

U.S. Geological Survey Spatial Data Access

John L. FAUNDEEN, Ronald L. KANENGIETER¹ and Michael D. BUSWELL¹

U.S. Geological Survey EROS Data Center, Sioux Falls, SD 57198-0001

Abstract

The U.S. Geological Survey (USGS) has done a progress review on improving access to its spatial data holdings over the Web. The USGS EROS Data Center has created three major Web-based interfaces to deliver spatial data to the general public; they are Earth Explorer, the Seamless Data Distribution System (SDDS), and the USGS Web Mapping Portal. Lessons were learned in developing these systems, and various resources were needed for their implementation.

The USGS serves as a fact-finding agency in the U.S. Government that collects, monitors, analyzes, and provides scientific information about natural resource conditions and issues. To carry out its mission, the USGS has created and managed spatial data since its inception. Originally relying on paper maps, the USGS now uses advanced technology to produce digital representations of the Earth's features. The spatial products of the USGS include both source and derivative data. Derivative datasets include Digital Orthophoto Quadrangles (DOQ), Digital Elevation Models, Digital Line Graphs, land-cover Digital Raster Graphics, and the seamless National Elevation Dataset. These products, created with automated processes, use aerial photographs, satellite images, or other cartographic information such as scanned paper maps as source data.

With Earth Explorer, users can search multiple inventories through metadata queries and can browse satellite and DOQ imagery. They can place orders and make payment through secure credit card transactions. Some USGS spatial data can be accessed with SDDS. The SDDS uses an ArcIMS map service interface to identify the user's areas of interest and determine the output format; it allows the user to either download the actual spatial data directly for small areas or place orders for larger areas to be delivered on media. The USGS Web Mapping Portal provides views of national and international datasets through an ArcIMS map service interface. In addition, the map portal posts news about new map services available from the USGS, many simultaneously published on the Environmental Systems Research Institute Geography Network.

These three information systems use new software tools and expanded hardware to meet the requirements of the users. The systems are designed to handle the required workload and are relatively easy to enhance and maintain. The software tools give users a high level of functionality and help the system conform to industry standards. The hardware and software architecture is designed to handle the large amounts of spatial data and Internet traffic required by the information systems. Last, customer support was needed to answer questions, monitor e-mail, and report customer problems.

1 Introduction

Access to the large spatial data holdings of the U.S. Geological Survey (USGS) is required by U.S. public law in some cases. The law states, "It is in the best interest of the United States to maintain a permanent, comprehensive Government archive of global Landsat and other land remote sensing data shall provide for long-term storage, maintenance, and upgrading of a basic, global, land remote sensing data set and timely access for parties requesting data."² The mission of the USGS's National Mapping Discipline is to meet the Nation's need for basic spatial data, ensuring access to and advancing the application of these data and other related earth science information for users worldwide.³

Online access to USGS data began in 1991 with the release of a system called the Global Land Information System (GLIS). Users could connect to the service with Telnet or dial-up modems. Queries using spatial, temporal, and dataset-unique criteria were allowed, resulting in a list of the files, photographs, or images that matched the user’s criteria. Graphics were employed to illustrate the spatial extent the data represented on the Earth. Some data were subsampled and provided as browse images as a means to preview the actual products. This capability was used primarily to assist in the determination of cloud cover and the quality of satellite imagery.

The GLIS was replaced in 2000 by Earth Explorer, a system with considerably more capabilities. The system still used metadata queries, but users could search multiple inventories from a single query, save the metadata in the Federal Geographic Data Committee (FGDC) Geo profile, and place orders securely using credit cards. Another new feature was the ability to establish standing requests that would query for data based on the user’s criteria and send e-mail messages whenever a match occurred. The user enters their criteria once and then receives metadata describing query matches entirely outside the system. Earth Explorer does not contain any data within the interface itself, but it portrays datasets and allows users to place orders for those data. Generally, Earth Explorer lists and describes remotely sensed and cartographic data. Examples of the remotely sensed datasets include Landsat Multispectral Scanner (MSS), Corona, and aerial photographs. Cartographic holdings include Digital Line Graphics and Digital Raster Graphics (DRG). **Table 1** is a complete list of the datasets represented and the number of records for each.

Earth Explorer is available at URL <http://earthexplorer.usgs.gov>.

Table 1. Datasets and their total number of records represented in Earth Explorer.

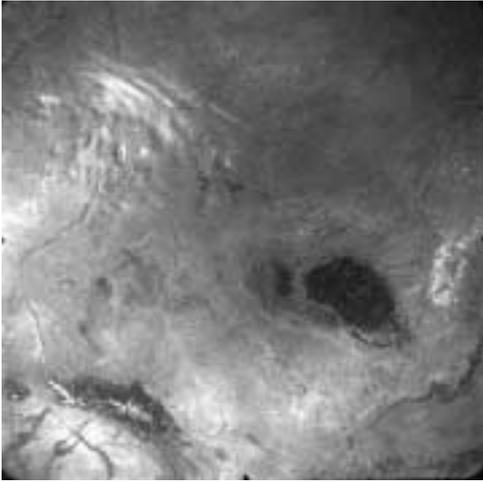
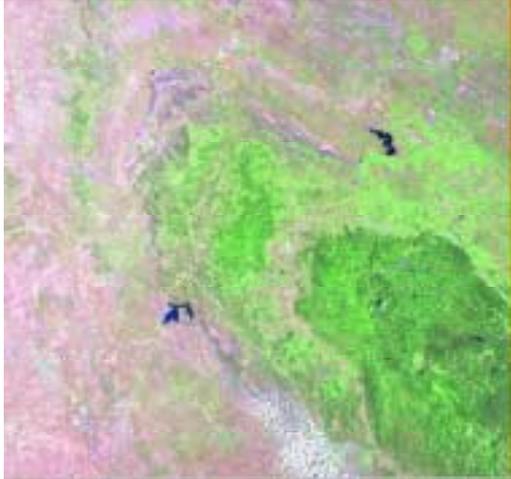
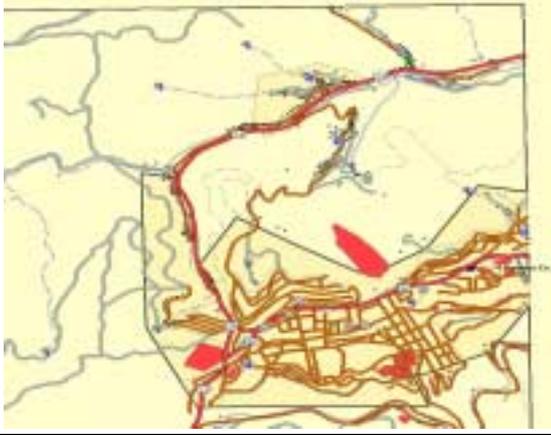
Dataset	Number of Records
CORONA	853,000
Landsat MSS	631,178
Landsat TM	480,222
Landsat ETM+	194,684
AVHRR	257,675
SPOT	720,000
Digital Orthophoto Quadrangles	221,337
National Aerial Photography Program	1,629,838
National High-Altitude Photography	511,844
Digital Line Graphs	195,356
Digital Elevation Models	90,231
Digital Raster Graphics	57,285
Paper Maps	56,000

Figures 1 through 4 illustrate examples of the CORONA, Landsat ETM+, Digital Orthophoto Quadrangle, and Digital Line Graph datasets (available through Earth Explorer). Figure 1 shows a CORONA image over Black Hills, South Dakota, August 21, 1964. (Image DS09066A063MC065). Figure 2 is a Landsat ETM+ image over Black Hills, South Dakota, August 2, 1999. (Image 7034029009921450). Figure 3 is the Digital Orthophoto Quadrangle image in South Dakota, June 16, 1995 (Image DI0000000042790). Figure 4 is the corresponding Digital line graph image in South Dakota (Image DL025210HY144244).

2 Earth Explorer

The resources to design, build, and operate Earth Explorer are different from those the USGS had used previously. After compiling the requirements, USGS staff did internal analysis to determine whether the USGS possessed the skills and experience to design and build an interface that would meet the requirements, which included a firm delivery schedule. The analysis indicated that the USGS possessed the skill to do the work and some of the experience, but the effort could not be accomplished within the schedule required. The requirements were advertised through an RFP

(request for proposal) and bids were solicited from commercial vendors. A company was selected and contractual arrangements were made regarding the design and building of this next-generation system. USGS staff representing database management, information systems, and software engineering held numerous meetings and teleconferences with the commercial vendor to ensure that the requirements were understood. A schedule was established that conveyed the periodic management and technical meetings, the initial deliverables, and an acceptance test period. After the first interface code was officially accepted, mechanisms were established to accommodate trouble reports and enhancements not covered within the requirements. As tracking the delivered code became more complex, a configuration management system was used to ensure proper code migration from a development, to a testing, and finally to a production environment.

	
<p align="center">Figure 1. CORONA</p>	<p align="center">Figure 2. Landsat ETM+</p>
	
<p align="center">Figure 3. Digital Orthophoto Quad</p>	<p align="center">Figure 4. Digital Line Graph</p>

The USGS learned many valuable lessons through its first use of a commercial vendor arrangement in a public information system. The commercial vendor was able to deliver new and patched code much faster than the USGS could. The USGS could extend personnel resources in its software engineering by leveraging the commercial vendor for much of the work rather than controlling all aspects of an information system development project from within the organization. The commercial vendor was able to keep the USGS abreast of developing international and ad hoc standards that benefited development. On the other side of the ledger, communication channels had to be refocused a few times regarding authorized work, the cost associated with those work pieces, and which elements were considered part of the original requirements and which were enhancements. Establishing regular communications at both the management and the technical

levels resolved most issues. In general, the USGS considers its overall return on investment to be quite high.

3 Seamless Data Distribution System

One of the key trends in geographic information systems (GIS) is the rise of seamless datasets. In the past, geographic data were partitioned because of physical, hardware, software, and political considerations. For example, when paper maps were the dominant form of geographic media, it made sense to partition a dataset so that the paper partitions would fit on a desktop. Today GIS users dominate the need for geographic data. Their computer processing, storage, and network capabilities continue to increase, making GIS more powerful and able to process larger datasets. Many researchers are using GIS to study features or trends that are global or national in scope. By removing arbitrary boundaries, seamless datasets provide these researchers a more natural and powerful mechanism to analyze and study large areas of interest. This is useful in emergencies or disasters, when data need to be available for immediate use. Furthermore, during emergencies there is not time to search for the best data available over a region or to edge match and process partitions of a dataset to a desired quality and consistency. The data need to be GIS ready, the best available, of consistent quality, and accessible. All these requirements have given rise to seamless datasets and their distribution on the Internet.

To meet these requirements, the USGS released the Seamless Data Distribution System (SDDS) in January 2001. The SDDS allows Internet users to determine areas of interest, established output parameters, and directly download actual spatial data. Currently, 30-meter resolution National Elevation Data (NED) and National Land Cover Data (NLCD) seamless datasets are available from SDDS. As the USGS continues to develop seamless datasets, more will be available to the public. Table 2 provides more information on the seamless datasets available for ordering and datasets used for reference layers. The SDDS is available at URL <http://gisdata.usgs.gov/seamless>.

Table 2. Seamless datasets provided by the USGS and map services used as reference layers.

Dataset	Coverage	Data File Size
National Elevation Data (NED)	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	60 GB – Demgrids, 20 GB - Shaded Relief, 1 GB – Metadata
National Land Cover Data (NLCD) *	Conterminous U.S.	16.25 GB Integer Data, .001 GB – Metadata
National Hydrography Dataset (NHD)	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	22 GB – Data 8.6 GB – Reference Layer
Geographic Names Information System (GNIS) - Condensed Version	Conterminous U.S., Alaska, Hawaii, Puerto Rico, and Virgin Islands	.015 GB – Reference Layer
Bureau of Transportation and Statistics (BTS) Transportation	Conterminous U.S.	4 GB – Reference Layer
National Atlas Reference Layers: U.S. State and County Boundaries	Conterminous U.S., Alaska, and Hawaii	.32 GB – Reference layer

* Datasets available from SDDS.

Greater resources are required to design, build, and operate a seamless information system than for a metadata-based information system. Seamless data required changes in the way data are stored, maintained, and accessed. The data and metadata are stored online, so a large amount of disk space is required. Special emphasis is placed on the maintenance of seamless data to ensure that they remain the most current, best available, and of consistent quality. The distribution of

seamless data over the Internet requires a large communications infrastructure with adequate server resources.

Because the seamless data concept is new, a different user interface was envisaged. Rather than focusing on the metadata, the seamless interface focused on the data. The user interface is much more graphical; an example is shown in Figure 5. The view of the data rendered in the interface is a map service. Each new display layer added or deleted, and each zoom, pan, or select command chosen by the user create a new map that is delivered live over the Internet. Map content and geographic coverage defined by the USGS can contain one or more data layers, depending on the purpose of the map. The SDDS uses NED and NLCD data layers to preview orderable data. Reference data layers (e.g., state and county boundaries, city names, streams, and roads) help define the area of interest. The map services displayed in the SDDS interface uses map pyramids and scale-dependent rendering. This provides faster data transfer rates to the user and allows browsing data at smaller scales without sacrificing detail at larger scales.

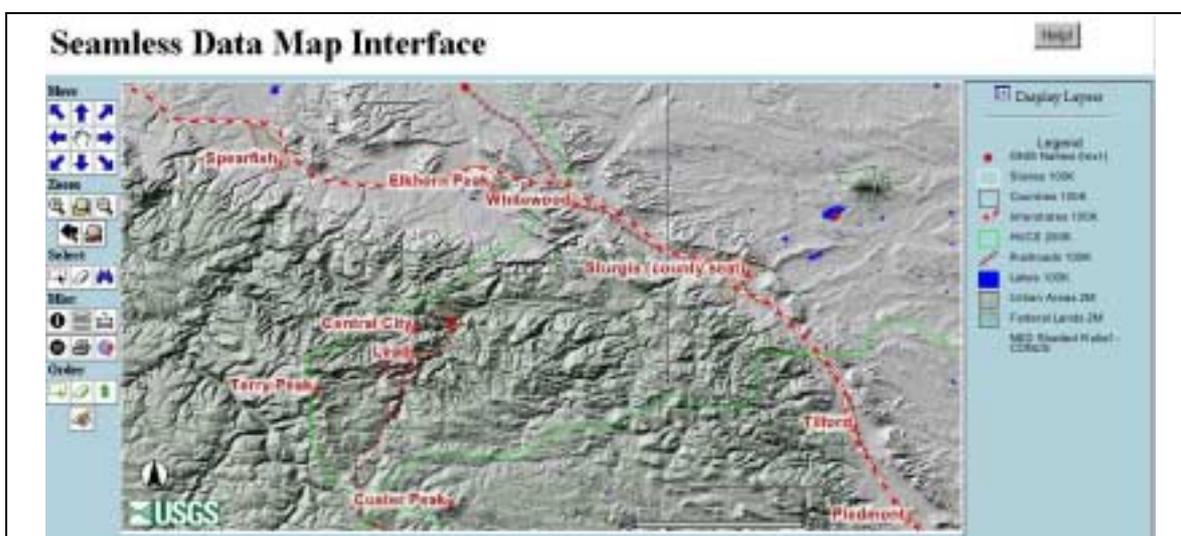


Figure 5. Screen capture of the Seamless Data Distribution System

The latest GIS technology was needed to create, store, display, and distribute the seamless data. The key commercial off-the-shelf tools used in the creation of SDDS include Oracle database software and Environmental Systems Research Institute's (ESRI) Spatial Data Engine (SDE) and Internet Map Service (IMS) software tools. The SDDS uses the latest versions of SDE and IMS to manipulate the large seamless raster datasets.

The ESRI technology was obtained by the USGS team under a Cooperative Research and Development Agreement (CRADA). The CRADA enables the USGS and ESRI to collaborate on developing seamless data technology, data analysis capabilities, and resource requirements. The Seamless Development Team leveraged the CRADA experience to create the production data delivery system.

4 Web Mapping Portal

Experiences gained were also used to publish GIS-ready information over the Internet, significantly cutting the cycle time from design to a production release. Concurrently with the portal development, ArcIMS map services were created for seamless data layers, updating the metadata, creating Web Mapping Server Specification (WMS) connectors, publishing datasets on the Geography Network, and establishing a mirror site maintained by ESRI. The USGS is working to extend the Internet interfaces for map and data delivery and build the USGS infrastructure for future data access.

The USGS EROS Data Center (EDC) provides access to national GIS-ready information through map services published on the Web Mapping Portal. The portal offers multiparticipant, distributed Web mapping services consisting of the NED, NLCD, and National Hydrography Dataset (NHD). In addition, the portal hosts a condensed version of geographic names from the Geographic Names Information System (GNIS) and vector reference-layer map services from the National Atlas.

The advantage of the use of different map services over the Web is that these services can be brought together in a mapping environment without loading all the data locally. For example, the map in Figure 6 was created using USGS-published map services for shaded relief, roads, geographic names, streams, and lakes, plus NLCD data downloaded from the SDDS. The map services were added directly from the Geography Network. The NLCD layer was rendered semitransparent and was overlaid on the shaded relief to give it a sense of depth.

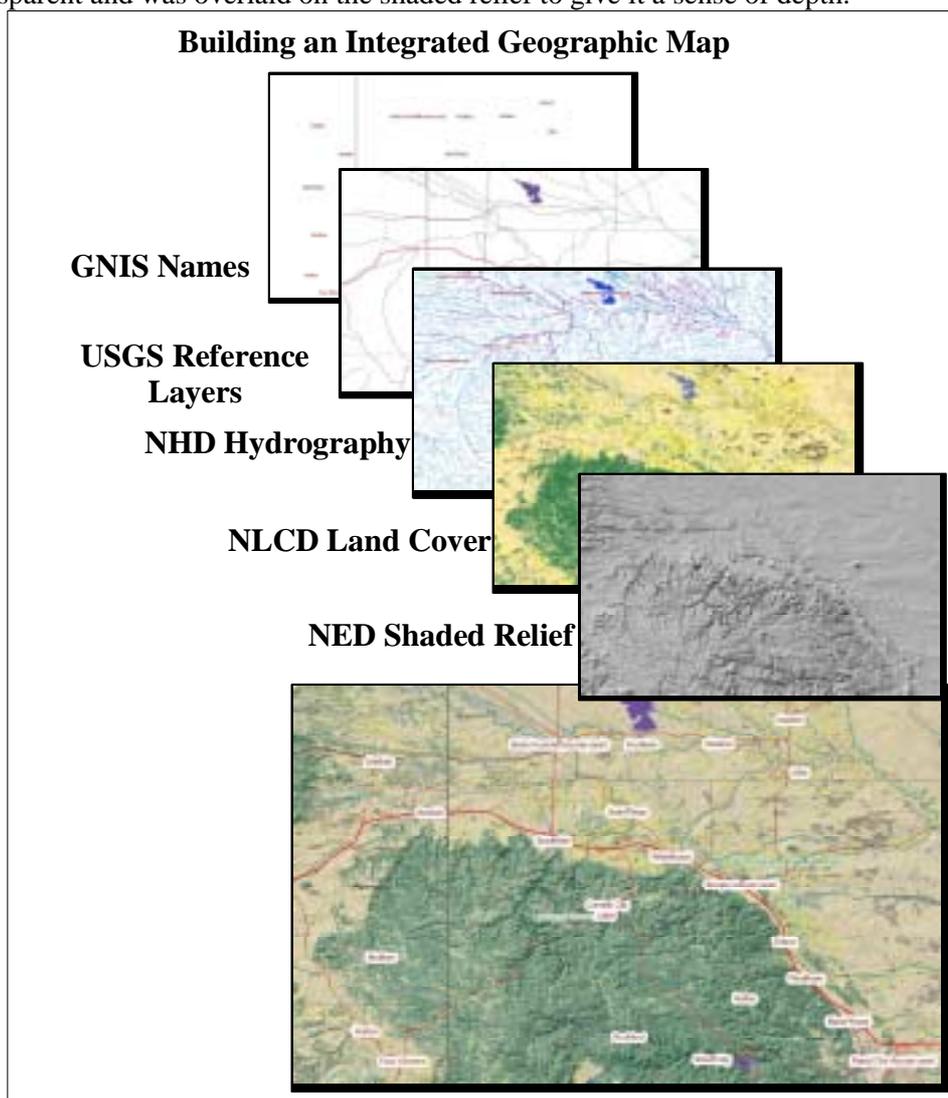


Figure 6. Building an integrated map using data services and datasets (Map built using ArcMap. GNIS, USGS administrative and reference layers, NHD, and NED shaded-relief layers downloaded directly from the Geography Network. NLCD ordered from the Seamless Data Distribution System and downloaded by means of FTP.)

The offerings consist of map services in ArcIMS, WMS format, and FGDC metadata. These services are accessible by using the GeoExplorer search facilities of the Geography Network site at <http://www.geographynetwork.com> or from the USGS site <http://gisdata.usgs.net>.

Also on the Web Mapping Portal are three map viewers that highlight the International Hydro 1K map services for Asia and Africa and *The National Map* Rapid Prototype datasets:

(<http://gisdata.usgs.net/HYDRO1K/ASIA/viewer.htm>),
 (<http://gisdata.usgs.net/HYDRO1K/AFRICA/viewer.htm>),
 (http://edcw2ks18.cr.usgs.gov/website/National_Map/viewer.htm)

The Web mapping services have been designed to be scalable in a multiuser environment that centers on the use of geodatabases that store vector and raster datasets, the interrelationships, and the metadata. **Figure 7** shows a schematic diagram of the portal's public architecture.

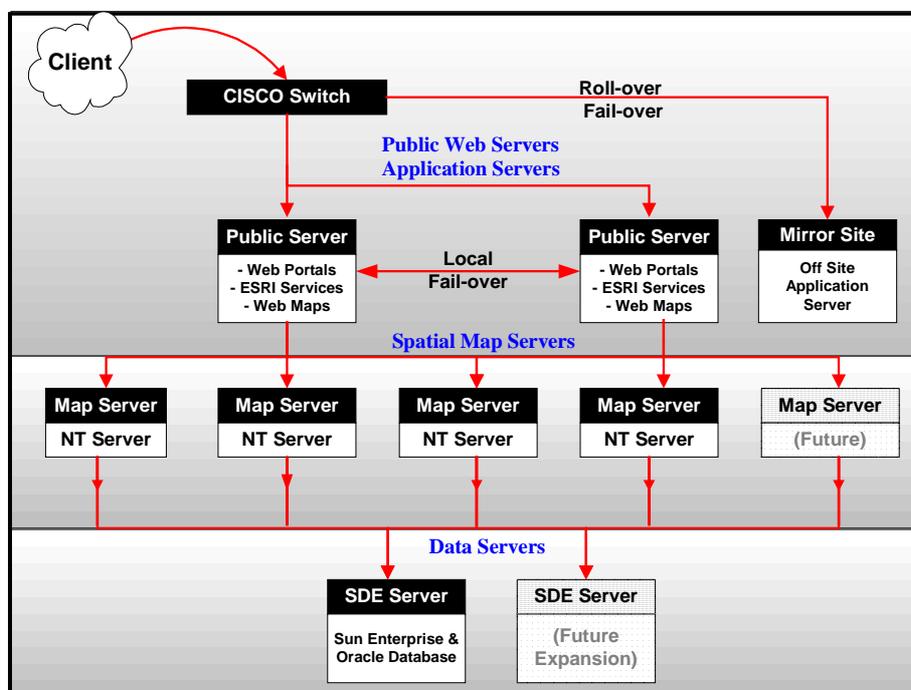


Figure 7. Web Mapping Portal public system architecture

When a user enters the site, the network router will direct the http Web traffic to one of the internal Web servers or to one located at an Internet Service Provider (ISP) provided by ESRI. If the USGS servers are down, requests will be directed to the ESRI mirror site. The network switch will direct the ESRI service requests to one of the ArcIMS that sends the client's request to the map servers for processing. The map servers can access and bundle maps and data into appropriate formats before returning the information to the user client either as JPEG, PNG, or GIF images. The foundation for the Web Mapping Portal is the SDE server. This is where all the data, georeferencing information, and metadata reside.

5 Summary

Access to the Nation's spatial data holdings is the foundational requirement on which the three Web-based interfaces described in this paper were built. Adding to this challenge is the fact that the USGS spatial data holdings continue to grow and evolve. The USGS archive at EDC manages more than 750 terabytes of satellite, aerial, and cartographic data. The Landsat 7 program alone acquires approximately 300 new scenes per day and coordinates activity across a network of 12 international ground stations.⁴

The USGS has a long history of providing spatial data access to the general public. In the last two decades, the task of customer order and inquiry has evolved from mail, telephone, and fax to Web-based interfaces. Furthermore, product delivery has evolved from physical media such as paper, computer tape, or CD to customers taking delivery of products through the Internet. The three Web-based interfaces are important steps forward for the USGS. By developing these interfaces using advanced Web technology, the USGS is now able to serve its customers better, has increased staff experience and skills with these tools, and has prepared the USGS for the National Map

requirements. The National Map is envisioned to be next step for the USGS in providing Web based spatial data access.

¹ Raytheon STX Corporation, under Government contract, No. 1434-CR-97-CN-40274. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

² "classification of the public lands * * *": The phrase is taken from the Organic Act of the U.S. Geological Survey, U.S. Statutes at Large, v. 20, p. 394 [The Statutes at Large are ordinarily cited in abbreviated form as 20 Stat. L., 394].

³ Land Remote Sensing Policy Act, October 28, 1992. USGS, National Mapping Discipline Strategic Plan, <http://mapping.usgs.gov/misc/strategic.html>

⁴ EROS Data Center Annual Report – Fiscal Year 2000.