World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:
http://www.ifla.org/IV/ifla71/Programme.htm

June 6, 2005

Code Number: **037-E**
Meeting: **82 SI - Government Information and Official Publications**

Developing the Digital World of Government Information and Official Publications: A View from the United States

Thomas F. Lahr (Presenter)
Deputy Associate Chief Biologist for Information
National Biological Information Infrastructure
US Geological Survey
12201 Sunrise Valley Dr. MS 302
Reston, VA, 20192 USA

Robin Haun-Mohamed
Development Project Manager
Program Development Service
United States Government Printing Office
732 North Capitol Street NW
Washington, DC 20401 USA

Eleanor G. Frierson
Deputy Director
National Agricultural Library
10301 Baltimore Avenue, Room 200
Beltsville, MD,20705-2351 USA

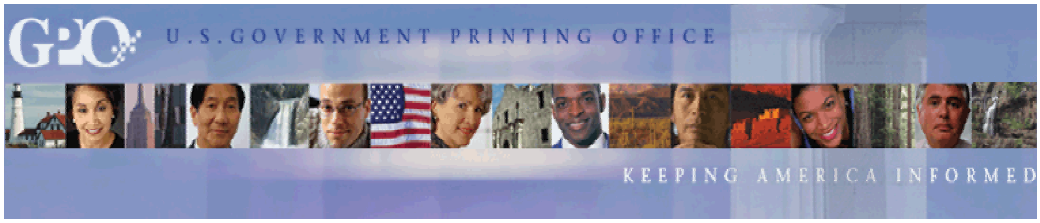
Abstract:
U.S. government agencies are creating new digital government information services, many through collaboration and partnerships within and outside the government. Massive digitization and digital preservation initiatives are proposed or in process. Support for eGovernment at the highest levels is creating opportunities for innovation and impetus for change in the way government information and official publications are created and made available. Initiatives discussed in the presentation will include two government-supported partnerships: the National Biological Information Infrastructure and the United States Government Printing Office national digitization plan followed by an overview of a cross-government partnership Science.gov. These innovative US Government systems and programs provide new value added access and retrieval options to U.S. government information which was previously unavailable or time consuming to locate.

The **National Biological Information Infrastructure** (NBII) a program of the U.S. Geological Survey, provides increased access to data and information on the nation's biological resources.

Science.gov, an inter-agency (including GPO and USGS/NBII), gateway portal, provides value added access to selected, authoritative U.S. government science and technology information.

Good Morning. My talk will give details on how U.S. government agencies are creating new digital government information services – often through collaboration and partnerships within and outside the government. I will discuss projects specific to the U. S. Government Printing Office (GPO), the National Biological Information Infrastructure (NBII) and a cross-agency information service - Science.gov.

The **U.S. Government Printing Office** (GPO), a U.S. Government agency since 1861 and a Science.gov partner, procures, produces, and disseminates tangible and electronic publications of the U.S. Government.



GPO has three essential missions:

- To provide all U.S. Government agencies with expert publishing and printing services, on a cost recovery basis.
- To provide permanent public access to the publications and other information products of the U.S. Government. This is accomplished in partnership with approximately 1,250 U.S. libraries in the Federal Depository Library Program (FDLP).
- To provide copies of printed and electronic publications and other U.S. Government information products to the general public.

Federal Depository Library Program

The FDLP ensures no-fee permanent public access to the publications from all three branches of the U.S. Government.

Public distribution of U.S. Government publications dates back to 1813, but the FDLP officially began in 1895 when it was transferred to the Government Printing Office.

The FDLP includes over 1,250 depository library partners throughout the U.S. and its territories.

- 53 are Regionals (100% selection rate)
- Others are Selectives (Select the publications that best meet the information needs of their communities)

U.S. Government Printing Office legacy digitization project

GPO plans to digitize a complete collection of tangible U.S. Government publications.

GPO is also identifying and acquiring digital publications through the printing process and from agency Web sites.

Together these initiatives will create a comprehensive electronic collection of past, present, and future U.S. Government publications.

Since the inception of the Federal Depository Library Program, the American public has had no-fee access to many of the information products produced by the Federal government. Until the early 1990s, this information was available largely as print, microfiche, maps, or other tangible formats. As GPO moved to a

predominantly electronic FDLP, steps were taken to identify, catalog, and archive online electronic resources.

The ease with which the public is able to access online electronic resources via the World Wide Web has generated requests for additional resources to be available in electronic format. These requests prompted many in the depository and the larger information community to begin digitizing the materials in the legacy collections.

GPO, in coordination with the depository libraries and other interested stakeholders in the information community, is developing a digitization plan for the material in the legacy collection. As the U.S. Government has an obligation to provide permanent public access to its information products, both current and historic, the objective is to ensure that the digital collection is available, in the public domain, for no-fee permanent public access through the FDLP.

The scanned images and associated metadata will be produced at a level of quality that is adequate to support preservation as well as future iterations of derivative products. The project has a target date of December 2007 to digitize at least 70% of the historic collection and GPO is evaluating ways to successfully achieve this goal.

Project scope: What is to be digitized and disseminated electronically?

U.S. Government publication - informational matter which is published as an individual document at Government expense, or as required by law (44 U.S.C.)

National Bibliography of U.S. Government Publications

National Collection of U.S. Government Publications

GPO is digitizing non-classified U.S. Government publications. A Government publication as defined by statute (Title 44 U.S. Code) as “informational matter which is published as an individual document at Government expense, or as required by law.”

As part of its mission to provide public access to U.S. Government publications, for over 100 years, GPO has provided bibliographic control over U.S. Government publications. It continues to do so through the National Bibliography of U.S. Government Publications.

All of the publications in the National Bibliography are part of the National Collection of U.S. Government Publications. The National Collection, currently in development, will consist of multiple collections of tangible and digital publications in light and dark archives located in multiple sites and operated with various partners within and beyond the U.S. Government.

Project scope: Current and legacy publications

In scope Federal publications for digitization –

- Legacy, or historic, publications
- Current paper publications that are not in electronic format
- Other tangible materials, including multi-media formats

Legacy, or historic, publications

GPO, in coordination with the depository libraries and other interested stakeholders in the information community, is developing a digitization plan for

the material in the legacy collection. The initial priority for legacy collection digitization is to provide the public with online versions of major U.S. Government publications, including the United States Code, Congressional Record, Bound Congressional Record, Federal Register, Code of Federal Regulations, Congressional Bills, Slip Laws, Statutes at Large, and United States Reports, to increase the retrospective coverage of these popular *GPO Access* databases.

Other publications that are of special interest to the public or endangered due to brittle paper are also being evaluated for early digitization. Depository libraries were surveyed in 2004, resulting in a list of individual titles libraries would like to see digitized early in the project.

Current publications

The majority of publications GPO will scan will be part of the historic collection; however, GPO may also digitize some current publications when it is unable to acquire an electronic format from the issuing Federal agency.

GPO's Digital Conversion Services, the unit responsible for digital services, will also convert a wide range of other materials, including microformat, maps, multi-media and tangible electronic publications. GPO continues to work with the information community to develop specifications suitable for the conversion of these specialized formats.

Authentication

Authentic Content: Content that is verified by GPO to be complete and unaltered when compared to the version approved or published by the publishing agency.

Official Content: Content that falls within the scope of the National Collection of U.S. Government Publications and is approved by, contributed by, or harvested from an official source in accordance with accepted program policy and procedures.

Authentication

As part of GPO's mission to provide permanent public access to official and authentic U.S. Government publications, GPO is currently implementing Public

Key Infrastructure (PKI) digital signatures to ensure the authenticity of its electronically disseminated content, including converted content.

Official Content is content that falls within the scope of the National Collection of U.S. Government Publications and is approved by, contributed by, or harvested from an official source in accordance with accepted program policy and procedures.

Authentic Content is content that is verified by GPO to be complete and unaltered when compared to the version approved or published by the publishing agency.

GPO's authentication initiatives enable GPO to certify documents as authentic and official, and provide the capability for users to determine that no changes have been made to files since they were authenticated by GPO. The certification process produces an integrity mark that is used to convey authentication information to users.

Version Control

Version control is the acquisition, bibliographic control, access, preservation, and authentication of all “unique manifestations” of a publication.

Version control

Version control is the acquisition, bibliographic control, access, preservation, and authentication of all “unique manifestations” of a publication. Digital content may change on a very frequent basis, and it is easy to transform content without identification or notification to users.

Version control may be accomplished by assessing various document attributes (e.g., structure, content, and format), creating metadata about these attributes, monitoring changes to the attributes, updating the metadata to indicate changes to the attributes, and creating links to related documents.

GPO is working with others to identify the appropriate triggers that necessitate the acquisition and preservation of a new version for various categories of

documents. For example, a version of a continuously updated digital map might be harvested on a recurring basis.

Access to digitized U.S. Government publications

Access to the digitized publications will be through *GPO Access* and the network of Federal depository libraries.

GPO Access
http://www.gpoaccess.gov

Federal Depository Library Program
http://www.gpoaccess.gov/fdlp.html

The digital preservation masters and the associated metadata will be preserved in GPO's electronic archives. No-fee public access to the content will be available through derivative files on *GPO Access*, a Web-based service that provides free electronic access to a wealth of important information products produced by the U.S. Government.

Federal depository libraries provide access to *GPO Access* for those otherwise without Internet access as well as expert reference assistance locating and using the publications.

The U.S. National Archives and Records Administration (NARA) recognizes GPO as an official archival affiliate for the electronic content on *GPO Access*. NARA assumes legal custody of the U.S. Government publications on GPO Access as part of the official Archives of the United States while GPO retains physical custody and is responsible for permanent public access and preservation of the records.

Current GPO Projects Associated with the Digitization Plan

- Registry of U.S. Government Publication Digitization Projects
- Web Harvesting
- Future Digital System

To enhance access to digital resources and to provide information about GPO and other digitization projects in process, GPO is developing a comprehensive registry of U.S. Government publication digitization projects.

GPO is proceeding with a pilot project for automated Web Harvesting of the Environmental Protection Web site to identify fugitive electronic publications.

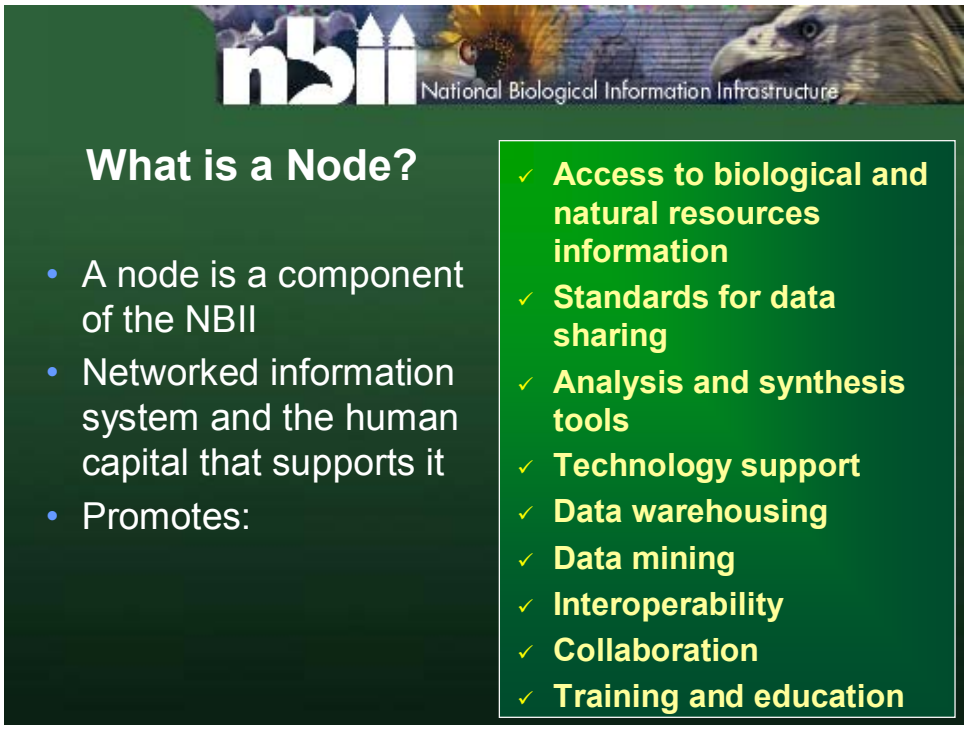
GPO plans to acquire and manage digitized and born digital publications in one content management system. This system, known as the Future Digital System, is currently in development, based on the Open Archival Information System model. The entire system, to be comprised of smaller interoperable systems including GPO's new integrated library system, is planned for full implementation in 2007.

Thank you very much for your interest in the U.S. Government Printing Office's project to digitize U.S. Government publications.

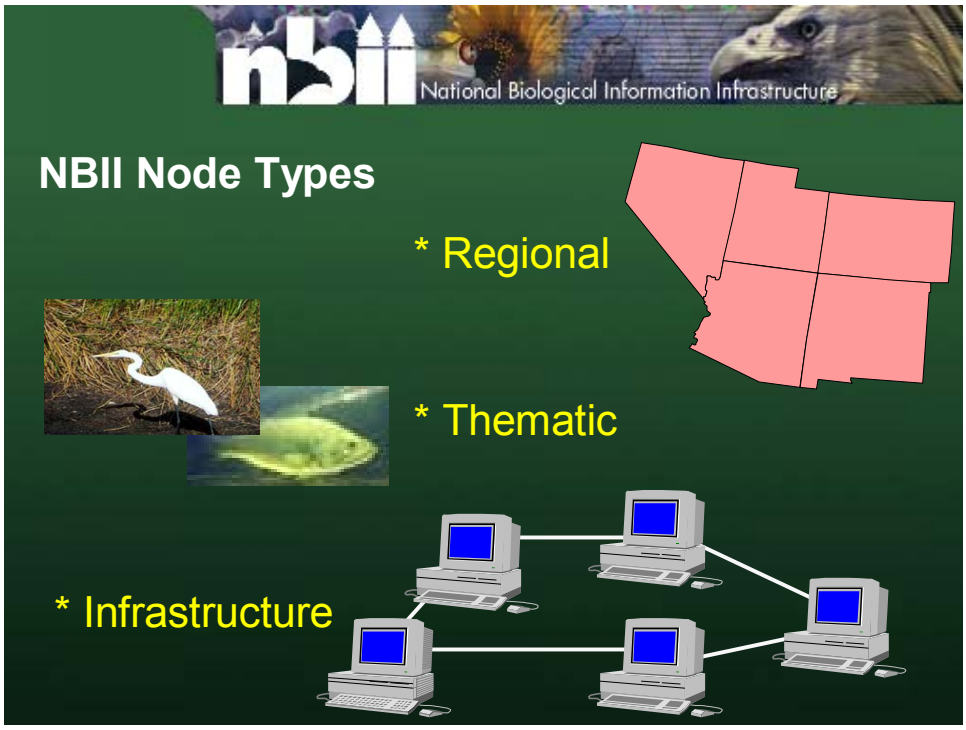
Another U. S. Government Program, The National Biological Information Infrastructure (NBII) <www.nbii.gov> was created to provide a comprehensive framework that allows information on biodiversity and ecosystems to be accessed readily and used effectively by a variety of audiences. The NBII Program is a broad, collaborative undertaking to provide increased access to data and information on the nation's biological resources. Coordinated by the U.S. Geological Survey (USGS), this user-friendly, Web-based, distributed system provides the public with natural science data and information that are scientifically reliable. The NBII makes it possible to collect and organize biological information from partners, add value, and provide products and services back to these partners and other users to meet their needs.

NBII end-users are those who manage, study, use, or simply enjoy biological resources. Typically, these users come from both the public and private sectors — scientists; planners; decision-makers in industry as well as federal, state, and local government agencies; international entities; teachers and students; and other private citizens. NBII end users come to the Program with a variety of information needs.

The Program's motto "Building Knowledge Through Partnerships" reflects its goal of uniting the intellectual capital of the private sector with the government's commitment to meeting the information needs of the country's natural resource managers and stewards. Data and information provided by and to these groups include diverse, high-quality biological databases, information products and services, and a variety of analytical tools. NBII partners and collaborators also work on new standards, tools, and technologies that make it easier for NBII customers to find, integrate, and apply biological resources information. The NBII's partnership structure leverages the resources of 200+ organizations.



One of the key components of the NBII is a “node”-based structure that is being developed to ensure broad partnerships and information from all sectors of society. The establishment of these nodes allows the NBII to provide the community of users with rapid access to data and information on biological resources as well as national -- and increasingly, international -- coverage on a range of major biodiversity and environmental issues. NBII nodes are focal points through which key elements for providing the information and services envisioned by the NBII are made available.



NBII nodes are of three types:

Regional -- Have a geographic orientation. By taking a regional approach, local data issues, data collectors, and owners are involved in the process. They also allow people closer to the issues and the partner groups to form active coalitions in addressing biological issues within their regions. For example:

- Offering information to evaluate strategies and practices associated with helping to stem the decline of salmon in the Pacific Northwest as well as the management of the region's forest ecosystems.
- Supporting information systems addressing interagency biodiversity and watershed assessments in California, the Pacific Coast, and southwestern desert ecosystems.

Thematic -- Focus on a particular biological issue, such as avian bird conservation, providing the support and infrastructure to help address these issues. Such issues often cut across multiple geographic areas and have national significance. An example:

- Invasive Species –the number one environmental challenge of the twenty-first century, they represent a significant economic challenge to our nation. This node provides access through a single Web portal to a vast array of information on invasive species throughout the nation.

Infrastructure: --The infrastructure and technical evolution the NBII is undergoing. Emphasis is placed on many varied infrastructure areas including: the networks and their security, standards, vocabulary, hardware/software, content management tools, geospatial applications, and

search & retrieval and data analysis tools.

The NBII is dedicated to helping us better understand and preserve our nation's living resources by unleashing the enormous power of the data and information resources that describe them. More specifically, the NBII is committed to ensuring biodiversity data are organized, accessible, and available to those who make decisions regarding the care, use, and conservation of natural resources. As technology changes, so too will the NBII's approaches to solving this important, ongoing challenge. But while this presentation only gives selected NBII successes, the Program's focus is on the future. We invite all interested parties from around the nation and the world to join us in helping to make the NBII an even more effective tool to facilitate environmental decision-making.

I would now like to move an overview of a United States inter-agency effort that links and combines data and information from both of the previous examples as well as much other scientific information, Science.gov

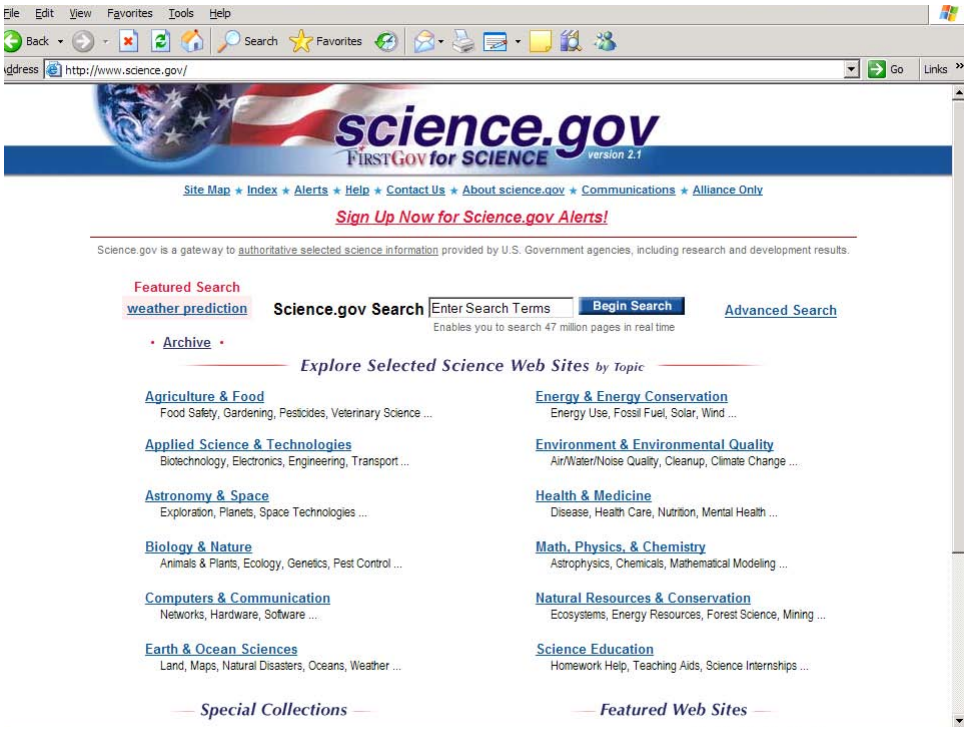
Science.gov: Providing Simple, Unified Access to Scientific and Technical Information through Collaboration and Innovation

You now have heard how two U.S. government agencies are creating new digital government information services. Both of these agencies also participate in Science.gov which provides new value added access and retrieval options to information which was previously time consuming or difficult to locate. This gateway to U. S. government science information allows simple unified searches across 30 databases and more than 1,700 science Web sites and accesses over 47 million pages of government science information.

Science.gov Vision

Provide cross-agency, gateway access to selected, authoritative U.S. government science and technology resources for the science-attentive citizen, including science professionals, students, teachers, business people, and members of the public interested in science

Science.gov was launched in December 2002, providing for the first time wide public access and a unified search of the government's vast stores of scientific and technical information as stated in the Science.gov vision.



This look at the Science.gov top page shows the key option available to all users, to enter search terms which will simultaneously search the web sites and databases. Science.gov has more than 1700 Web sites selected and cataloged by agency content managers which are organized for browsing under one or more of 12 topics and 175 sub-topics. Free-text searching of the selected Web sites is based on the text of these Web sites and their linked sites (to one level). Science.gov also allows the user to explore selected Web sites by topic, or to select a featured search or featured Web sites on current "hot topics." It also highlights new features as they appear, such as the new "Alerts" feature. Science.gov allows users to search the surface Web as well as the deep Web, where traditional search engines cannot go. It provides unified simultaneous searching of 30 databases that are generally not accessible to Web search engines. This special invisible Web searching was developed for science.gov. There is a single search interface for Web sites and databases. All information is free, with no registration required.

Science.gov Alliance

- **Dept. of Agriculture**
- **Dept. of Commerce**
- **Dept. of Defense**
- **Dept. of Education**
- **Dept. of Energy**
- **Dept. of Health and Human Services**
- **Dept. of the Interior**
- **Environmental Protection Agency**
- **National Aeronautics and Space Administration**
- **National Science Foundation**
- **Government Printing Office**

- **Support and coordination by CENDI, including support from the National Archives and Records Administration**

The Science.gov Alliance is a voluntary group of United States government agencies with science missions. These agencies are involved in the development, operation and enhancement of Science.gov. The Alliance includes the major U. S. science mission agencies and departments. Administrative support and coordination is provided by CENDI (www.cendi.gov), an interagency working group of senior U. S. government scientific and technical information managers. Agencies contribute software, servers, and in-kind services, fund usability testing, contribute content and perform cataloging functions. Each agency selects its best science information for inclusion in science.gov. The Alliance co-chairs serve as representatives to FirstGov, www.firstgov.gov.

Challenges

- **Broad scope of Federal research and development missions in science and technology**
- **Wide ranging interests of potential audiences**
- **Taxonomy issues given the broad scope and audience**
- **Blending resources from different agencies into cohesive functionality and page design**
- **Resources, funding and sustainability**

We have identified several challenges, and many are inherent in the nature of the enterprise -- the breadth of federal Research &Development and the scope of the audience that might be interested in this information. These issues are highlighted in the decisions of what URLs to select and the way the taxonomy should be expressed. We are still working to present technical information to a wide audience. Of course, the major challenge has been the blending of resources from different agencies into a cohesive product for the users.

What's Next?

- **Development of versions 3.0 and 4.0 is funded**
- **Evaluate customer satisfaction**
- **Refine and enhance design and search capabilities**
- **Continue adding content and databases**
- **Continue reviewing the taxonomy**
- **Continue linking to other Firstgov portals such as kids.gov**
- **Continue to add members to the Alliance**
- **Continue to develop support and sustaining resources**
- **Continue what we do best – collaborate and innovate**

By the end of 2005 Version 3 of Science.gov will be available which will provide Enhanced precision searching and MetaRank - an enhanced ranking system using metadata and Boolean searching. Version 4 which is in the planning stages for 2006 and beyond will have enhanced relevancy ranking (DeepRank) and full-text relevancy ranking for precision search. As we produce technical enhancements, we will evaluate customer satisfaction and continue to refine and enhance design and search capabilities. Since version 1 in 2002, we have routinely added content and databases, reviewed the taxonomy, and linked to other First.gov portals such as kids.gov. We will work on adding members to the Alliance and developing support and sustaining resources. Very importantly, we will continue what we do best – collaborating and innovating.

We have heard today about three innovative U.S. Government systems and programs which provide new value added access and retrieval options to government information which was previously unavailable or time consuming to locate. Thank you very much for your time and attention. Questions?

References

Authentication, Council Briefing Topic, Spring 2005, http://www.access.gpo.gov/su_docs/fdlp/council/briefing_topics/index.html

Baseline Requirements for Digital Reformatting and Delivery of Legacy Federal Document Collections, November 29, 2004, http://www.access.gpo.gov/su_docs/fdlp/pubs/crl_digital_baseline.pdf

Committee on Environment and Natural Resources Subcommittee on Ecological Systems (2001) *Ecological Forecasting: Agenda for the Future*, Reston, VA: U.S. Geological Survey.

Concept of Operations for the Future Digital System, May 16, 2005, http://www.gpo.gov/projects/pdfs/FDsys_ConOps_v2.0.pdf

Essential Titles for Public Use in Paper or Other Tangible Format, http://www.access.gpo.gov/su_docs/fdlp/pubs/estitles.html

Information Dissemination Implementation Plan, FY 2005-2006. April 14, 2005, http://www.access.gpo.gov/su_docs/fdlp/pubs/id_plan/index.html

Memorandum of Understanding (MOU) Between the Government Printing Office and the National Archives and Records Administration, August 12, 2003, http://www.gpoaccess.gov/about/naramemofinal.pdf

National Research Council (1993) *A Biological Survey for the Nation,* Washington, DC: National Academy Press.

National Research Council (2001) *Future Roles and Opportunities for the U.S. Geological Survey*, Washington, DC: National Academy Press.

NBII Metadata Clearinghouse (1997) Home page of the NBII Metadata Clearinghouse. Retrieved September 12, 2003 from the National Biological Information Infrastructure, http://metadata.nbii.gov

President's Committee of Advisers on Science and Technology (1998) *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.

Report from the Meeting of Experts on Digital Preservation: Digital Preservation Masters. Revised. June 18, 2004, http://www.gpoaccess.gov/about/reports/preservation.html

Report on the Meeting of Experts on Digital Preservation: Metadata Specifications. June 14, 2004, http://www.gpoaccess.gov/about/reports/metadata.html

Strategic Vision for the 21st Century, U.S. Government Printing Office, December 1, 2004, http://www.main.gpo.gov/pub_print/STRATEGICPLAN.html

Versioning: The Future of Managing Changes in Digital Content at the U.S. Government Printing Office, Council Briefing Topic, Spring 2005, http://www.access.gpo.gov/su_docs/fdlp/council/briefing_topics/index.html