Library Services for a Digital Future

Libraries' large data collections and varied digital assets demand tools and facilities for easy user access, data analysis, and storage

By **Duncan M. Aldrich** and **Greggory Stefanelli**

The proliferation of digital libraries over the past decade has produced so varied an array of digital collections and services that the term digital library defies a precise definition. And while what users expect of a library hasn't changed much—easy access to multiple resources, useful tools for working with the resources, and comfortable facilities—what it takes to meet these requirements has changed substantially with the advent of digital resources.

While most libraries have mastered the provision of basic digital services such as remote access to bibliographic databases, a growing number of libraries are venturing into home-grown digital collections. Others are experimenting with online user services such as reference chat, interactive FAQ databases, and institutional repositories. The direction each digital library effort takes depends largely on the perceived need of the library and its user community; on the extent to which the projects operate in collaboration with other digital library initiatives; and on the availability of resources to support each venture.

The University of Nevada, Reno (UNR) Libraries initiated its spin on digital libraries as a partner in a W. M. Keck Foundation grant awarded to the university in 1997. The overall grant ($2,250,000) supported a variety of earth science–related projects at UNR. The UNR Libraries's portion ($450,000) funded establishment of the W. M. Keck

Earth Sciences and Mining Research Information Center (the Keck Center), a series of digital collections and services accessible on the Web.

The Keck Center was an integral part of a proposal submitted to the W. M. Keck Foundation to create a center focused on the geosciences. Initial content collections included digitized and georeferenced U.S. Geological Survey (USGS) maps. The Keck Center also hosted interactive map services on an ArcIMS mapserver and CD-ROM–based bibliographic databases on a Citrix server.

Keck Center staffing has evolved since its implementation to meet the challenges specific to managing a growing digital library providing both services and collections. The group that operated the Keck Center was reorganized as the library's DataWorks work group in 2002. The DataWorks group subsequently became a library department in 2003,
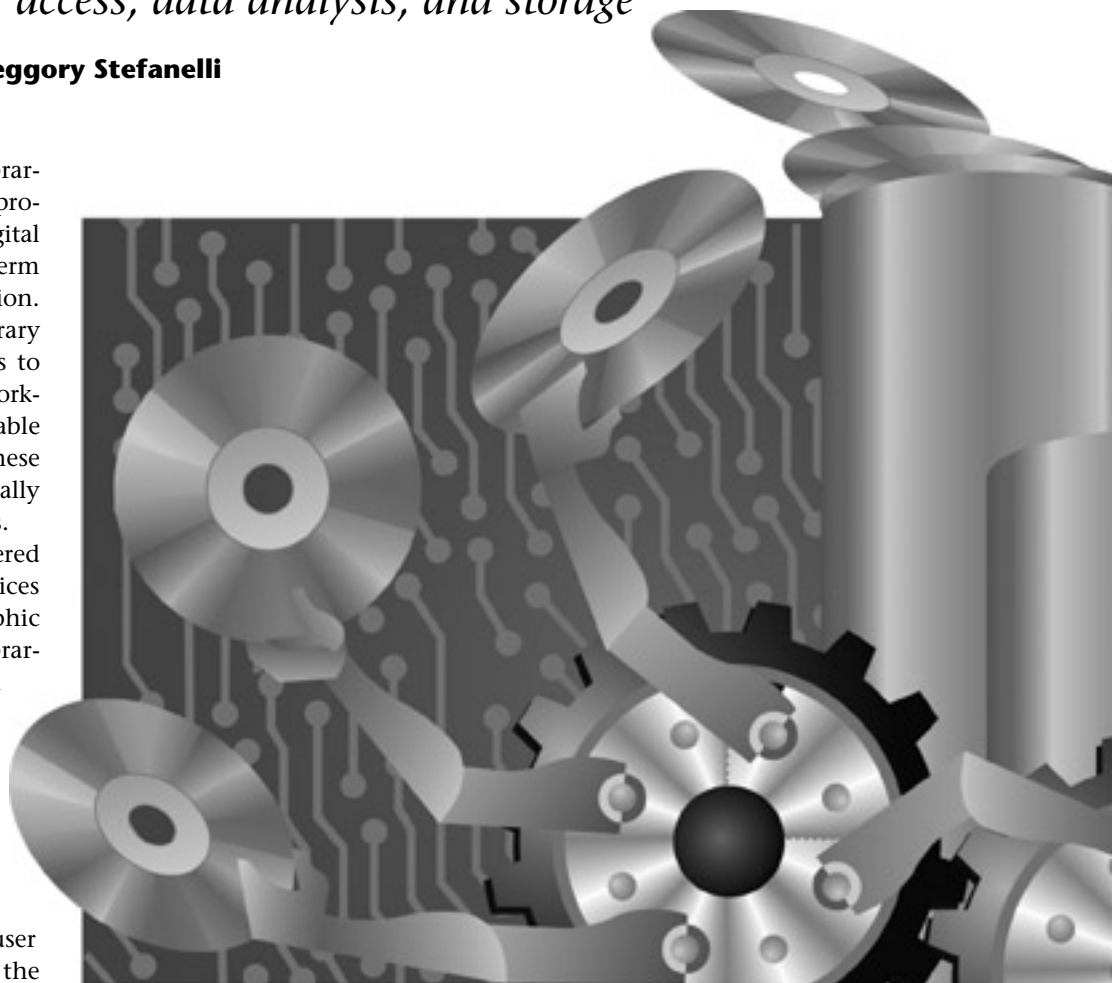
and in 2004, the library administration established the KnowledgeWorks committee, which functions as an umbrella advisory group over DataWorks and other library digitization efforts. (See <http://www.library.unr.edu/>.)

Keck Center Development

The DeLaMare Library, the university's earth sciences and engineering branch library, was initially approached by faculty from the Mackay School of Mines at UNR with the idea of submitting a proposal to the Keck foundation to create a Keck Center of Excellence in the Geosciences. The geosciences and engineering librarians collaborated with Mackay School faculty to draft the grant proposal, describing the Keck Center as a virtual rather than physical location providing a portal to information and data in support of geosciences research.

DeLaMare Library faculty and staff, in collaboration with the University Libraries' computer systems office, began developing the Keck Center in May 1998. Implementation of the Keck Center provided an opportunity to work through preliminary problems associated with building the technical infrastructure for supporting a sophisticated Web-based research platform that library patrons could use in accessing and manipulating the collections and resources.

Initial development of the Keck Center was accomplished by a group of three library faculty (each having other primary responsibilities), a part-time adjunct faculty hired for the project, and a small amount of student assistance time. No one was specifically in charge—a factor that ultimately led to managerial reorganization.

Equipment originally purchased for the Keck Center consisted of:

■ An FTP server connected to a 100-slot, dual-drive DVD jukebox providing Web access to a small DVD-based collection of digital maps and geographic images. These initial contributions to the collection were acquired from various public agencies such as the Bureau of Land Management and the USGS.
■ A Citrix Metaframe server farm to run CD-ROM–based bibliographic

databases such as IMMAGE the Merck Index, quickly adapted to provide Web access to a variety of data-analysis and visualization software packages, including ArcView, Maple, Mathematica, ChemDraw, ArcGIS, and ENVI.
■ An interactive mapserver running ESRI ArcIMS server software hosting an initial set of interactive maps created by our partners in the Mackay School of Mines and Nevada Bureau of Mines and Geology.
■ A server providing a node in the National Geospatial Data Clearinghouse (NGDC). The NGDC is a national network of metadata servers usually administered by state or federal agencies or universities. Users can query any server in the network for specific geographic metadata or data. The Keck Center received subsequent funding from the Federal Geographic Data Committee to create FGDC-compliant metadata for the Keck Collection.

The early work on the Keck Center succeeded in establishing a technical framework, collecting substantial amounts of data, and working through licensing arrangements with software suppliers, which have generally benefited digital library services. Equipment and software costs associated with developing the center were largely covered by the initial grant, though the library incurred related expenses, the most notable being the approximately $60,000 annually committed to staff and faculty salaries.

Problems were inevitable, of course. The FTP-DVD jukebox model proved to be slow, inefficient, and unreliable under high loads. Additionally, hardware failures combined with a lack of consistent backup and archival procedures resulted in extensive downtime and sometimes difficult data recovery. Some data had to be restored from poorly indexed CD-ROMs.

Under New Management

The limitations of management by committee, which had become increasingly apparent over the first two years, approached the crisis point

during a hardware failure. The library administration determined that better coordination of the development team was needed and might be achieved by bringing in a neutral party to lead the group.

The DataWorks work group (http://www2.library.unr.edu/dataworks/) was established in December 2001 to coordinate and provide direction for the Keck Center committee and to establish a unit within the University Libraries for dealing with future electronic information projects. Consisting of the library's Web Master, Server Manager, Director for Research Services, and newly created DataWorks Coordinator, the DataWorks group would manage and continue development of the Keck Center and, looking to the future, provide a management context for coordination of future digital projects.

The DataWorks coordinator also assumed general data services responsibilities for the libraries, such as overseeing the university's Interuniversity Consortium for Political and Social Research (ICPSR) membership, providing outreach to faculty and students and training library staff. These duties were previously distributed among various librarians.

The reassignment of the DataWorks Coordinator from a line librarian position was symbolic of the libraries' need to reallocate resources to new technologies. The librarian who volunteered to fill the position had a strong background in statistical resources but only limited experience with data analysis and visualization tools. He invested a substantial amount of time getting up to speed on geographic information system (GIS) and statistical analysis software. The Library's Server Manager, who is directly responsible for managing DataWorks systems, has a background in the physical sciences and mathematics and experience with many of the visualization and modeling software packages. This staffing arrangement assures that at least a minimal level of support could be provided on all software that DataWorks supports. Since the Keck Center is virtual, there was no need to physically relocate staff.

The DataWorks Group identified initial tasks to plan and deploy a more robust and fault-tolerant data storage implementation:

- Develop and deploy adequate back-up and archival procedures consistent with the unique nature of the Keck Collection (extremely large data sets).
- Contract for an off-site location to maintain archival copies of all resources on the Web site and develop an aesthetically pleasing Web interface.
- Create or procure a robust search application for the collection.

These projects were accomplished within the first year, giving DataWorks time to expand the Keck Center's scope of operations beyond the geosciences.

From Virtual to Physical

Over the past three years the Keck Collection—initially containing multi-spectral satellite imagery, digital elevation models, orthophotoquads (a type of georeferenced aerial photograph), and digital topographic maps—has grown significantly. A variety of other data collections have been added, including ASTER images, Mars laser altimetry data, and Venus radar data.

To accommodate this growth and provide redundancy, a second NAS was added and the original NAS capacity greatly increased. The possibility of downtime due to hardware failure was essentially eliminated by employing a distributed file system across the two NAS devices. Incremental backup occurs nightly, full backups weekly, and archival backups monthly. Three multiprocessor Citrix servers provide Web access to data visualization and modeling software.

The current hardware configuration consists of two 8-terabyte enterprise network attached storage (NAS) devices that provide substantial digital storage, file redundancy, and structured, well-managed backup and archival procedures. The NAS devices provide infrastructure for a Windows Storage Server Distributed File System (ntDFS). An LTO2 tape library provides 34 terabytes of backup and archival storage.

Software selection for the Citrix server farm depends on the type of data in the various collections and on faculty and student requests. Currently the Citrix server farm hosts approximately 70 applications, application extensions, or special modules providing a wide range of data analysis, modeling, and visualization capabilities. In addition, a large-capacity database server (MS SQL Server 2000) with 1.6 terabytes of storage, provides a database server backend for special projects and applications. Individual faculty and departments are invited to use space on DataWorks data storage servers for scholarly projects.

A DataWorks lab located in the main library is open during hours of library operation, though staffed only during selected hours. The lab consists of 15 multiprocessor workstations that host all applications available on the Citrix server farm. Additionally, SAS and SPSS software packages were purchased for social science–oriented statistical analysis. Textual analysis software, QSR NU*DIST N6, was also procured. The addition of social sciences statistical analysis software reflected a new emphasis for DataWorks—support for data needs on campus regardless of discipline. Other standard PC software (such as Microsoft Office and Jasc Paint Shop) is also loaded on the machines. All of the software products support both instruction and research.

The lab gives students, faculty, and staff a location to use high-end workstations replete with large-format HP DesignJet 500PS 42 and HP Color Laserjet 4550N color printers. Many of these applications are also available on 10 dual-processor Mac G5s in the Dynamic Media Lab, located adjacent to the DataWorks lab. The Dynamic Media lab works closely with the DataWorks lab in supporting users and applications.

User authentication for all DataWorks platforms and services is handled via the campus Active Directory known as UNR NetID. DataWorks and the UNR Libraries were early adopters, providing a test bed for later campus development. NetID now serves as the authentication method for all services provided by the Information Technology Division. In addition, many administrative and academic departments employ NetID for authentication on their departmental systems. Active Directory has proven to be a reliable and stable mechanism for providing authentication services for a variety of platforms, including Windows, MacOS, and Unix.

Moving Ahead

In 2003, DataWorks transitioned from a work group to a library department. The library pays the salaries of the DataWorks department's full-time librarian and half-time server manager from its personnel budget. Salaries for the two graduate assistants and student workers are paid from the library operations budget. As a department, DataWorks continues to receive considerable assistance from the previous members of the DataWorks group, particularly the library Web Master.

The short-term future for DataWorks was recently defined by a series of projects done in collaboration with UNR academic faculty. DataWorks's participation consists of the following capabilities:

- Interactive mapping services based on ArcIMS for serving and preserving such data as Great Basin geothermal energy resources and earthquake maps of Nevada and eastern California. A major project to provide Web-based access to real-time earthquake data acquired by the Nevada Seismological Laboratory (NSL) seismograph network will be finalized in early 2006.
- A SharePoint portal server hosting collaborative space for several campus initiatives. DataWorks staff manage the SharePoint servers and provide outreach to campus entities that might benefit from using the Web-based collaborative platform SharePoint provides. The main target to date have been committees needing to share documents, such as search and promotion and tenure committees.
- Deployment of an SPSS SmartViewer Web server for publication of interactive SPSS OLAP cubes to the Web. This allows the university to publish its institutional data to the Web and faculty to publish statistical research in support of publications or conference presentations.

- Storage and support for analysis of Mars Orbiting Laser Altimetry (MOLA) data from the Mars Global Surveyor.
- Storage and support for analysis of a very large (500 gigabyte) set of molecular modeling data.
- A Web-based bibliography of dissertations pertaining to Borneo (in collaboration with the Borneo Research Council).
- Development of an interactive Web database of the complete works of the Spanish novelist Benito Perez-Galdos (which provides for contributed annotations to Web-based Perez-Galdos texts).

These later projects engendered considerable discussion among the Data-Works Group as to what the unit's focus should be—exclusively numeric data or digital information generally—and about setting up a digital repository for the university's creative output. Subsequent library and IT discussions have been carried out under Knowl-edgeWorks, a library committee that provides an umbrella for three units tentatively named DataWorks, Image-Works, and WordWorks.

DataWorks will focus on data collections and services; ImageWorks on digital image, video, and audio files; and Word-Works on digitized and native digital texts. The expectation is that Knowledge-Works will position the libraries to meet the following strategic planning goals:

- Anticipate a largely digital future that encourages disintermediated information retrieval and remote and more immediate (virtual) access to large collections of information.
- Provide a data hosting and man-agement service to all university researchers.
- Work directly with teaching and research faculty to shape not only the content of these services but also their presentation.

Conclusion

The name KnowledgeWorks resonates well with the University Libraries's change of focus from information to knowledge, demonstrated by construc-tion of the Knowledge Center to replace the library in 2008. Just as the Keck Cen-ter is much more than a collection of digital objects, the new Knowledge Cen-ter will be more than a library. Within the Knowledge Center, the Keck Center will continue to provide a virtual plat-form on which visitors not only locate warehoused data resources but also manipulate materials with analysis and visualization software and store their own data for personal, collaborative, or public use. In this sense the tools and resources that DataWorks makes avail-able enable Knowledge Center users to transform information into "small k" knowledge. A simple example is down-loading census data from the Web and creating a digital map project using GIS software. This is the sort of added value the new Knowledge Center will make possible in a variety of ways. *e*

Duncan M. Aldrich (Duncan@unr.edu) is Data Services Coordinator and Greggory Stefanelli (stef@unr.edu) is Server Systems Manager at the University of Nevada, Reno Libraries.