

Proceedings of the 6th Annual Federal Depository Library Conference

April 14-17, 1997

LC/G&M's Scanning Program: Where We Are and How We Got Here

Elizabeth Mangan, Library of Congress
Washington, DC

When the staff of the Geography and Map Division (G&M) at the Library of Congress (LC) began assessing how digital forms of geographic and cartographic information could be integrated into its collections, we realized the new technologies would require knowledge and skills that we did not possess. It was also clear that the complex and expensive array of computer equipment that was required would be difficult, if not impossible, to acquire through the normal budgetary channels for appropriated funds.

In November 1993 the James Madison Council, a private sector advisory group to the Library, provided \$30,000 to investigate the establishment of a corporate support group for the G&M Division. That same month, Alan M. Voorhees, a long-time friend and supporter of the Geography and Map Division and the Library, volunteered to lead the effort in attracting industry support.

The first meeting of this corporate support group, which was named the Center for Geographic Information, was held at the Library on January 12th, 1995, when eight firms committed to being charter members: Autometric, Inc., Environmental Systems Research Institute, Harvard Design and Mapping Company, H.M. Gousha Company, Intergraph Federal Systems, MAGELLAN Geographix, MapInfo Corporation, and Tangent Engineering. Mr. Voorhees, who is chairman of the board of Autometric, a firm engaged in many aspects of the latest geographic technologies, agreed to be chairman of the Center for Geographic Information.

The purpose of the Center is to coordinate the contribution of various resources and knowledge from a wide spectrum of the geographic information and cartography industry. These donations will:

1. Assist the Geography and Map Division in making the transition to the age of electronic maps and digital forms of geographic information through advice, training, and financial support for acquiring hardware, software, and data sets;
2. Facilitate sharing the rich cartographic resources of the Division electronically;

3. Promote the use of electronic forms of geographic information by many sectors of the nation, including libraries, academia, industry and commerce, education, and the general public;
4. Encourage the deposit of digital spatial data sets by American and foreign governments, industry, and academic producers; and
5. Advance the library's publication, education, and exhibition programs in geographic information and cartography.

The charter members of the Center established full membership at \$5,000 annually and associate membership at \$500 annually. Support from industry members also includes in-kind assistance, as necessary, in providing the Division with appropriate equipment and software to begin developing expertise in the scanning of maps and the use of software and digital forms of geographic data; deposit of data sets; and participation on committees to accomplish the goals of the Center.

Members of the Center benefit in a number of ways:

1. The Division's expertise in cataloging cartographic materials is shared with developers and users of digital forms of geographic information;
2. Because of its unique position within the Library, the Center can sponsor programs that address specific needs of the cartographic and geographic information communities and provide useful links among these communities and with Congress and other institutions;
3. By working with producers and users of geographic information and digital cartography the Division can ensure that digital forms of geographic information are systematically collected and preserved for the future use of the nation; and,
4. Most importantly, members have enhanced access to the Library's vast collections of cartographic materials, which they are encouraged to use as resource material and to distribute in a variety of value-added formats.

For those of you who might not know, the Library of Congress collection of cartographic materials is the largest in the world, containing approximately 4.6 million maps; more than 60,000 atlases, which contain another 8 to 10 million maps; approximately 300 globes; as well as relief models, puzzles, fans, powder horns, and almost anything else that might have a map on it. In the early 1990s, digital files of geographic data began appearing among cartographic materials deposited by Federal mapping agencies and cartographic software through copyright deposits. The Division's collections now contain approximately 4,000 CDs and computer software packages which are controlled in an online ProCite database restricted to Division staff.

The Center for Geographic Information has met five times since the organizational meeting in January 1995: June 1995 at the Library; October 1995 at MAGELLAN Geographix in Santa Barbara, California; May 1996, back at the Library; September 1996 in Bellevue,

Washington, sponsored by Corbis; and January 1997 in Denver, Colorado, sponsored by Tangent.

As of the first of April the Center's membership has grown to 14 members and 10 associate members. The new members are Autodesk, Inc., Corbis Corporation, Digicolor, Inc., the Hewlett-Packard Company, LizardTech, Inc., Microsoft Corporation, Mindscape, Inc., Rand McNally and Company, and Tactician Corporation. The new associate members are ADC The Map People, CommuniVision, EDR/Sanborn, Inc., Adrian B. Ettlinger, Macromedia, Inc., MapLink, Inc., Navigator Publishing, Spatial Data Institute, and Systems Planning and Analysis

We have made great strides in moving into the new world of digital geographic information in the last three and a half years, largely because of the efforts of our staff and their success in convincing private sector companies that they have a lot to gain by becoming involved in this enterprise. Direct cash contributions to the Center total approximately \$100,000, and the value of hardware and software contributed to date is over \$700,000.

The first meeting resulted in the first donation. Robert Garber, Chief Operating Officer of Tangent Engineering, now Tangent Color Systems, spearheaded the indefinite loan of a large-format, flatbed, color scanner, a Sun Sparc workstation, and a Hewlett Packard 650 plotter. This system can scan flat items up to 24 by 34 inches, in 24-bit color, at resolutions up to 600 dots per inch (dpi). Of course the plotter only prints at 300 dpi.

As a result of this donation, we acquired the technology to scan maps, and the library's overall National Digital Library Program (NDLP) adopted our proposal to establish a National Digital Library (NDL) Program for Cartographic Materials, which it has agreed to support through the funding of four positions in the Division to run the scanning program.

In a ceremony in the division in April 1995, Dr. Billington, the Librarian of Congress, and John Kluge, President of the James Madison Council, cut the ribbon on the scanner and participated in the scanning of the first image from the Library's cartographic collections, George Washington's A plan of my farm on Little Huntg. Creek & Potomk, which the first President drew in 1776. The result is so fine that it is difficult to distinguish the scanned map from the original manuscript.

During the first year after the scanner was installed, we scanned several hundred maps in a trial or testing phase and have been working with the members of the Center on technical standards and workflow design in preparation for large-scale scanning projects for the NDL.

Shortly after the donation of the scanning system, the Hewlett Packard (HP) Company made a major donation to the Division and we found ourselves in the enviable position of having the full infrastructure necessary to accomplish our long-range goals. With the HP donation of computer equipment worth nearly \$600,000, the Center for Geographic Information and the Division's related GIS Facility both took a giant leap forward.

During a reception in May 1996 to thank Hewlett-Packard for its generous donation in support of the Library's NDL Program for Cartographic Materials, Dr. Billington remarked on the importance of this gift. He observed that, "In helping the Geography and Map Division

adapt to the modern world of geography and cartography, the executives of HP who made this donation possible have also embellished the vision of their company's co-founder, David Packard," adding that the NDL Program was started with \$13 million in seed money of which \$5 million was from the Lucille and David Packard Foundation.

In preparation for the installation of this new equipment, the Division's Reading Room was re-configured to construct a secured area to house the servers and optical storage jukebox as well as the scanning equipment. Access to this area is restricted to selected staff members through the use of a swipe card system to unlock the doors. Additionally, an alarm system with both motion and infrared-heat detectors is used to secure the area when we are closed.

The HP donation will assist in the development of the Division's GIS Facility as well as its NDLP scanning effort. It consists of a Series 9000, K400 Server, featuring four 100 MHZ processors, 1.2Gb system memory, 8Gb internal hard-drive, running HP/UNIX and a SureStore 165ST optical disk jukebox, featuring slots for 128 Write-Once-Read-Many (WORM) or Rewriteable optical disks for a total of 165Gb of storage. HP also donated three 712/100 workstations, featuring 100 MHZ, PA-RISC processors and 192Mb of system memory, running HP/UNIX and a Series 9000 J-200 workstation, featuring two 100 MHZ processors, 256Mb RAM, 2Gb hard-drive, and special graphics enhancement, to replace the Sun Sparc workstation on loan from Tangent for the NDL Program.

Recognizing our need for equipment to provide access to the digital images created for the NDLP and to allow patrons to use Geographic Information Systems, HP donated three HP Vectra PCs, featuring 120 MHZ processors, running Windows95; three HP Pavillion PCs, featuring 166 MHZ processors and full multi-media capabilities, running Windows95; three Envisex P Series X-Window Stations; a 715/100 workstation, featuring a 100 MHZ, PA-RISC processor, 256Mb of system memory, and an enhanced graphics display, running HP/UNIX; two DeskJet 1600CM color printers; a LaserJet 4MV, 11" x 17" printer, and a DesignJet 755C, 36" roll feed, color plotter, with 72Mb of memory.

Following the installation of the equipment we were able to appoint the remainder of our NDL team. We now have five full-time staff members responsible for creating digital images, a Processing Technician, two Digital Conversion Specialists, a Digital Conversion Coordinator for Production, and a Digital Conversion Coordinator for Systems, who also serves as our WebMaster.

Ten core historic Americana aggregations from the Division's collections have been designated to be scanned for the National Digital Library Program. These include Civil War maps, county land ownership maps and atlases, late 19th and early 20th century panoramic maps of U.S. cities, Sanborn fire insurance maps, and maps and atlases of the District of Columbia. During the first year after the arrival of the scanner we tested the scanning of a variety of materials from the collections, and even though most all of the items scanned have been done as on-demand requests, we have been able to scan examples from most of the aggregates designated for the NDLP.

You have probably noticed that I have been using the word "scanning" rather than "digitizing" in describing these activities. The Division is heedful of this distinction because the images we are creating are raster images that have not been geo-referenced.

Since Tangent scanners had formerly only been used by the military for scanning current maps, the scanning of the Library's collections is their first opportunity to have this equipment used with older materials which present a whole new range of problems, including: discoloration; scanning through Mylar™; atlas bindings which can't be flattened; brittle, disintegrating maps, to name just a few. The handling of our valuable, rare, and often delicate items presents problems all of its own. Once these items are scanned we don't want to expose them to the process unnecessarily. Unfortunately the scanner doesn't seem to understand this concern. While scanning the Division's Agnese atlas the two lenses went out of calibration, resulting in a mis-alignment on each image, which wasn't discovered until the entire atlas had been scanned! Tangent has continued to work closely with us while we are learning the scanning software and to make the necessary modifications to both the scanner and the software when problems arise and to upgrade the software when improvements are made.

The Scanning Committee of the Center for Geographic Information is advising us in setting up our NDLP production of images. The Committee was established during the Center's meeting last May to discuss the workflow and technical issues concerning the images for the NDLP. The Committee has advised us to scan at 300 dpi and to save the images in TIFF format.

In establishing a production workflow we have encountered a number of stumbling blocks in getting the equipment networked using the Library's standard twisted-pair token ring network and processing the images in preparation for making them available on the Web. Moving the images around in this environment has proven problematic. Finding software to manipulate and edit the images which is compatible with the HP/UNIX has been hopeless since most software of this type is written primarily for Solaris/UNIX. When we discovered that Claris had developed Macintosh emulation software for HP/UNIX, we were confident that we would be able to use the Macintosh version of xRes on the workstations provided by HP. Macromedia, one of the Center's members, donated two copies of xRes, the Macintosh version and the WindowsNT version. Unfortunately, we have been unable to successfully load the Claris emulation software.

We have loaded the WindowsNT version of xRes on an Intergraph workstation which is on loan to the Library for evaluation. Since this machine is located one floor above us, we must ftp the image to the Intergraph machine in order to perform quality review and post-processing to rotate, crop, adjust brightness or contrast, and stitch together the TIFF images. Through working with these large images, approximately 180Mb each, and conferring with Macromedia, we have determined that the ideal computer for this task would be a single processor Pentium Pro, at least 166 MHZ, with 500Mb of RAM, graphic accelerator card, 21 inch monitor, and an 8Gb hard-drive running WindowsNT; or, alternatively, a Macintosh 9600 PowerPC with a 300 MHZ single processor.

Following the xRes processing, the images are temporarily stored on HP Surestore jukebox magnetic optical platters which can each store 1.3Gb. We then compress the image, using

a wavelet-based image compressor software called Multi-Resolution Seamless Image Database, or MrSID, which was developed by LizardTech, another corporate partner in the Center for Geographic Information. This software integrates multiple resolutions of an image in a single file which means that when users zoom in they get better and better resolution. Since MrSID stores the images in a seamless manner the user can zoom in and out. A Netscape plug-in which will enable a user to pan, providing immediate access to any portion of a large image as quickly and easily as another, is scheduled to be released in the next few months.

Although MrSID is a "lossy" image compressor, we can compress the image at a ratio of nearly 22:1 and not experience any loss of information, because MrSID keeps track of every pixel from every image. One of the unique features of this compression method is its ability to decompress only that portion of the image requested by the user, which means that images of any size can be decompressed by a user with as little as 1Mb of RAM. Additionally, MrSID builds on each successive resolution by using data already decompressed and loaded, so the user gets immediate access to any location in the image at any resolution.

This software has four components:

- MrSID Compress, to compress images;
- MrSID Retrieve, to decompress MrSID files for use with other software;
- MrSID Viewer, a stand-alone application for viewing MrSID files; and,
- MrSID Distributed Image Database Server for Internet access to MrSID files.

MrSID doesn't require any special hardware to operate. MrSID Retrieve and Viewer operate on any platform, MS Windows 3.1, 95, NT, and MacOS and requires about 1Mb of RAM. MrSID Compress and Distributed Database Server operate in WindowsNT, UNIX, and MacOS and require 32Mb physical RAM plus enough storage for the original image and for the compressed file. Input files can be raw data or in a number of popular raster formats including TIFF. MrSID Viewer allows exporting in MrSID compressed format and in non-MrSID formats.

The first aggregate of items to be scanned as a G&M Division project is our collection of panoramic views. Following the listing of items in *Panoramic Maps of Cities in the United States and Canada: A Checklist of Maps in the Collections of the Library of Congress...*, published in 1984, we have completed Alabama through Minnesota. Permission to include facsimiles and items for which we have a photographic reproduction from another institution will be sought, so that those items can be made available in electronic form.

While we expect to include images of maps from atlases in the NDLP, we have encountered a number of problems with these. Even with access to a book scanner, the size and weight of many atlases would prohibit the use of this type of scanner. We have tested using a digital camera suspended above an atlas in a book cradle, but the distortion inherent in the variable distance to the center of the cameras lens, especially in the gutter, has thus far proven unsolvable.

In addition to the bibliographic data in the USMARC record, which will serve as the primary access to the images in the Library's system, we will be building a metadata database using Paradox. The metadata database will include non-bibliographic data, including the number of images required for each scanned map, the number of versions of each image, the date the item was scanned, the scanner used, the reason or project for which the item was scanned, etc.

An unexpected benefit of our scanning program has been a new avenue for acquisitions. During the visit of the Madison Council in May 1995, the Chief of the Division was showing a facsimile of View of the University of Virginia, Charlottesville & Monticello taken from Lewis Mountain which was drawn by Casimir Bohn and published by E. Sasche & Co. in 1856 when Mr. Voorhees mentioned that he owned an original of this view. Mr. Voorhees said that he would donate the original to the Library if he could have a print from the scanned image to replace it, since it was hung in his hotel in Richmond. When a former staff member, who is extremely knowledgeable in the history of cartography and printing, was visiting the hotel for a meeting and saw and examined the framed view he remarked, "I thought Al gave the original to the Library."

We have also created a number of plotted images to be presented by Dr. Billington as gifts to Senators, Congressmen, and other dignitaries, including a facsimile atlas which will be presented to the Pope during a visit this month by the Madison Council to Rome.

The Division's home page is now available at
<http://lcweb.loc.gov/rr/geogmap/gmpage.html>.

and the images we have created will soon be accessible through our home page or the Library's American Memory home page.