The New Space and Earth Science Information Systems at NASA's Archive

JAMES L. GREEN*

The National Space Science Data Center (NSSDC), established in 1967, serves as a long-term archive and distribution center for data obtained on NASA space science flight investigations and provides a variety of services to enhance the overall scientific return from NASA's initial investment in these missions. NASA science data at the NSSDC cover the disciplines of astrophysics, Earth science, planetary physics, and space plasma physics.

Over 6,000 gigabytes of digital data (125,000 magnetic tapes) and 91 million feet of film products from NASA science missions have been acquired by the NSSDC since it was established in 1967[1]. To handle the requests for both digital and film products, the NSSDC has a variety of computer systems, both interactive and batch; dedicated photo laboratory facilities; large online database management machines; and optical mass storage devices. It also manages NASA's largest computer-to-computer wide area network.

Based on current agreements concerning future NASA missions, the NSSDC data holdings will increase dramatically, nearly doubling every 2 years, reaching nearly 40,000 gigabytes by 1995. Innovative ways of managing the information about such large volumes of data and implementation of large mass storage systems are necessary to provide users with better archive access while the NSSDC effectively manages the ever-increasing volumes of data that are coming into the archive[2].

With the ease of electronic access dramatically increasing over the last few years, the NSSDC has created a major new thrust by developing online computer information systems accessible to remote users 24 hours a day. Currently, not all the information about the NSSDC archive is accessible to remote users, and less than 2 percent of the NSSDC's total digital data archive is online[3], but these systems are already a major achievement in providing rapid access to NASA-acquired science data that is unprecedented in archive data management.

*Direct all correspondence to: James L. Green, NASA Goddard Space Flight Center, Code 630.2, Greenbelt, Maryland 20771.

Government Information Quarterly, Volume 7, Number 2, pages 141–147.
ISSN: 0740-624X.

This article focusses on the new NSSDC online interactive systems that are used extensively over international computer networks. These systems are typical of how NSSDC is responding to user demands for rapid access to archived data.

WORLDWIDE NETWORK CONNECTIONS

Many computer network connections have been made to provide remote access to the NSSDC by all its diverse users. Figure 1 shows a breakdown of the major network connections by communication protocol. The Bitnet connection supports only mail communication among many universities in the United States and NSSDC. Selected science computer network nodes worldwide and the general public primarily use the X.25 international packet networks to gain access to the data center. There are two major wide-area NASA networks that are used extensively by NSSDC: The Space Physics Analysis Network (SPAN)[4] and the NASA Science Network (NSN). SPAN contains more than 2,800 nodes in the United States and is internetworked with more than 10,000 nodes in the United States, Europe, Canada, and Japan (through other networks such as HEPNET). SPAN is managed by NSSDC and used exclusively by space and Earth scientists working primarily on NASA-related missions and projects.

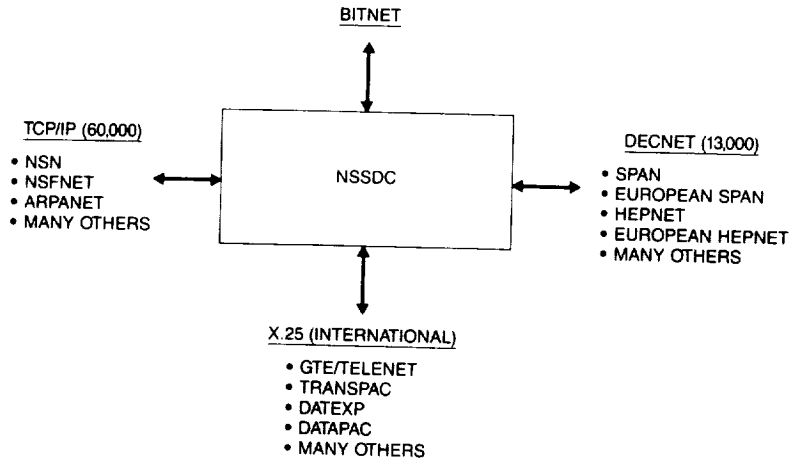

Figure 1. The wide-area network access to the NSSDC. NSSDC manages the SPAN computer network, which supports many connections to other wide-area DECnet networks such as HEPNET. Another major NASA network is the NSN, which provides TCP/IP connectivity to many other computer networks and the NSSDC. The Bitnet connection supports only mail communication between many universities in the United States and NSSDC. Selected science nodes throughout the world, in addition to the general public, use the X.25 international packet networks primarily to gain access to the data center.

NSN (which uses the TCP/IP protocol) is internetworked with other wide-area networks, such as ARPANET and the NSFnet, and can reach many thousands of computers, primarily at universities. In general, these wide-area networks are of relatively low speed but are providing a tremendously valuable service for remote users to gain access to NASA computer resources and to communicate with fellow researchers across the country.

The bulk of the wide-area network traffic is for informational purposes such as remote logon and mail; however, data transfer (in limited amount) is also supported. The wide-area networks provide the pathways for remote users to access the NSSDC facilities at any time, day or night[5].

"NEW TECHNOLOGY" DATA AND INFORMATION SYSTEMS

The NSSDC is responding to an ever-increasing number of user requests by putting more of the data and information about the data in its archive online for direct user access. With the ease of electronic access dramatically increasing over the last few years, the NSSDC's new online computer information systems can now be accessible to remote users 24 hours a day. This allows the NSSDC to "remain open" past normal working hours, providing scientists and students the ability to "browse" through the online information to look for an important data set.

The new online data and information systems currently operational at the NSSDC are shown in Table 1. These systems have been a tremendous success, handling more than 2,500 accesses by remote users annually and growing rapidly. The systems shown in Table 1 provide a variety of services, depending on the desires of the community of scientists they serve. The online systems provide information about data holdings, with several levels of complexity. For instance, the Master Directory (MD) contains a high-level overview of data held in the NSSDC and at a number of other NASA centers, established U.S. science research institutions, and other U.S. government agencies such as the National Oceanographic and Atmospheric Administration (NOAA) and the U.S. Geological Survey (USGS). The MD, therefore, is the first reference system to point to where the data are held[6]. More detailed information about data holdings, such as the processing history, quality, time resolution, etc., must be found in the other data systems, such as NASA Climate Data System (NCDS), to which the MD will refer a user.

The online information systems MD, personnel data base, Total Ozone Mapping Spectrometer (TOMS) data, Solar Wind data (OMNI), and the International Ultraviolet Explorer (IUE) request system are all accessible from one computer account called the NSSDC Online Data and Information System (NODIS). The interactive traffic to the NODIS account over the last year is shown in Figure 2. Since 1987, many new systems have been added to the NODIS account. Figure 2 shows the tremendous popularity of the NODIS system, with the average user accessing NODIS more than twice a month.

Another important example of the ready access to the data in the NSSDC archive is the IUE Interactive Request System. This system allows scientists to order observations taken by the IUE spacecraft.

The IUE Interactive Request System became operational in November 1987 through the NODIS account. The request system[7] consists of a large online mass storage device, menu-driven interactive information access software, a high-speed local-area network connecting the online storage with the interactive front ends, and the wide-area networks access to the system.

Rapid access to selected data has been frequently requested by scientists. Since it is not known ahead of time what sections of any one data set will be requested, the NSSDC loaded all the IUE data into the NASA Space and Earth Science Computer Center's IBM 3850 Mass Store in order to better accommodate the large user demand through faster request

Table 1. NSSDC New Technology Online Systems.

SCIENCE DISCIPLINE	SERVICE	INFORMATION	DATA*
All			
	Master Directory	X	
Astrophysics			
	IUE Request System	X	X†
	ROSAT Information Management System	X	
	Astronomy Catalog System	X	X
	STARCAT with SIMBAD access	X	X
Atmospheric Science			
	NASA Climate Data System	X	X
	Ozone TOMS Data	X	X
Land Sciences			
	Crustal Dynamics	X	X
	Pilot Land Data System	X	
Space Plasma Physics			
	Central Online Data Directory	X	
	Omni Solar Wind Data System	X	X
	Plasma and Field Models	X	X††
	Coordinated Data Analysis Workshop	X	X
General			
	SPAN-Network Information Center	X	
	Personnel database	X	

NOTES:
* Only partial data sets are available
† All available data is on line
†† Only software is being distributed

and delivery response. It is important to note that the NSSDC typically manages its archive offline. Storing all the IUE data online was done with full IUE Project cooperation to gain valuable experience with highly requested online data sets. The IUE data that are currently online consist of over 70,000 unique star images and spectra and total approximately 90 gigabytes in volume.

The IUE interactive request system software runs on the NSSDC VAX computers, which allow for a remote SPAN user to log on and order IUE data from the electronic IUE Merged Observer Log.

Once the exact data segment requested has been identified, the NSSDC request coordinator networks the IUE data from the Mass Store system. For requesters desiring a small number of spectra, the NSSDC request coordinator can network the data through SPAN to the target computer of the requesting individual within approximately 24 hours or create a magnetic tape to be mailed. Requests for IUE data sent on magnetic tape are handled easily by this system. Requests for IUE data also come to the NSSDC through letters and phone calls (not all users are on computer networks).

Figure 3 shows the yearly number of IUE images requested by individuals from 1979 to 1989. NSSDC also sends large amounts of IUE data to other archives; these requests are not included in this figure. The solid colored bars in Figure 3 show the number of images sent out on magnetic tape to individual requestors; the size of the cross-hatched bar represents the average monthly number of IUE images that have been networked to remote users using SPAN. From 1979 to nearly the end of 1987 the only service the NSSDC offered was an offline service in which a tape copy of the data was produced and sent to the requester. The bar graph also shows the yearly number of IUE images requested in 1988 and 1989.

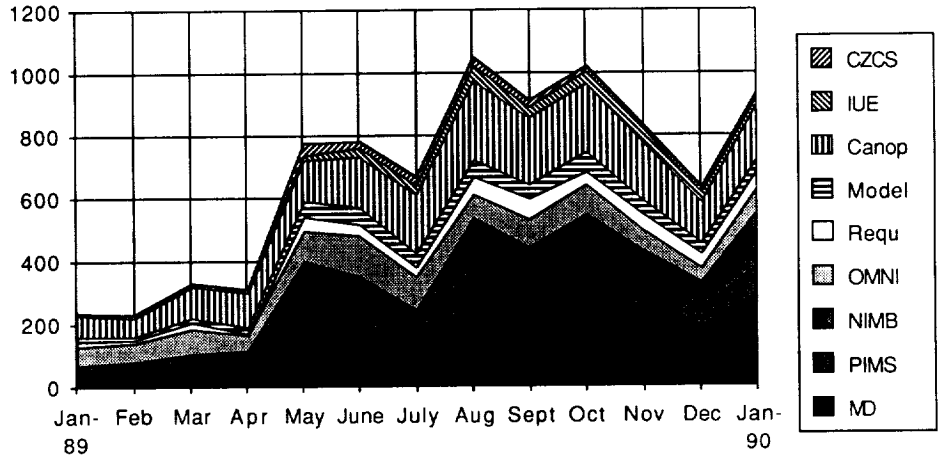

Figure 2. Interactive traffic from remote users, accessing several key NSSDC systems. Using the NODIS System, number of sessions (open circles) and users (filled-in points) is plotted against time. The dramatic increase seen after April 1989 results from several Information Systems becoming operational. It is important to note that the average number of accesses by scientists is more than twice per month.

Figure 3 also shows the dramatic increase in the amount of IUE data requested in 1988 and 1989, reaching approximately 3500 images and spectra a year. The computer networks were used to deliver, about 38 percent of the data, in 1988, and 60 percent of the data in 1989, while the remainder, are satisfied by sending magnetic tapes. Currently, the trend in the use of the IUE Interactive Request System, as well as all the other systems shown in Table 1, continues to climb.

The networked IUE images satisfied requests from many scientists at 15 institutions in the United States, Europe, and Canada (locations serviced by SPAN). In addition, care is taken to use SPAN for networking of the IUE data at times of non-peak network usage. Tests are currently under way in which some IUE data are compressed before being

networked to the user's remote computer and are then decompressed, therefore reducing the communication load on the wide-area network.

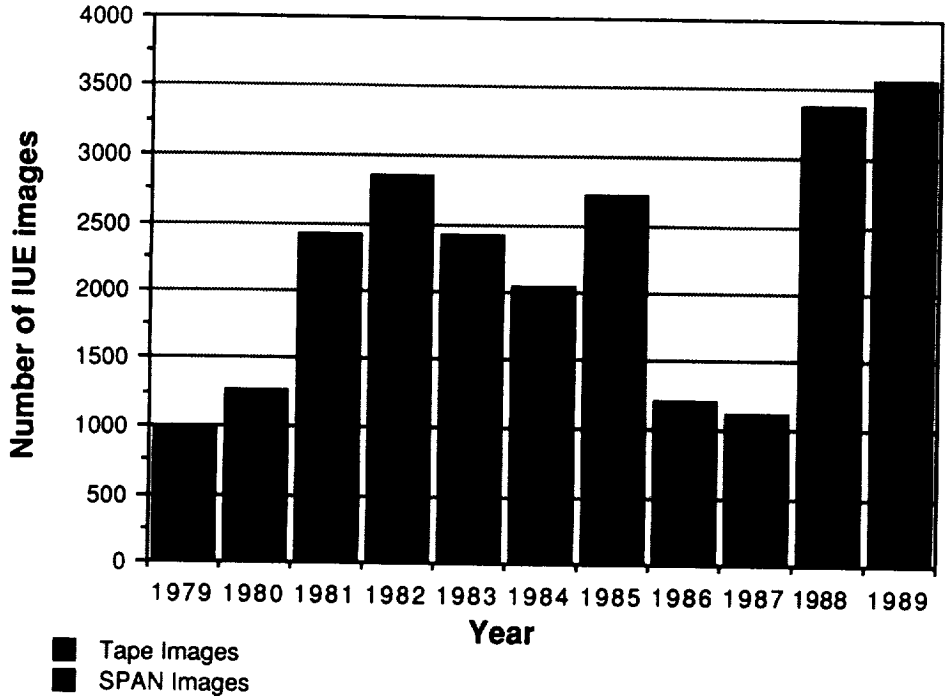

Figure 3. Number of IUE images requested (per year) since the archive was opened in 1979. Although the IUE Interactive Request System became operational in November 1987, it was not until January 1988 that remote users routinely accessed the new interactive system. The huge increase in the number of images distributed in 1988 and 1989 can be attributed to the better service that is now provided electronically. The cross-hatched bar in 1988 shows that nearly 40 percent of all images requested are delivered over computer networks. In 1989 it was 60 percent.

The request results for 1988 and 1989 (Figure 3) clearly show that the tremendous increase in requested data results from the convenience the interactive request system provides to the user. The following factors are a major part of the user convenience provided by the IUE Interactive Request System:

- immediate ordering of needed spectra/images;
- rapid turnaround providing the desired data while the scientists are interested;
- data loaded to the target system (no tape handling);
- data arriving in the desired format; and
- no need to send replacement tape to the NSSDC (currently the SPAN and NSN networks are a "free" service to users).

The IUE example is typical of all the interactive data and information systems that allow NSSDC to disseminate as much information as possible in a timely manner.

FUTURE INFORMATION SYSTEM ACTIVITIES

Based on current agreements concerning future NASA missions, the NSSDC data holdings will increase dramatically, approximately doubling every 2 years, reaching nearly 40,000 gigabytes by 1995. This is a staggering amount of data. Within the next year, the Hubble Space Telescope (HST) will be launched and will produce approximately 2,500 gigabytes (2.5 terabytes) of data per year. HST is a new type of mission for NASA; it is being designed as a "nearly permanent" observatory in space (lasting at least 15 years) and is one of NASA's several Great Observatories in space. Because of the huge amounts of data and the importance of the mission to the astrophysics community, the Space Telescope Institute in Baltimore, Md., has been created to be responsible for the science mission management and data archiving for the HST. It plans to make all HST data readily available through an online data archive.

The ability to electronically access and query the contents of a remote archive is of tremendous importance, greatly facilitating research in the space and Earth science fields. These online information systems must continue to grow in capability and complexity in order to accommodate the huge data volumes NSSDS expects to manage in the near future. It is clear, from examples like the Master Directory and IUE Interactive Request System, that online interactive information and data retrieval systems do provide a better service to the science research community than the offline letter requests for highly requested data sets.

Through the new online information systems, the NSSDC is striving to support active archive research on the individual scientist level at any time of day or night convenient to the researcher. The NSSDC will continue to aggressively pursue the "electronification" of its information about data and, to the extent reasonable, its archived data[8]. Much of the new data coming into NSSDC will be managed by the online interactive systems, but much more work remains to be done with the existing archived data.

NOTES AND REFERENCES

1. James L. Green, "NSSDC Data Holdings: Scope, Complexity, and Future Requirements," in *Proceedings of the Workshop on Geophysical Informatics*, Viniti Publishing House, Moscow, 1990.
2. Joseph King, "Mass Storage Technologies for NASA's Future Data Environment," in *Proceedings of the Workshop on Geophysical Informatics*, Viniti Publishing House, Moscow, 1990.
3. James L. Green, "The NSSDC Information Systems: Status and Plans," in *Proceedings of the Workshop on Geophysical Informatics*, Viniti Publishing House, Moscow, 1990.
4. James L. Green, "The Space Physics Analysis Network," *Computer Physics Communications*, 49 (1988): 205-213.
5. James L. Green, "NSSDC Applications for Computer Networking," in *Proceedings of the Workshop on Geophysical Informatics*, Viniti Publishing House, Moscow, 1990.
6. Joseph King, "Data Finding in a Distributed International Environment," in *Proceedings of the Workshop on Geophysical Informatics, Viniti Publishing House, Moscow, 1990.*
7. *Charleen M. Perry, "The National Space Science Data Center and International Ultraviolet Explorer: 1978#Present," in Decade of UVAstronomy with the IUE Satellite*, ESA Special Publication (ESA SP-281) Vol. 2: 357-360, June 1988.
8. James L. Green (editor), The National Space Science Data Center: Users Guide, NSSDC Technical Publication, (January 1989).