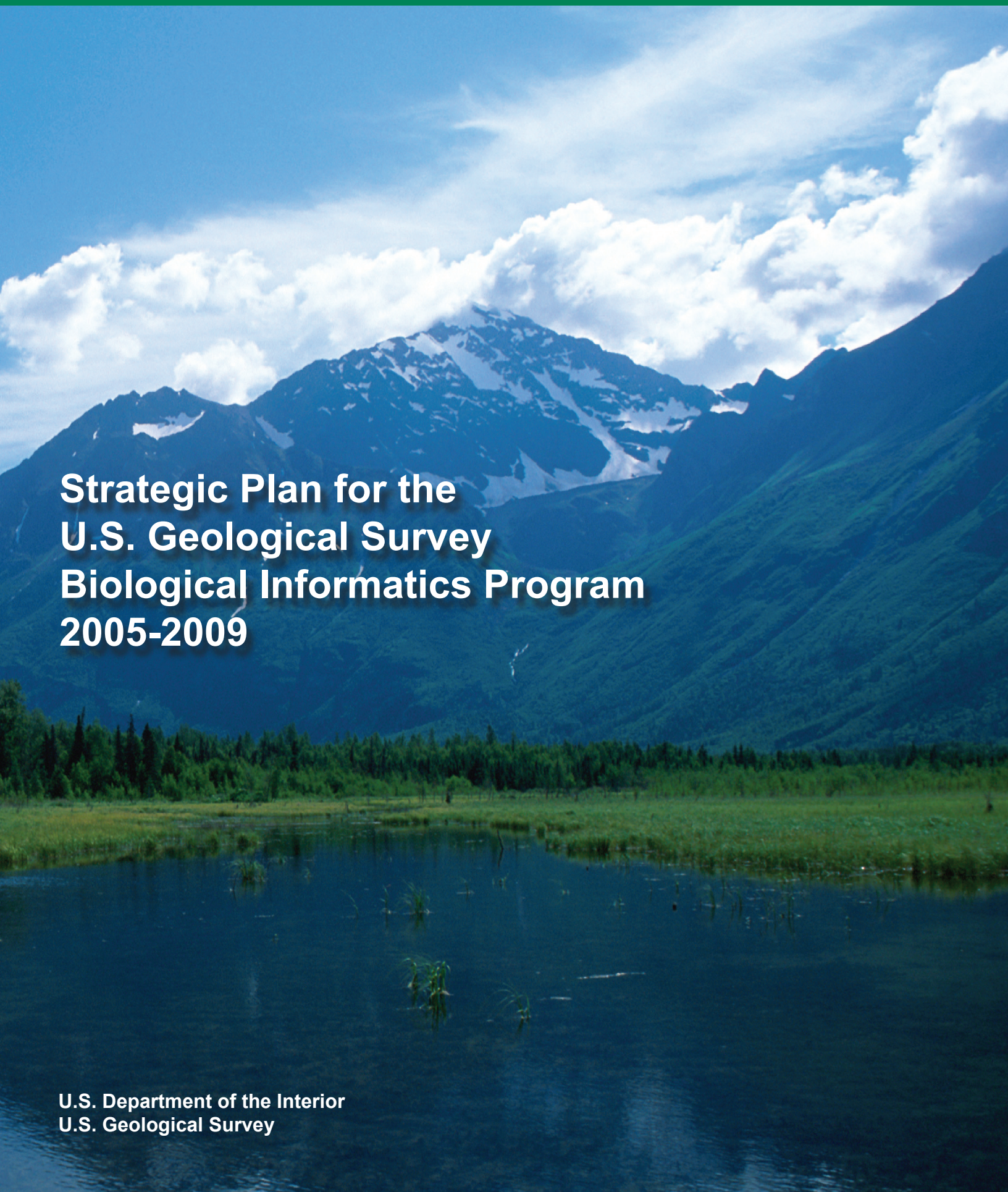
USGS
science for a changing world



Strategic Plan for the U.S. Geological Survey Biological Informatics Program 2005-2009

U.S. Department of the Interior
U.S. Geological Survey

Strategic Plan for the U.S. Geological Survey Biological Informatics Program: 2005-2009

By Michael Ruggiero, Marcia McNiff, Annette Olson, and Ben Wheeler

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
Gale A. Norton, Secretary

U.S. Geological Survey
P. Patrick Leahy, Acting Director

U.S. Geological Survey, Reston, VA: 2005

Suggested citation:

Ruggiero, Michael, Marcia McNiff, Annette Olson, and Ben Wheeler. 2005. *Strategic Plan for the U.S. Geological Survey Biological Informatics Program: 2005-2009*. U.S. Geological Survey, Biological Resources Discipline, 20 pp.

Photo on front and back covers by John J. Mosesso/NBII

CONTENTS

EXECUTIVE SUMMARY 1

BIOLOGICAL INFORMATICS: THE INTEGRATION OF INFORMATION FROM GENOMES TO BIOMES 3

The Call for Biological Informatics

Challenges

BIOLOGICAL INFORMATICS IN THE ORGANIZATION 4

Department of the Interior

U.S. Geological Survey

Biological Informatics Program

Added Value of Integration

Partnerships

THE ROAD AHEAD 12

Goal 1: Increase the availability and usefulness of biological resources data and information (Content)

Goal 2: Implement technologies and tools to integrate, analyze, visualize, and apply biological information to natural resource issues (Tools)

Goal 3: Develop, apply, and promote the adoption of standard practices, protocols, and techniques to enhance knowledge discovery and retrieval from various resources (Infrastructure)

Goal 4: Facilitate information science and technology research that supports the advancement of biological informatics capabilities (Research)

Goal 5: Apply innovative technologies and best practices to improve the development, description, and dissemination of biological information to customers (User Services)

IMPLEMENTATION 13

Table 1: Strategies, Outcomes, and Measurements Necessary to Meet the Biological Informatics Program Mission, Goals, and Objectives

APPENDIX 1 Biological Informatics Program Subjects of Particular Interest 19

APPENDIX 2 List of Abbreviations 20

Acknowledgments

We want to thank all of those who have contributed to the production of this plan. We would especially like to thank the following people for their valuable comments and contributions: Gladys Cotter, Frank D'Erchia, Mike Frame, Tony Frank, Sharon Gross, Tom Gunther, John Hill, Lief Horwitz, Kate Kase, John Mosesso, Doug Muchoney, and Phaedra Ponds.

Finally, we acknowledge Ron Sepic and Michelle Davis of the NBII National Program Office who handled, respectively, the editing and design of this publication and who oversaw its production.

Photo credit: John J. Mosesso/NBII

EXECUTIVE SUMMARY

Biological resources, from the genetic information within organisms to the life-sustaining services provided by ecosystems, are vital to human well-being. In order for researchers, resource managers, decision-makers, and interested citizens to understand these resources and assure their maintenance, it is important that the capacity exists for accessing, sharing, and utilizing pertinent data and information. The Biological Informatics Program is building this capacity in support of the Serving Communities mission area of the U.S. Department of the Interior (DOI) Strategic Plan.

The mission of the U.S. Geological Survey (USGS) Biological Informatics Program is to create the informatics framework, provide the scientific content, and develop the public and private partnerships needed for the understanding and stewardship of our Nation's biological resources. This mission is achieved through the following goals:

- Increase the availability and usefulness of biological resources data and information (Content),
- Implement technologies and tools to integrate, analyze, visualize, and apply biological information to natural resource issues (Tools),
- Develop, apply, and promote the adoption of standard practices, protocols, and techniques to enhance knowledge discovery and retrieval from various resources (Infrastructure),
- Facilitate information science and technology research that supports the advancement of biological informatics capabilities (Research), and
- Apply innovative technologies and best practices to improve the development, description, and dissemination of biological information to customers (User Services).

Governments and institutions are calling for the collection of and access to a wide variety of critical information about our biological resources. This spans information on species, habitats, and ecosystems, and also includes impacts of human activities, best practices for management, and more. The following core integrated components of the Biological Informatics Program have been created to design, develop, implement, and update the capabilities needed to accommodate its mission:

Gap Analysis Program (GAP): provides State, regional, and national conservation assessments of native vertebrate species and natural land cover types.

Vegetation Characterization Program (VEG): characterizes vegetation on public lands using a consistent methodology supported by national standards.

National Biological Information Infrastructure (NBII): implements a fully digital, interactive, distributed system that provides access to scientifically reliable natural science data and information.

Integrated Taxonomic Information System (ITIS): provides an authoritative source of species names and their hierarchical classification.

USGS Science Center-Based Informatics Projects: supports USGS Science Center-based knowledge-sharing projects, planned for and executed at the Center level, that address Center priorities.

Strategic direction is set through consultation with Program stakeholders, partners, and customers throughout Federal, State, tribal, and local governments; non-governmental organizations; academic institutions; the private sector; and the global community. Progress toward these goals will be measured by means of a set of indicators that express the extent of the biological resources knowledge base made accessible through Program activities.

Information on swamp sparrows (*Melospiza georgiana),* juvenile shown here, and many other North American birds is available through the NBII Bird Conservation Node <http://birdcon.nbii.gov/> and the Gap Analysis Program <http://gapanalysis.nbii.gov>.

Photo credit: John and Karen Hollingsworth/U.S. Fish and Wildlife Service

Biological informatics, or bioinformatics, seeks to integrate these diverse data and information in order to enhance our knowledge of our environment, from genomes to biomes.

Photo credit: USGS Forest and Rangeland Ecosystem Science Center

BIOLOGICAL INFORMATICS: THE INTEGRATION OF INFORMATION FROM GENOMES TO BIOMES

A diverse environment underpins human survival, health, and the economy. Similarly, changes caused by human and natural forces can have large ramifications. In order for researchers, resource managers, decision-makers, and interested citizens to further understand and maintain the environment and its multitude of biological resources, it is important that the capacity be improved for accessing, sharing, and utilizing the biological data and information resulting from research. Biological data and information, however, vary in format and content: they may pertain to genetics or other characteristics of a particular species, or they may pertain to an entire ecosystem. They can contain land-use characteristics, or geospatially-referenced information. Biological informatics, or bioinformatics, seeks to integrate these diverse data and information in order to enhance our knowledge of our environment, from genomes to biomes.

The term *bioinformatics* includes the collecting and linking, storage, organization, integration, analysis and synthesis, delivery, and application of biological data and information. By applying information science and technology, as well as scientific principles, bioinformatics provides new gateways into vast, accessible warehouses of information.

The Call for Biological Informatics

Biological informatics is rapidly growing in response to increasing needs by users and by the capabilities of technology. Researchers are asking both broader and more fine-tuned questions, examining the influence of numerous potential factors at multiple points along continuous temporal and spatial scales, often requiring access to diverse data from other sources. In response to ecological problems such as unsustainable resource use and invasive species, resource managers and decision-makers, from local to global, are calling for organizations to not only collect and organize as much biological information as possible, but to understand, model, and forecast environmental conditions. Finally, students and the general public are seeking increased access to both raw data and to compiled information and knowledge. For all users, it is crucial that the information is accurate and scientifically credible, relevant, timely, well-organized, easily accessible, and sufficiently documented. In addition, tools for visualization, synthesis, and analysis of raw data are needed to enhance understanding.

Biological informatics research and development has been identified as a critical national priority. The President's Committee of Advisers on Science and Technology (PCAST) Panel on Biodiversity and Ecosystems[1] and the President's Information Technology Advisory Committee[2] have recommended significantly increased activity in this area. The National Research Council's Committee on Grand Challenges in Environmental Sciences has recently identified the need for fundamentally increased understanding of biodiversity and ecosystem functioning as one of the eight "Grand Challenges in Environmental Science" facing our Nation and the world today.[3]

[1] President's Committee of Advisers on Science and Technology (PCAST) Panel on Biodiversity and Ecosystems, *Teaming with Life: Investing in Science to Understand and Use America's Living Capital* (Washington, DC: Executive Office of the President of the United States, 1998), xv.

[2] President's Information Technology Advisory Committee, *Report to the President. Information Technology Research: Investing in Our Future* (Arlington, VA: National Coordination Office for Information Technology Research and Development [NCO/IT R&D], 1999).

[3] National Research Council, *Grand Challenges in Environmental Sciences.* Committee on Grand Challenges in Environmental Sciences, Oversight Commission for the Committee on Grand Challenges in Environmental Sciences (Washington, DC: National Academy Press, 2001), 106 pp.

Challenges

Critically important to the Nation's ecosystems, economy, and security are significant biological issues, such as invasive species, emerging diseases, unsustainable resource use, and climate change. Although the Federal government invests more than $600 million per year in biological data collection, it is difficult to address these issues because of limited accessibility and lack of standards for data and information.

There is still a lack of access to scientific data of known quality, despite the widespread availability of Internet connectivity. This deficit is primarily due to a lack of interoperability across hardware, software, and data systems; and a lack of applications and tools to meet user needs. Challenges relating to the integration of data and information reflect the complex nature of biodiversity and ecosystems data. Species can be studied at scales ranging from genomes to biomes, from microscopic to global, and from days to years. Variable quality, sources, methods, and formats (e.g., observations in the field, museum specimens, and satellite images) present additional challenges.

The biological informatics field must be able to adapt quickly to anticipate user needs and to incorporate rapidly emerging technologies that change the way data are collected, stored, managed, and disseminated. As technologies such as wireless communications proliferate, and computing platforms become smaller and more mobile, the ability to seamlessly extract, analyze, and process data in the field can be realized. Even though these technologies offer new solutions, they also create new informatics challenges.

The biological informatics field must be able to adapt quickly to anticipate user needs and to incorporate rapidly emerging technologies that change the way data are collected, stored, managed, and disseminated.

Photo credit: USGS Vegetation Characterization Program

BIOLOGICAL INFORMATICS IN THE ORGANIZATION

The USGS recognizes the value of information and its role in improving the quality of life and the environment. It established the Biological Informatics Program in 1996 to focus on this role. The Program is a national-level capability within the USGS. The Biological Informatics Office (BIO) of the Biological Resources Discipline (BRD) administers the Biological Informatics Program and responds to the needs and requirements of both DOI and the USGS.

Department of the Interior

The DOI mission is to protect and manage the Nation's natural resources and cultural heritage and provide scientific and other information about those resources. The DOI Strategic Plan further states that Department bureaus manage a wealth of accessible, online science data and information systems and are committed to maintaining and improving these assets for the benefit of the public.

The Biological Informatics Program supports the DOI mission area of Serving Communities Goal 2 (Advance Knowledge Through Scientific Leadership and Inform Decisions Through the Applications of Science) and Serving Communities Goal 2, Strategy 1 (Improve Information Base, Information Management, and Technical Assistance).[4]

Close collaboration with a variety of partners is crucial to serving communities. The Program works with these partners to assemble the network that delivers the benefits of good science to planners, decision-makers,

[4] U.S. Department of the Interior, Strategic Plan, Fiscal Year 2003–2008 (Washington, DC: U.S. Department of the Interior, 2003), 92 pp. Available at <http://www.doi.gov/ppp/strat_plan_fy2003_2008.pdf>.

DOI STRATEGIC PLAN ELEMENTS PERFORMED BY THE BIOLOGICAL INFORMATICS PROGRAM

Mission Area: Serving Communities

End Outcome Goal 2.
Advance knowledge through scientific leadership and inform decisions through the applications of science.

Strategy 2.1
Improve information base, information management, and technical assistance.

Performance Measure 2.1.1
Inform Decisions through the Application of Science: Improved stakeholder access to needed science information and stakeholders reporting that information helped achieve goal.

Performance Measure 2.1.2
Content and Expanse of Knowledge Base: Percent of surface area with temporal and spatial monitoring, research, and assessment/data coverage to meet land-use planning and monitoring requirements.

researchers, educators, and the general public. Through activities at the local, regional, national, and global levels, the Biological Informatics Program works to make the biological resource knowledge of the United States available to the world and the world's biological resource knowledge available to the United States. In addition, these activities support the U.S. Government's international obligations relating to the sharing of biological information. Examples of these obligations include the Convention on Biological Diversity (CBD), Cartegena Biosafety Protocol, Summit of the Americas, and the Summit on Earth Observations.

U.S. Geological Survey

As the Department's principal science bureau, the USGS "serves the Nation by providing reliable scientific information to describe and understand the Earth; minimize loss of life and property from natural disasters; manage water, biological, energy, and mineral resources; and enhance and protect our quality of life."[5] The USGS Strategic Plan has as one of its long-term goals to "ensure the continued availability of long-term environmental and natural resource information and systematic analysis and investigations needed by customers … for informed decision-making about natural systems."[6] The long-term goal of Environmental Information Science, one of the USGS Future Science Directions, is: "To increase understanding of natural systems in terms of the information that is contained and processed through these systems. To be a leader in development and application of advanced scientific capabilities for the acquisition, analysis, understanding, and communication of environmental science information."[7] The Biological Informatics Program directly supports the Bureau mission and specifically addresses these long-term goals.

Biological Informatics Program

The mission of the USGS Biological Informatics Program is to create the informatics framework, provide the scientific content, and develop the public and private partnerships needed for the understanding and stewardship of our Nation's biological resources. The Biological Informatics Program provides credible, applicable, unbiased information for science-based decision-making, particularly as it pertains to the

[5] U.S. Geological Survey. "About USGS." At U.S. Geological Survey Web site: <http://www.usgs.gov/aboutusgs>.

[6] U.S. Geological Survey, Strategic Plan: 2000–2005 (Reston, VA: U.S. Geological Survey, 2000), 25 pp. Available at <http://www.usgs.gov/stratplan/stratplan_rev.pdf>.

[7] A. F. Frondorf, D. R. Boldt, D. R. Hutchinson, W. G. Miller, D. R. Posson, and S. A. Sipkin, U.S. Geological Survey Open-File Report 2005-1180 (2005)—ONLINE ONLY. Available at <http://pubs.usgs.gov/of/2005/1180/>.

conservation, management, and use of the Nation's natural resources. The program is committed to making available the data and information that are critical to scientific discovery and application. Databases, maps, and publications are vital mechanisms for conveying this information to users. In addition, the Program is committed to the development and easy accessibility of tools, models, visualizations, and applications to aid policy and resource managers in the analysis and synthesis of scientific data to support decision-making. The Program works in cooperation with many organizations throughout the United States and the world to provide critical information to partners, stakeholders, customers, and the general public. Through electronic infrastructures, the Program delivers relevant data and information faster and in more interoperable formats than in the past, leading to better stewardship of the Nation's natural resources.

Added Value of Integration

The core and interdependent components of BIO have been specifically designed to integrate information across geographic and political scales (local to global) and biological levels of organization (genomes to biomes).

The core and interdependent components of BIO have been specifically designed to integrate information across geographic and political scales (local to global) and biological levels of organization (genomes to biomes).

- **Landscapes, Stewardship, and Species Distributions.** GAP generates databases on native vertebrate species distributions and natural land cover types to provide State, regional, and national conservation assessments. It also serves as a standard geospatial and biological foundation upon which other biological information from the NBII can be analyzed.
- **Vegetation Communities.** The VEG component characterizes vegetation using a consistent methodology supported by national standards. The products are specifically intended to serve natural resource managers by providing them accurate, fine resolution vegetation land cover assessments that are then used by the land managers for managing biological resources under their stewardship. VEG supports the calibration of remote sensing methodologies and verification of land cover types used by GAP.
- **Genomes to Biomes.** The NBII provides GAP, VEG, ITIS, and others with a fully digital, interactive, distributed system that provides scientifically reliable biological data and information. It also provides network-wide methods for organizing content to enhance the identification, retrieval, and use of information. These include the use of standard metadata and schema for documenting data sets and information products.
- **Biosystematics and Nomenclature.** ITIS provides an authoritative source of species names and their hierarchical classification. It serves as a taxonomic standard for other BIO components and the global community, enabling the comparison of biodiversity data sets at all biological levels.

The Program works collaboratively with others to ensure that it has high quality data and information that can be used to address issues of resources management. To that end, BIO engages USGS science centers, non-governmental organizations, museums, universities, international organizations, and other partners in the creation of data content and resources to address resource management needs.

Descriptions of the major activities of the Biological Informatics Program are provided below:

National Biological Information Infrastructure. The NBII is the premier partnership for making biological data, information, and associated tools and technologies more accessible for customers and partners to use in making cost-effective, informed decisions regarding resource management, environmental considerations, disease vectors, control of invasive species, and other issues. The NBII uses the capabilities of the World Wide Web and other advanced technologies to establish a distributed "federation" of biological data and information sources, through which users can find biological information, retrieve it, and apply it to resource management questions. Partners and customers who take part in this effort include government agencies at all levels, private sector organizations, natural history museums, libraries, academic institutions, international scientific organizations, and the public.

Gap Analysis. GAP provides broad geographic information on the status of species and their habitats. This involves identifying the degree to which native animal and plant species are represented in the present-day mix of conservation lands (those species not

Figure 1. How Biological Informatics Program Components Are Integrated

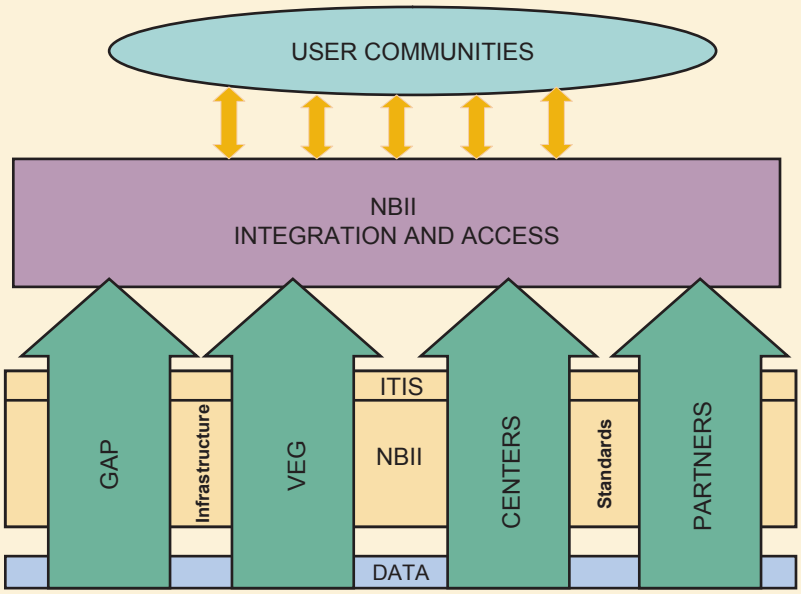


Figure 1 illustrates how data and information are served to customers through a careful integration of BIO's components. Data are generated through major programs such as GAP, VEG, Science Centers, and Partner initiatives. These data are passed through standards and protocols like ITIS and Federal Geographic Data Committee (FGDC) profiles to assure interoperability and quality. The NBII infrastructure provides the computing system foundation upon which these processes operate. Additional NBII functionalities serve the needs of the user community through data and information integration and dissemination.

adequately represented constitute conservation "gaps"). Currently, GAP products are available for most of the country, including digital databases describing Statewide land-cover assemblages, vertebrate distributions, and characterizations of land stewardship. The current emphasis of the program is on completing GAP projects in the few States where data are not available, developing regional data sets, updating selected regions of the country with state-of-the-art methods and technologies, organizing new aquatic projects, and developing partnerships with data users to facilitate the use of GAP information in land-management decisions.

Vegetation Characterization Program. VEG focuses on creating national vegetation standards, technologies, and products. It is a key component of the cooperative effort by the USGS and the National Park Service (NPS) to classify, describe, and map vegetation communities in more than 270 national park units across the Nation. This landmark program is both the first to provide national-scale descriptions of vegetation for a Federal agency and the first to create national vegetation standards for its data products. It is meeting specific information needs identified by NPS, with additional cooperative projects for the U.S. Fish and Wildlife Service (USFWS) and the Bureau of Land Management (BLM). Current efforts include dovetailing protocols for invasive species inventory and fire fuels related to vegetation to achieve efficient and integrated field data collection protocols.

Integrated Taxonomic Information System. The USGS leads and works with other Federal agencies (including the Environmental Protection Agency [EPA], U.S. Department of Agriculture [USDA] Agricultural Research Service [ARS], USDA Natural Resources Conservation Service [NRCS], National Oceanic and Atmospheric Administration [NOAA], Smithsonian Institution [SI], National Science Foundation [NSF], and National Park Service [NPS]), organizations, institutions, and taxonomic specialists across the United States and internationally to operate the largest taxonomic thesaurus and database of its kind in the world. ITIS provides an accepted scientific name (with

a unique Taxonomic Serial Number) as the "common denominator" for accessing information on such topics as biodiversity, invasive species, declining amphibians, migratory birds, fishery stocks, pollinators, agricultural pests, and emerging diseases. ITIS is also the source of biological names for the NBII. ITIS supports the only comprehensive national taxonomic database that provides free access (directly over the Internet) to standard scientific names for all U.S. plant and animal species. Currently, ITIS is coordinating efforts with the NBII Invasive Species Information Node in the development of an early detection and rapid response system, building taxonomies for invasive species as a key component of that system.

USGS Science Centers. While the Biological Informatics Program is a national capability, the USGS Regions and Science Centers are important partners. A portion of the Program's appropriated budget is distributed to the Centers to support informatics activities planned and executed at that level.

Partnerships

Fundamental to the structure of the Program is its capacity to build and strengthen partnerships across Federal, State, local, international, academic, non-profit, and private sectors. In order for the Program to succeed in providing qualitative and quantitative natural resources data and information to decision-makers, it must incorporate all of these interests to ensure a complete picture drawn from the best possible sources. Strategic partnerships are established to:

- Increase content,
- Develop and implement standards,
- Promote the development and implementation of advanced technology, and
- Leverage resources among partners.

The Program has built partnerships with organizations through which the major biological resource data and information management activities of the country are linked. Such partnerships include:

State agencies and organizations holding state data and information. States and associated organizations (such as the International Association of Fish and Wildlife Agencies) have key responsibilities for managing the Nation's natural resources. In line with these responsibilities, they gather and maintain large quantities of natural resources data and information.

Figure 2. Biological Informatics – Local to Global

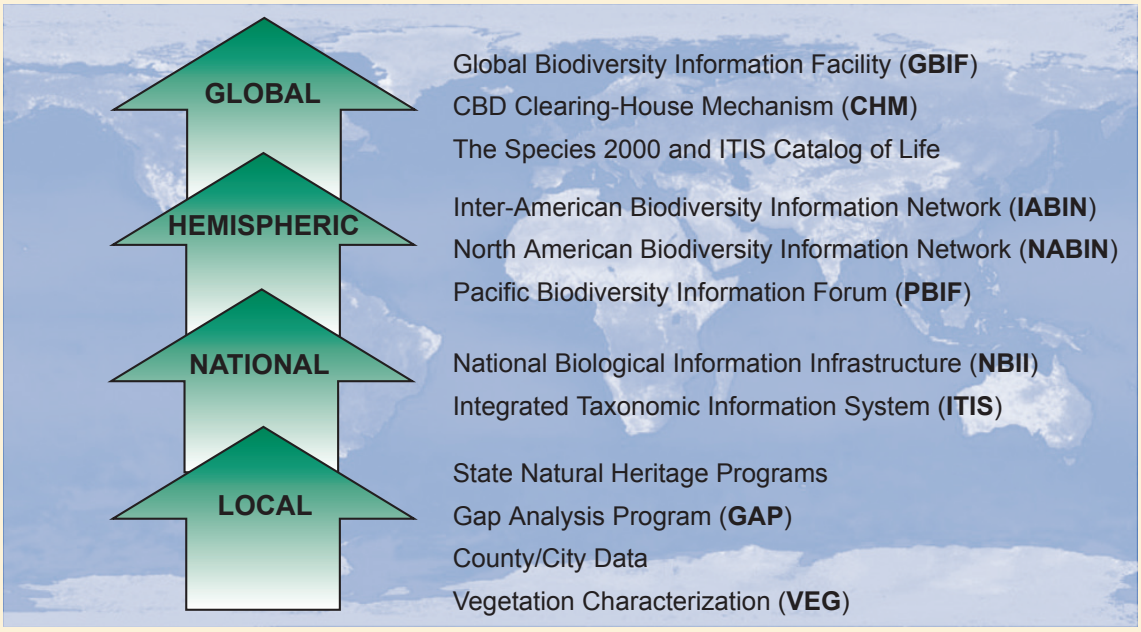


Figure 2 illustrates biological informatics components that flow from local to national to hemispheric to global levels. This approach avoids duplicative investments in both technology and content and expands that content base available to the user community.

Biological and ecological research organizations. Numerous organizations across the Nation and around the world are engaged in scientific research in the biological/ecological realm. In order to access this vast source of data and information, partnerships are needed that facilitate unifying and synthesizing science information to address critical biological issues.

Federal and other land management organizations. The key objective of the Program is to provide information and tools to resource managers for decision-making. To do that, the Program develops national-level partnerships with agencies responsible for land and resource management to understand their information needs, to make their natural resources information accessible, and to leverage resources.

Leaders of biological networks and systems. The Program partners with other networks to facilitate complete content coverage, data access, standards development, and technology deployment.

Leaders of information science research and development (R&D) activities. The Program partners with organizations leading or conducting information science R&D that can be leveraged and applied to the biodiversity and ecosystems domain.

Stewards of natural science collections. Natural history museums, herbaria, botanical and zoological gardens, and other natural science collections are important contributors to the Program, providing both taxonomic expertise and specimen data.

International initiatives. Biological resource issues cross many political boundaries and, therefore, must be addressed at regional, hemispheric, and global levels. Organisms including disease vectors and invasive species can be transported anywhere in the world within 24 hours. Serious and costly issues such as invasive species and zoonotic diseases require the sharing of knowledge with other countries. Participation in continental, hemispheric, and global biodiversity networking initiatives provides the opportunity to make U.S. biological information and data resources available to the world and to make the world's biological knowledge available to the United States.

Coordination for NBII efforts to develop an avian influenza detection system is centered at the USGS National Wildlife Health Center, which houses the NBII Wildlife Disease Information Node (WDIN). Today, WDIN is busy developing a Web-based system <wildlifedisease.nbii.gov/ai/> that may assist in projecting the spread of a highly pathogenic strain of avian flu (Asian H5N1). Injecting the virus into eggs (shown here) demonstrates the Center's commitment to gain an understanding of the disease from a variety of perspectives.

Data and information about invasive species, such as multiflora rose (*Rosa multiflora*), are available through a variety of Biological Informatics Program components.

Photo credit: Elizabeth Sellers/NBII

Examples of Recent Accomplishments and Ongoing Work of the USGS Biological Informatics Program

National Biological Information Infrastructure

To date, over 250 agencies and organizations have partnered with the NBII, providing content, expertise, and in-kind contributions. Currently, the network has eight regional and four thematic nodes. The NBII, through the FGDC, developed the FGDC Biological Profile to support the creation of national metadata standards that are unique to biological data and information. The NBII supports the national Chronic Wasting Disease repository and is responsible for Web and digital map enabling of the Breeding Bird Survey, which is designed to track the status and trends of North American bird populations. The NBII is also helping to develop the DOI-wide Early Detection and Rapid Response capability.

Global issues such as emerging diseases and invasive species warrant the Program's participation in leading international biological informatics initiatives. The Program, with direction from the Department, was instrumental in establishing IABIN, an effort to initiate the biodiversity information network for the Western Hemisphere. The Program also serves as the U.S. Focal Point(s) to IABIN and for the CHM of the CBD. It is also the U.S. Node of GBIF.

Gap Analysis Program

GAP has either finished or is completing its standard geospatial data layers for all states (approximately 83 percent coverage of the U.S. landmass), with the exception of Alaska. GAP's statewide data products include: 1) land cover assemblages, 2) vertebrate species distributions, and 3) characterizations of land stewardship. ReGAP (Regional GAP) projects are currently updating state projects and combining them into multi-state regions. Current ReGAP projects include the Mid-Atlantic (three states), Southeast (ten states), and Southwest (five states) regions of the United States, with plans for additional regional updates to provide national coverage over the next five years. Aquatic GAP presently supports projects in both large and small watersheds, including the Great Lakes; and the Missouri, Upper Tennessee, and Lower Colorado rivers. The USGS also continues to emphasize GAP research and the development of applications to serve the needs of DOI's land management agencies, including the USFWS and the BLM, as well as other Federal agencies such as the U.S. Forest Service.

Vegetation Characterization Program

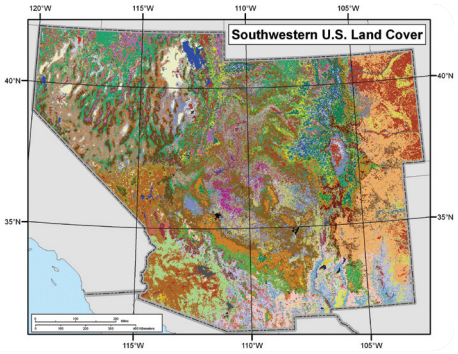
VEG, in cooperation with the NPS, has developed vegetation products for more than 30 national parks. This program is both the first to provide national-level descriptions of vegetation for a Federal agency and the first to create national vegetation standards for its data products. VEG results have provided input to fire fuel models at the national level. In cooperation with the USFWS, VEG products have also been generated for select National Wildlife Refuges.

Integrated Taxonomic Information System

ITIS is the comprehensive, standardized reference for the scientific names – as well as synonyms and common names – for all the plants and animals of North America and the surrounding oceans. The Biological Informatics Program took the lead in the creation and implementation of ITIS, a cooperative effort among many Federal agencies (e.g., EPA, USDA NRCS, NOAA, SI, and NSF). ITIS operates the largest taxonomic thesaurus and database of its kind in the world. It currently contains over 463,000 names for more than 213,000 species. In addition, ITIS and Species 2000 are also key partners in the creation and implementation of the Catalog of Life, the global taxonomy authority.

The Vision

The Biological Informatics Program will provide the premier enterprise system for accessing and disseminating integrated and reliable biological information.



Map credit: USGS Southwest Regional Gap Analysis Project

THE ROAD AHEAD

We will move forward to address the challenges with a broad federation of partners. Together, we will provide access to integrated data and information about the Nation's biological resources. This will maximize the return on the sizable investment in data generation and collection by improving their access and usability. Overcoming the challenges will require a fully digital, Web-based, distributed system; scientifically reliable data and information; local to global data coverage; and new information technologies for science-based and effective decision-making. The following goals and objectives provide a pathway to success.

Goal 1. Increase the availability and usefulness of biological resources data and information (Content). The scope of content of interest to the Program includes data, information, and knowledge of or about biological resources. Subject areas of interest include: biosystematics and nomenclature; landscape, ecosystems, and habitat; biodiversity and ecosystem assessment; species and populations; genes and ex-situ diversity; environmental drivers and ecological processes; and social and economic drivers. A complete list of the biological topics on which the Program focuses is included in Appendix 1.

Objectives:

1A. Obtain the broadest possible participation of both public and private sector stakeholders and partners in identifying high priority data and information needs.

1B. Increase access to data and information for USGS Science Centers and partners, as appropriate.

1C. Develop or select, implement, and promulgate standards and protocols to promote interoperability and information integration capabilities.

1D. Increase the variety of data and information available through the NBII.

Goal 2. Implement technologies and tools to integrate, analyze, visualize, and apply biological information to natural resource issues (Tools).

Tools developed, or acquired and adapted, by the Program include those necessary for the creation, discovery, retrieval, and management of appropriate content as well as those for manipulating and displaying selected content. Tools include catalogs and directories, decision support systems, visualization applications including mapping capabilities, models, standards, and protocols for the collection and reporting of data.

Objectives:

2A. Develop methods and techniques to improve information organization and expand discovery capabilities.

2B. Implement and enhance mechanisms to develop and support the creation and management of metadata for all types of data and information.

Goal 3. Develop, apply, and promote the adoption of standard practices, protocols, and techniques to enhance knowledge discovery and retrieval from various resources (Infrastructure).

The Program's infrastructure activities address the requirements of maintaining a distributed Internet-based network of resources. It includes the development or adoption of standards and protocols to increase data accessibility, integration, and usability. In some instances, it includes the installation, maintenance, and security of hardware and software resources to support specific requirements of the Program.

Objectives:

3A. Develop the hardware and software systems required to support the Program and all of its components.

3B. Support standards development to integrate core functionality of partner infrastructure.
3C. Support the computational capabilities required to apply models and other tools to available data.

Goal 4. Facilitate information science and technology research that supports the advancement of biological informatics capabilities (Research).

Research includes investigating new informatics techniques and technologies, as well as exploring how existing techniques and technologies might be applied in novel ways for possible application within the Program. Applied research areas of interest include data collection, mining and access, analysis, modeling, simulation, forecasting, interpretation, and visualization of biological and ecological data and information.

Objective:
4A. Conduct research in bioinformatics to support the discovery, management, visualization, and application of biological data and information.

Goal 5. Apply innovative technologies and best practices to improve the development, description, and dissemination of biological information to customers (User Services).

User services includes activities such as outreach, training, and technical assistance. These services target both users and contributors of the Program's data and other information resources. This goal also includes Web site usability protocols and information diffusion assessments to understand customer needs.

Objectives:
5A. Increase user knowledge and application of tools, techniques, and capabilities supported by the Program.
5B. Implement feedback mechanisms, periodic evaluations, and other techniques to improve user capabilities.

IMPLEMENTATION

This program will be implemented through a series of steps, carefully coordinated among the major Program components and partners. Table 1 (next page) provides a description of the strategies, outcomes, and measurements necessary to meet Program mission, goals, and objectives.

The activities and progress of Biological Informatics Program components are reviewed through an ongoing dialog with groups of interested stakeholders. These groups – comprising individuals from both the public and private sector and representing various disciplines including biology, library and information science, and computer science – have included the Biodiversity and Ecosystems Informatics Work Group (BioEco), a working group under the Office of Science Technology and Policy; the NBII Science Committee; the BRD Watchers; the GAP review committee; and working groups coordinated by ITIS, among others.

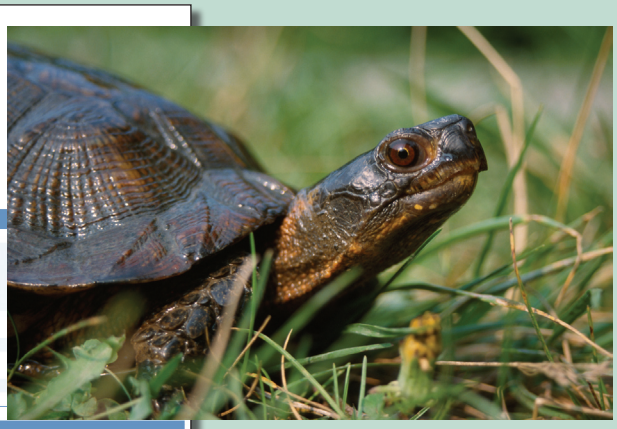


Photo credit: John J. Mosesso/NBII

This excerpt of a standard taxonomic report for the wood turtle (*Glyptemys insculpta*) is only a portion of the wealth of information available through ITIS.

Table 1. Strategies, Outcomes, and Measurements Necessary to Meet the Biological Informatics Program Mission, Goals, and Objectives

Goal 1: Increase the availability and usefulness of biological resources data and information (Content).		
Objective 1A: Obtain the broadest possible participation of both public and private sector stakeholders and partners in identifying high priority data and information needs.		
Strategy	Outcome	Measure
1A1. Develop and implement a process to engage new and existing partners to identify high priority data and information needs.	Process is in place to identify, acquire, and distribute highest priority data and information, as defined by the global biological community. Highest priority needs of stakeholders are identified.	• Increase in number and variety of partners engaged in priority-setting process (2005-2009), and • Number of workshops, consultations, and other mechanisms held in order to help set priorities (2005-2009).
Objective 1B: Increase access to data and information for USGS Science Centers and partners, as appropriate.		
Strategy	Outcome	Measure
1B1. Integrate USGS Science Center and other USGS Discipline activities into the development, implementation, and management of program priorities.	Close collaboration occurs between USGS Science Centers, other USGS Disciplines, and Biological Informatics Program components on biological information generation, management, and dissemination activities.	• Number of USGS Science Center and other USGS Discipline activities that collaborate in BIO priority-setting (2005-2009), • Number of USGS Science Center and other USGS Discipline activities that utilize and contribute to BIO products (2005-2009), and • Percentage of Geography, Water, and Biology customers satisfied with the ease and timeliness of access to information products (GPRA/PART Measure; 2005-2009).
1B2. Integrate partner activities into the development, implementation, and management of program priorities.	Close collaboration occurs between partners and Biological Informatics Program components on biological information generation, management, and dissemination activities.	• Number of partner activities that utilize and contribute to BIO products (2005-2009).
1B3. Develop multilingual capabilities to discover, retrieve, and use biological resource data and information.	International and non-English-speaking users gain access to previously English-only biological data, tools, and applications.	• Multilingual versions of search software for Inter-American Biodiversity Information Network and other program components (2005).

Strategy	Outcome	Measure
1B4. Increase content in NBII regional and thematic nodes in response to the priority biological issues of concern to our customers and stakeholders.	Content in NBII regional and thematic nodes is increased in response to the priority biological issues of concern to our customers and stakeholders; including expanded availability of GAP and VEG data and products.	• Increased content in regional and thematic nodes reflecting program and partner priorities (2005-2009), • Regional updates to nationwide GAP data (2005-2009), • Aquatic GAP projects to update data on U.S. watersheds (2005-2009), • Percentage of U.S. land with land characterization and species distribution information available for resource management decision-making updated in the last 5 years (PART Measure; 2005-2009), • Percentage of North American migratory birds for which scientific information on their status (species distribution and number) and trends are available in a standardized and exchangeable format, to improve conservation plans of Federal and State agencies (PART Measure; 2005-2009), • Percentage of North American amphibians and reptiles for which scientific information on their status (species distribution) are available in a standardized and exchangeable format, to improve conservation plans of Federal and State agencies (PART Measure; 2005-2009), • Percentage of North American mammals for which scientific information on their status (species distribution) are available in a standardized and exchangeable format, to improve conservation plans of Federal and State agencies (PART Measure; 2005-2009), and • Percentage of surface area with temporal and spatial monitoring, research, and assessment/data coverage to meet land use planning and monitoring requirements (DOI One Plan Number of square miles assessed by GAP, GPRA/PART Measure; 2005-2009).

Objective 1C: Develop or select, implement, and promulgate standards and protocols to promote interoperability and information integration capabilities.		
Strategy	Outcome	Measure
1C1. Expand the Biocomplexity Thesaurus, ITIS, and other authorities required to promote biological and ecological information discovery, retrieval, and integration.	Authorities are accepted and used by the biological informatics community.	• Number of terms added to authority files (2005-2009), and • Increased use of authority files (2005-2009).
1C2. Develop and support the use of metadata standards and exchange protocols for all types of data and information.	There is widespread use of metadata standards and exchange protocols for all types of data and information.	• Number of NBII Clearinghouse metadata records (GPRA/PART Measure; 2005-2009), and • Increase in number of people trained in use of metadata standards and exchange protocols (2005-2009).
Objective 1D: Increase the variety of data and information available through the NBII.		
Strategy	Outcome	Measure
1D1. Target information types and sources underrepresented among currently accessible resources.	Greater range and depth of data are served by Biological Informatics Program, components and tools.	• As U.S. node to GBIF, serve data from worldwide natural history museums, herbaria, zoological and botanical gardens, and aquaria (2005-2009), • Percentage of Natural History Museum specimen data records available online via the NBII, to assist researchers in identifying and addressing threats to human and animal health (PART Measure; 2005-2009), • Amount of invasive species data and information available online via the NBII, to assist in modeling and forecasting the spread of invasive species (PART Measure; 2005-2007), • Amount of fire-related data and information available via the NBII, to assist land managers in fire management decision-making (PART Measure; 2005-2009), • Percentage of U.S. Federally-listed threatened and endangered or indicator fish species for which scientific information on species status is available in a standardized and exchangeable format to improve conservation plans of Federal and State agencies (PART Measure; 2005-2009), and • Number of cumulative gigabytes managed (GPRA/PART Measure; 2005-2009).

Goal 2: Implement technologies and tools to integrate, analyze, visualize, and apply biological information to natural resource issues (Tools).		
Objective 2A: Develop methods and techniques to improve information organization and expand discovery capabilities.		
Strategy	Outcome	Measure
2A1. Identify and prioritize, in consultation with partners and users, various tools that will advance the functionality of biological information services.	Applications and tools are developed and implemented for a wide variety of geospatial, temporal, and trends-based biological data.	• Increased use of new tools and applications (2007-2009).
2A2. Implement technologies to locate, access, and analyze large data sets and present data and information in forms appropriate to the user.	Technologies are implemented to locate access, integrate, analyze, and present data and information in forms appropriate to the user and the application.	• Number of systematic analyses and investigations delivered to customers (GPRA/PART Measure; 2005-2009).
Objective 2B: Implement and enhance mechanisms to develop and support the creation and management of metadata for all types of data and information.		
Strategy	Outcome	Measure
2B1. Develop suites of integrated tools (i.e., "scientific desktops") targeted to the information needs of selected user communities.	Tools are available for user communities that meet their specific needs.	• IABIN Biobot search utility refined for multilingual users in Latin America, Caribbean (2006), • Portal communities created and managed for stakeholders (2005-2009), and • IABIN Invasives Information Network (I3N) cataloging tool (2006).
2B2. Provide applications that increase the efficiency of content management.	Content is managed efficiently.	• Average cost per gigabyte of data available through servers under program control PART (Measure; 2005-2009).
Goal 3: Develop, apply, and promote the adoption of standard practices, protocols, and techniques to enhance knowledge discovery and retrieval from various resources (Infrastructure).		
Objective 3A: Develop the hardware and software systems required to support the Program and all of its components.		
Strategy	Outcome	Measure
3A1. Develop innovative hardware and software systems required to support the Program and all of its components.	Program needs are met by new hardware and software.	• Increased use of new hardware and software (2005-2009).

Objective 3B: Support standards development to integrate core functionality of partner infrastructure.		
Strategy	Outcome	Measure
3B1. Provide standards, protocols, and techniques to support accessibility, integration of data, and usability among partners.	Partners use standards, protocols, and techniques provided.	• Increased use of standards, protocols, and techniques by partners (2005-2009)
Objective 3C: Support the computational capabilities required to apply models and other tools to available data.		
Strategy	Outcome	Measure
3C1. Ensure adequate computational and storage capacity to apply models and other applications.	Program has adequate computational and storage capacity to apply models and other applications.	• Growth in storage and computational capacity devoted to modeling and applications (2005-2009).
Goal 4: Facilitate information science and technology research that supports the advancement of biological informatics capabilities (Research).		
Objective 4A: Conduct research in bioinformatics to support the discovery, management, visualization, and application of biological data and information.		
Strategy	Outcome	Measure
4A1. Support exploratory research workshops and symposia to identify new areas of research.	New areas of research are identified at workshops and symposia.	• Increased grant proposals on topics identified in workshops and symposia (2005-2009).
4A2. Work with the broader scientific community to conduct and leverage innovative research through collaborative plans and programs.	Innovative research is conducted through collaborative plans and programs.	• Increased funding or leveraging for collaborative research with the broader scientific community (2005-2009).
Goal 5: Apply innovative technologies and best practices to improve the development, description, and dissemination of biological information to customers (User Services).		
Objective 5A: Increase user knowledge and application of tools, techniques, and capabilities supported by the Program.		
Strategy	Outcome	Measure
5A1. Provide technical assistance and training to users.	New users are trained on new tools, applications, and metadata.	• Number of formal workshops or training provided to customers (instances/issues/events) (GPRA/PART Measure; 2005-2009).
5A2. Expand outreach and education activities.	Outreach and education activities are expanded to a wider audience.	• Increased volume of updated materials such as teacher and student tool kits and online biological resources (2005-2009), and • Increased presence at community events (2005-2009).
Objective 5B: Implement feedback mechanisms, periodic evaluations, and other techniques to improve user capabilities.		
Strategy	Outcome	Measure
5B1. Conduct periodic evaluations and solicit feedback to improve user satisfaction.	User feedback is effectively collected and integrated into program planning and activities.	• Increase in user feedback for planning process and activities (2005-2009).

APPENDIX 1
Biological Informatics Program Subjects of Particular Interest

Biodiversity and Ecosystem Assessment

- Aquatic GAP
- Intensive monitoring
- Man and the Biosphere Program
- NSF Long Term Ecological Research Program
- Survey networks
- Terrestrial GAP
- Vegetation Characterization

Biosystematics and Nomenclature

- Ecoregions
- Life zones and bioregions
- Taxonomic authorities (ITIS)

Collections and Reference Institutions

- Bibliographic services
- Botanical gardens
- Libraries
- Museums
- University collections
- Zoological gardens

Environmental Drivers and Ecological Processes

- Atmosphere
- Climate change
- Contaminants
- Nutrient cycling
- Soils
- Water

Genes and Ex-situ Diversity

- Plant and animal registries
- Seed and germ plasm banks
- Tissue and gene banks

Landscape, Ecosystems, and Habitat

- Coastal
- Desert
- Ecoregions
- Forest
- Freshwater
- Habitats
- Land and water cover
- Marine
- Montane
- National map
- Prairie
- Protected Areas
- Reef
- Riparian
- Riverine
- Tundra
- Vegetation classification
- Wetlands

Social and Economic Drivers

- Biocomplexity
- Human demographics
- Land use planning
- Sustainable development
- Urban sprawl

Species and Populations

- Anadromous fish
- Culturally important species
- Endangered species
- Endemic species
- Harvested species
- Invasive species
- Marine mammals
- Migratory birds
- Wild relatives of crop species

APPENDIX 2
List of Abbreviations

ARS	Agricultural Research Service
BIO	Biological Informatics Office
BLM	Bureau of Land Management
BRD	Biological Resources Discipline
CBD	Convention on Biological Diversity
CHM	Clearing-House Mechanism
DOI	Department of the Interior
EPA	Environmental Protection Agency
FGDC	Federal Geographic Data Committee
GAP	Gap Analysis Program
GBIF	Global Biodiversity Information Facility
IABIN	Inter-American Biodiversity Information Network
ITIS	Integrated Taxonomic Information System
NABIN	North American Biodiversity Information Network
NBII	National Biological Information Infrastructure
NOAA	National Oceanic and Atmospheric Administration
NPS	National Park Service
NRCS	Natural Resources Conservation Service
NSF	National Science Foundation
PBIF	Pacific Biodiversity Information Forum
PCAST	President's Committee of Advisers on Science and Technology
PITAC	President's Information Technology Advisory Committee
R&D	Research & Development
ReGAP	Regional GAP
SI	Smithsonian Institution
USDA	U.S. Department of Agriculture
USFWS	U.S. Fish and Wildlife Service
USGS	U.S. Geological Survey
VEG	Vegetation Characterization Program