rest out. Better use of and support to the users of Enterprise GIS tools.
	An NSDI that supports the USGS Science Strategy would include relevant base and thematic data that are refreshed at an appropriate rate and yet are maintained as time-accessible snapshots to allow change and context evaluation. The SDI would provide a geographic framework for the publication of most scientific data of the USGS, allowing for easy visual analysis of geographic and temporal phenomena.