National Oceanic and Atmospheric Administration (NOAA) 2011 Report to Congress on Data and Information Management

Introduction

NOAA's mission within the Department of Commerce (DOC) is to understand and predict changes in climate, weather, oceans, and coasts, to share that knowledge and information with others, and to conserve and manage coastal and marine ecosystems and resources. In support of this mission, NOAA provides critical end-to-end data management (including observation, data archiving, stewardship, access, and assessment) for much of the Nation's environmental data.

In October 1992, Public Law 102-567, Section 106[1] directed the Secretary of Commerce to perform a needs assessment on the adequacy of NOAA's environmental data and information systems looking forward ten years. The law specifically called for the Secretary to take into consideration the need to: i) provide adequate *archive capacity*, ii) establish, develop, and maintain *information access*, iii) develop effective *interagency interfaces*, iv) develop and use nationally-accepted *formats and standards*, and v) integrate and interpret data from different sources to produce *information for policymaking*.

In its subsequent reports, NOAA has documented significant maturation of its digital archives, enterprise architecture for information bases, internet interfaces for ready access, and standardized formats and organizational structures for integration and interpretation. This document summarizes key activities in these areas and serves as a capstone to this reporting requirement. The progress over the past two decades is summarized in the November 2010 NOAA policy on the management of environmental data and information[2], which simply states that *environmental data will be visible, accessible, and independently understandable to users*.

Archival Capacity

Since 1992, NOAA archives have faced tremendous growth from incoming environmental data requiring safe storage and stewardship (**Figure 1**). Such growth has been driven by increasing and diverse large-array data streams from satellite sensing systems, Next-Generation Weather Radar (NEXRAD), climate, weather, and ocean models, multi-beam and side-scan sonar systems, and other *in situ* and remotely-sensed sources. All of these sources have increased data density and resolution, reflective of large advances in technology over the past twenty years, and are projected to continue to grow exponentially, contributing to an archive size of over 250 petabytes (250,000 terabytes) by 2030.

[1] NOAA Authorization Act of 1992
[http://thomas.loc.gov/cgi-bin/query/F?c102:7:./temp/~c102GqzQN7:e13865:#]
[2] NOAA Administrative Order 212-15
[http://www.corporateservices.noaa.gov/ames/NAOs/Chap_212/naos_212_15.html]

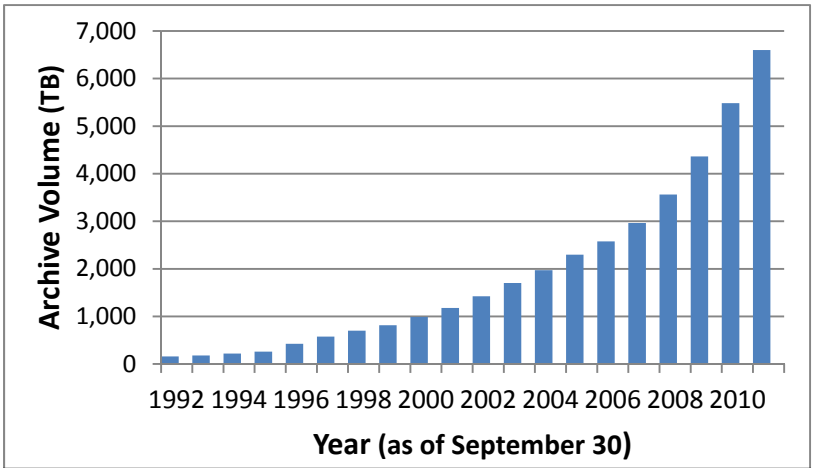

Figure 1. Growth in NOAA Environmental Data Archive Holdings, 1992-2011

NOAA has responded to this explosion in archive volume through continuous improvement in its information capacity, driven both by technological innovation and responsive governance and policymaking. A groundbreaking *What to Archive* policy[3] for environmental data was developed and formalized in December 2008. In articulating four broad steps, including identification of records, appraisal of records, decision and approval, and implementation, the policy serves as a standard for information acceptance into NOAA's archive. In May 2009, the procedure was accepted by the National Archives and Records Administration (NARA) as a best practice and was added to the NARA Toolkit for Managing Electronic Records web portal.

Information Access

As part of its 2011 Scientific Integrity Policy[4], NOAA outlined principles for ensuring transparency and reproducibility of its scientific work, including the need to provide ready access to the resultant data streams. The ability to implement such a policy reflects NOAA's position as a trusted manager of scientific information covering a breadth of formats and disciplines and ensuring consistency with privacy and classification standards and data sharing policies promoted by the Department of Commerce and NOAA.

NOAA's progress in providing data access dates back to the 1990s. Work during the latter part of that decade culminated in 2001 with the development of the interactive data access systems called the NOAA National Data Center Virtual Data System (NVDS), which made available priority data sets for users via the Internet. NVDS transformed customer service at NOAA, which led to cost reductions for servicing user requests with the transition from offline (*e.g.*, telephone, e-mail, fax, and letter requests) to online (*e.g.*, online ordering systems) access (**Figure 2**). As web access has become available to most citizens, customers have increasingly accessed NOAA data online at no charge as the cost savings from automation have been passed along to the public.

The substantial increase in online data delivery reflects NOAA's response to concerns dating back to the mid-1990s, specifically, that the lack of adequate Internet connectivity would be a major limiting factor and that stretched network resources would dictate which new technologies could be incorporated into NOAA's operational systems. In 2001, a plan for improving management of

[3] What to Archive Policy [http://www.joss.ucar.edu/daarwg/feb09/NOAA_Records_Brochure_4_pages_Dec_9.pdf]
[4] NOAA Administrative Order 202-735D, [http://www.noaa.gov/scientificintegrity/PDFs/DRAFT_NAO_202-735_FINAL.pdf]

NOAA's information technology resources was developed that sought improvements in network environments with adequate bandwidth, technologies for customer service, information technology security, management of web servers, and the utilization of high performance computing resources for emerging model-generated and remotely-sensed data streams. The subsequent spin-up of the Comprehensive Large Array-data Stewardship System (CLASS) has done much to provide the agency with adequate capacity for the future growth in environmental data.

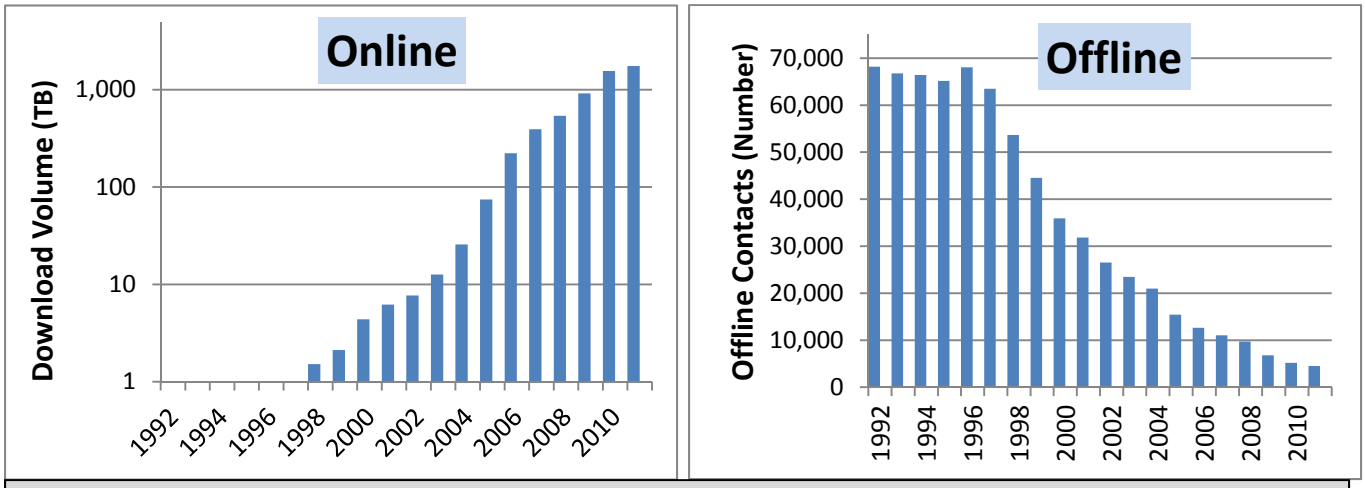

Figure 2. Transition from Online to Offline Data Access, NOAA National Data Centers, 1992-2011

Interagency Interfaces

As online access to NOAA's data has expanded, so too has the typical user's level of technical sophistication and scientific expertise. Online users increasingly have been searching for information and answers to specific questions rather than requests for bulk data. The needs of business and industry have become more complex, with queries for interrelated data and supporting information and documentation, rather than one particular type of data. The development of integrating approaches, such as the Meteorological Assimilation Data Ingest System (MADIS), the Integrated Ocean Observing System (IOOS) Data Integration Framework, and the National Integrated Drought Information System (NIDIS) exemplify thematic responses.

When technically feasible, NOAA has extended a number of innovative data services through Internet-based interfaces over the past decade. In 2003, an enterprise-wide geographic information system, the comprehensive NOAA Observing System Architecture system, was developed to enable NOAA to document and manage its multiple observing systems in an integrated manner. Through its contributions to both GeoSpatial One-Stop (GOS) and *data.gov*, coupled with its more recent stand-up of web portals (including *drought.gov* and *climate.gov*), NOAA has been a pacesetter for interagency portal development that leverages web mapping services. Such services allow inventory, documentation, data collection, data discovery, and access to framework layers covering the breadth of NOAA's organization. Since 2009, NOAA's contributions to semantic web and Service-Oriented Architecture (SOA) approaches have allowed users to combine environmental information in so-called "mashups" – or rapid dataset mergers. To this end, *data.gov* now includes over thirty major geographic datasets from NOAA that can be readily combined with other federal,

state, tribal, and local data spanning physical and socioeconomic topics. NOAA is also assisting in in the migration of Geospatial One-Stop to *geo.data.gov*, as well as implementing the next generation government-wide geospatial data and services portal, Geospatial Platform.

Central to NOAA's plan to modernize and improve its environmental data and information systems is the Global Earth Observation-Integrated Data Environment (GEO-IDE) initiative. GEO-IDE leverages SOA and a standards-based data and information infrastructure that will provide access to the full range of underlying data system capabilities – a prerequisite to integrated portal development.

Since 2005, NOAA has been committed to the development of international interfaces for data management and sharing, primarily through the Intergovernmental Group on Earth Observations (GEO), but also through numerous initiatives of the World Meteorological Organization. Management of data and information on a global scale is being addressed as part of the Global Earth Observation System of Systems (GEOSS), which is being developed by GEO.

Formats and Standards

A cornerstone of NOAA data management is adherence to appropriate open-standard formats. For its web-transmitted information, NOAA relies on common standards organizations for geospatial data access, data publication, and data formats, including the Open Geospatial Consortium (OGC), the Federal Geographic Data Committee (FGDC), the International Organization for Standardization (ISO), and the World Wide Web Consortium (W3C).

Standardization of NOAA data dates back to the use of punch cards before 1980. Since the 1990s, NOAA has contributed to the development of the National Spatial Data Infrastructure (NSDI) to ensure a common geospatial framework for the Nation and to minimize duplication. At the same time, changes in federal law (**Table 1**) underscored the need for NOAA to revamp its organizational structure to better coordinate operational and strategic decisions relative to budget and acquisition of information technology. By 2005, NOAA had developed a baseline observing systems architecture and had begun a portfolio and strategic investment analysis. These analyses, and others, allowed NOAA to prioritize and target investments and program development.

A program-oriented structure consisting of four Mission Goals and a Mission Support Goal allowed NOAA to corporately approach data issues in an interrelated manner. NOAA created an enterprise-wide method for planning and evaluating observation and data requirements, and established the NOAA Observing Systems Council (NOSC) to oversee observing systems, data, and information management and planning. The NOSC formed the Environmental Data Management Committee (EDMC) to coordinate the development and implementation of data management policy across NOAA, while at the same time the NOAA CIO Council created the NOAA GIS Committee. Since 2006, NOAA has managed archive, access, and stewardship requirements definition and end-to-end management plans for NOAA data and products through its Data Archive and Access Requirements Working Group (DAARWG) and the Data Management Integration Team (DMIT). These structures are essential as NOAA responds to OMB enterprise architecture and data consolidation initiatives.

Table 1. Federal Legislation and NOAA's Data Management Response, 1992-2011

Legislation	Year	Agency Mandate	NOAA Data Management Response
Government Performance and Results Act	1993	Improve government project management	5-year strategic planning with annual performance plans focused on data management operations and lifecycle
Clinger-Cohen Act	1996	Improve federal acquisition, use, and disposal of information technology	Aligned strategic mission and information resource planning with capital planning and investment
Telecommunications Act	1996	Leverage private-sector deployment of advanced information technologies	Utilize commercial solutions for data accessibility
Federal Electronic and Information Technology Accessibility and Compliance Act (Section 508)	1997	Provide accessibility of computer systems to all people, regardless of disability or severity of impairment	Adherence to web standards compatible with assistive technologies for cognitive, visual, hearing, and motor impairments
Information (Data) Quality Act	2001	Provide policy and procedural guidance for ensuring and maximizing the quality, objectivity, utility, and integrity of information disseminated	Adherence to formats and standards to promote accessibility, transparency, reproducibility, and timely, low-cost data delivery
E-Government Act	2002	Improve the management and promotion of electronic government services to improve citizen access to information	Development of interagency interfaces and a corporate strategy for portal development
Federal Information Security and Management Act	2002	Provide information security for both information and systems that support operations and assets of the agency	Secure enterprise solutions for data storage and stewardship coupled with access controls for development and operations
Government Paperwork Elimination Act	2003	Increase use of electronic forms, filing, and signatures to conduct official business with the public	Data conversion efforts to electronic formats, including scanning/imaging and data keying, along with automated observing systems

Information for Policymaking

The 1997 NRC report on global access to scientific data[5] notes:

> (The) challenge is to develop data management and archiving infrastructure and procedures that can handle the rapid increases in the volumes of scientific data, and at the same time maintain older archived data in an easily accessible, usable form. An important part of this challenge is to persuade policymakers that scientific data are indeed a precious resource that should be preserved and used broadly to advance science and to benefit society.

NOAA has risen to this challenge through the nexus of improved access, customer interfaces, and standardization of formats, which all contribute to the agency's ability to respond efficiently and in a timely manner to such requests. NOAA also continues to emphasize the need for good data and metadata for data discovery.

As an agency committed to science, stewardship, and service, NOAA is often called upon by policymakers for environmental data for decision support. The need for data sharing covers not only the path of data through observation, ingest, archive, and access/assessment in NOAA, but also

[5] NRC, *Bits of Power: Issues in Global Access to Scientific Data*. National Academy Press. Washington, DC, 1997, p.62.

intersects with scientific researchers at all stages. The NOAA response to the Deepwater Horizon incident in 2010 underscored the value of data sharing using common nomenclature and metadata. In January 2011, the National Science Foundation mandated that all proposals must have a data management plan. Such a plan is in keeping with NAO 212-15, and advances the convergence of technology and policy in support of a national data sharing environment wherein environmental information (including data, samples, physical collections, and other supporting materials) can be shared with others at no more than incremental cost and within a reasonable time. To this end, NOAA's EDMC is finalizing Procedural Directives that will require data sharing by NOAA Grantees and the development of Data Management Plans by significant NOAA projects that produce data.

The touchstone for providing environmental information for policymaking is robust national and international assessments of environmental conditions. From its inception, NOAA has played a leading role in supporting the International Panel on Climate Change (IPCC). Simultaneously, it has been a key contributing agency to the U.S. Global Change Research Program's (USGRCP) National Climate Assessment (NCA).[6] National climate assessments provide a status report on climate change science and impacts, and serve as a resource for decisions related to social, ecological, and policy systems, and integrated analyses of impacts and vulnerability. NOAA is committed, as part of the ongoing NCA activity under USGCRP (as well as through the IPCC effort), to act as a data manager for a new generation of Integrated Assessment Models (IAMs). The stewardship of these data will ensure transparency and reproducible results relative to the science applications of assessment activities.

Conclusion

Today, NOAA defines data management in the context of two major synchronized activities: *data management servic*es and *data stewardship*. According to NAO 212-15, these activities constitute a comprehensive end-to-end process that includes movement of data and information from the observing sensor to the user. Along this path, acquisition, quality control, metadata cataloging, validation, reprocessing, storage, retrieval, dissemination, and archiving of data with scientific integrity and attention to detail are required.

As the agency enters the second decade of the 21st Century, it confronts the continued challenge of leveraging technological advances and nimble organizational approaches to extend quality environmental data and information to U.S. citizens and global partners. From policy decisions related to coastal zone management to fisheries conservation to reducing storm prediction lead times, NOAA has a critical mission that demands stewardship of environmental data at every step. Reporting through Public Law 102-567, Section 106 over the past two decades has done much to position NOAA to move forward in partnership with other agencies, academia, international groups, and the private-sector to meet these challenges. Both NOAA and the Department of Commerce look forward to informing the Congress and the Nation on its broad data management responsibilities.

[6] National Climate Assessment
[http://www.globalchange.gov/what-we-do/assessment]