REPORT UAG-99

WORLD DATA CENTER A
for
Solar-Terrestrial Physics

PROCEEDINGS
of the Workshop on Geophysical Informatics
Moscow, August 14-18, 1988

January 1991

NATIONAL GEOPHYSICAL DATA CENTER

International Council of Scientific Unions

Panel on World Data Centres

Academy of Sciences of the USSR

Soviet Geophysical Committee

PROCEEDINGS

of the Workshop on Geophysical Informatics

Moscow, August 14 – 18, 1988

Edited by

J.H. Allen
V.A. Nechitailenko

Moscow 1990

CONTENTS

THE WORLD DATA CENTER SYSTEM. A CURRENT AVENUE FOR INTERNATIONAL DATA EXCHANGE

J.H. Allen

ICSU Panel on World Data Centres
Liaison Sub-committee for IGBP

Introduction

The World Data Center (WDC) system is a dynamic international network of 27 currently active centers linking data contributors to data users. It has operated for some 30 years under ICSU guidance. The WDCs were created in 1957 as archival centers for geophysical and solar data collected during the International Geophysical Year (IGY). In 1960 ICSU requested that they be continued as a non-governmental mechanism for international data exchange in their respective disciplines. Overall guidance is provided by the ICSU Panel on World Data Centres, taking account of advice from associated international scientific bodies.

Principles and responsibilities of WDC operation are published in the GUIDE TO THE WORLD DATA CENTER SYSTEM: PART I, THE WORLD DATA CENTERS (December 1987). The GUIDE also includes historical information about the WDCs, a concise description of each discipline center, information about other international data centers and data exchange networks outside the WDC system, and a brief account of several new major international scientific programs for which WDC services are needed, including GLOBAL CHANGE.

In 1978 the 17th General Assembly of ICSU recommended that "prior to approving the initiation of new projects in the fields of geophysics and solar-terrestrial physics, the Executive Board should ensure that the planning for these projects includes clear provision for data collection, archiving and distribution and that such plans have been developed in consultation with the ICSU Panel on World Data

Centres." From a WDC viewpoint, the International Geosphere-Biosphere Program (IGBP), "Global Change", is an especially challenging new scientific program. Much of it is outside traditional WDC areas of "geophysics and solar-terrestrial physics", yet IGBP emphasises identifying key global data bases and determining decadal and longer-term reference values for its disciplines, clearly tasks which call for organized data management.

The ICSU Panel on WDCs has appointed a liaison sub-committee to work with the Special Committee on IGBP in meeting the requirement for sound data management planning. As Co-Sponsor of this Study Conference and the following Workshop on Geophysical Informatics, the Panel hopes to inform the IGBP community about the existing WDC system, to identify specific data management roles that may be played by present centers, and to learn whether there are needs and resources for new facilities to be created in order to organize IGBP data management in disciplines not presently within the WDC framework. In these meetings we seek to identify available mechanisms for efficient global exchange of data and information, to anticipate new needs that will develop, and to learn how best to adapt new technology within realistic resource limits to perform the necessary tasks for the global scientific community.

The World Data Center System In 1988

Today 27 active centers comprise the operating elements of WDCs A, B and C that together form the World Data Center system. They are described in detail in the most recent GUIDE and are listed below. WDC-A is a distributed (or "virtual") center located in the United States and divided along scientific discipline and platform lines into nine (9) different sub-centers. WDC-B is in the Soviet Union and comprises two(2) centers that also span all discipline areas for which the WDC system was originally created. Thus, WDCs A and B are known as complete or "complex" centers. WDC-C is distributed among several countries and is divided into C1 and C2 sections. WDC-C1 is in Europe (8 centers) and WDC-C2 is in Japan (8 centers). Together all WDCs engage in routine and special exchanges of data and information on film, on magnetic tape and floppy disks, in published or manuscript form, as data maps, and by computer mail, telex and facsimile machine. Lately WDC-A centers have begun to experiment with preparation of data collections on optical discs and other mass storage media for more efficient data exchange, improved user access to data, and more compact

archival storage for digital data. The WDCs are also sources of information about international scientific programs, sometimes distributed worldwide through discipline or program newsletters, and of scientific expertise in their respective discipline areas. Some WDCs function as analysis centers, producing indices or models and most distribute derived products as well as primary data.

Many WDCs are co-located with national centers in their sponsoring countries and share staff who must be alert to distinguish whether, in a given instance, they are acting in a national role or on behalf of the international community. Original principles from the IGY era still are the basis for day-to-day center operations although some are no longer so important as they were during the earlier epoch. For example, less emphasis now is given to the IGY practice of duplicating all data received among the several WDCs so as to assure their survival somewhere in the event of a disaster at one location. Duplicate records, particularly of extensive digital data files, are not seen as the only means to assure practical (affordable) accessibility to a given data set by scientists in each region of the world.

Fast and effective electronic communications today link several WDCs, joint catalogs are regularly prepared and individual WDCs have key data catalogs on-line. However, in some disciplines valid reasons are recognized to continue routine copying and exchange of selected data sets and this still is called for in the "GUIDE". Usually such data are only taken daily or less frequently and they often are in analog chart format or film images. Other fully duplicated data sets include both analog and digital data bases obtained during major scientific programs and summary indices or models derived at an analysis center. In either case, users benefit from improved communications among centers and better common catalogs assure efficient access to data. The risk of catastrophic loss at a single center is acceptable when resource limitations are balanced against high costs to copy large digital data bases.

It is still thought to be worthwhile to have multiple WDCs worldwide for the same disciplines. Each center provides a focus in a major global region for the collecting, archiving and dissemination of the type data for which it is responsible. Through cooperation with scientists of the region served, these centers identify and copy those key data sets that are most important to preserve for future reference. They are convenient regional nodes in the global network for deposit into the system of data collected by scientists and institutions in that area. They encourage working visits by scientists who take advantage of hands-on access to open data files, and who use computer and other specialized analytical equipment needed for the operation of the center. As regional nodes in the worldwide communication and information network connecting the WDC system and other data sources and collections, these centers are also natural collection and dissemination points for information about programs. Finally, such centers are especially suited for training younger scientists and are valuable sites where staff expertise is developed for the many different types of data handled. Resident WDC scientists are a resource for cooperation in program planning and research within the larger scientific community of government agencies, academia and industry.

Principles of WDC Operation

A continuing WDC operating principle from the inception of the system is that data are completely accessible to all scientists in all countries without exception. The concept of "classified" or "restricted distribution" data sets, while realistic for national centers, is foreign to WDC principles. If data are held by a WDC, they are available to anyone requesting them although sometimes charges for the cost of copying and handling are applied to requests from groups that do not supply data to the system and hence are not entitled to equivalent exchanges without charge.

WDCs are expected to fulfill data exchange requirements set out in the GUIDE, Part 1 section on general principles and in the discipline and program sections either as published in 1979 or in new editions which become available to supercede the older documents. They respond to directions from the ICSU Panel and to resolutions and recommendations from appropriate international organizations. WDC-to-WDC data exchanges are made without charge as part of the continuing routine exchange process. Special requests that would cause unusual impact on the staff or budget resources may bear a charge. Also, there may be special charges for data obtained by a WDC from a national or regional data center in response to a request from another WDC or from an individual scientists. In the case of requests for data not kept at a WDC, the centers are expected to assist in obtaining the data or to forward the request to another center or institution that may be able to respond directly.

WDCs are to maintain data collections with proper facilities for their safe long-term retention and for efficient retrieval and accurate reproduction and dissemination to users. International standards of data accuracy, clarity and durability are to be employed at WDCs and they are expected to maintain a continuing effort to explore the use of modern technology to improve techniques of data storage, communications and user access. The centers are to be open to visiting scientists from all countries and regular reports on operations and catalogs or inventories of holdings are provided that describe available data and services at WDCs. Where multiple WDCs have data of the same type, they are expected to prepare joint catalogs and make these available to potential users. They must endeavor to coordinate their activities, standardize data formats (possibly maintaining separate and different internal and external/exchange formats), and cooperate in international data-gathering projects through advance planning and participation in international scientific meetings. While the data sources are ultimately responsible for the quality of data provided to WDCs, each center is expected to cooperate in reasonable efforts to assure data reliability, accuracy and quality.

World Data Center – A

Coordination Office, US National Academy of Sciences, Dr. P.J. Hart.

(Titles reflect types of data serviced by each WDC-A discipline center)

[sponsoring institution given in brackets]

The nine discipline centers of WDC-A are:

GLACIOLOGY (SNOW and ICE), Boulder, Colorado, USA
[University of Colorado, supported by NOAA/NGDC]

MARINE GEOLOGY and **GEOPHYSICS**, Boulder, Colorado, USA
[NOAA, National Geophysical Data Center (NGDC)]

METEOROLOGY, Asheville, North Carolina, USA
[NOAA, National Climatic Oceanography, Washington, DC, USA Data Center]

Seismology, Golden, Colorado, USA
[U.S. Geological Survey]

ROCKETS and **SATELLITES**, Greenbelt, Maryland, USA
[NASA, National Space Science Data Center]

ROTATION of the **EARTH**, Washington, DC, USA
[U.S. Naval Observatory]

SOLAR-TERRESTRIAL PHYSICS, Boulder, Colorado, USA
[NOAA, National Geophysical Data Center]

SOLID EARTH GEOPHYSICS, Boulder, Colorado, USA
[NOAA, National Geophysical Data Center]

World Data Center–B

Operated under the auspices of the Academy of Sciences of the USSR, Soviet Geophysical Committee, Prof. V.V. Beloussov.

WORLD DATA CENTER–B1, Obninsk, USSR
[USSR State Committee for Hydrometeorology and Control of the Environment]
Data held for disciplines; Meteorology, Oceanography, Marine Geology and Geophysics, Glaciology, Rockets and Satellites, Rotation of the Earth, Tsunamis, Means Sea Level and Ocean Tides.
WORLD DATA CENTER–B2, Moscow, USSR
[Soviet Geophysical Committee, Academy of Sciences USSR]
Data held for disciplines; Solar-Terrestrial Physics and Solid Earth Geophysics.

World Data Center–C1

Represented on the ICSU Panel by Dr. E. Friis-Christensen.
EARTH TIDES, Brussels, Belgium
[Royal Observatory of Belgium]
GEOMAGNETISM, Copenhagen, Denmark
[Danish Meteorological Institute]
GEOMAGNETISM, Edinburgh, UK
[British Geological Survey]
GLACIOLOGY, Cambridge, UK
[Scott Polar Research Institute]
RECENT CRUSTAL MOVEMENTS, Prague, Czechoslovakia
[International Centre for Recent Crustal Movements]
SOLAR ACTIVITY, Meudon, France
[Observatoire de Paris]
SOLAR-TERRESTRIAL PHYSICS, Chilton, UK
[Science and Engineering Research Council of UK]
SUNSPOT INDEX, Brussels, Belgium
[Royal Observatory of Belgium]

World Data Center–C2

Represented on the ICSU Panel by Dr. M. Sugiura.
AIRGLOW, Tokyo, Japan
[Tokyo Astronomical Observatory, Ministry of Education]

AURORA, Kaga, Japan
[National Institute of Polar Research, Ministry of Education]
COSMIC RAYS, Tokyo, Japan
[Institute of Physical and Chemical Research]
GEOMAGNETISM, Kyoto, Japan
[Kyoto University, Ministry of Education]
IONOSPHERE, Tokyo, Japan
[Ministry of Posts and Telecommunications]
NUCLEAR RADIATION, Tokyo, Japan
[Japan Meteorological Agency, Ministry of Transportation]
SOLAR RADIO EMISSIONS, Toyokawa, Japan
[Nagoya University, Ministry of Education]
SOLAR-TERRESTRIAL ACTIVITY, Tokyo, Japan
[Institute of Space and Astronautical Research, Ministry of Education]

Some WDCs have unique scientific and technical expertise that qualify them to provide unusual services. For example, the Auroral Electrojet (AE) magnetic activity index was for years derived at WDC-A for STP and now is produced by WDC-C2 for Geomagnetism. WDC-C2 also derives the DST (Disturbance Storm Time) equivalent equatorial index of globally symmetrical magnetic storm effects. WDC-C for Sunspot Index in Brussels, Belgium produces the International Sunspot Number, continuing the series begun much earlier in Zurich.

Roles of WDCs in Major International Programs

World Data Centers cannot serve only as passive repositories of data obtained from contributors who are isolated from contact with the system and in isolation from active users of their data and products. In order to fill their roles as collection centers for both monitoring data and data taken in experimental campaigns, as sources of data copies, standard indices, derived summary compilations, and internationally adopted global and spatial models, the WDCs must also take an active part in planning and conducting international scientific programs and in analyzing the results from these programs.

During the International Magnetospheric Study (IMS: 1976–79), the WDCs developed new services in providing information, data management and analysis that were unique to IMS. They achieved such success that these became models for similar services sought for future programs. The IMS Satellite Situation Center generated timely lists of satellite conjunction intervals, times when constellations of working satellites were well placed to perform coordinated observations. They alerted scientists worldwide to plans for new satellite instrumentation, produced valuable orbit foot-track maps used to coordinate joint programs of ground-based and space observations, and assisted in the orbit rescue mission for a mislaunched satellite.

The IMS Central Information Exchange (IMSCIE) Office was internationally staffed and provided a clearinghouse for information about ground-based, rocket, balloon, airborne, shipboard, and satellite observing campaigns. The monthly IMS Newsletter was a air-mailed to some 3,000 scientists worldwide. It contained articles about recent observations, maps and descriptions of instrument networks, proposals for special analysis topics, early reports of results from studies that should be taken into account in planning further operations, and a calendar of scheduled launches and expeditions for the next six months. This information resulted in many instances of voluntary cooperation between experimenters who increased the impact of their programs by coordination with others. Telex messages were used to pass timely information to national contacts and groups of scientists seeking to coordinate programs or take account of special conditions, e.g. the occurrence of sunspot cycle minimum.

Finally, the Coordinated Data Analysis Workshop (CDAW) was developed to provide a means for researchers to combine many different data sets on a dedicated central computer system with data from monitoring arrays and other research projects. Those contributing data met as a group to study their special topic by intensive analysis of the combined on-line data using a family of custom data retrieval, analysis and reproduction programs. The results were published and continued access to the assembled data base was available to participants and often to the larger community of interested scientists.

In addition to providing these new specialized activities for international programs, the WDC system continued the basic routine of collecting, processing, archiving, and disseminating services for data from the global arrays of surface monitoring sites and from space.

World Data Center Applications of New Technology

WDCs and their affiliated national data centers are working together to apply new technological tools to solving data and information exchange, archiving and analysis problems. Data sharing networks

such as SPAN and the SELDADS-II provide rapid access to large amounts of digital data. WDCs communicate by telex, facsimile and electronic computer mail links connecting Australia, Canada, Denmark, England, Japan, the USA, and the USSR and exchange data and information daily. Planning this meeting was made possible in the time available only through extensive use of these communications links. Mass storage devices are making major changes in the ability of centers to keep and efficiently retrieve digital data. Personal computers are making possible the efficient maintenance of mailing lists, directories of program participants and scientific instruments, and with word processing and data base management software and laser printers, are giving unprecedented capability to translate information into published formats.

Some centers are testing uses of Compact Disc-Read Only Memory (CD-ROM) and Write Once, Read Many (WORM) optical discs. In mid-1987 the National Geophysical Data Center (NGDC), which operates WDC-A for STP, produced CD-ROM "NGDC-01", Geomagnetic and Other Solar-Terrestrial Data of NOAA and NASA. This CD contains 530 MBytes of digital data including two catalogs and 15 different data bases. It is provided to users with accession software to retrieve, copy and display the contents using relatively inexpensive personal computer equipment. This is the first time these related but usually separate data bases have been collected on a single, random access medium so that it is practical to combine them in a variety of new ways. Before the CD, many magnetic tapes would have had to be mounted and sequentially searched to obtain the data and there was no way to efficiently scan them. Now, with direct random access and a data transfer rate of 3 Mbytes/min, one can quickly review changes in global magnetic activity (since 1868), compare them with sunspot number counts (since 1700) and solar flares (since 1955), and examine conditions in interplanetary space (since 1963). A PC, CD-reader and software are to be available during these meeting for anyone interested in a demonstration or a hands-on trial of this new data compaction and exchange medium.

Several agencies are cooperating to place spatial data onto CD-ROM and display false-color images of the ocean bottom and bathymetric contours for large regions. Accession software for this CD was developed at NGDC and we look forward to soon having the new CDs available for distribution. Seismological data are routinely distributed worldwide on CDs. Another project is in progress to place geological data for a major global region on CDs.

A PC-compatible WORM optical disc unit now is used at WDC-A for STP to store geomagnetic and ionospheric data. We are systematically transferring data from archival tapes onto this more stable optical medium. Files of 1-minute resolution Canadian and US magnetic observatory data are transferred in 15 Mbyte blocks onto 400 Mbyte capacity 2-sided WORM discs. Files of 35 Mbyte digital ionosonde data are also being moved onto WORM discs. These data are copied directly in ASCII coded standard international exchange formats. Much more data can be compressed onto the optical media by use of binary or other compression formats and experiments are in progress to evaluate these options.

A procurement action is under discussion (in spite of diminishing budgets) to obtain one of the new helical scan VHS cassette digital magnetic tape units to explore how inexpensive high-capacity tapes can be combined with optical discs to provide a mixed strategy for servicing digital data. These technological innovations are either available now or soon will be available at WDCs and other centers to meet the need of new data-intensive international scientific programs.

Among the current and planned ICSU programs that are likely to involve WDCs, perhaps the largest and most ambitious is the "International Geosphere-Biosphere Programme" (IGBP), often called "GLOBAL CHANGE". The Panel on World Data Centers has established a sub-committee for liaison with the ICSU Special Committee on IGBP. Panel members cooperated with the IGBP Special Committee's Working Group on Data Management and the Soviet Geophysical Committee to plan two international meetings in August 1988 in Moscow: the Study Conference on IGBP and the Workshop on Geophysical Informatics.

The main goal of the Study Conference is to make significant progress toward identifying data management problems of IGBP disciplines and to begin establishing a plan to meet the different needs. This will be particularly challenging because of the diversity of disciplines involved, many without a history of organized, systematic international data exchange.

The main goal of the Workshop will be to identify how the WDC system can meet specific IGBP needs through application of available technology and where new systems of data and information exchange must be provided. Emphasis will be on the mechanics of handling large data bases to achieve program requirements within resource limitations.

In late 1981 a delegation from WDC-A visited WDC-B in Moscow and other scientific institutions. Prior to this time the most rapid exchange of small amounts of data and of information had been via telex exchanges. During the visit it was possible to establish a computer-to-computer link to an account established in the United States

by NOAA on a leased computer system. Already WDC-A in Boulder had been using this link to quickly exchange data and information with WDCs for Geomagnetism in Kyoto, Copenhagen and Edinburgh, with WDC-C for STP in Chilton (UK), and with sources of data in the US, Canada and Australia. Addition of WDC-B2 to this network greatly improved the speed with which we could exchange information and small amounts of data. Emphasis is deliberately placed on the greater utility of this network for exchanges of information rather than digital data. We find that use of the postal services to exchange data on paper, film, floppy disk, magnetic tape and Compact Disc still is the most efficient and inexpensive method.

In preparation for these meetings we have had regular and effective communications using this network but also using TELEMAIL, OMNET, SPAN, BITNET, NASAMAIL, and GSFCMAIL electronic mail systems as well as telephone, Telex and written communications. The main problem is the number and variety of systems which seem often to require different protocols to enter, create messages, and establish the link to the intended destination. It usually is not possible to enter the Directories of systems other than the base system used locally as a gateway to the others – there should be a common Directory, an electronic telephone book organized in simple alphabetic order, that cuts across all these various networks and that can be quickly searched by target name and/or institution or address. Another problem is that expensive commercial systems seem to operate in parallel with large subsidized systems and access to any one system by a group is more a matter of economics than any other reason.

International Data Exchange Outside the WDC System in 1988

Many organizations that are not part of the WDC system provide international data services for data bases of global or large regional extent. Reference was made above to "analysis centers" among which are included the "permanent services" for astronomy, geodesy, geophysics and related sciences provided by centers of the Federation of Astronomical and Geophysical Services (FAGS). These include:

- Bureau Gravimetrique International
- Centre de Donnees Stellaires
- International Centre for Earth Tides
- International Earth Rotation Service

- International Service of Geomagnetic Indices
- International URSIgram and World Days Service
- Permanent Service for Mean Sea Level
- Quarterly Bulletin on Solar Activity
- World Glacier Monitoring Service.

The World Meteorological Organization (WMO) is an inter-governmental body to facilitate planning between nations for the organization of meteorological programs. WMO does not operate networks or centers but serve a facilitation role to help nations plan and organize their activities and to determine what will be done and by whom in order to assure that the system works. They have guided establishment of a comprehensive data collection, exchange, and archiving system between national weather services worldwide. This includes the Global Observing System (GOS), Global Data Processing System (GDPS) and Global Telecommunications System (GTS). National Meteorological Centers (NMC), Regional Meteorological Centers (RMC), and World Meteorological Centers (WMC) exchange weather data in near real-time through the GTS and use these data in forecasting and tracking weather. They cooperate closely with World Data Centers for Meteorology which serve in a long-term archival, dissemination and analysis role.

The Intergovernmental Oceanographic Commission (IOC) of UNESCO performs a function in respect to marine sciences similar to WMO's in meteorology. It includes the International Oceanographic Data Exchange (IODE) system to enhance marine research, exploration and development by facilitating exchange of oceanographic data and information between participating Member States. An interconnected system of Responsible National Oceanographic Data Centers (RNODC) and National Oceanographic Data Centers (NODC) receive, process, quality check and exchange specified oceanographic data according to the IOC Manual on International Oceanic Data Exchange, currently under revision. The IOC provides guidance to the Panel on WDCs and to the two WDC for Oceanography in matters of oceanographic data management.

Other groups exist to coordinate or centralize data collection for global and regional seismic data, e.g. International Seismological Center and regional seismic centers for the Mediterranean area, for Southeast Asia, and for South America. Seismic sea wave data are collected at the International Tsunami Information Center. There is a World Ozone Data Center in Canada and various other specialized centers in many countries that provide data collection, processing, analysis and archiving services.

In many countries there are national data centers, often part of a government agency or formed as part of a special project, that hold global data or that assemble data collections from sensors in several nations. As mentioned above, some are co-located with and sponsor WDC centers. Others cooperate in international data exchange in order to forward their programs. Access to data in these centers may sometimes be restricted to participants in the program or their use may be subject to other "rules of the road" developed by participants in forming the data collection. An example in this category in the US is the incoherent scatter radar data collected at the National Center for Atmospheric Research (NCAR). This effort is funded by the National Science Foundation because they sponsor several incoherent radar facilities as research programs to study the upper and middle atmosphere. Similar data from other groups, e.g. the European Incoherent Scatter (EISCAT) Radar, are joined with those from the NSF sites to create a multi-national data collection. Catalogs and information about these data and how they may be obtained are provided by NCAR.

Future Prospects for the WDC and Alternative Systems

The description of the WDC system given above is based on experiences gained through years of working within WDC-A for STP, in close association with other WDC-A discipline centers, with staff of WDC-B2 and WDC-C centers, and with members of the Panel who were active in creating the WDC system for the IGY and with shaping it to continue serving other programs since the IGY. Much of the structure and content reflects the current GUIDE: Part I. I hope it is reasonably objective although the process of selection of a few topics from among many possible lends opportunity for imparting a personal viewpoint which might change after review by others. These concluding observations are more highly personal and speculative than other material covered here.

One possible future scenario for the WDC system is that it could remain relatively fixed in scope, trying only to continue serving those disciplines for which it now has centers and providing collection, processing, archiving and dissemination services only for those types of data specified in the latest GUIDE. This would be a future in which the WDC system continues to be of great utility to those disciplines served but it would mean neglecting a present opportunity to grow by entering new fields and providing new services. Also, it would present increasing problems of interaction with competing data and information handling system that would arise to meet new emerging needs. It would also be to ignore some consequences of economic, technological and institutional dynamics now acting on the WDC system.

Since the IGY there has been a persistent trend of changing scope and size of centers in the WDC system; a trend that probably will affect future alignments as fewer but larger centers cover more disciplines. This has resulted in the present distribution of disciplines served by WDC centers which no longer replicate exactly the disciplines of the IGY era. In 1957 there were more sub-centers with many located in academic institutions that were specially funded for the IGY period. After the active data gathering and research phase was completed, funding for these more restricted centers became increasingly difficult to obtain and competing interests in new programs saw other tasks take precedence over continued data center activity. As a result responsibility for data archival and dissemination (and sometimes analysis) in these disciplines began to collect at a few larger centers, usually in government agencies having national responsibility for related data operations. This trend was assisted by the growth of digital data collection, the expense of processing large amounts of "raw" digital data, and the cost of large main-frame computers used to manipulate the data. Government centers with their requirements for major investment in data management support facilities and a relatively stable staff environment providing continuity of experience in working with specialized types of data became attractive centers toward which the data collections and responsibilities of smaller IGY centers gravitated. Also such centers had the "inertia" (pejorative) or "stability" (compliment) attributed to government operations and so were capable of taking a longer view about the continuing importance of data management, archiving and dissemination than were institutions forced to achieve broader or more immediate goals such as student education or commercial profitability.

Within a given discipline, even narrowly defined, it is evident that the capability is growing for modern international programs to generate larger amounts of data than can be dealt with at one or a few locations if these are already tasked with a full time job to provide data services for continuing, older programs. The need to perform specialized data processing and analysis at facilities of the institutions that collect the original data probably will lead to formation of an increasing number of specialized data centers outside the WDC system. The value of collected and processed data obtained from an expensive array of surface instruments or from a satellite program is in part the potential offered to purchase entry for the

responsible project scientists (or their sponsoring agencies) into active participant roles in international scientific programs or into the organized scientific life of their community. These combined forces make it inevitable that specialized centers will arise.

For some disciplines there are such immediate and important needs for access to global data that governments are compelled to be directly involved in the process of data collection, processing and exchange. Clear examples are weather and oceans. Operational needs and other needs do not preclude these data from entering the WDC system but it implies the existence of a more formal array of governmental centers and exchange networks before transfer, usually, of processed and summary data to WDCs.

It is important to recognize these several reasons why WDCs cannot hold "all the data" and why there must be other centers. But, they should know of the existence of these other centers, have access to on-line inventories and published catalogs of their data holdings and services and be able to provide a "referral service" to WDC users. Even to the extent of sometimes taking an active part in obtaining and transmitting the needed data.

Solar-Terrestrial Physics in the United States is a loosely coupled, somewhat vaguely defined multi-disciplinary scientific area. In the report on Geospace Environment Modeling (GEM, May 1988) just prepared for the National Science Foundation, a case is made for the creation of "Discipline Data Centers" (DDCs) to be established as part of active research efforts within universities and independent research institutes. Such centers would focus on providing community-wide access to processed data from particular instruments or arrays for which they are responsible. Emphasis would be mainly on digital data although efforts would be made to provide display capabilities to users through common graphical software. However, it is worth nothing that the GEM report calls for the continued provision of basic monitoring data, indices and other summary data products, and information from World Data Centers. The proposed growth of DDCs is to complement the existing system, not replace it.

One hazard associated with a rise in number of specialized data centers is the increasing likelihood of different data formats specialized to match the local computing facilities so as to achieve the most economical access or efficient use of the data. Often such a center will have special equipment either not available elsewhere or not within the budgets of other institutions worldwide. Even among the present WDCs there are some problems with differing formats and facilities but overall they have worked to focus the international scientific community's attention on the need for accepted standards and for providing data and information in a variety of formats suitable to many classes of users. Where WDCs have special equipment for data analysis, their relative ease of accessibility to scientists from all countries makes these resources generally available.

Another problem of smaller centers, particularly those located in universities or small research institutes, is that such centers frequently form around the interests and programs of a particular scientists, often a world leader in that discipline. However, careers change and research interests shift for many reasons, sometimes with the result that the work of maintaining such a data collection is effectively abandoned. Even if these are offered to a WDC, unless care has been taken to plan for eventual transfer of the responsibility it is likely that no useful transfer can occur. The long-term WDC commitment to data management, collection and dissemination makes them logical final repositories for even specialized data collections but there must be sufficient cooperation and advance planning between WDCs and specialized centers.

Finally, specialized centers tend to develop around the most advanced technological research groups. They necessarily employ the latest computers, display devices, electronic networks and other tools to the performance of their work and to the provision of their services to other users. Typically, they cooperate most closely with other groups worldwide that have the same or similar levels of ability to access and manipulate data. There are outstanding examples of leading research scientists who have shared their results widely and taken care to make them available in a variety of formats suitable to a range of user skills but these are not commonly encountered. Since their inception, WDCs have been noted for trying to match data and information products and services to the needs of all levels of users.

Conclusions

There exists a system (network) of World Data Centers performing a vital task of collecting, archiving and disseminating global data and information for the international scientific community within specific geophysical and solar disciplines.

Their data and services are accessible to scientists of all nations and efforts are made to make them available without charge through exchanges or for minimal cost.

WDCs operate in close cooperation with national centers for the same disciplines and cooperate with varying success with inter-governmental and specialized data systems.

WDCs have provided essential general support and innovative new types of services for international scientific programs.

WDCs have unique qualifications as facilities for special analysis programs by resident scientific staff and visiting scientists and are well suited for training students in data management and applications.

WDCs are nodes in local, national, regional and global electronic mail and data transfer networks.

WDCs are active in the adaptation of new technology to meeting data management needs for providing prompt and comprehensive information.

The **WDC** system provides at least a role model for emulation in creating a new organized data management mechanism for GLOBAL CHANGE and, by expansion to cover new disciplines, are the preferred basis for efficiently implementing data and information management plans.

WDCs will be central elements in a future mix of large centralized data and information management facilities and specialized data/analysis centers.

THE NSSDC INFORMATION SYSTEMS, STATUS, AND PLANS

James L. Green

National Space Science Data Center
Greenbelt, MD 20771, USA

Abstract

The National Space Science Data Center (NSSDC) is NASA's largest archive for processed data from spaceflight missions and was established in 1966. Over the years, the NSSDC has built an extensive system to manage its archive. This system is called the SIRS, or System for Information Retrieval and Storage. By today's standards, SIRS is an older technology and is used exclusively by the NSSDC operations personnel to satisfy letter or phone call (offline) requests for data.

Within the last several years, the NSSDC has begun to develop and place into operation remotely accessible online information systems. The new technology online information systems provide a variety of services, depending on the desires of the community of scientists it serves. One of the major characteristics of these systems is that they are all accessed by remote science users (24 hours per day) and the NSSDC operations personnel (satisfying letter or phone call requests). The new online information systems have been a tremendous success, logging over 2500 accesses by remote users per year and growing rapidly. The common theme of the new online systems is to first provide information about data holdings with several levels of complexity. In addition to the information about data, several of the NSSDC systems, such as National Climate Data System (NCDS) and Coordinated Data Analysis Workshops (CDAWs), support manipulation, browsing, and display of selected data sets that each of these systems manage.

This paper will provide an overview of all the NSSDC information systems and will discuss NCDS in more detail as a typical example of one of the NSSDC's more advanced systems. Future plans at the NSSDC, call for a full migration of information from the old technology system into the advanced online database management systems. Since the NSSDC utilizes several database management systems, it will be essential that these systems interoperate in the future.

Introduction

The National Space Science Data Center is NASA's largest archive and had been accumulating information and data since it was established in 1966. The following list shows the type of information that the NSSDC continually compiles to accompany the data archive.

- Individual spacecraft and their instruments
- Information on data sets from individual spacecraft instruments
- Data sets built with data from many instruments and/or many flights
- Software packages held at NSSDC and specific to individual data sets
- Ground-based, rocket, and balloon data sets of potential interest to NASA researchers
- Geophysical computer software models
- Software packages independent of individual data sets (e.g., AIPS, IRAF, LAS, NGS)
- Rocket launches (Spacewarn requirement)
- Scientific papers and other documents relevant to specific data sets, instruments, facilities, etc.
- Addresses of requesters of NSSDC services and data
- Requests being processed at NSSDC
- Data set granules (often called inventory)

This paper will provide a brief overview of the major NSSDC information systems. A distinction will be made from the older NSSDC information system, which has been in use for over 15 years and is only used by the NSSDC staff, and the new technology systems, which accommodate the operational and remote user queries for information. The NCDS will be discussed in detail as a typical example of how the NSSDC is responding to ever-challenging user demand for rapid access to NASA-acquired data. The reader is referred to Green, 1988c, for more details of all the NSSDC systems and capabilities.

"Old Technology" Information Systems

Over the years, the NSSDC has built an extensive system to manage its archive. This system is called the SIRS, or System for Information Retrieval and Storage. By today's standards, SIRS is an older technology and is used exclusively by the trained NSSDC operations personnel to satisfy letter or phone call (offline) requests for data. SIRS uses a database management system that was developed at the NSSDC, since commercial systems were not available at the time it was needed.

SIRS runs on a MODCOMP classic computer (see Figure 1) and manages over 850 fields of information. The total size of the SIRS is approximately 70 megabytes. SIRS is a hierarchial system that has 12 files describing:

- Spacecraft, instruments, data sets (over 4000)
- Associated bibliographies (29,000 entries)
- Inventories (tapes, microfilm reels, etc.)
- International science community personnel (mostly NSSDC users)
- Data requests (e.g., 2,300 in 1985)

The type of output obtained from SIRS consists of hard copy data catalogs, documentation to mail with data, project bibliographies, and management statistics.

"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, on line 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 per day. This allows the NSSDC to "remain open" past the normal working hours, providing scientists and students the ability to "browse" through the online information to look for an important data set.

Currently, not all the information about the NSSDC archive is remotely accessible to users, and less than 2% of the NSSDC's total digital data archive is on line (see Green, 1988a), but these systems are already a major achievement in providing rapid access to NASA-ac-

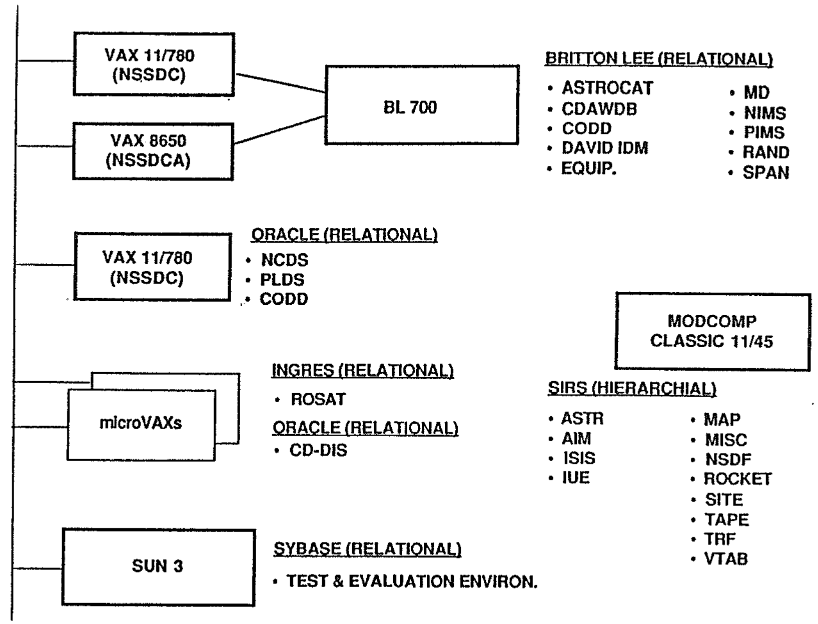


Figure 1 - The present hardware and software database management environment of the NSSDC is illustrated in this figure. The software systems ORACLE, INGRES, and SYBASE execute on the interactive systems that are remotely accessed. The Britton Lee system is a hardware database management system that is directly connected to the interactive VAX computer front ends and off loads the query processing from the front ends. The Hierarchial SIRS system executes on the MODCOMP Classic and is used exclusively by the NSSDC staff in satisfying offline letter or phone requests.

quired science data that is unprecedented in archive data management. The new online data and information systems currently operational at the NSSDC are shown in Table 1. These systems are accessible by a wide variety of computer networks as described in detail by Green, 1988b. The new online information systems have been a tremendous success, handling over 2500 accesses by remote users per year and growing rapidly.

The systems shown in Table 1 provide a variety of services, depending on the desires of the community of scientists it serves. The common theme of the online systems is to first 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 NOAA and USGS. The MD, therefore, is the first reference system to point to where the data are being held. 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 NCDS) to which the MD will refer a user. The reader is referred to King, 1988b, for additional details about the Master Directory.

Figure 1 shows the total software and hardware information systems environment of the NSSDC. The new information systems are using a variety of commercially available software database management systems such as ORACLE, INGRES, and SYBASE. The Britton Lee system is a commercially available database management machine. Currently, the ORACLE, INGRES, and Britton Lee systems are operational. The SYBASE system is a recent acquisition and is under test and evaluation.

In addition to the information about data, several of the NSSDC systems such as NCDS support manipulation, browsing, and display of selected data sets that each of these systems manage. In the next section, the NCDS system will be discussed to illustrate the full range in capabilities of these new online information systems.

NASA Climate Data System

The NASA Climate Data System is an interactive operational software system that manages, manipulates, and displays key climate data in the NSSDC archive. NCDS provides these capabilities by linking together software subsystems that perform specific functions. NCDS is one of the most technically advanced systems at the NSSDC

TABLE 1: NSSDC NEW TECHNOLOGY ONLINE SYSTEMS

SCIENCE DISCIPLINE	SERVICE	INFORMATION	DATA*
All			
	Master Directory	X	
Astrophysics			
	IUE Request System	X	X†
	ROSAT Info. Manage. Sys.	X	
	Astronomy Catalog Sys.	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 Dir.	X	
	Omni Solar Wind Data Sys.	X	X
	Plasma and Field Models	X	X††
	Coordinated Data Anal. Wkshp.	X	X
General			
	SPAN-Network Info. Center	X	
	Personnel database	X	

NOTES:

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

and is electronically reachable by users from any of the major computer networks (see Green, 1988b). The initial concept behind the development of NCDS, which started in 1980, was to test proposed solutions to some the overwhelming problems of managing large amounts of useful earth, oceanographic, and atmospheric data coming from NASA spaceborne and (non-NASA) surface-borne measurements. Not only has NCDS been used by scientific researchers, but over the last couple of years it has been used in the classroom at two universities, as a tool in the education of climatologists.

The structure of NCDS, shown in Figure 2, consists of a user interface and catalog, inventory, data access, data manipulation, and graphics subsystems. The user interface is via a software package called the Transportable Applications Executive (TAE), which is tailored to transparently integrate and accommodate the NCDS subsystems. The NCDS/TAE interface is a menu driven system that hides all the details of the operating system from the user.

The catalog subsystem contains descriptive information from over 150 climate-related data sets. Within the inventory subsystem, NCDS directly manages over 20 key climate data sets. These data sets can be accessed through the data access subsystem, with the ability to create data files or data tapes. In addition, the capability exists for the reformatting of requested data into the Common Data Format (Treinish and Gough, 1987), which can then be easily displayed and manipulated. The graphics and data manipulation subsystems are used for browsing and, in some cases, analyzing the data.

Future Information System Activities

The NSSDC will continue to aggressively pursue the "electronification" of its information about data and, to the extent reasonable, its archived data (see Green, 1988b, and King, 1988a). Much more needs to be done to bring the offline directories, catalogs, and inventory archive information, currently managed by SIRS, into the new information system environment. Through the new information systems, the NSSDC is striving to support active archive research on the individual scientist level during any time of day or night convenient to the researcher. Although not extensively discussed in this paper (see King, 1988b), the Master Directory is envisioned to be a major resource for the international community, which will greatly speed the identification and location of desired data. The continued building of the MD capability will be a major effort at the NSSDC.

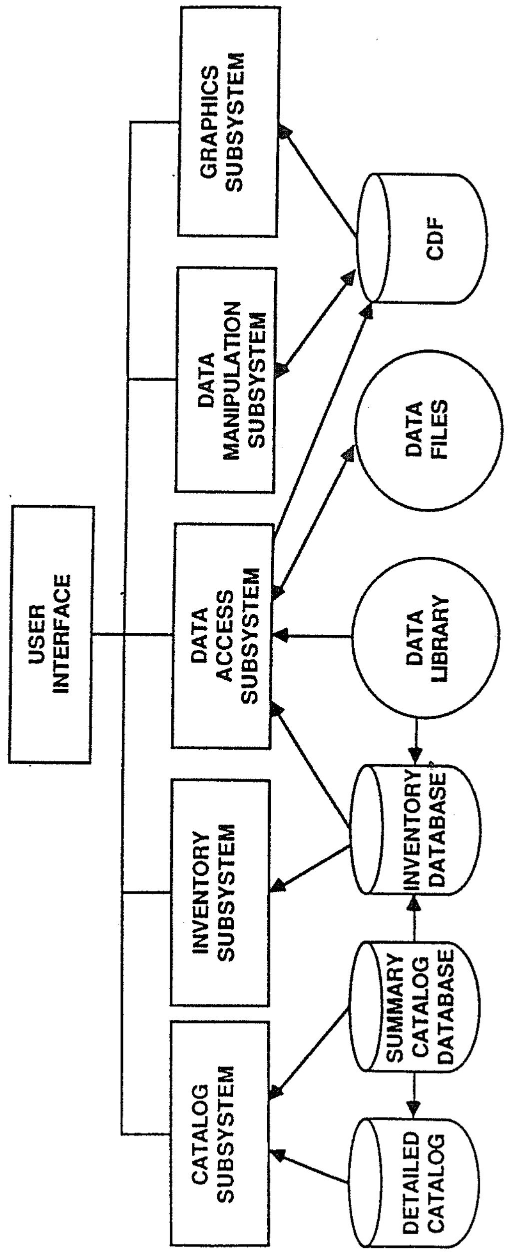


Figure 2 - An overview of the structure of the NCDS is provided that illustrates the subsystems of this capability. NCDS is a completely automated data and information archive, distribution, and display system for over 20 major climate and ocean data sets. This system provides several levels of descriptive information about the data that it manages and provides for data manipulation and graphics tools for browsing and elementary data analysis.

Within the next few years, the NSSDC will be establishing an information management infrastructure to facilitate the rapid ingest of newly acquired data. This infrastructure will involve standardizing many of the interfaces that the NSSDC has with the NASA missions supplying data and information for archiving.

Acknowledgement

I would like to gratefully acknowledge Karen Satin, who has done an expert job in technically editing this article, and my secretary, Marjorie Layfield, for her accurate and rapid typing of the manuscript.

References

Green, J.L., NSSDC data holdings: scope, complexity, and future requirements. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988a pp. 205–215.

Green, J.L., NSSDC applications for computer networking. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b pp. 168–175.

Green, J.L. (editor), The National Space Science Data Center, NSSDC Technical Publication, August 1988c.

King, J., Mass storage technologies for NASA's future data environment. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988a pp.

King, J., Exchange of data and information between WDC-A-R&S and WDC-B2. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b pp. 281–284.

Treinish, L. A., and M.L. Gough, A software package for the data-independent management of multi-dimensional data, EOS, Vol. 68, 28, pp.633-635, July 14, 1987.

INFORMATION SYSTEMS AT WDC-C1 FOR GEOMAGNETISM, COPENHAGEN

Eigil Friis-Christensen

WDC-C1 for Geomagnetism, Danish Meteorological Institute Copenhagen, Denmark

Introduction

The WDC-C1 for Geomagnetism in Copenhagen, Denmark was, as many of the existing World Data Centres, established as part of the International Geophysical Year (IGY: 1957/58). The Danish Meteorological Institute (DMI) has a long term tradition in serving international scientific programs involving collection of data from a number of places around the world, as was the case during one of the first major international projects, namely the International Polar Year, 1932/33. The director of DMI at that time, D. la Cour, had been the key figure in the establishment and operation of the magnetic observatory in Godhavn and had gained extensive experience regarding the design and operation of magnetic instruments used in arctic regions. He became President of the Commission for the International Polar Year, and instruments designed by him and built in Denmark were used to a great extent and DMI was selected as the central archiving institution regarding data from this international program. But even before that, DMI had been involved in experiments on expeditions to the arctic regions already during the 1st Polar Year 1882-83, and additional expeditions to the arctic during the following years. Facilities to calibrate the large number of instruments put up during the International Polar Year 1932/33 and during the following years were established in Denmark and DMI is still a host of IAGA's service on comparisons of magnetic standards.

The WDC-C1 for Geomagnetism is hosted by the Division of Geophysics which in its scientific and monitoring programme has focused on Solar Terrestrial Physics (STP). In particular the traditional mutual connection between Greenland and Denmark has influenced the selection of research projects and of the observational programme. In 1972-73 a dense chain of temporary geomagnetic stations was operated in Greenland, and this chain was revived during the next major international effort, the International Magnetospheric Study (IMS: 1977-79). Since the dedicated staff of the WDC-C1 was rather limited, the close contact to a research group within the same field was essential for its continued operation and for its gradual renewal. The close connection between experimental research and the service of a broader scientific community also provided a basis for being able to look at the more principal problems of international data exchange from the data centre's point of view as well as from the data supplier's and the data user's.

Information Retrieval System

Today the computer system used at the WDC-C1 for Geomagnetism is based primarily on a Novell local area network (LAN) consisting of about a dozen IBM-PC or -AT compatible personal computers within the Division of Geophysics, connected to the Danish Meteorological Institute central mainframe computer, a UNISYS 1180/82 APS. On the LAN the file server is equipped with a hard disk with a capacity of 2*160 Mbytes. One of the AT's on the LAN is connected to a 1600 bpi tape drive for transfer of data to and from the LAN. Recently a CD-ROM reader was included on the system. A HP compatible laser printer/plotter is connected to the file server and is available for all LAN users. The connection to the central computer is presently achieved through asynchronous serial lines through a 9600 baud multiplexed line, but is intended to be replaced by a TCP/IP connection. The central computer is connected through a network to a UNISYS 1190/92 computer at the Danish University Computer facility, which is further connected to the European Academic Research Network (EARN) and BITNET. Electronic mail to institutions on most international networks, including SPAN is available from EARN. In addition to this, the commercial networks, Telenet and Uninet, may be used via DATAPAK.

The information about the holdings of the data has since 1982 been updated and accessed using an computerized data base management system. First on a WANG 2200 minicomputer, but after the introduction of the PC's the databases were transferred to the network using a 4th generation database programming tool DATAFLEX. This system was selected because unlike most commercial database management

systems, it provided good possibilities to tailor a system to the more specific scientific needs without becoming too complex to use. Another advantage of the system is that it generates true multi-user facilities on a network allowing a number of operators to access and even update in the same record simultaneously. Finally the system is available for a number of different computers, not only PC's, which means that the programs and databases are nearly unlimitedly transportable.

Although the system does not in principle differ from other database information retrieval systems, it may be of interest to give a brief review of its components.

DATAFLEX is a commercial application development system. Its structure is described as an extended relational DBMS with data independent utilities and command language. Some main figures: Maximum number of DBMS files is 250, maximum indexes per file is 9, maximum file size is 2 gigabytes, maximum records per file is 16.7 million, maximum record size is 16k bytes. The command files have a source version, which can be edited with a normal text editor. A source file is compiled separately and the compiled file is run within the dataflex operating environment which supplies all the run-time utilities.

One of the major objectives when designing the system was that it should be easy and safe to use, even for an operator with limited experience. Of course, and this may seem trivial when looking at today's abundant supply of high quality programs for the personal computer, it was a necessity that the user interface should be screen oriented in contrast to most main frame computer programs at that time which were line oriented, and demanded rather experienced, or at least dedicated, operators. Another main objective was that the system should be integrated in the data information systems used in the Division of Geophysics. For example the necessary information about a magnetic observatory should be stored and updated at only one place regardless if data from this observatory belonged to the WDC-C1 holdings or not. This meant that the observatory database should be independent of the data catalogs, which, on the other hand, should be able to incorporate information from the observatory database. This is a classical feature of a relational data base system which has been extensively used outside the scientific world, but nevertheless seemed not really a common element in scientific databases in use at that time.

With the observatory database as a primary database, which contains all the necessary information about the observatories used in the World Data Centre C1, as well as in research projects carried on independently of the WDC-C1, all the other databases of the WDC-C1 system relates to this. The other existing databases are: Geomagnetic recordings and hourly values, geomagnetic indices, earth currents, and data from different temporary magnetic chains which have been operated during IMS and at other times, and from which data have been transferred to the WDC-System.

A typical database like for example the geomagnetic records and hourly values database contains information on a monthly basis of the WDC-C1 holdings. It is further specified in the database on which media the data are available, microfilm or -fiche, reports, magnetic tapes, or the latest introduced medium, the CD-ROM. The databases regarding temporary magnetometer chains contain information about the holdings on a daily basis, but without detailed information about which stations are actually contributing data for a specific day.

For all these databases standard report programs are available for output either in printed form to floppy disks or magnetic tapes for exchange of information with the other World Data Centers. This has been used for instance in the preparation of the joint catalog for geomagnetic data prepared by WDC-A in 1985. In addition it is fairly easy to design a report program to provide a specific output.

Every user connected to the LAN has access to the database system. The system is menu driven, and each menu may be assigned a different password so that selected databases may not be updated or accessed without authority. This design also permits a relationship between WDC and non-WDC databases, which means that the marginal development costs can be kept relatively low compared to a situation, where the data centre was a completely independent function.

Data Bases

The main part of the holdings of WDC-C1 consists of geomagnetic records in form of microfilm or microfiche. The total number of reels of microfilms is about 16000. These microfilms contain data from 392 geomagnetic observatories mostly from IGY (1957) onwards. The geomagnetic archive from the 2nd Polar Year 1932–33 has recently been included in the catalog. Besides this the data centre contains magnetic records from single stations for an extended interval. Godhavn in Greenland, for example, contributes with data from 1926 till now. The records are either copies of the original records or computer generated plots from digital data, supplied from the organizations which operate the magnetic observatories.

Hourly values and a number of other geomagnetic and solar data which previously were stored on magnetic tapes are today available on

the CD-ROM "NGDC-01", Geomagnetic and other Solar-Terrestrial Data of NOAA and NASA, produced by the National Geophysical Data Centre (NGDC) in USA.

In addition to the main holdings the centre has a collection of rapid-run magnetograms, tellurigrams and earth current data. Digital one-minute data for selected observatories have also recently been added to the data base.

Apart from its own databases, the WDC_C1 acts as a referral centre regarding specific project data as for example the Greenland Magnetometer Chain. Limited data in form of single events are normally available from these projects through the principal investigator.

Near Term Future Directions

The WDC-C1 for Geomagnetism in Copenhagen is a fairly small centre with only one full-time employed staff member dedicated solely to the daily WDC-C1 activities. For more specific tasks like development of new systems and the handling of specific requests, the WDC-C1 relies on its integration in the geomagnetic research group at DMI.

During the recent years, when major emphasis was placed on large central computer systems, a small centre had difficulties in being able to keep the same development speed as the larger data centers. This trend, however, seems to have changed with the appearance of small, low-cost personal computers with a throughput which for a number of tasks is superior to many main frame computers. In particular connection to the international electronic networks will make it possible for the small centers to be an integral and vital part of the global WDC-System. The individual data centers may rely on joint efforts where this is needed, and may develop specialized services where capacity and know-how is available locally. A distributed WDC-System seems to be a efficient way for smaller centers to continue to serve the scientific community with their specific possibilities.

So it is certainly one of the major goals for the WDC-C1 in Copenhagen to be connected as directly as possible to international networks on which the other data centers are connected, and to networks, to which the users of the data are connected. This means that small centers should not necessarily duplicated all the facilities at the large centers. On the other hand it also means that the larger centers with advantage could leave some specific tasks to the smaller centers. If this direction is shared by the WDC community it implies that the smaller centers could take up tasks where the interest of the host organization makes it natural to concentrate on a specific aspect of the data centre activities.

WDC-C1 for Geomagnetism in Copenhagen is integrated in a research group specialized in polar ionospheric and magnetospheric research related to the coupling of the earth's environment to the Sun and the interplanetary medium. In consequence of this the following near-term future plans have been discussed:

1) 'True' networking, i.e. directly from the LAN to the international networks including SPAN.

2) In addition to the present activities to specialize as a centre of polar geomagnetic data and other related data as for instance all-sky camera data, ionograms, and riometer data from Greenland.

3) Calculate and distribute a Polar Cap (PC) index of geomagnetic activity based upon near polar stations, preferably one from each hemisphere. This project is planned in cooperation with the Arctic and Antarctic Research Institute (AARI), Leningrad and WDC-B2, using data from polar cap stations Thule in the northern and Vostok in the southern hemisphere. The index could be derived with less delay than the AE-index and would complement this index.

4) Increase its role as a "referral" centre regarding polar data obtained during special campaigns.

DATABASE SERVICES AT WORLD DATA CENTRE C1 FOR SOLAR-TERRESTRIAL PHYSICS

M.A. Hapgood

WDC C1 for STP, Rutherford Appleton Laboratory,
Chilton DIDCOT, Oxfordshire OX11 OQX
United Kingdom

Introduction

The World Data Centre C1 for Solar-Terrestrial Physics was originally established in 1957 as the World Data Centre C1 for the Ionosphere. It was located at the Radio Research Station, Slough, and administered by the Department of Scientific and Industrial Research (DSIR). In the intervening thirty-one years there have been numerous organisational changes. However, the only change to have a major effect on the operation of the Centre was its move in 1982 from Slough to Chilton, where it now forms part of the Rutherford Appleton Laboratory (RAL). This laboratory is operated by the Science and Engineering Research Council – the successor of DSIR.

The move to Chilton stimulated a review of the Centre's operation, following which a "Modernization Programme" was initiated. This had the objective of increasing the range of modern datasets available from the Centre: (a) by providing more data in digital form; and (b) by providing new types of data that were needed by the scientific community (for example, solar wind data). The increased remit of the Centre, implied by objective b, lead to the change in the title of the Centre.

In 1984, RAL also established the "Geophysical Data Facility" – as a national facility for the UK scientific community. This facility has the task of handling the large geophysical datasets from modern instruments, especially from spacecraft experiments. Since the WDC and the GDF have many objectives in common the two services have developed in a co-ordinated fashion and now make use of much common software for user interfaces, database management, data manipulation, graphics and networking.

Present status

Databases

The WDC operates a number of database services as listed in Table 1 below. Most services make use of modern database management systems: either (a) a commercial system called INFO; or (b) a locally developed system called R-EXEC (Read, 1986a and 1986b). A few services make use of customised Fortran programs. Both database systems are relational and thus provide the flexibility that is needed for scientific applications. Moreover, they permit users to choose a subset of data through a sequence of queries, rather than by a single complex query (as in SQL-based systems). The former approach is strongly preferred in scientific applications, since it permits verification at each step in the sequence. The use of INFO is now being phased out in favour of R-EXEC, which provides a greater range of facilities needed for scientific data management, e.g. trigonometric and other mathematical functions, array handling.

Table 1. Database services at WDC-C1 for Solar-Terrestrial Physics

1. Solar-geomagnetic indices:

a. daily values – of Sunspot Number, Solar Radio Flux, Mean Ap and Sum Kp; three-hourly values of Kp and Ap. Uses the R-EXEC database system.

b. monthly values – of Sunspot Number, Solar Radio Flux and IG and IF2, the ionospherically derived indices of solar activity. Uses the INFO database system; migration to R-EXEC is planned.

c. provisional daily values – of Sunspot Number, Solar Radio Flux and mean Ap. Uses the INFO database system.

2. Provisional hourly fo F2 values from ionosondes at Slough & South Uist. Uses the INFO database system.

3. Full set of scaled characteristics from six UK-operated ionosondes. Uses a customised Fortran program.

4. MSIS86 model of upper atmosphere temperature and composition. Extracts of indices from daily values database (service 1.a); uses customised Fortran program.

5. Catalogue of all vertical incidence ionosonde data held in the WDC at RAL (in analogue and digital form). Uses INFO.

6. Catalogue of incoherent scatter radar data held by the EISCAT group at RAL. Uses R-EXEC.

7. Solar wind plasma and interplanetary magnetic field data. Planned database, will use R-EXEC.

In addition, the WDC provides two on-line information services:

1. Solar Daily Forecast. Report and forecast of solar-geophysical activity received by daily telex from the Space Environment Services Center, Boulder.

2. Meeting programme. Information about meetings organised by the MIST (Magnetosphere, Ionosphere and Solar-Terrestrial) section of the Royal Astronomical Society.

The WDC service also includes an option to send electronic mail messages to members of WDC staff. This may be used to make comments about the existing services or to order data not available online.

Database access

There are three ways in which data may be retrieved from the WDC Databases:

a. **Via a menu system.**

This provides a means by which unskilled users can retrieve data online. The menu system first allows users to define their selection criteria, and then to retrieve the data and put them into a computer file. This file may then be viewed on the screen, printed at the data centre (and forwarded by post), or transferred over the computer networks to the user's own computer (e.g. for analysis by local software). In the future, we plan to provide graphical display of WDC data. In these cases we will provide a similar set of options, i.e. on-screen plotting, printing at the data centre, and file transfer (as GKS metafiles).

b. **via regular data products.**

The WDC has a number of applications in which the databases are used to generate standard products. For example, the database of provisional foF2 values from Slough and South Uist (service 2) is used to produce a monthly bulletiin, which is disseminated world-wide.

c. **ad-hoc queries.**

The database management systems (INFO and R-EXEC) used by the WDC both have extensive query languages. Thus there is considerable opportunity to process specialised queries in an ad-hoc manner. At present, such queries are handled by WDC staff who have the necessary expertise so users must contact the WDC when this facility is required. In the future, it is planned to give users access to the R-EXEC query language and to provide education in its use.

Network Access

The WDC services are operated on the IBM 3090 mainframe at RAL, which is connected to the UK scientific computer network (JANET – the Joint Academic Network). Thus JANET users can gain full access to WDC services (interactive terminal and file transfer). Users outside JANET (and, in particular, users outside the UK) can also gain access but only by a more complex route. Interactive terminal access is available via the public X.25 networks (e.g. DATEX-P in West Germany, TRANSPAC in France, TELENET in the USA), from which JANET can be reached via gateways at RAL and London. File transfer to other countries is available via EARN (the European Academic Research Network) and its US counterpart, BITNET.

Usage

Use of the databases has risen steadily since the service started in February 1984 and now amounts to about 2000 queries per year. At present, the great majority of this use is from the UK – from 17 sites on JANET and two on the UK public X.25 network (PSS). International use amounts to only 2,5% – mainly from Norway and West Germany – and so the WDC would like to encourage greater use from outside the UK.

Future plans

Data

Over the next few years the WDC plans to introduce several new services, for example:

a. Solar wind plasma and interplanetary magnetic field. This database is designed and is awaiting implementation.

b. Interplanetary scintillation (IPS) data. Data from the IPS array at Cambridge, with which it is hoped to detect shocks and high speed streams in the solar wind, thereby allowing the better prediction of geomagnetic activity.

c. Auroral electrojet indices (AE, AU and AL).

Technology – hardware and software.

The WDC will co-ordinate its work with other RAL space science data services, e.g. the Geophysical Data Facility, to make good use of

available resources, especially by sharing software for common applications. Future developments will include greater use of databases and networks, and the exploitation of new media such as optical discs (e.g. CD-ROM systems). An important objective will be to support the PC workstation concept by ensuring that users of the machines can easily obtain WDC data – on physical media and by network file transfer.

References

Read, B.J., 1986a, "A Relational Data Handling System for Scientists", RAL Report 86-053.

Read, B.J., 1986b, "Scientific Data Manipulation in a Relational Database System", RAL Report 86-084.

SOME USER EXPERIENCES AND PROPOSALS CONCERNING WDC INFORMATION SYSTEMS FOR SOLAR–TERRESTRIAL DATA

C.-U. Wagner

Central Institute for Astrophysics
DDR–1590 Potsdam, GDR

Abstract

Experiences will be decribed gained during the building up of data bases for studies of magnetospheric storms especially in regionally limited areas (e.g. European sector). Some proposals for providing additional informations (hopefully in an interactive way) concerning satellite-born measurements and digitized data sets from ground-based measurements are deduced. Another note is made concerning the necessity of a hierarchical information systems about distributed data bases and data banks.

Inside the magnetosphere and the upper ionosphere most of the processes are steered by the magnetic field. In order to investigate e.g. the coupled system magnetosphere–ionosphere it is necessary to use simultaneous measurements in different altitudes from the same magnetic flux tube, i.e. measurements done simultaneously by different satellites crossing the same flux tube in different altitudes and by ground-based stations. For a given time interval the two conditions–high-density ground-based networks and simultaneousness of measurements from different satellites–can be fulfilled only in a limited region. Therefore for investigations of this kind it is necessary to build up complex data bases for regionally limited areas. Experiences gained during this work will be described, and some proposals concerning improvements of the information systems of WDC's will be made.

Some magnetospheric disturbances have been selected for a common cooperative study in the framework of the cooperation of socialist countries. Fig. 1 shows one of these intervals, and Table 1 shows the

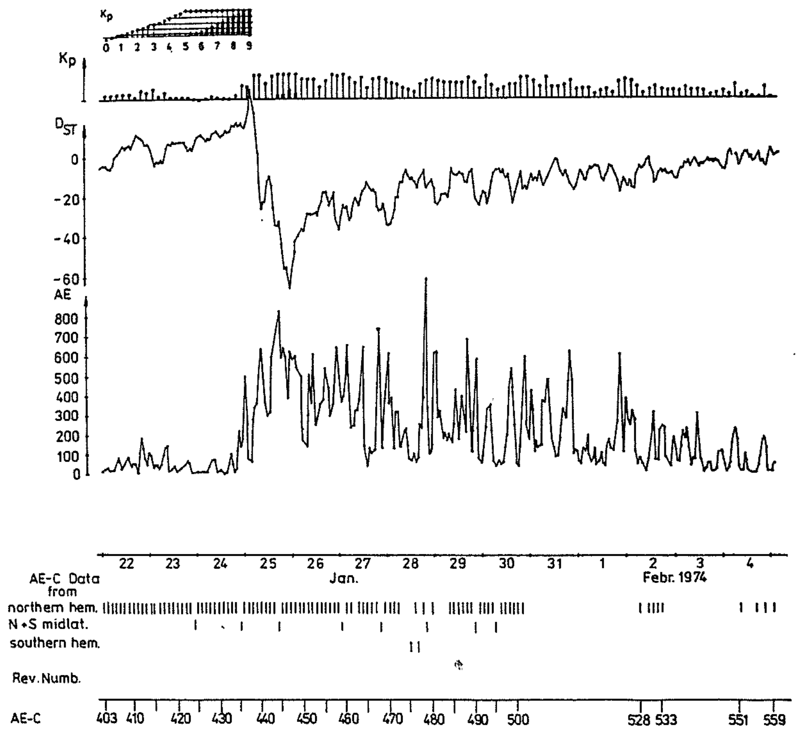


Figure 1 - Sample of magnetospheric disturbance selected for a common cooperative study. January - February, 1974.

A. Satellite-born Measurements

Satellite / height range	parameter	Instrument / principal investigator
IK-10 750-1450 km	electron density	CLP C.-U. Wagner
	electron temperature	CLP C.-U. Wagner
AE-C 155-700 km	neutral composition	NATE N.W. Spencer; NACE D.T. Pelz; OSS A.O. Nier; UVNO C.A. Barth
	neutral temperature	NATE N.W. Spencer
	neutral density	MESA K.S.W. Champion
	ion composition	BIMS H.C. Brinton; MIMS J.H. Hoffman
	ion temperature	RPA W.B. Hanson
	total ion density	RPA W.B. Hanson; CEP L.H. Brace; BIMS H.C. Brinton
	ion drift	RPA W.B. Hanson
	electron density	CEP L.H. Brace; RPA W.B. Hanson; BIMS H.C. Brinton
	electron temperature	CEP L.H. Brace
	photoelectron spectra	PES J.P. Doering
NNSS ca. 1000 km	total electron content vertical ionosph. content	Beacon R. Leitinger
ISIS 2	electron density profiles	Topside sounder Whitteker

B. Groundbased Measurements

Typ of measurement	source of data	deduced physical parameter
Ionosonde	European Ionosonde net	maximum electron density (19 stations) electron density profiles (12 stations)
Ionospheric Absorption	Finnish Riometer Network GDR Absorption meas.	Precipitation of high-energy electrons
Geomagnetic variations	Northern Hemisphere Observatories	Distribution of electric fields, field-aligned currents, conductivity, Joule heating

Table 1 - Database 22.1. - 5.2.1974

European longitudes

complex data base for the European sector. The author wants to use this possibility to thank the WDC B2 in Moscow, the WDC for Rockets and Satellites in Greenbelt and the WDC A in Boulder for providing the group with data and for their continuous support.

Different problems have been met for satellites and for ground-based data:

How to find and to identify measurements made simultaneously by different satellites in the same region? From the catalogues (e.g. NSSDC Data Listing 88-01, January 1988) it is easy to identify which satellites have been principially in operation, but it needs already more time to receive informations which allow to estimate at what time the satellite has been inside the limited region. There are no informations available from the WDC which of the instruments onboard of a given satellite have been switched on during a given interval of some minutes and whether the instruments have given useful data. So, the only way for the user is to request all the data available for a given interval and then to spend a lot of time to check whether data have been included useful for this special study. In order to cut down the time for identifying and receiving satellite data for a special investigation it would be useful to improve the information system of the WDC for Rockets and Satellites in the following direction:

To provide the user in an online mode with orbit informations (e.g. the osculating elements) for a given time sequence and with informations about the working regime of the satellite and the instruments giving the possibility to identify which instruments have been switched on during a given time interval along a given orbit. It would be still better to have online access to a table saying which instruments have produced valuable data along a given orbit and for which section (UT-interval) of the orbit. If this information cannot be stored at the WDC it would be useful to have an information how to contact the corresponding Principal Investigator through communication lines in order to get this informations. In all our investigations data summaries and "geophysical data files" have been proven to be of very high value for selecting intervals of special interest as well as for investigating large-scale phenomena. This kind of information unifying in a clearly defined manner the information from instruments with quite different time regimes has been available e.g. for the AEs, DE, GEOS and VIKING satellites. It is strongly recommended that in future "geophysical data files" should be also generated for other complex satellite missions e.g. in the frame-work of Interkosmos.

Some proposals concerning data sets from ground-based stations in a regionally limited area: Building up a data base for a given region means that measurements from a great number of ground-based stations using different instruments and recording systems have to be included in a unified form. This process of deducing and unifying the data (e.g. by digitizing analogue records) is very much time consuming. Therefore, this work should be done in such a manner that the resulting data base may be used by different scientists for various investigations. The data should be stored in a standardized self-descriptive data format. If secondary information is deduced from primary data as e.g. n(h)–profiles from digital or digitized ionograms this information should be given in addition to but not instead of the primary data.

These regional data sets or at least the information that such a data base exists and how it can received by interested scientists should be included into the online inventories. For some kind of measurements internationally agreed standards for data are still missing. IAGA, URSI, CODATA and other responsible organizations should be encouraged to continue their work in that direction.

During recent years introducing new equipments and modern techniques lead to a strong increase of data flux at ground-based observatories, too. Introducing more effective modern storage media at the WDC, in principle, seems to give the possibility to store also these data at the WDC. But the question crises whether this would be necessary and useful. Wouldn't it be enough to provide the user with the information that such data exist and how to get it (e.g. by including this information in an inventory)? National or regional organizations may play its part in collecting and distributing this information in an online mode. The trend seems to go towards a more hierarchical system. This is further supported by the generation of regional centers working in an operative scheme e.g. for monitoring the weather in cosmos or for seismology.

In preparing data projects for the complex international programs of the nineties (e.g. IBGP,STEP, ISTP) it seems to be useful to think over the whole system for distributing data and information.

THE NATIONAL CLIMATIC DATA CENTER INFORMATIONAL SYSTEMS AND PLANS

Kenneth D. Hadeen

National Oceanic and Atmospheric Administration Asheville, North Carolina 28801 USA

The World Data Center-A Meteorology is part of the National Climatic Data Center (NCDC), and my brief discussion this morning will cover various aspects of NCDC which are also applicable to the WDC-A Meteorology. Figure 1 shows the specific holdings of WDC-A, but the data holdings in the national center are also available through request to WDC-A Meteorology.

The National Climatic Data Center was established in 1951 and serves as the national archives for environmental data having enduring value to the nation. There are in excess of 400 data sets in the center, some dating back over 100 years.

The volume of data is between 83–85,000 gigabytes. We have approximately 105,000 6250 bpi digital tapes, 120,000 cartridges, and over 1 million microfiche. The NCDC works closely with the World Meteorological Organization (WMO) and collects and publishes the Monthly Climatic Data for the World (Surface) and publishes the Decadal Weather Summaries for the WMO Regions.

In addition to archiving the standard meteorological parameters of temperature, humidity, pressure, winds, precipitation, clouds, etc., the NCDC collects turbidity and chemistry data from the Background Pollution Monitoring Program (BAPMON), rocketsonde data for upper troposphere and lower stratosphere, ozone data from satellite SBUV sensors, vegetation indexes, etc.

The NCDC is currently involved in three areas which may be of interest to this meeting. The first area is the development of computerized catalogs and inventories. In addition to developing and maintaining catalogs for general use for all our data holdings, we have

NATIONAL CLIMATIC DATA CENTER

WORLD DATA CENTER-A FOR METEOROLOGY

HOLDINGS

- FOREIGN DATA LIBRARY
  - DATA PUBLICATIONS
  - RESEARCH PUBLICATIONS
- IGY AND IGC METEOROLOGICAL DATA - JULY 1957 TO DECEMBER 1959
- GLOBAL ATMOSPHERIC RESEARCH PROGRAMS (GARP) DATA SETS
  - GARP ATLANTIC TROPICAL EXPERIMENT (GATE)
  - FIRST GARP GLOBAL EXPERIMENT (FGGE)
  - WINTER MONSOON EXPERIMENT (WMONEX)
  - SUMMER MONSOON EXPERIMENT (SMONEX)
  - WEST AFRICAN EXPERIMENT (WAMEX)
  - ALPINE EXPERIMENT (ALPEX)
- HIGH ALTITUDE ROCKETSONDE DATA - SEPTEMBER 1959 - DECEMBER 1976

Figure 1 - The specific holdings of the WDC-A for Meteorology

specialized catalogs of WDC-A holdings, and data for Experimental Climate Forecast Centers. Perhaps we will develop a special catalog of data applicable to IGBP. Some of the satellite data inventories are on-line and accessible by a local phone call to TELENET. These inventories have a browse capability, and data orders can be placed on-line.

A second area is the development of networking programs using PC-LANs. Currently, NCDC has implemented one PC-LAN of 29 personal computers of the IBM-XT class with two servers of 286 class, and is in the process of implementing a second PC-LAN with 16 work stations and two additional servers which will tie together the functions of data entry, data quality control, and data applications to the VAX 11-780 and the UNISYS (Sperry) 1100-62 mainframe computers. The communications protocol used is TCP/IP.

The third area is data dissemination. The NCDC works on the principle of providing data on the media and in the format required by the user. We have limited data sets on-line, but disseminate most of our data from publications, microfilm, microfiche, and images, as well as on magnetic tape, floppy disk, and by MCI-Mail (electronic mail). The NCDC has three programs to put climate data on CD-ROM media. One completed project was by a joint venture with the private sector. US West has put on CD-ROM daily maximum and minimum temperatures, precipitation, snowfall, and evaporation data as reported by nearly 25,000 past and present stations in the United States. Some of the most frequently requested statistics have been precalculated. The ease and speed with which data can be identified and viewed, or exported into other programs for further analysis, sets a new standard for the professional user of climate data. We have two additional CD-ROMs under development with other organizations.

A look at our near-term future activities indicates that we will be expanding out efforts in the communication/networking areas to interact with other data centers, the major data processing centers, and to be able to access the major networks like NSFNET, UNIDATA, SPAN, etc.

We are also exploring and using various geographical information systems to see which we should concentrate on to better display our data.

THE WDC B1 INFORMATION SYSTEM AND HOW IT MAKES DATA AND INFORMATION AVAILABLE TO THE USER

V.I. Smirnov, V.Ye. Korotkov, V.I. Melnikov

All-Union Research Institute of Hydrometeorological Information – World Data Centre,
Obninsk, Kaluga Region, USSR

The All-Union Research Institute of Hydrometeorological Information – World Data Centre (AURIHI-WDC) has the responsibilities of a national centre for oceanographic, meteorological and water account data; it is also the centre collecting scientific and technical information for the State Committee for Hydrometeorology, the sectoral source of scientific and technical information in the international system INFOTERRA; moreover, it functions as the national centres for IGOSS and MEDALPEX. The WDC B1 is a part of the Institute and it is responsible for eight disciplines. To fulfil its tasks, WDC B1 uses all computerized systems available at the Institute, which are:

– Computerized system of recording data transmitted via the GTS in real time;

– Computerized data base system using a special (national) data description language as well as international formats such as the FGGE format, GF-3, etc.;

– Computerized data quality control system;

– Computerized scientific and technical information system with remote access to data bases;

– Computerized satellite hydrometeorological data reception system.

To render services to users, non-real time data bank system designed at the Institute is operated.

The computerized systems are based on the most advanced home-made computer and communication facilities: ES and CM computers,

display stations, personal computers, computer networks (oriented computer network between the Hydrometcentre (Moscow) and the AURIHI-WDC). Powerful computer and data transmission resources of the Institute can be used to fulfill WDC B1 tasks whenever required.

For example, the AURIHI-WDC takes part in the integrated international experiments as a World Data Centre and as one of level UB Centres in such experiments as FGGE, ALPEX, and some others.

DISTRIBUTED INFORMATION SYSTEM FOR PLANETARY GEOPHYSICS

V.A. Nechitailenko

Soviet Geophysical Committee

Molodezhnaya 3, Moscow 117296, USSR

Tasks of Information Backing of Planetary Geophysics Research, the Role and Place of the WDC System

Information backing is used here in the sense of a unity of closely related means and techniques, such as:

- information entities of scientific research (data, metadata, information in broad sense, models);
- controlled entities (data and metadata bases, knowledge bases);
- means of managing and access to data (DBMS, MDBMS, hard- and software for local, tele- and network access, DD/DS information-retrieval thesaurus, etc.).

Its four main aims are:

- information support of activities under international geophysical projects;
- creation of disciplinary- and problem-oriented data bases and DBMSs;
- solution of standard tasks of Planetary Geophysics using local and telecommunication access to information resources;
- solution of distributed tasks using a network access to distributed data bases.

The policy of data gathering and international data exchange, as it was defined nearly 30 years ago when the WDC system came into being, is diverging more and more from the actual exchange practice. The future of the WDC system, its improvement and ever increasing role in international geophysical research is, to a great extent, associated with the development of up-to-date policies in the field of gathering, exchange and managing of data. These tasks of Planetary Geophysics are

of a global nature, greatest part of them being practically the same, especially for those countries which cover great distances and are interested even though for that only reason, in the global approach to data analysis.

The present-day Planetary Geophysics research is marked by:

- a sharp increase in the volume of incoming data, mainly in the machine-readable form, which has become possible due to high productivity measuring equipment (satellite and rocket measuring complexes, ground-based systems, e.g. MST-radars, etc.) The generated data fluxes suprase by far the capacity of the existing centres for its collection, storage and effective backing;
- a greater emphasis on holding of problem-oriented observation campaigns backed by a high density of temporal and spatial coverage of the sphere under study and the strive to restrict the access to gained data through their distribution only to projects participants, as a rule before their analysis is completed. For this very reason the demand for identity of available information resources handled by different countres becomes practically unattainable.

The development of up-to-date hard- and software for gathering, backing, processing and exchange of data and information using the modern information technology based on computer networks, makes it possible to set up a "user-friendly" interface. Not only would this serve to play down the difficulties caused by the above two tendencies, but also make it expedient to diversity information resources keeping up their actuality in different centres for discipline- or problem-oriented data bases. Some basic elements of modern information technology have been put to use in Geophysical information systems, for instance, the use of data dictionary/directory concept in the MIAS system (M.T. Jones, T. Sankey, 1979), a system of distributed data catalogue in the NSSDC, USA.

It must be pointed out that the centres, first and foremost the WDC of geophysical data, are losing their significance as "passive" storages, their sponsor or/ and mediator role being brought to the fore in maintenance of information resources and their international exchange. The increasing role of information tasks in Geophysics calls for the necessity of establishing a permanent analyzing of data, with a view to developing data indices, compatible conceptual data models, standard algorithms of transforming the low level data into that of higher levels, constructing and verifying multi-purpose models.

These centres must have geophysicists to carry out various kinds of analysis securing for the purpose the necessary means of access to and work with the data, as well as guidance by specialists in Geophysical information technology.

Of great importance also is the problem of outlining the prospects of developing the data centres system as well as a model of international geophysical data exchange.

Even though some of these aims now appear somewhat far-away and difficult of attainment, their clear-cut definition can, we are positive, facilitate our purposeful advance in this direction, lessening the entropy arising from the differences of purposes and insufficient coordination of activities on the part of the sides involved in the international geophysical data exchange.

Data & Information Exchange Model for WDC System

Information resources

One of the important tasks of data centres is creation, development and keeping in actual mode of various information resources which are numbered here according to their descending priorities:

Information bases, **metadata bases**, inventories containing concise, clear, uncontradictory description of all types of data included in the funds of a centre, the WDC system or associated centres; description of data formats and infrastructure, access methods and conditions under which data can be obtained, etc.

Entities for description included into the metadata bases will be any data set that is of independent interest and can be a subject of international data exchange unconditionally or under defined conditions. Adherence to one of the basic principles of the WDC system namely the equivalence of their funds, is loosing its popularity now for it actually leads to curtailing the system's activity.

The WDC-A and -B universality should be defined in relation to the description of any data (metadata) which are accessible to users under known conditions.

Routine data bases in hardcopies (publications, microfilms, etc.) for permanent use.

Routine data bases in computer-readable form associated with special software or any kind of DBMS.

Problem-oriented data bases (PODB) and DBMS developed under agreement between WDCs concerning rational division of labour, or submitted to centres (on temporal basis) by Projecty Coordinator or other centres (national, regional, institutional, etc.)

Models of geospheres, regions, global processes etc. (knowledge bases).

Other information resources supported by centres on temporal basis as part of their activity in backing research under international geophysical projects.

Exchange Entities

- data, metadata and associated information;
- models;
- applied software.

Exchange Media

- in form of hardcopies;
- on movable computer-readable carriers (discette, magnetic tape, etc.) in formats recommended by Guide to international data exchange through the WDCs (5-th Editon)
- using telecommunications.

The latter is probably impracticable for transmitting great data volumes but is indispensable in supporting of remote access to catalogues, metadata bases, forming of requests on data in remote mode, for transmitting operative data and information, managing complex geophysical experiments as well as in solving distributed tasks of Planetary Geophysics using the network access to distributed data bases

Exchange Channels

WDC – WDC

IDC – NDC WDC – main channel of filling up and renewing WDC information resources;

PIs – WDC users – information backing of international geophysical projects;

WDC – PODB – WDC participation in creating and maintenance of PODB.

Language Means for Data Exchange

Development of "user-friendly" interfaces calls for the necessity to make an easier access for users to information managing systems through unification of language means for data and information exchange.

The aims of unification are : (i) to make easier the work of user-non-programmers minimizing the number of language means, and (ii) to cut short request preparation time.

The unification of language implies at present:

- use of identical or compatible means to form requests and to make effective retrieval in information bases;
- use of authentic bi-and multilingual dictionary/directory of data;
- use of united or compatible information retrieval thesaurus on Planetary Geophysics (in future, on IGBP disciplines);
- use of unified or compatible sets of parameters and formats for describing information resources, data sets, software and DBMS as well as projects, observational networks, acquisition systems and any other means of data obtaining, processing, archiving;

and in future:

- use of software means to support the global distributed conceptional data model to provide unified access to data managed by different DBMS (for example, DAVID system concept, being developed at NSSDC, USA);
- use of agreed protocols and formats (level-to-level and end-to-end) for application and presentation levels of the ISO computer network architecture oriented on solving the network and distributed tasks of Planetary Geophysics.

Architecture of the Distributed Information System for Planetary Geophysics (DIS PG)

Classes of Tasks DIS PG is Called to Solve

The numerous tasks of Planetary Geophysics need for their solution a wide variety of ways and means, including algorithms, hard- and software. Depending on the kind of request to the System, all the tasks can be divided into four classes:

- local tasks, i.e. the tasks that can be solved without the use of remote resources (data, information, programs, specialized computer systems, etc.);
- mail tasks, solved using the E-mail (sending of messages, small files exchange);

- teleprocessing of data (using remote terminals); sending of messages and files, remote initialization of tasks, interactive data, information and programs exchange;
- network processing: dialogue and batch modes of information exchange, remote access to data bases and banks, specialized resources of the network and software, solution of distributed tasks, (i.e. the tasks calling for inclusion into the integral handling and analysis of data on distributed information, soft- and hardware resources of the network);
- soft- and hardware of Coordinated Data Analysis Workshop (CDAW).

The System Components

The DIS PG comprises the following components:

1. Central Metadata Base (CMDB) (Central directory/catalogue created and maintained in actual state by the Committee's host computer. The CMDB must include the following documents:

- the Committee and WDC-B's funds data description;
- the WDCs funds data description;
- description of data belongging to other centres and institutes available for the DIS PG user;
- description of the DIS PG-associated information resources;
- description of problem- and discipline-oriented information retrieval system (IRS) and DBMS, ways of access and use.

The level of detailed description of documents in CMDB depends on the availability of associated information bases for separate disciplines or projects and must be representative enough to provide the potential user with adequate response to his request about the availability and accessibility of the necessary information or data.

2. Documents retrieval is effected using data dictionary/directory system and information-retrieval thesaurus. Factographical retrieval (especially in the case of similar documents and associated information bases) must also be made available.

3. Associated information bases and catalogues realized by computers of the Committee, data centres and institutes linked up with the Committee's host computer through the AKADEMSET transport network.

4. DBMS realized by the Committee's host computer and host (or terminal) computers belonging to other centres or institutes.

5. The central monitor of the system controls the processes of interaction between the user and local or remote resources of the system, including:

- the authorized request-type control;
- registering of requests, specifically fulfilled ones for looking through and copying of data and information;
- making up of tasks to answer the delayed requests;
- initialization of sessions for fulfilled delayed requests.

6. Information and data bases, knowledge bases (local and remote) of general type, discipline- and problem-oriented bases.
7. Software packages, including SP for network processing of geophysical data.
8. Specialized soft- and hardware resources.

In terms of a computer network the above components, at least the first five of them, should be viewed as the network addressed units (NAU). Since open systems of network teleprocessing presuppose the possibility of independent and free-and-easy interaction of NAUs, the corresponding unit must be supplied with means of request authentification for session initialization.

References

M.T. Jones, T. Sankey. The MIAS Oceanographic Database – An Integrated Database/Data Dictionary System. – In: *Database Achievements*. ed. by G.J. Baker, London: A.P. Publ., 1979, pp. 69-95.

J. Thieman. NSSDC's On-line Data Catalog System. – *NSSDC Newsletter*, 1986, vol. 2, No 1, pp. 1, 10, 11.

B. Jacobs. An Overview of the DAVIS System. – *NSSDC News*, 1986, vol. 2, No 2, pp. 10.

UNIFICATION OF DESCRIPTIONS OF DATA COLLECTIONS ON THE ENVIRONMENT SITUATIONS IN METADATA CATALOGUES

Ye. V. Andreyeva, Ye. P. Ryzhykh

All-Union Research Institute of Hydrometeorological Information – World Data Centre

Obninsk, Kaluga Region, USSR

At present the metadata catalogues, containing descriptions of collections of data proper, are being compiled in a number of centres. The collections contain information (knowledge) on the subject area (SA) "Situations in the environment".

Here are some typical examples of messages which are collection descriptions.

1. Monthly means of water temperature for the USSR river basins for the period 1885-1977;
2. A list of active meteorological stations for the territory of the USSR which carried out meteorological observations at 8 synoptic hours in the period 1966–1976.

The topic of the first message is information on mean monthly water temperature, the second informs about meteorological stations carrying out observations according to a described programme.

The entries of the catalogue contain information on the archive data medium and structure, depository, the contact point, etc.

Such descriptions perform a communication function in the course of environmental study and their structure must therefore be either conventional or even unified. This is only possible if the structure of the collection description is based on a conventional conceptual scheme of SA. The conceptual scheme must completely cover the SA in question acting as a classifier for it. Also, the conceptual scheme must be adaptive with respect to new objects for classification, i.e. collections of data on environmental situations. It implies that new objects for classification must fit into the scheme, the latter left unchanged. The structure of the scheme must also determine the structure of the collection descriptions.

The main principles of constructing a scheme meeting the above requirements are examined in the paper.

Principle 1. The choice of classification characteristics (properties) of the objects for classification is based on a general idea of the process of knowledge which is exhaustively described 1). by the object of knowledge (categories: MATTER, MOTION, SPACE and TIME) and 2). by the instrument of knowledge (categories: MEANS, METHOD). The categories mentioned are known as general science categories of Ranganathan and are used in the facet-category analysis when designing conceptual models of SA.

The use of the categories provides for a complete coverage of SA and the adaptivity of the conceptual scheme with relation to new objects for classification.

In a specified SA, the categories are interpreted with an account of specific features of the field of knowledge in question. In a given SA these categories are: Phenomena, Medium, Area, Time, Means and Methods (Table 1).

In a general case, the description of the collection in terms of the categories may take the following form: “Information on PHENOMENA going on in a certain MEDIUM on a given AREA at a specified TIME, obtained by a specific METHOD using specified MEANS.”

It is obvious that the collection description in terms of these categories is characterized by a high measure of generality.

Principle 2. Further structurization of the classification conceptual scheme of the SA is based on the property of the object as a totality of the general and the particular. Due to this, the two independent subfacets are introduced in each facet-category.

Principle 3. A possibility of describing the properties of thé object by different means is also provided by introducing nominative (naming), ordinal (ordering) measuring (ascribing a value) and cumulative (counting the number of objects) scales. The process of describing collections is treated as measuring (in a broad sense, rather than in a narrow physical sense only) values of various properties of collections and recording these values in various scales.

E.g., a part of the AREA observed can be described as a “Region in the North Atlantic” (nominative scale), as an area with the coordinates: Nxxx xxx, Wxxx xxx (measuring scale), and Marsden squares $a_1b_1 g_1$, a b g (ordinal scale). In the category TIME, the measuring scale is equivalent to the chronological, and the nominative scale contains

Table 1 - Comparison between special and universal (general science) categories for SA "Situations in environment"

Universal categories	Special categories	Definition of special categories
1. Matter	Medium	A section of human habitat and activity in terms of which phenomena and processes are considered
2. Motion	Phenomena	Changes in noosphere characterized by a specific set of parameters
3. Space	Area	An area of the Earth surface on which a section of the medium is projected
4. Time	Time	Temporal characteristics of phenomena observed in noosphere
5. Method	Method	A system of regulative principles of practical or theoretical activity in the knowledge process
6. Means	Means	Intermediary systems (Platforms, devices, measuring instruments) in the process of phenomenon parameters registration

names of different time fragments: “winter season”, “flood period”, “the International Geophysical Year”, etc.

In the category MEDIUM, the nominative scale is used for naming the layers and levels of the observed media: troposphere, baroclinic layer, isothermal level, etc; and the measuring scale is used for recording heights and depths describing the fragment in question.

In the category METHOD, the nominative scale contains the method’s name in terms of generalization (observed, average, variance, etc.), while the generalization measure (for example, time averaging) is recorded in the measuring scale.

In the category PHENOMENA, the nominative scale is used for recording parameter names, while the information on measurement accuracy is recorded in the measuring scale.

For recording the names of MEANS (observing platforms, instruments) the nominative scale is commonly used. But often the meteorological stations are identified by numbers in the ordinal scale.

Recording the number of measurements in a certain layer, such as setting the resolution in a medium indirectly or the number of observational platforms as a means of estimating the resolution in an area, may serve as an example of using the cumulative scale.

Principle 4. Hierarchic structures are used for describing constant connections within the facet-categories independent of a particular situation.

Successive specification of independent subfacets and hierarchies, in accordance with the above principles, yields the conceptual scheme. This scheme is an upper basic level of our SA thesaurus. The lower level of the thesaurus is a subject index in which the terms are a number of values of the respective slots (Fig. 1) arranged by terminal tops-slots of the classification scheme.

The terms used for describing situations in different aspects are unified with the help of the subject index.

However, not all the aspects of situations are unified in the collection description. The medium and area boundaries as well as time period resolution values are not unified, and the corresponding slots in the conceptual scheme remain unfilled.

The description of any collection, or rather the corresponding situation in the environment is ascribing to its properties either certain terms from the subject index or numeric values for corresponding slots.

The structure of the collection’s description is defined by the facet section of the thesaurus. Hence, the description acquires a unified form of a questionnaire.

The questionnaire forms of collections descriptions, compared with the catalogue descriptions of the same collections in such known system

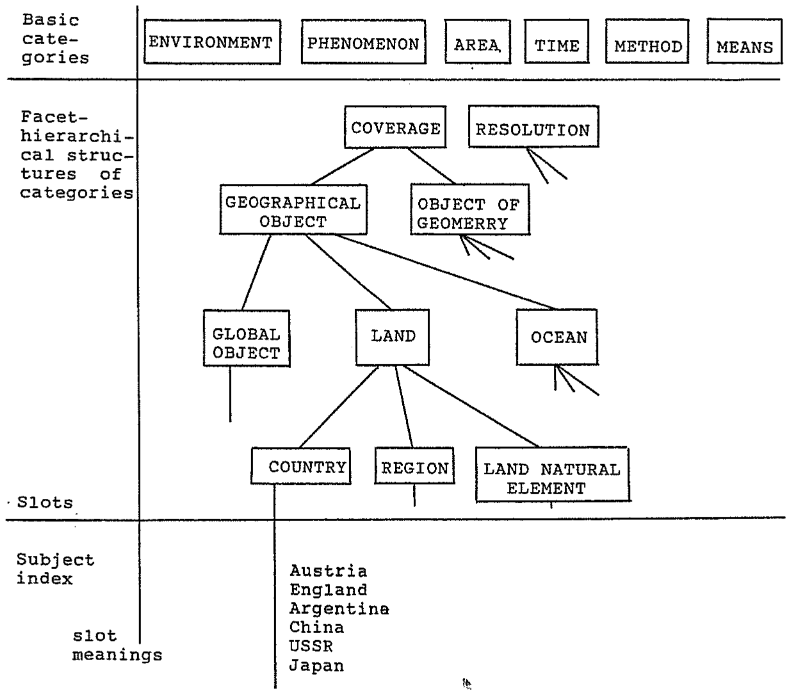


Figure 1 - Structure of the catalogue thesaurus

as NEDRES, MEDI, and INFOCLIMA, WDC-A1 and WDC-B1 Catalogues, are given in Annex.

The questionnaire form of collection description is convenient for several reasons.

First, the situational (i.e. not constant) character of relations between the object categories and sub-categories in the conceptual scheme of SA allows for the absence of any of them by removing the respective items in the questionnaire.

Second, the items of the questionnaire in the computerized catalogue in the "menu" both when describing (abstracting) collections and when drawing up a request for retrieval.

The items of the questionnaire in the published catalogues can be arbitrarily arranged. For example, in the NEDRES Catalogue (USA) the following items are combined in one cluster: parameter with dimensions, instrument, method and medium.

To describe a collection, not only the terms of the subject index but also the terms of the conceptual scheme can be taken as items of the questionnaire. E.g., it is allowed to use the term "Sea" which is the name of the conceptual scheme slot and a generalization of all the terms-meanings "attached" to the slot as the name of the item "Name of the Geographical Object".

Having defined the composition and the structure of the questionnaire, i.e., the description of collections, we shall try to explain the meaning of the term "situation" as it is used in the definition of SA.

All the above examples of filled in questionnaires can be treated as messages whose topic is "information on PHENOMENA".

In terms of the same conceptual scheme messages of such type can be imagined: "information on the MEANS used for obtaining values of parameters of phenomena going on in a certain medium with the help of a certain method", etc. Here the topic is "information on the MEANS".

One can similarly imagine messages whose topic is information on the MEDIA, AREAS studied and METHODS used. The implicit use of all six categories is not at all mandatory in each specific case.

All the diversity of the described aspects of real world is defined by a single notion, "situations in the environment", since there are situational connections between the category objects of subject areas.

The above-mentioned messages are nothing but catalogues of geographical areas, observational platforms, catalogues of methods and instruments used.

As is well known, catalogues along with information from the subject area "Situations in the Environment" contain additional information on the collections: on what data media and in what format the data

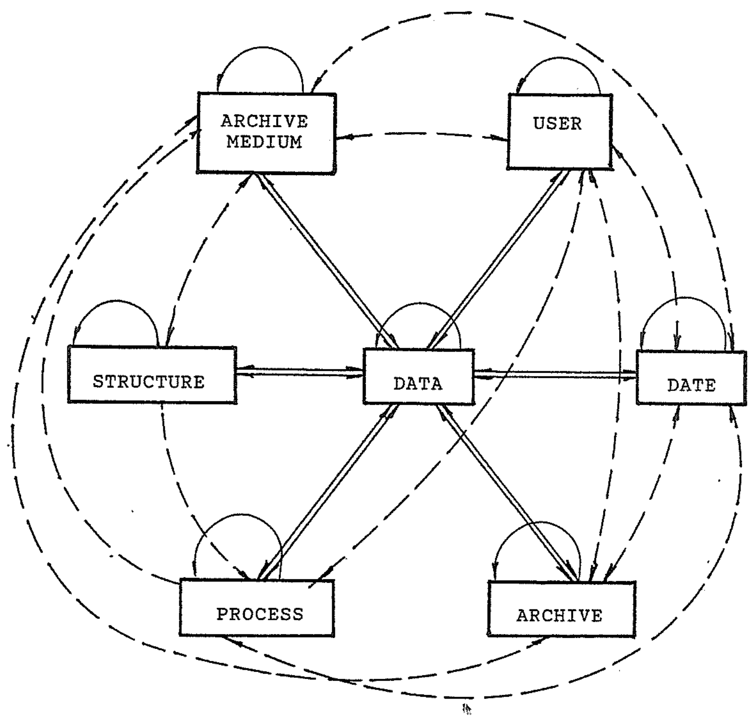


Figure 2 - Conceptual model of the subject area "Natural Environment State Data Holdings"

Thick arrows show external relations between objects of the type "Data" and other objects, generating features, the meaning of which are given in some additional items of the questionnaire.

proper are recorded, who the owner is (contact address), description of data control and processing, etc.

In its turn, all the additional information can be considered in terms of the conceptual scheme of the SA "Data Holdings on the Environment State". The conceptual model of this subject area (Fig. 2) has the following object categories: USER, ARCHIVE, PROCESS, STRUCTURE, ARCHIVE, DATE and DATA. The last category is the one that represents the SA "Situations in the Environment".

The object subspace of the SA, which is described by a model with its own thesaurus and subject index, corresponds to each of the categories.

Some individual aspects describing the above-mentioned object subspaces, related to the space DATA, are included as items in the extended questionnaire describing collections.

A list of additional items in the questionnaire-description of collections of the items is well known from the structure of various catalogues which are published.

Conclusions

1. The conceptual scheme shown in the paper serves as a basis for unifying the structure of catalogues for various disciplines in the SA "Situation in the Environment". It allows for a ready transfer of data collection descriptions from catalogue to catalogue.
2. The structure of the questionnaire, based on the given conceptual scheme, involves structures of all the known catalogue systems as a particular case.
3. The unification of the terminological base not only at the conceptual scheme level but also at the level of subject index, which allows the transfer of collection description from one catalogue to another without any changes. The array of NEDRES can be taken as a basic terminological array for unification.
4. The additional advantage of the scheme suggested is a possibility of creating and maintaining "inverted" catalogues (e.g., catalogues of observational platforms, measuring instruments, geographical areas, etc.) in unified conceptual and terminological contexts.

NEW/EMERGING TECHNOLOGIES FOR LONG-TERM DATA CENTER APPLICATIONS

Carl C. Abston

World Data Center-A

325 Broadway, Boulder, Colorado 80303 – USA

Data Center

What are the primary functions of a data center? Although each of us has rather different ideas as to the definition of a data center, most of us will agree that its most basic functions are the collection, archiving and dissemination of data, both digital and analog.

What operations are necessary to achieve these basic functions?

The collection of digital data is usually performed using magnetic tape, although more recently small data sets or updates are being exchanged via floppy disk. Collecting data on magnetic tape is, however, only a small part of the problem. Although international data formats in the various disciplines have been adopted, there still remain variations in these formats. To create a worldwide collection of data for one narrow discipline, it is usually necessary to reformat the entire data base, and probably sort the data as well. The data should be reviewed for consistency and quality. In the past few years, scientists have been requesting that the data centers do more than these simple functions, that they in fact provide more analysis, graphics, compaction, and so on.

Archiving digital data over the past 30 years has meant storing the data on magnetic tape although in the early days of the World Data Center system, some data were stored on punched cards and even punched paper tape.

To disseminate data, magnetic tape is used as the exchange media, though dissemination also takes the form of publications like Solar-Geophysical Data or information about the data such as the various catalogs which are created and distributed periodically.

Although each of you has different opinions as to the exact nature of data center responsibilities, most will agree that these are indeed some of the most basic functions of a data center located anywhere in the world.

Brief History of Computer Development

What are the new emerging technologies which can be effectively used for long-term data center operations?

Let's look back a few years and examine the changing technologies.

In the 1960's we were all using mainframe type computers in a batch tape/punch card/paper-type environment.

In the 1970's, some of the data centers were using main frame and minicomputers. Batch-type operations were shifting to interactive terminal operations. The use of paper tape died out. Massive tape handling operations were slowly being replaced with random access disk operations. Time-sharing and telecommunications became popular. The minicomputer systems were quickly growing up and brought to the data centers the first real alternative to a mainframe computer.

In the early 1980's, specifically 1981, IBM introduced the IBM personal computer. This single event has probably had the most profound impact on computer science since the introduction of the transistor. For the first time in history, we have a real world-wide computer. In the early 1980's, the PC began finding its way into telecommunications, into areas previously not automated such as accounting, small data bases such as inventories or reserved for word processors, data center operations were not greatly affected. The primary trend was away from batch tape-to-tape operations on mainframes and into interactive terminal operations. Punch card technology died. Time sharing was still expanding and international telecommunications became a common reality. Mainframe computers were also changing. They were improving interactive operational modes and total disk storage capabilities. Minicomputers were continuing to improve in speed, flexibility and cost performance.

In the mid 1980's, specifically August 1984, IBM introduced the IBM AT personal computer. The AT PC brought with it enough CPU (central processor) power to be useful for some data handling operations such as data editing, reformatting and limited graphics. These machines have become a standard as they were sold by the millions throughout the world.

They were cloned by many companies worldwide. Software and hardware add-ons were produced by thousands of companies world-

wide. For the first time, true world-wide international competition emerged creating ever improving products and falling prices for a standardized machine.

In the late 1980's, specifically April 1987, IBM introduced the PS/2 PC system. Until the introduction of the PS/2, all of IBM's PC's were hardware and software compatible. Changing the hardware standard for the internal buss means that all hardware add-on cards cannot be used on this new system. When a company like IBM, with 70% of the world-wide computer market introduces anything it usually becomes a standard. However, two years after the PS/2 introduction, it is still unclear what the long-term trend will be for this system. Until that trend can be determined, it is probably a good idea to avoid extensive use of this machine.

The beginning of the IBM PC literally revolutionized the computer world. A whole new industry was born. And with it has emerged many products which are applicable to the data center operations.

From all of this activity and new products, let's look at what can really be used by the World Data Center System.

Personal Computer

Central Processing Unit (CPU)

First, we have the IBM type personal computers (XT and AT). They are standard throughout the world. Although the price varies greatly between countries, they are nevertheless inexpensive to buy and maintain. They have enough CPU power to perform most data center applications. Specifically, the AT with a 80286 processor running at 12 mhz offers 1 mip of power. Of course the AT-type clones using a 80386 processor chip running at 25 mhz offers 5 mips (about 5 times the cpu performance of a VAX-780). And within two or three years the Intel 80486 running at 40-60 mhz will be available with 3 times the cpu power of the 80386. However, it is really unclear that such a fact processor can be effectively utilized for most present data center applications.

Input-Output

More important than super fast CPU speeds is the whole question of input-output requirements.

First, no computer system is very useful in processing data until it has access to the data. Most international exchanges of data are via standard IBM-type 9 track 1600 bpi magnetic tape. Literally every country in the world can read and write on this media. For a computer to get access to these data, a 9-track tape drive must be available. You can connect a tape drive directly to an IBM XT or AT-type computer or you can connect the PC to another computer which has a tape drive. We will discuss this approach under networking later in this paper. Assume we have a tape drive directly connected to a personal computer. To process the data, we normally read a file from tape, storing these data on magnetic disk. It is necessary to have a disk large enough to hold a copy of the input data, a copy of the output data, plus all the necessary software. A 1600 bpi tape can contain as much as 40 million characters of data; therefore, a 90 million character disk or two 44 million character disks are a minimum disk configuration for effective data processing.

The disk random access speed and data transfer speeds are also very important for effective data processing. The most popular random access speeds are 64, 38, 28 and 18 milliseconds per access. Of course, the faster the access speed the faster processing can proceed. The standard XT and AT disk drives have a maximum data transfer rate of 125 kilobytes/second. Recently introduced disk controllers (SCSI or ESDI) increase the data throughput from 125 kilobytes per second to 925 kilobytes per second. For example, processing a given amount of data using the standard AT disk controller with a 28 millisecond access disk is three times slower than using a ESDI controller with an 18 millisecond access disk.

Data Compaction (software)

Various Data Compaction techniques have been available for a number of years, however, they have generally been inconvenient to use and inefficient for general purpose use. They have been very successful in disc storage compression such as the RLL controllers and image files generated by optical scanners and used in disk top publishing systems.

Recently however, I have discovered a public domain compaction routine which offers considerable possibilities for data center functions. The routine is named PKARC (compaction) and PKXARC (decompaction). The routines run only on an IMB-PC XT or AT-type machines. A megabyte of data can be compacted or decompacted within 30 seconds.

The amount of compaction depends of course on the data file. For example, a program in binary object code was reduced from 290KB to

109KB whereas a catalog text type file was compacted from 800KB to 96KB.

With this compaction routine it is possible to exchange floppy discs with large amounts of data, increase the telecommunications speed of data transmission, or even further compact archive data. The latest CD-ROM (GLORIA) from USGS/NOAA uses this compaction routine for much of the data. For example, the original Gloria data was over 2.5 gigabytes in size.

By converting 3-digit ASCII data values into 8-bit binary (1 digit), the data base was reduced to 750 megabytes, by using PKARC on some of the data, the data base was further reduced to 570 megabytes and by mastering the data onto a CD-ROM, the data base was compacted into a single 12 cm disc. In this example, data compaction took several forms. The first compaction was the conversion from ascii into binary, the second was the PKARC type of compaction, and the third was archiving the data onto a single optical disc. The original 2.5 gigabytes of data required over 83 1600-bpi magnetic tapes or about 25 6250 bpi tapes.

Data Compaction (hardware)

Data Compaction based hardware is discussed under disk storage.

Graphics

Screen/Monitor

As data are being processed, some effort is usually made to check for quality or consistency. Although this can be done by printing out millions of characters, it is frequently more effective to display the data in graphical form. In the early 1980's most graphics were monochrome (black and white) using hercules resolution. However, by the mid 1980's good and inexpensive color monitors became available. The CGA (300x200 point resolution) was available. By 1987, multisync color monitors were available supporting a variety of resolutions in both TTL and Analog modes. EGA (640x350 16 colors) became widely available, VGA (640x480 16 colors) was introduced with the PS/2 machines in 1986, and enhancements to VGA became available in 1988. These include resolutions of 640/480 with 256 colors, 800x600 with 256 colors, and 1024x768 with 16 colors. With a new multisync color monitor ($650US) and a new enhanced graphics card ($300), it is now possible to support all of the above graphic standards. Time series data can be effectively presented on a color monitor using any of these standards.

However, image data, especially from satellites, require the enhanced VGA color resolution. Contour or track-line data can be effectively displayed in EGA, VGA or preferably 1024x768 resolution.

Hardcopy printer

Although displaying graphics on a monitor is useful in checking and correcting data, a hard copy is frequently required for more detailed analysis, publication work, or for end users. In the past various types of plotters have been used, however, their usefulness is generally limited to vector-type data.

The PC revolution has stimulated very rapid development in printer graphic devices. For example, the 24-pin dot matrix printers produce near letter quality character output as well as excellent quality 180 dot per inch graphic output. Most of these devices are black and white although a few use tri-color ribbons to produce a variety of colors. The cost varies widely depending on quality and speed. However, in the US costs range from $500-$1100.

More recently, the laser-type printers have become popular. These black and white devices are typically 300 dots per inch (12 dots per millimeter). When combined with a postscript controller, they are capable of reproducing raster images in 256 levels of gray shade. Again, costs of these systems vary widely but in the US market a good system can be procured for $4000-5000. The HP Laserjet with additional memory is about $2500.

A newly introduced technology which may become very popular is the thermal wax printer/plotter. They can reproduce images at 300 dots per inch resolution with 256 colors. They are available starting at $8000 US.

New types of technologies for printing/plotting are being introduced almost monthly and it is not clear which technology will prove to be the most desirable over the long term.

Data Storage

In every data center in the world, long term data storage is absolutely critical. Therefore it is important that we discuss and understand the good and bad points of different storage media.

9-Track Magnetic Tape

The most popular storage media is 9-track 1600 and 6250 bpi tape. The 1600 bpi mode is now in use world-wide. The 6250 bpi mode is extensively used in the developed countries. Most data centers now have libraries of both types of tapes numbering from a few thousand into the millions. Magnetic tape, although a good exchange media, is a poor long-term archive media. The shelf life of tape is generally quoted at 5 years although under good storage conditions you may be able to read tapes up to 20 years.

Magnetic Tape Cassettes (3480)

These 3480 tape cassettes were introduced by IBM in the mid-1980's and are extensively used at many large IBM mainframe installations although they are hardly used outside of this environment. This system will probably never be popular in the world data center system.

Hard Disc Drives

Most PCs have some type of hard disk drive. In 1981, the 5 megabyte disk drive was the most popular type. By 1988, 44 mb drives have almost become a standard in the US market place, with 79, 90 and 150 mb drives commonplace. As the price of these drives has dropped, the popularity has increased. As the popularity has increased the cost has dropped. Therefore, we have seen prices fall from thousands of dollars to hundreds of dollars in a year or two. For example, a Micropolus 70mb 28ms disk is today selling at $525 instead of $2000 a year ago.

At WDC-B in Boulder, we have recently installed in a PC AT, a 340 mb 18ms disk drive with an ESDI controller at total cost of $2500. This permits a full 6250 bpi tape (150mb) to be copied to the hard disk, the data reformatted and quality controlled and an output file of 150mb created, and finally the resulting data copied back onto an archive tape. The results of using this system has meant improvements in computer performance, human performance, and quality control.

Removable Hard Disc Cartridges

Real removable hard disc cartridges, not high density floppies, are available from several sources. One such type is made by SyQuest and holds 44 MB (formatted) of data with a 40 ms access and a 20,000 hour

MTBF (mean time before failure). The total system (drive, cable, controller, and software) costs $1200 US and $100 for each disc cartridges. It is directly compatible to all IBM XT and AT type systems.

Although there are no standards for these devices, they are never-the-less inexpensive enough to be an interesting method for exchanging data. When erasable optical disc prove to be available, inexpensive, reliable, and responsible, they may prove to be a better investment.

Optical Disc (WORM)

Optical Disc Write Once Read Many Times (WORM) system is a very different approach to recording data as compared to the magnetic media. A laser gun actually changes the surface reflectivity for each bit of data written by one of several methods. The laser also understands these changes in reflectivity when reading the data. This system provides a user with the ability to write his own media on his own machine and thereby create a long term (10-40 years) non-erasable, removable, random access data storage facility.

Three different sized media are popular, 5-1/4, 12 and 14 inch. The 5-1/4 inch disks are being connected to personal computers. They are available from many companies and range in storage capacity from 200 to 400 million characters per side with 2-sided media available. They cost $3000 to $4000 US and they vary greatly in performance. The 12 inch WORM is also available from several vendors, holds approximately 1-2 gigabytes per side and is mostly connected to mini and mainframe computers. Costs vary widely from $10,000 to $20,000. The 14-inch units are only available from Kodak and hold over 3 gigabytes per side.

This non-erasable media has been developing slowly throughout the 1980's, but still, unfortunately, suffers from a total lack of standardization. WDC-A in Geomagnetism is using two 5-1/4 inch WORM drives connected to personal computers to archive high resolution time series data. At WDC-A, the experience has been generally favorable although a number of problems were encountered during startup of the project. Each side of a disc holds 200 million characters of data (400 megabytes for a two side disc).

The WORM system offers data centers a very compact non-erasable media, large amounts of removable storage and random access. The worst concern is the lack of standardization. Writing archive data onto WORM media carries a risk that as a standard finally emerges all previously written WORM discs will have to be copied to the new standard. However, this approach may be preferable to continuing to hold the world's archives of various data bases on erasable and volatile magnetic tape.

Optical Disc (CD-ROM)

CD-ROM optical disc is also a non-erasable optical-type recording. The disc is 12 centimeters in diameter and can hold up to 660 megabytes on a single side. Only single-sided media exists. The recording format is standardized on the ISO 9660 definition. Therefore, all CD-ROM drives can read all CD-ROM discs. The lifetime of a correctly mastered disc is estimated at between 10 and 40 years.

However, media must be created at a central mastering facility. The creation of the master glass copy is expensive ranging from $2000 to $4000. After the glass master has been used, it is discarded after a few months as the glass is not stable over a long period of time. Duplication of each additional copy, however, is $4.00 or less. Therefore, CD-ROM media is an expensive way to make single copies but an inexpensive way to make and distribute hundreds of copies.

WDC-A for Geomagnetism began working on their first CD-ROM in December 1986. This project was completed in June 1987. During the following year 500 copies were distributed to 20 different countries. The CD-ROM readers easily fit into or may be connected externally to any IBM PC XT or AT computer at a cost of $800 in the US market. Interfaces for both the Apple Macintosh and the entire Digital Equipment Corp. (DEC) line of VAX's will be available this year.

The WDC-A experience in CD-ROM technology has been very positive. The first CD-ROM contains the total world's collection of digital geomagnetic hourly values data available as of 1987 along with 14 other data bases. With hundreds of copies distributed world-wide, it is unlikely that these data will ever be lost.

This collection was made over a 30 year period, reformatted and processed by many different people at the World Data Centers and written on hundreds of tapes. Bringing together this collection in preparation for mastering a CD-ROM, numerous problems were discovered. For example, long-term magnetic tape storage involving hundreds and even thousands of tapes is not always reliable. Many people involved in preparing data in many formats or variations of "standard" formats cause many misunderstandings. Changing project managers at the data centers create additional problems. Solving these problems required a half staff year of effort. It meant getting copies of some of the original data from other national or international data centers, quality control and correction, reprocessing and reformatting most of the data base.

The WDC-A (Boulder) experience with long term archives in magnetic tape is not excellent. We believe, however, that the long-term survivability of data on CD-ROM is excellent.

CD-ROM Quality Control Problems

Reports have been surfacing in Europe, that CD-ROMs are failing after four years. The latest information available indicates that the problem is oxidizing of the aluminum substrata. This oxidization changes the surface characteristic and deteriorates the readability of the stored data. Does this mean that all CD-ROMs are failing after four years? No. What this does indicate, however, is a quality control problem at some mastering facilities. Properly mastered CD-ROMs still have an estimated lifetime of 10 to 40 years. Research is also continuing on alternate metallic substrata. It has been found for example, that Gold and Silver substrata do not oxidize although the cost is prohibitive.

Those of you who still remember the early days of magnetic tape also remember the numerous quality control problems with that medium.

It isn't surprising to discover quality control problems with any mass produced product, especially a revolutionary new technology. What is surprising with CD-ROM technology is that it has taken four years to discover any quality control problems.

Optical Disc (Erasable)

Erasable optical drives for PCs are being announced presently by several companies with shipment of products to begin 1988. Depending on the performance characterisitics and price, this may provide the means for handling multi-gigabyte data files from a PC. It could even prove a handly method of exchanging data between World Data Centers. If the claims are accurate, the performance is very quick, and the cost is low, then the impact on PC systems will be enormous and the data centers may wish to consider compatible optical drives in each center for data exchange and archiving. However, before we do much planning, it is necessary to remember the exaggerated claims which have been made over the past 10 years regarding optical disc technology.

It is a good idea to wait and see if these advertisements prove to be accurate. This item may prove to be a good candidate for discussion at the next such meeting on emerging technologies.

Helical Recording Units

Perhaps the newest recording method for digital data is Helical video tape cassette recording. The technology is the same as used in your home video cassette recorder. Two types are currently available, the standard VHS and 8 millimeter cassette size. They use exactly the same cassettes as you use in your home recorder. The recording units are available at $6000-$7000 for a PC or VAX or $50,000 for a very high performance mainframe type system.

The cassettes are available for $5-8 each. Each cassette holds between 1.2 and 2.2 GB of data providing sequential data access. Retrieving data from the end of a fully recorded tape is relatively slow up to 10 minutes on most systems. Very rapid hardware and software changes are being made this year. Reasonable systems for under $5000 should be available by the end of 1988 from many companies.

This media offers a very compact method of storing data as well as exchanging data when recording formats are standardized. For example the World Data Center/National Data Center complex in Boulder hold's less than 200 gigabytes of actual data.

If the entire archive was recorded using the Helical recording method, it could be stored on a small book shelf and be directly accessible by a PC with a helical drive unit.

Helical Scan recording devices can also be used very effectively to back up the magnetic disk data and optical disc data from PC's and minicomputers.

The disadvantages are the non-standardization and volatility of magnetic storage.

WDC-A Boulder will begin testing these devices by the end of 1988, budgets permitting.

Digital Audio Tapes

Perhaps Digital Audio Tape (DAT) recording will become an important archive method. However, this technology is very new and directed toward the commercial market at this time. Within the next couple of years this technology may be a serious consideration for data storage.

Networking

The PC revolution solved a lot of problems, however, it also created a few new ones. A personal computer, although it has adequate CPU power for data processing applications, is absolutely worthless without access to the data bases. This access can be achieved by connecting tape drives, printers and optical disc units onto each separate PC, or the access can be achived by networking all of these units together. Networking is a much less expensive method than equipping each PC with all of the necessary peripherals.

Ethernet LAN

Although a number of networking systems exist, the most common and standard method used today is the Ethernet Local Area Networks (LAN). This requires that each computer on the network be connected with coax cable, fiber optic, or twisted pair wire and have an Ethernet controller card (approx $400 US for one PC). The Ethernet physical network is capable of moving 10 million bits per second.

LAN Protocol

The most commonly used protocol on Ethernet LANs of multivendor computers is TCP/IP. TCP/IP is used for both local area networks and wide area networks. This protocol permits any computer on the system to become a host.

The host can receive and transmit data upon request from any other host on the system. TCP/IP runs on most computer systems today including IBM-type PC's, VAX's, Apollos, SUN's and UNISYS.

Another widely used PC LAN protocol system is Novell. The Novell operating system is installed on a PC. This PC now becomes a server to all other PCs on the LAN. The server PC's disc and printer peripherals are now available as shared devices by all PC's on the network. For example, the server disc may be identified as F: and any PC in the network can enter F: and begin talking to disc F exactly the same way as any other local disc. With Novell many users can have access to the same data base at the same time. The server provides disc access; however, the PC's which are accessing the server are actually providing the CPU processing power. Therefore, the server can handle a considerable load before reaching saturation.

Another protocol system used at WDC-A's in Boulder is the Network File System (NFS) by SUN Computer System. This system permits a SUN or VAX with an UNIX Operating System to become an intelligent server to any PC connected to the network. For example, it becomes possible to use all of the PC DOS commands when accessing data files on a VAX minicomputer. Any PC in the network can run a program to read and write data a record at a time on either the VAX or his own PC.

Wide Area Networking

At the WDC-A's in Boulder, Washington, and Asheville a pilot project is underway to interconnect via 56 kilobaud lines all three centers' Ethernet LAN using TCP/IP protocol. This is fairly inexpensive, only requiring communication lines, a 56 kbps modem, and an IP Router at each center. This will permit any PC from any of the three centers to transfer files to any server in any center.

Operating Systems

DOS

Dos 1.0 was introduced with the IBM XT in 1981 and has remained the primary operating system for personal computers up until today. Over the years, the system has been updated with increasing capabilities. The newest release of DOS is version 4.0 which is available now. This version permits formatting of 2 gigabytes of disk storage, use of the Expanded-Memory Specification (EMS), plus an optional SHELL with pull-down menus and popup windows, scroll bars and icons for use with both keyboards and a mouse.

OS/2

OS/2 1.0 was introduced with the new PS/2 series April 1987. It is a multi-tasking system with the capability to directly use very large discs and very large memory sizes.

Unfortunately, the system makes an AT 80286 based computer run as slow as an XT and uses 2 million characters of memory in the process. However, OS/2 may become the most common operating system for the 80386 systems with their real 32-bit processors and data channels. We hope to experiment with OS/2 on an 80386 machine within the year.

Currently available DOS software can run as a task under OS/2; therefore continuing software compatibility.

UNIX/XENIX

UNIX or XENIX is another operating system which has been adapted to the PC. This operating system was created in AT&T Bell Labs in the early 1970's along with the C programming language. Today UNIX runs on more different types of machines than any other operating system. Perhaps only DOS is more widely used because of the millions of personal computers in use. UNIX is a multitasking, multi-user, virtual memory system designed to handle many users. Some of the bad points of UNIX are its need for system resources, speed of operation in comparison to DOS, the lack of compatibility with present DOS designed software and its user unfriendliness. Actually, UNIX isn't user unfriendly but rather cryptic. In other words, it uses commands like "ls" for listing out the file names in a directory or "cd" to change directories. DOS is actually a subset of the Unix operating system. It is, in fact, a powerful operating system and I think UNIX will dominate the non-IBM operating system market in the future. DOS and OS/2 will probably dominate the PC market for the next ten years.

Again, WDC-A in Boulder plans to test UNIX from Santa Cruz Corp. on an 80386 PC this year. The worst part of UNIX for the PC is that not much commercial application software is available to run on Unix but this situation is improving. The available Unix applications are also more expensive since they are priced for multiple users, DOS applications can also run under UNiX, but not as efficiently as in a straight DOS environment.

My personal thoughts regarding the use of these operating systems, other than DOS on personal computers, is *don't*, unless you have a very good reason. For example, if you have a custom program requiring more than 640 kilobytes of memory such as a contouring program, image processing or GIS (Geographic Inventory System), either OS/2 or UNIX is, of course the best answer without reprogramming.

Why Use These New Technologies

After several pages of discussion regarding these new emerging technologies, the real question is why? Why use them if the present system is working? Do they really solve all of my problems? Of course the answer is no, they will not solve all problems. They will, however, solve a lot of problems and create a few new ones.

Cost/Performance

Cost/performance is always a critical point in any data center operation. Data centers are always short on money because available funds are generally spent in creating new experiments, new data recordings or new observatories but seldom spent in processing and preserving these new data.

Why can PC-based technology lower costs? There are several important reasons.

First, the marketing cost recovery philosophies with the personal computer are very different from the traditional mainframe approach. IBM, with perhaps 70% of the world's total computer market, always sold computers based on what the market would bear, in other words for as high a price as possible without any regard to the actual production costs. However, the personal computer market changed a lot of things. For the first time, we got true international competition in an open and uncontrolled market. The American manufacturers would price their products just below the IBM price; the Japanese companies price their products low enough to try to eliminate competition frequently below manufacturing costs, the other Asian companies, priced their products just a few percent above manufacturing and shipping costs. When all of these marketing and pricing philosophies collided, the prices on PCs and PC products tumbled. This has not happened in the mainframe computer area, although prices have been dropping in the minicomputer area.

Second, with the large number of PCs being sold, (12 million in US corporate offices and 23 million worldwide), research and development cost could be divided by millions of units rather than only a few thousand, thereby further reducing PC costs.

Third, the success of the PC provided a huge market for other peripheral and software companies. In fact, half of the available computer R&D money is being directed toward the PC. This has meant drastic and rapid improvements in many technologies such as printers, discs, memory, optical discs and software. In the early 1980's peripherals designed and developed for large machines were downsized to be used on PCs. Today, peripherals developed for the PC's are being directly interfaced with the larger machines because of costs.

Fourth, it is inherently much less expensive to build one processor chip than the traditional method of mainframe CPU design.

The above few paragraphs are discussion about costs, not necessarily performance. Perhaps the most informative method to illustrate cost and performance is the following WDC-A example.

Cost/Performance Example

WDC-A in Boulder shares a mainframe with other National Data Centers and operates both a minicomputer (VAX-750) and a distributed network of IBM XT and AT-type computers (63 PC's). The following is an example of the costs involved in both systems.

We are using a VAX 750 with 1.3 gigabytes of disc storage, 8 megabytes of memory and two 1600/6250 bpi tape drives, 600 line per minute impact printer, Unix operating system, editor, Fortran-77 and C compilers, TCP/IP communications and other utilities. This system was purchased three years ago for $170,000. The system runs fairly well with eight users depending on each user's system resource requirements. Maintenance, both hardware and software, is another $20,000 per year.

To create an eight user PC-based network at today's US prices, we need eight user PC's, two network servers and two tape drives. An 80286 AT type PC running at 12 MHZ with 60 megabyte 28 millisecond disc, 640 kb of memory, an EGA graphics card, color monitor and a 3-COM Ethernet interface card costs about $2500. The server PC's should include 350 million character disc (16.5 millisecond access) with an ESDI controller and 1600/6250 bpi tape drive. The large disc with controller costs $2500 and the tape drive $12,000 or about $17,000 per server. Printers are available beginning at $200 up to $1100 for 24 pin dot matrix printers, $ 1800 and up for Laser printers, $8000 and up, for color thermal wax printer. Lets say we add one top-of-the-line dot matrix printer for each user PC at $1000 each. To this add $1000 per machine for software with each computer customized for a different purpose. The total cost for this system is $70,000.

These are very different systems and it is important to understand that not only are the costs very different but the performance and capabilities are also very different. With a single user on the VAX, that user has all resources available and the system will run very quickly. One user of the PC-based system will run at a fixed speed. If eight users are on the VAX, the total VAX CPU and disk I-0 power must be divided by the eight users. Performance is much slower than before. On the PC system with eight users, each PC is running at the same CPU speed as before. Actually a 12 mhz 80286 processor is faster than the VAX 750 processor and the language compilers are generally more efficient for

the PC than the 750. In fact, our comparisons of applications written in Fortran-77 running on a PC and on the 750 with one user indicates the PC runs a little faster. Application written in C however, indicate performance on the PC at about 3 times that of the 750.

With eight users on both the VAX and PC networked system, applications programs will run many times faster on the PC. However, there are also other differences. For example, on the PC with a DOS operating system, the user's program cannot exceed about 640 kilobytes of memory. The VAX user can get up to 6 megabytes of memory for a single program. On the VAX, many users can be updating the same data base at a time. On the PC network, only one user can update a data base at the same time. However this problem can be solved by adding a $5000 Novell server (hardware and software package) to the network allowing many users access to the same data base at the same time.

Graphical display on a terminal connected to the VAX is very slow because the connection speed is usually less than 9600 baud (960 characters per second) whereas the PC graphic display is running at I-0 channel speed (125,000 characters per second). An EGA raster image requires 640x350 or 224,000 pixils. At 960 pixils per second, an image requires 233 seconds minimum time of a terminal connected to a minicomputer or mainframe computer. The same image on the PC actually will run as fast as 8 seconds. Therefore, graphical presentations especially for quality control, become a very practical tool.

The same problem exists with the tape drives. With one user on the VAX system, 4-6 megabytes of data can be transferred per minute. With a heavily loaded system of 12 users, 1/4 megabytes will be transferred per minute. The PC-based tape drive will read and write data at 2 megabytes per minute in all cases.

Example Analysis

The above example shows two different systems with different performance characteristics and different costs. Each system has its own advantages and disadvantages. In the given example, it seems quite logical that the PC network offers superior performance. However, it is difficult to actually comprehend the full PC network power. I think this is true because as one considers personal computers, one is only thinking of a single PC and its capabilities, rather than the overall network capability.

In this example, there are eight user PC's, two server PC's, two 6250/1600 bpi drivers, eight printers, eight color monitors with graphic controllers, $10,000 in software and all the associated networking.

There is a total of eight 70 MB discs plus two 350 MB discs, or a total of 1.26 gigabytes of disc storage. There are eight user PC's and two server PC's providing 5-10 MIPS of CPU power (depending). The total combined memory is 6.4 million characters.

The eight printers can print up to 1160 lines per minute in draft quality, 360 lines per minute of near letter quality, or draw eight graphs at a same time. The eight color monitors and graphic cards are capable of displaying both vector raster images as fast as 2-8 seconds per image per system.

Only by adding up all of these statistics does the total system power begin to emerge. Now it becomes more obvious that the PC network offers many performance advantages over the minicomputer in this example. However, there are other examples where the minicomputer would be superior.

Modularity

Looking at the above VAX versus PC network systems, several points become obvious. When more users are added to the VAX system, the CPU time must be divided by more users; therefore, the original users are getting less performance. Without replacing the entire minicomputer system, CPU power cannot be upgraded.

The networked PC system however, can be expanded with additional user PC's for $4500 each. If the network becomes overloaded, it can be split into two LAN's. If a common data base server running NOVELL becomes unresponsive, another server can be added to the network and some data bases moved. If more processor power is required for some application, a 80386 system can be added. If one application requires more than 640 kb of memory, Unix or OS/2 operating system can be added to a user PC.

Reliability

The demonstrated reliability of the PC network is many times better than our present VAX system. The VAX will experience a hardware failure once every couple of months. A PC may run a year or two without a failure.

When a failure does occur, the mainframe and minicomputers require very skilled maintenance people. The PC failure however can be determined by inserting a few spare parts, and the defective component discarded rather than repaired.

Single Point of Failure

When a minicomputer or mainframe computer stops working, the entire center also stops working. When the repair takes days or sometimes a week to fix, the center suffers a lot of wasted time. When one PC stops working, only one persons time is wasted and the repair or replacement is usually quicker.

Data Survivability

Holding data which cost hundreds of millions of dollars to collect on thousands of magnetic tapes is about analogous to an accident waiting for a place to happen. The optical technologies, especially CD-ROM, assure long term data survivability. Helical recording, even though it is also magnetic, reduces the archive to a bookshelf size space and reduces the media costs to a point where duplicated archives become economically feasible.

Improved user services

Using a network of PC's has proven at WDC-A in Boulder to generally improve the quality control of the ingest data. The laser printers have improved the publication quality. Distributing data on CD-ROM with accession software has proved to greatly enhance the usability and understandability of these data.

Summary

The above text has very briefly discussed the good and bad points for each of the emerging technologies. I think we could spend a week just discussing any one of these different technologies.

Perhaps the most important question is why should we use any of these new technologies? The simple answer is "improved cost/performance, improved staff productivity, improved quality control, improved data survivability, and improved user services'.

Does this mean that mainframe computing is out of date? Not necessarily. Every type of computer today can be used effectively when given the proper mix of jobs. For example, if a data center mounts 10,000 tapes per day, then a large mainframe is probably the best choice of computers. If a data center mostly does modeling applications requiring thousand of computer hours per day, then a CRAY type computer is the best choice today. However, if a data center mostly receives data, performs quality analysis (QC) on the data, reformats and edits, archives and finally sends the data to customers upon request, then a mainframe computer is probably not the most effective type of computer configuration today.

Once again, I must point out that a single stand alone personal computer has very limited capabilities. And it cannot become a powerful tool without access to a lot of peripheral devices.

DYNAMIC MODEL OF THE GEOSPHERES IN GEOPHYSICAL INFORMATICS AND THE APPROACH TO THE KNOWLEDGE BASE ON PLANETARY GEOPHYSICS

Natalia E. Papitashvili

Soviet Geophysical Committee, Moscow 117296 USSR

Vladimir O. Papitashvili

IZMIRAN, Troitsk, Moscow Region 142092 USSR

Abstract

The concept of a dynamic information model of the geospheres is formulated in this paper. The logical structure and some definitions are described. This will allow the development of a common approach to the knowledge base on planetary geophysics used for dedicated analysis and modelling.

Let us define the concept of "geophysical informatics" as a volume of data and information about any events, entities or processes, which have had or have continued in the geospheric system. From the planetary geophysics point of view, we could outline the magnetosphere as an "application domain" (AD), which is limited by near-Earth space environment and the infinitely thin conductive ionosphere. At any moment of time a number of processes are occuring in the magnetosphere which characterize interactions with the space environment. These processes appear in the ionosphere as an "image". In this way the "information flow", generated by the magnetosphere, characterizes the dynamics and conditions of the physical system.

Let's assume that there is an "observer" watching the information status of the physical system at any moment of time and stores this status in his memory. Therefore, it is possible to suppose that his memory contains "data" which reflects the physical world. These "data" could be interpreted as an "information system" or "information model" of the geosphere. The process of mirroring the physical system into an information system is a relationship between the two realities of the world, the generator and the receiver of information. It could be called an "information process" /1, 2/.

The main atributes of information are structuring, meaning and value. The structuring of geophysical information closely connects to the model of the information generator, i.e. the model of the geosphere. The structuring of information is conducted in accordance with the entity rules, which connect the data atributes. The value of geophysical information is defined mainly by its substance, completeness, reliability, responsiveness and a number of other characteristics /3/.

The above mentioned properties and operations with geophysical data are the subject under investigation for geophysical informatics. Therefore, geophysical informatics is a science about the principles and the transformations of the Earth sciences information of the entire planet.

As in cybernetics, geophysical informatics could be divided into a few disciplines; e.g. ionospheric informatics, geomagnetic informatics, etc. Let's concentrate our attention on the last one: geomagnetic informatics. Methodologically we should conduct conceptual modelling of geomagnetic data and should form a number of requests to the information system of geomagnetism and estimate the kinds of movements in the information process (IP) for the perception of the interaction between the magnetosphere and it's environment, including the ionosphere. Also, we should analyze possible schemes of conceptual elaboration of the knowledge base, which can describe our knowledge about the real world through the formalized entities.

It is possible to consider the global network of magnetic observatories and different geomagnetic surveys as a generator of information about the geomagnetic field. Let's assume that all geomagnetic data in the magnetosphere are a multidimensional array. These dimensions are three spatial coordinates and time samplings (annual, daily, hourly values, etc.) of three components of the geomagnetic field. After selection of the time sampling (e.g. hourly values) the multidimensional array divides into three 4-dimensional arrays which are possible to describe by "flat" files of a relational data base.

Thus, the elaboration of the information system of geomagnetism could be connected with definitions of entities: geomagnetic field components and their atributes (value, coordinate system, time and place of observation, type of averaging and variability). Of course, in the magnetosphere it is possible to define other entities, characterizing the plasma component of the system. Also, it is possible to define other entities which are outside of the magnetosphere (solar wind and IMF) and ionospheric parameters (electric fields and currents, conductivity, Joule heating, etc.). The list of all entities will characterize the physical system as a whole. There are different relationships between entities which are characterized by their similar and dissimilar properties.

Let's simulate the information movements about the interactions of the magnetosphere with it's environment (near-Earth space). Using the regressional analysis of ground-based geomagnetic data from the worldwide observatory network, the interplanetary magnetic field and the solar wind data, we can combine the model of geomagnetic disturbances along the different points of the Earth's surface /4, 5/.

Taking into account some assumptions about the ionospheric conductivity, it is possible to calculate the current function, potential, electric field and currents along the ionosphere and the geomagnetic field lines relative to the region of investigation. In this way, we can simulate geomagnetic disturbances above the region of investigation in real-time and spatial fashion using only near-magnetosphere parameters. Such data processing in our information system permits us "to compress'" the initial data. Now, there is no necessity to collect initial geomagnetic data in full, from the complete observatory network, for the end-point picture of modelled geomagnetic disturbances. This model can be normalized by the real-time measurements in the key (from the point of view of physics of interaction processes) observatories.

Thus, users, except initial geomagnetic data, can be supplied by a "compressed image" of possible (simulated) processes of interaction between the solar wind plasma and the magnetosphere which is projected on 9 the "ionospheric screen". Serial consequences of such "images" (or "frames") is a "movie", showing the feasible information status of the magnetosphere in each moment of time.

Adequacy of this "dynamic information model" (DIM) of the magnetosphere, compared to the real events, is defined by two main assumptions:

– physical consistency and assurance of our theoretical and empirical knowledge about interactive processes which rely not upon physical relationships, but also rely on statistical regressional relationships;

– sufficient distribution of global or regional observing networks which put data in the DIM.

From this point of view, it is easy to estimate the method for optimization of the global magnetic network. It is necessary to derive from the global network, a few subnetworks or chains for main geomagnetic indices calculations. Only 5–6 equatorial observatories will be enough for the Dst-index derivation, 5–10 middle-latitude observatories can provide Kn and Ks (or Kp) indices, 10–12 stations along auroral zone will be enough for the AE-index. Two observatories near both north and south geomagnetic poles can provide us the polar cap (PC) index of the "inferred IMF direction" (so called Svalgaard-Mansurov Effect). Not more than 30–40 magnetic observatories, optimumly distributed on the Earth's surface, can provide us the main indices and normalize our "dynamic information model" to real events.

Let's consider by which method of operation it will be possible to compress geomagnetic data. From our point of view, the concept of the Pilot Data System on Solar–Terrestrial Physics (PDS/STP) is most suitable for the Dynamic Information Model /6/. In the scope of this concept, the next items to be elaborated are:

Data Management
Networks and Communications
Users Interface
Massive Data Processing
Elements charged with the responsibility of overall system engineering and special processes, artificial intelligence or expert systems, etc.

Elaboration of the DIM prototype at least carries out the notion of a "knowledge base" on planetary geophysics, i.e. a set of data, algorithms, models and images which reflects our knowledge about the real world. Therefore, a declarative imagination about the real world in computer memory will be considered as a knowledge base and knowledge in the computer memory can be presented with the help of some symbolic system and semantics /7/.

Let's consider the problem of knowledge in the sense of computer understanding. This understanding is based on the simulation of the application domain and knowledge about it and the language with the help of which the connection between input message and real subject (described in this message) is established. It is possible to pick out a system based on the operations of pattern recognition. But it is clear that such systems can solve only the information–reference tasks comparing the input message with the document in computer memory (search criteria). Joining application and item-oriented descriptions of application domain and using the sequent-claster analysis technique, it is possible to teach our information system to combine different logical views of the real world and select only one pattern from the majority; in other words, to forecast possible development of the physical system. Therefore, we believe that computer understanding of the problem can correct input messages. A higher degree of computer understanding, of course, is a natural language interface.

As mentioned above, the modelling approach is widely used in geophysics and seems to be the direction of the future. As a rule, some simple algorithmic models can be applied. But such a description demands the full knowledge about the physical system and represents the closed model of the real world.

An approach, established on incomplete knowledge about the physical system is more general. Such type of models are defined as open ones. They permit one to specify the application domain in more natural form than the algorithmic one. They are: classificational systems, relational systems, semantic networks, frames, productions and reduction systems.

Information-search thesaurus which all based on boulean logic ("whole-part", "part-whole", "functional likeness", etc.) are related to classificational systems /8/.

The system, established on the relationships, is presented as a limited set of tables which correspond to a mathematical relationship. The multidimensional array, described above, is an example of a system based on these relationships. The main advantage of relational systems is a simplicity of design technology using the help of serial normalization of the relationships. For relational data bases, the logic conditions were elaborated. Using this logic, it is possible to build formula which express the meaning of the statement and, therefore, create a natural language search criteria.

Semantic networks have recently been introduced for the tasks joined with elaboration of knowledge systems. Semantic networks consist of the directed graphs which have been marked as nodes and arcs. Entity corresponds to a node and semantic relation between entities corresponds to an arc. Such an aggregate of the relationships can be transformed into a semantic network but complicated techniques for comparison of graphs do not always permit the implementation of semantic networks for the knowledge base. For geophysics, it makes sense to use a semantic network for the declarative description of information only.

Frame is a logical unit of information of a system-structuring description of the application domain containing (on the basis of semantical properties of this subject area) the empty positions (slots) which, after filling with definite information, convert the frame into the medium of definite knowledge. For geophysical purposes, the frame systems are generally used for the spatial presentation of the subjects and figures, etc.

A more convenient method for the realizing the knowledge presentation, is based on the production systems. Productions are the mathematical analog of the logic description and their acceptance permits the building of more complicated descriptions. However, not every model in the real world can be accepted and brought to productions.

Reductional systems belong to the methods of intuitive (heuristic) search. Here, in the multidimensional space, there is a sequence of the operators with automatic construction of the reductional model which carries out the initial situation to the end-point. For example, computer simulation using the free search algorithms, selection and qualification of results from a previous step which conformed the boundary conditions should be related to the reductional systems.

Therefore, we can conclude that attempts have been made to build a scheme of possible knowledge base in planetary geophysics which have been conducted by scientists around the world. Some approaches are being developed in the National Space Science Data Center (NASA), as well as other centers and institutes /9/.

From our point of view, the most important objective is to formulate tasks which are necessary to be solved in close cooperation with all World Data Centers. As an initial phase, it is necessary to create:

1) Common Dictionary/ Directory and Thesaurus on data and knowledge for Planetary Geophysics with the purpose of defining the common terminology and unifying all types of data.

2) Also, it is necessary to realize that serial adaptation of different algorithmic procedures into a common knowledge base are needed for the initial data processing, i.e. "data compression" and geospheric simulation.

3) Two above mentioned topics automatically lead us to develop PDS/STP which will have features of the dialogue expert system for geomagnetic diagnosis and forecasting of magnetospheric processes. On this basis, the elaboration of the International Automated Geophysical Informatics System for WDCs will be possible.

4) More complicated elaboration of the knowledge bases using frames, semantic networks or reductional systems will be possible at each Center separately.

Acknowledgements

We wish to express our gratitude to Carl C. Abston (NGDC/NOAA) for helpful discussions and assitance in the preparation of final version of paper.

References

1. Kulikovskiy, L.F. and V.V. Lyutov, 1987, *Teoreticheskie osnovi informatsionnikh protsessov. (Theoretical background of the information processes)*. "Visshaya shkola" Publ. H., 248 p., Moscow.

2. Chetverikov, V.N., G.I. Revunkov and E.N. Samokhvalov, 1987, *Bazi i banki dannikh (Databases and databanks)*. "Visshaya shkola" Publ. H., 248 p., Moscow.

3. Teorey, T.J. and J.P. Fry, 1985, *Design of database structures.* – Part 1. "Mir" Publ. H., pp. 28-163, Moscow.

4. Kamide, Y., H.W. Kroehl, B.A. Hausman, et al., 1983, *Numerical modelling of ionospheric parameters from global IMS magnetometer data for the CDAW-6 intervals. WDC-A for STP*, Report UAG-88, NGDC/NOAA, 197 p., Boulder CO.

5. Levitin, A.E., Ya. I. Feldstein, R.G. Afonina, et al., 1984, *Large-scale electric fields and currents in the high-latitude ionosphere*. Parts 1-6. "Gidrometeoizdat" Publ. H., Moscow.

6. Abston, C.C., N.E. Papitashvili and V.O. Papitashvili, 1986, *Concept and structure for an automated solar-terrestrial physics data base system*. Preprint No. 11a (625), IZMIRAN, 21 p., Moscow.

7. Shemakin, Yu. I., 1985, *Vvedenie w informatiku (Introduction in the informatics)*. "Finansi i statistika" Publ. H., 190 p., Moscow.

8. *NASA Thesaurus*. 1985, Hierarchical listing (Vol. 1) and access vocabulary (Vol. 2). NASA SP-7053, STIB/NASA, Washington D.C.

9. Campbell, W., L.H. Roelofs and N.M. Short, Jr. 1987, *The development of a prototype intelligent user interface subsystem for NASA's scientific database systems*. TM-87821, GSFC/NASA, 32 p., Greenbelt MD.

SOME CONSIDERATIONS CONCERNING DATA SETS FOR GLOBAL CHANGE

Gregory W. Withee

National Oceanographic Data Center
NESDIC/NOAA

Washington D.C. 20235, USA

August 4, 1988

Background

Proper data and information handling will be critical to the International Geosphere Biosphere Program (IGBP). The assembly of data sets is certainly a major part of the overall data and information handling problem. While the specific properties of data sets required by the IGBP have not been established,. some general attributes are known. For example, the data sets need to be global in nature and have to span long timeframes; they must be dense enough in time and space to contain, in a non-aliased way, the scales of activity desired (a few months to tens of years in time); they must be immune from calibration problems, i.e., any offset due to sensor technology or processing technique cannot mask the global change signals; and finally the data sets must be accessible.

The above properties of global change data sets shape the methodology used for their assembly. In general, the steps toward assembling such data sets may be viewed as: data observation; data collection; data assembly into global data sets; quality control; and data set assimilation into 4-dimensional (4-d) grids.

The purpose of this paper is to highlight some international considerations when assembling these data sets. Such a discussion raises questions that hopefully will be discussed at the IGBP Study Conference.

Introduction

Ptrinciples of data management for scientific programs such as IGBP have been discussed in many studies and reports (e.g. 1,2). The guiding principles used for the following discussions are:

- Data management is end-to-end, i.e., it includes all steps from observation to model output, both real time and non-real time.
- Data management includes management of information about data calibration, validation, principal investigator, etc. Documentation such as information on instruments and calibration, observation procedures, data reduction algorithms, and quality control is essential. Wherever possible this documentation should be available with the data.
- Scientific advice must be sought on variable prioritization, data requirements, etc.
- Scientists and data managers must work closely together. For example, each data set being prepared should be processed through a scientific application(s) for quality control.
- Data handling standards should not be ad hoc—they must be actively sought and coordinated.
- All data bases should be easily accessible.

In the paragraphs below a few key points are presented for each step identified in the data set assembly process. Where appropriate, considerations of an international nature have been provided to stimulate discussion at the workshop. Note that not all of these thoughts originated with the author.

Data Observation

General Remarks

Observations from a single purpose research investigator may still be useful for gloval change researchers. As in the case of meteorology, members in the global change community will need to share global data. Observations also need to be timely—having a data set sit on a researcher's desk for 10 years hurts the global change researcher's progress.

Quality control starts at the observation planning stage, and should be thought about from the sensor itself to the overall space/time network design. Observations/network planning require an ongoing monitoring of observations being taken, and an analysis of all observations that have been taken. Planning for in situ and satellite observations must be done in concert.

International Considerations

1. Establish an international clearing house where information can be obtained on the observations that have been, are, and are planned to be taken. This is being recommended not only for subprograms of IGBP, but for all related experiments, such as the World Ocean Circulation Experiment (3). This clearing house(s) must be long term. The World Data Centers, operating in conjunction with ICSU and UN bodies such as WMO, UNESCO, and UNEP, are candidates.

2. Small, inexpensive personal computer systems, called CLICOM, have been developed by World Meteorological Organization (WMO) to gather, analyze, and make available meteorological data particularly as a tool for developing countries. The CLICOM model could be followed for other data types of use in IGBP. This could be a valuable link to allow developing countries to participate in and benefit from IGBP.

Data Collection and Assembly

General Remarks

Perhaps the most difficult task is to identify, collect, and assemble a global data set. For some variables, such as atmospheric pressure international mechanisms, in this example WMO, have been established for timely data exchange. This makes for easier data assembly. International organizations such as WMO and IOC are essential to data exchange. The World Data Centers play a strong role here.

When building data sets, care must be taken to ensure that older data can be used along with new data obtained from new techniques (again, documentation is critical).

Assembling all the data is important. Conclusions based on, for example, one third of the total possible data population, may be misleading or invalid.

International Considerations

3. One or more intergovernmental organizations, such as UNEP, UNESCO, or WMO in cooperation with nongovernmental bodies such as IGBP, need to adopt standards for data exchange of all pertinent global change variables. We must try to ensure that global assembly of data sets includes as much data as are available.

4. An international network should be established for the effective knowledge of and access to data sets. This global change network might be viewed as analogous to the Global Telecommunication System in WMO.

Data Processing and Quality Control

General Remarks

Data sets going into global files are checked for internal consistency (i.e., within the project) and for climatological consistency (i.e., with known conditions). The checks should be made under the guidance of, or directly by, a researcher who is using the entire data set for an application. Errors are detected through use of the data. Such an effort has been started under TOGA (4).

A note of global change caution: recently we are finding that out-of-bound checks with known climatological variations may exclude real variability. We must be careful not to throw away real anomalies (5).

Validation/consistency checks need to be performed in a timely way. It has been argued, however, that we need less, not more, quality control. Sometimes one data set will be checked many times, in different ways, so that the end user has no way of knowing what was done to the data. There is a need for better documentation about corrected data.

International Considerations

5. An international body, probably ICSU or IGBP itself, needs to establish some basic quality control guidelines for each variable of interest. International in situ calibration exercises are in order. Validation meetings such as the recent meeting in Holland are essential.

6. A pilot or demonstration project using the quality standards should be financed internationally, and include co-participation by a research institution and a data center.

Data Set Distribution/Access

General Remarks

For many variables where a low density of observations exists, or where models not been developed, there is value in distributing the actual observations. This can be done on tape, networks, or CD-ROM. The Comprehensive Ocean-Atmosphere Data Set (6) is an example of a data set which should be put on CD-ROM, and made widely available.

An online system providing knowledge of data sets, their availability, and a simple browse capability on part of the set would be very useful.

Larger satellite data sets have special problems. Directory and inventory information are still very important, but easy access to the entire file is diffucult. We must try to find ways of developing and using high-information content, low-bit volume data sets instead of low-information content, high-bit volume data sets from satellites.

International Considerations

7. Establish and maintain an international directory of available data sets. Again, the World Data Centers might be candidates for this activity.

Data Set Applications

For some variables there are models which can be used to fill in a regular 4-dimensional grid. These models do not produce unique 4-d data sets, so that many such 4-d data sets may exist over the same time and space domain.

For some researchers these 4-d data sets will be the stepping off point for their research, rather than the original data sets themselves. Thus, we need to include these assimilated data sets in the IGBP data and information management problem.

International Considerations

8. An international data laboratory, perhaps structured like the European Center for Medium Range Weather Forecasting, would ensure that some of these global data sets would be built and used. Furthermore, such a laboratory would foster a global interdisciplinary research community. This center or laboratory would be a blending of good science and good data management and could house many of the international activities recommended in this paper. It would provide a chance for talented scientists to work on global problems which may not be financed in his or her country. The UNEP GRID facility, in Geneva, if expanded in concept, may be an example of such a laboratory setting. All data relevant to IGBP would be available at this center.

Conclusions

Data sets for IGBP will be important cornerstones for IGBP science. But in many areas, the required data sets are years from being available. The author provides eight international recommendations which might help accelerate the process of data set assembly and availability. The recommendations are not precisely defined, but may serve to focus some discussion during the workshop.

References

1. National Research Council, 1982: *Data Management and Computation*, Washington, D.C.: National Academy Press.

2. Committee on Data for Science and Technology, 1986: *Directions for Internationally Compatible Environmental Data*, International Council of Scientific Unions.

3. U.S. Science Steering Committee for WOCE, 1988: *U.S. WOCE Implementation Plan*, First Draft, U.S Planning Office for WOCE, Texas A&M University, College Station, TX.

4. White, W.B., G. Withee, S. Pazan, and C. Noe, 1988: *Joint Environmental Data Analysis (JEDA) Center*, American Geophysical Union, Eos, 69:9, March 1, 1988

5. S. Levitus, *NOAA/Geophysical Fluid Dynamics Laboratory*, private communication with author, June 1988

6. Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph., P.M. Steurer, and J.D. Elms, 1985: *Comprehensive Ocean-Atmosphere Data Set; Release 1.* NOAA Environmental Research Laboratories, Climate Research Program, Boulder, Colorado, 268 pp.

DATABASING OF SCIENTIFIC DATA

Robert R. Gamache and Bodo W. Reinisch

University of Lowell Center for Atmospheric Research

450 Aiken Street, Lowell, MA 01854, USA

Introduction

This report addresses the collection, storage and distribution of scientific data. We do not intend discussing in detail storage devices, archiving techniques or the present state of storage media. Nor will we talk about relational database systems. Although these systems are very valuable and useful for managing, constructing and using databases, their transportability is a problem. Distribution of scientific data should not be dependent on the type of computer system or operating language at a users site. Much of what is discussed below is based on experience of managing a database and using other databases. The driving force of such work is the dissemination of scientific data for others to use, question, contribute, etc. To this end, we believe that the ease of use of the distributed data is a most important characteristic.

Some basic philosophies for creating a scientific database are discussed. An examples is given, based on data from the University of Lowell Center for Atmospheric Research (ULCAR) Digisonde 256 system.

Creating a Database Structure

The creation of a structure for a distributed database should be done with the eventual user in mind. For databases with a large circulation, this implies users at all levels with varying degrees of familiarlarity with the data and computer expertise. For a database to the successful it must be simple enough for the beginner and sophisticated enough for the experienced user. This leads to the question of size vs. ease of use. Clearly all indicators are now pointing to "user

friendly" applications as being the way of the future. New methods of data storage, such as optical discs, are making the use of "tricks" for storing data no longer necessary or fruitful. Some argue that the rate of data collection and the increase in storage capabilities are growing on par. Our feeling is that a database must be in the "user friendly" category if it is to be successful. This almost necessitates abandoning clever storage techniques.

One should then ask "What criteria yield a user friendly database".

The first criterion that must be addressed is the transportability of the database. If it is to be used by many, it must work on many types of systems. This fact makes the use of a relational database structure (one which acts on fields and keys, etc.) such as dBASE III2 or VAX-Rdb3 not applicable for this task, since these type of packages are written for particular operating systems. They are however excellent for database management at the primary site of a database, and may be a necessity for very large databases. To be operable on many types of systems (VAX, CYBER, UNIVAC, IBM, etc. mainframes, workstations such as SUN, Appolo, Micro-VAX, etc. or any one of the many personal computers) the database must be created from *standards*. This suggest the use of ASCII characters for the data rather than binary or one of the many data packing schemes. The resulting increase in the volume of the database is offset by the ease of its use. If any programs are to be included with the database (this is discussed later), they should be written in a standard language. The suggestion here is the use of FORTRAN which is supported by all systems.

Another data size reduction scheme is the employment of only integers to represent the data. This clearly does not fit into a user friendly scheme since translation to the actual numerical values requires many different scale factors, an undesirable feature. Thus floating point numbers should be written as floating point numbers, integers as integers, characters as characters, ...

This brings up the issue of format. Unfortunately, a format must be selected for the data in order for the database to be read. The suggestion is to keep the format simple and well defined. If many of the data fall into the same format, they should be grouped together into a record (a single read or write) and read with the same format. If this is not possible, a format should be chosen such that the data can be visually understood and written in that form. This format should then be carefully described so that any user can read the data. The length of a line of data should be chosen so that the database can be visually scanned, printed, and edited by most available equipment. In the example given below a line length of 120 characters was chosen. This length line can be printed and viewed by most printers and monitors. Most modern editors can also handle lines of this length rather easily.

The database should have a standardization of units. For example, one and only one unit of length should be chosen, one unit of energy, one unit of mass, etc. Deviations from this are sometimes necessary to avoid unnecessary use of exponentials, e.g. in the example below we use MegaHertz for sounding frequencies and Hertz for Doppler frequency shifts. The database should also be accompanied by a report that completely and clearly describes the variables of the database, the format which they were written with, the blocking structure and any other information needed to use the database.

Currently, the most common form of distribution of databases is still 9-track magnetic tapes. For large databases this often requires the blocking of many records in order for the database to fit on a single magnetic tape. When this is done the blocking structure should be documented. This is very important for users when they first try to read the database.

One of the best aids to the user is to distribute the database with a program that can be used to read it. We recommend this to be in a standard language and we suggest FORTRAN, since FORTRAN is supported by any system used for scientific studies. The program should be "structured", well commented and easy to understand. In the simplest form, the program can be a subroutine to read the data. Potential users add this subroutine to their program to select certain data. Depending on the type of database the program can be made "user friendly" i.e. questions appear on the monitor asking for a selection of data, etc. An example of this is the program SELECT which accompanies the HITRAN database4. This is a molecular absorption database and the program SELECT allows a user to select data for particular frequencies, molecules, isotopic species, bands and intensities. Checks are made to ensure that the selections are correct, e.g. a molecule and isotopic species designation agree; when incorrect the question is repeated.

The last point to address is the use of a structure that can be extended for future additions. This applies to databases where a block of data is repeated, e.g. the results of a single experiment make up a block and the database is the results of many experiments. The problem that can arise is that the experiments may not always yield results for all variables in the database. Rigidly fixing the structure means that many of the data elements will be zeroes. In this type of repeating data base, one can add an index to each block (experiment) which indicates what data were taken and how many elements of

each will occur. This can be used to eliminate all null elements in the database. This also gives one the flexibility to leave room for variables in the database that are not yet obtained but are expected in the near future. With this, the database does not change structure with time and it maintains flexibility without the increased size.

The ULCAR ADEP Database

The example given is for the Digisonde ADEP5 (ARTIST Data Editing and Printing) output. In this database, ionospheric data are stored for various stations as a function of time. The structure of the ADEP database file is shown in Table 1. Each block of data corresponds to data from one ionogram and is preceded by a number of integers called the Data Block Index. The structure of the file is such that only the information that was recorded for the ionogram need be stored. We define 37 Codes that correspond to particular groups of data (a Group is all lines of data for a single code), a Line is a sequence of Elements, terminated by a carriage return/line feed, and an Element is a single datum in the specified format. The Data Block Index indicates which Group (Codes) are present and how many Elements are in each Group. Thus if the 11th Data Block Index value is 0, there are no Elements for the Code 11 (amplitudes for the ARTIST scaled F1 0-trace); if it were 23 there would be 23 of these Elements. Thus, the Data Block Index completely describes what and how much data is to be found in the block.

The ADEP database file is written as a standard ASCII text file with a maximum line length of 120 characters. It can easily be printed or edited with standard text editors, read in a programing environment such as FORTRAN, C, Basic or PASCAL, or read and processed by relational databases such as dBASE III or VAX-Rdb.

The format of the data corresponds to the particular data and does vary from one Code Group to the next. All data within and one Code Group, however, are for the same type and format. This simplifies the reading of the data. The number of characters in a given Code Group can easily exceed the 120 characters per line limit. In this case, the output overflows to succeeding lines, thus a Code Group may extend over several Lines of data.

Most of the data can be used without further scaling or manipulation. Because of the nature of the data, the Doppler numbers for the traces (Codes 8, 12, 16, 20, 24 and 28) index an array given by the Doppler translation table (Code 5). Other than this, the data can be understood by visual inspection. In the data, all heights are given in

Code	Format	Description	
	40I3	DATA FILE INDEX	
1	60Z1	IONOGRAM SOUNDING SETTINGS (PREFACE)	
2	50F8.3	SCALLED IONOSPHERIC PARAMETERS	
3	20I2	ARTIST ANALYSIS FLAGS	
4	10F7.3	GEOPHYSICAL CONSTANTS	
5	16F7.3	DOPPLER TRANSLATION TABLE	
6	400I3	ARTIST O-TRACE POINTS - F2 LAYER	VIRTUAL HEIGHTS
7	400I2		AMPLITUDES
8	400I1		DOPPLER NUMBER
9	400F6.3		FREQUENCY TABLE
10	150I3	ARTIST O-TRACE POINTS - F1 LAYER	VIRTUAL HEIGHTS
11	150I2		AMPLITUDES
12	150I1		DOPPLER NUMBER
13	150F6.3		FREQUENCY TABLE
14	150I3	ARTIST O-TRACE POINTS - E LAYER	VIRTUAL HEIGHTS
15	150I2		AMPLITUDES
16	150I1		DOPPLER NUMBER
17	150F6.3		FREQUENCY TABLE
18	400I3	ARTIST X-TRACE POINTS - F2 LAYER	VIRTUAL HEIGHTS
19	400I2		AMPLITUDES
20	400I1		DOPPLER NUMBER
21	400F6.3		FREQUENCY TABLE
22	150I3	ARTIST X-TRACE POINTS - F1 LAYER	VIRTUAL HEIGHTS
23	150I2		AMPLITUDES
24	150I1		DOPPLER NUMBER
25	150F6.3		FREQUENCY TABLE
26	150I3	ARTIST X-TRACE POINTS - E LAYER	VIRTUAL HEIGHTS
27	150I2		AMPLITUDES
28	150I1		DOPPLER NUMBER
29	150F6.3		FREQUENCY TABLE
30	20I2	MEDIAN AMPLITUDE OF F ECHO	
31	20I2	MEDIAN AMPLITUDE OF E ECHO	
32	20I2	MEDIAN AMPLITUDE OF ES ECHO	
33	20E9.4E1	TRUE HEIGHT F2 LAYER COEFFICIENTS	
34	20E9.4E1	TRUE HEIGHT F1 LAYER COEFFICIENTS	
35	20E9.4E1	TRUE HEIGHT E LAYER COEFFICIENTS	
36	20E9.4E1	TRUE HEIGHT MONOTONIC SOLUTION	
37	20E9.4E1	VALLEY COEFFICIENTS	

N O T E S

Nomenclature:

- Block - All data for one ionogram.
- Group - All lines of data for a single Code.
- Line - A sequence of Elements, CR/LF terminated.
- Element - A single datum in the specified format.

Table 1 - ARTIST Data Editing Program (ADEP) Block Format

kilometers, all sounding frequencies are in MegaHertz, all Doppler frequencies are in Hertz, Amplitudes are given in Decibels and angles such as latitude, longitude and dip angle are in degrees. All Digisonde times are in Universal Time (UT).

The structure of the database separates the E, F1 and F2 layers whenever possible. When there is no f_0F1, separation of F1 and F2 is not possible and all information for the F layer as a whole is contained in the F2 data Groups. The Ordinary and Extraordinary traces are treated separately. Although the current ARTIST does not scale the X-traces this information will be provided in the near future. When they do become available the database, and all programs that use the database, will not have to be changed at all. This is due to the flexibility gained by use of the Data Block Index.

An ADEP database file may contain one or more Blocks of data, with each block corresponding to one ionogram. A Block must contain one or more Groups of data, the Data Block Index being mandatory and all other Groups (Codes) being optional. A Group may have its data spread across one or more lines. Each Line has one or more Elements.

Distributed with the ADEP database is a FORTRAN subroutine which reads the data. For users who wish not to use Fortran, the routine can be used to explain how to read the data. The Edit Descriptor repeat factors shown in Table 1 represent a maximum number of fields in that Code Group. Most Code Groups will be of variable length and read by loops which require only a single non-repeated Edit Descriptor.

Summary

The database structure that was described here and implemented on the ADEP database has many good qualities. It is very flexible and will not need to be changed as the ULCAR ARTIST system improves (inclusion of X-traces, decreased frequency step sizes, etc.). Perhaps its best feature is the ease of using the database. The structure is well defined, the format fixed yet flexible enough so that several years down the line programs used to manipulate it now will still be useful. It is also easily maintainable which becomes increasingly important as the database grows in size.

References

1. R.J. Sica, A.D. Richmond, B.A. Emery, S. Chakrabarti and V.B. Wickwar, "*The Frist CEDAR Data Base Report*", NCAR/TN-308+STR, NCAR Technical note, National Center for Atmospheric Research, Boulder CO, 1988.
2. *dBASE III*, Ashton-Tate, Culver City, CA, 90230, USA.
3. *VAX-Rdb*, Digital Equipment Corp., Maynard, MA, USA.
4. L.S. Rothman, R.R. Gamache, A. Goldman, L.R. Brown, R.A. Toth, H.M. Pickett, R.L. Poynter, J.-M. Flaud, C. Camy-Peyret, A. Barbe, N. Husson, C.P. Rinsland and M.A.H. Smith, "*The HITRAN database: 1986 edition*", Appl. Opt. 26, 4058(1987).
5. R.R. Gamache, T. Bullett, Z. Zhang, and B.W. Reinisch, "*ADEP Database Report*", The University of Lowell Center for Atmospheric Research, August, 1988.

COMPATIBILITY PROBLEMS IN COMPUTER-TO-COMPUTER COMMUNICATIONS

An Overview on Network Protocols

E.Jabs

ESA/ESOC Robert-Bosch Strasse 5 D-6100 Darmstadt FRG

Introduction

Electronic data exchange (communication) between different data centres, or more specifically between computers of different data centres, requires a certain degree of compatibility of computer and communications facilities.

The problem is trivial in case if computers of a single range of a single supplier are to be interconnected and if this supplier provides adequate facilities which match existing communication circuits and networks of PTTs or of other basic communication services suppliers.

In reality this trivial case is rarely existing and communities like geophysical data centres have to face a situation, that almost each data centre has different type of noncompatible computer facilities with more or less developed communications facilities, which are conforming to various standards. Even the basic communication networks, available in the different host countries e.g. by the PTTs or other communication services suppliers, may be not compatible.

The need for supplier independent and international communications standards is therefore obvious and has also been recognised by a variety of international standardisation bodies, the most important of which are subbodies of ISO (International Organisation for Standardisation) and of the CCITT (Comité Consulatif International de Téléphone et de Télégraphie).

The first step was to develop a common language and a reference model for computer system interconnection, which was achieved by the OSI (Open System Interconnection) committee of ISO. The resulting 7-layer-model will here be briefly presented as an important reference.

The reference model does not yet provide detailed protocol and service specifications for the different layers which could be a suitable basis for implementation. They had to be developed by other subcommittees. Not all of these specifications are currently completed and the implementation of completed standards as marketable products takes time. ISO communications protocols are currently therefore only available to a limited extent.

Almost all real computer communications of today are therefore still based on vendor specific protocols or other private protocols. The most important of these, the one from IBM and DEC, are therefore here also briefly reviewed.

Todays computer networks are therefore in most cases pragmatic solutions, applying often mixtures of vendor specific and private protocols.

Incompatible computer networks may also be interconnected by means of "gateways", which are protocol converters, translating protocol elements, services and addresses of one network into those of the other.

The paper concludes with a brief description of relevant computer network solutions currently applied by the European Space Agency.

The OSI-Reference Model

The OSI reference model, also known as "7-layer-model", which was issued and is maintained by ISO, is logically composed of an ordered set of layers, through which users (application processes) of different systems communicate with each other by exchange of meaningful messages. The logical structure of this model is shown in Figure 1.

Essentially the lowest three layers or levels (1-3) are concerned with the communication protocols associated with the network through which two intercommunicating computers are connected. The upper three layers (5-7) are concerned with the protocols necessary to allow machines with heterogeneous operating system and different internal data formats to interact with each other. The transport layer (4) has an intermediate role and masks the higher layers from the lower, network dependent layers.

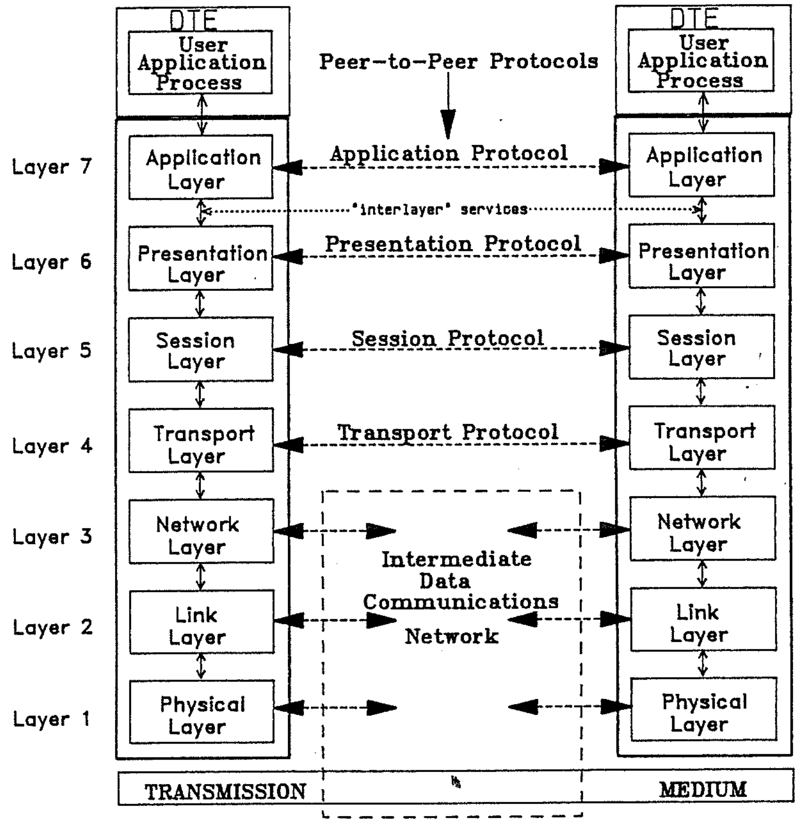


Figure 1 - OSI Reference Model

The function of each layer is specified in the form of a protocol which defines the set of rules and conventions which are used by the layer in order to exchange information with a peer layer in a remote system. Each layer provides a defined set of services to the next higher layer and in turn uses the services provided by the layer immediately below it while transporting message units associated with the layer specific protocol.

The functions of the seven layers can be summarised as follows:

Physical layer: provides the physical network interface (plugs, sockets, signal levels, signalling rates etc.), associated with the transmission of a "bit stream".

Link layer: provides a reliable, error free data packet transmission facility across the physical link.

Network layer: is concerned with the routing and switching of data and with the combination of different network topologies or addressing structures. It provides to the transport layer the service, which enables that layer to exchange Transport Protocol Data Units (TPDU) with a remote transport layer.

Transport layer: provides both connection management and data transfer services. The connection management service allows a user of the transport layer (i.e. the session layer) to establish and maintain a logical connection to a correspondent transport user in a remote system (i.e. host-to-host control). The data transfer service provides the means for the exchange of data between two corresponding users over this connection and performs error correction, independent of any network used.

Session layer: provides dialogue control and synchronisation and maintains an end-to-end communication path between two application processes for the duration of each complete Application Layer Activity or Transaction.

Presentation layer: ensures, as primary function, that the information communicated between two application processes is meaningful to both processes (encoded in understandable form). A secondary function can be encryption of information, which might be required to increase communication security and to maintain the confidentiality of transmitted data.

Application layer: is the highest OSI protocol layer and is concerned with the provision of application services to network users (end-user application processes), such as file transfer, terminal access, job transfer, electronic mail.

Progress of ISO Layer Standardisation and Relevant Implementations

The first international standard which is largely conforming to ISO layers 1-3 was initially issued by CCITT in 1980 and is known as X.25 standard. It was revised by CCITT in 1984 to be compatible with the ISO network service (ISO 8348). It forms the basis for existing implementations of most European public packet switched networks (e.g. TRANSPAC, PSS, DATEX-P), but also for large US networks (e.g. GTE-Telenet) and for packet switching equipment available on the market. Many large companies and organisations operate a private X.25 network, using commercial switching equipment.

These X.25 network will also be capable to support the higher ISO layer standards, if they are used in accordance with ISO 8878, an additional ISO specification, defining how X.25 (84) can be used to support the ISO Network Service.

ISO standards of the transport layer (ISO 8072, 8073) and of the session layer (ISO 8326,8327) were issued in 1986. First implementations of these protocols by IBM and DEC are since about one year on the market, whereas most other major computer suppliers have not yet released relevant products. Announcements are however expected in the near future.

The ISO presentation layer standard (layer 6) was approved in April 1988 (ISO 8822, 8823).

The first ISO application service (layer 7) standard, released also in 1988, is FTAM (File Transfer, Access and Management) and is specified by ISO 8571/1-4.

Further layer 7 protocols followed in May and June 1988:

- basic class virtual terminal access (ISO 9040, 9041)
- basic class job transfer and manipulation (ISO 8831, 8832)

Approval of the interpersonal messaging (electronic mail) standard (DIS 9065, 9066) is expected to take place very soon.

First commercial implementations of standards released in 1988 can be expected to be available in one or two years time.

This means that communication between end-users of heterogeneous systems using a full ISO protocol stack will start to become feasible only in 1990 and following years.

It should here be noted that there exist CCITT recommendations for start/stop terminal access (X.3, X.28, X.29) which predate the definition of OSI. These recommendations, although widely supported and compatible with X.25 networks are incompatible with the ISO protocol stack and are expected to loose importance in the future. X.29

(version 84) can cooperate with the ISO network service and will therefore remain in use much longer.

Special attention should be paid to the X.400 series of recommendations (the electronic mail standard of CCITT issued in 1984). Several commercial products are now on the market, but CCITT and ISO have agreed that a revised version will be jointly published in 1988. The above referenced DIS 9065, 9066 specifications are based on this revised version. The current X.400 products may therefore have only a limited lifetime.

Important National Network Standards

There are two major national network protocol standards, which are still largely in use and have, to some extent, influenced the ISO standardisation process. These are TCP/IP in the US and Coloured Book protocols in the UK.

TCP/IP

This acronym stands for Transmission Control Program/Internet Protocol, referring to the two major layers defined by this protocol set, which was developed as part of the ARPA research network, interconnecting many US universities and research centres.

The Transmission Control Protocol (TCP) corresponds to the OSI transport layer (4) and the Internet Protocol (IP) to the OSI network layer (3). Communication via X.25 networks is supported.

There exists also a set of applications services, termed user level protocols, comprising:

- file transfer
- mail transfer
- virtual terminal

These user level protocols correspond principally to the OSI applications layer (7), but include also functions of the session and presentation layer which are not separately defined. TCP/IP offers also the ability to define and implement unique user level protocols.

It appeared that the standards for user level protocols, operating on top of TCP/IP leave room for interpretation. Implementations are often end-user or vendor specific and therefore not really compatible, and in most cases also not commercially supported.

Although the network was and is successfully used, it was decided some time ago that there will be no further development, in particular no refinement of higher layers, and that users will eventually migrate to the ISO protocols.

There are two circumstances, which make TCP/IP still interesting to new users:

a) TCP/IP is currenly the native network protocol standard of UNIX machines

b) TCP/IP supports Local Area Networks (LANs) efficiently, in addition to WANs (Wide Area Networks)

Coloured Book

This particular set of protocol standards, called after the distinctive cover of the different reference hand books, was developed and is still widely used by the UK academic community within JANET (Joined Academic NETwork).

The major books are listed below:

The Yellow Book defines a network independent transport service and the protocols to support it. It provides functions covered essentially by the ISO network and transport protocol.

The Green Book defines how to use CCITT recommendations X.3, X.28, X.29 for terminal access applications.

The Blue Book defines a versatile File Tansfer Protocol (FTP).

The Red Book describes the network independent Job Transfer and Manipulation Protocol (JTMP).

The Grey Book defines an electronic mail protocol.

The Fawn Book defines a protocol providing efficient operation of full screen interactive terminals over a packet switched network.

Other books concern local area network protocols.

The use of these Coloured Book specifications is now widespread throughout the UK academic community and there exist implementations on almost thirty different computer systems. The protocols are used in a number of other countries and are the basis of several international collaborations.

Now that the first phase of the definition of the ISO-OSI protocols is nearing completion, the UK academic community is planning transition to these OSI protocols. It must therefore be expected that the Coloured Book protocls will not be subject of further development.

Major Vendor Specific Communications Network Protocols

As long as the complete ISO protocol stack is not yet available, the established native communications protocol of the major computer suppliers will remain of great practical importance. But even after full availability of ISO compatible protocols it has to be assumed that products like DECnet from Digital Equipment Corporation (DEC) and SNA from IBM will be maintained. They are, therefore, here briefly reviewed.

IBM's RSCS networking software, which is used within EARN (European Academic Research Network), should also be mentioned in this context.

DECnet

DECnet is a family of software products enabling to interconnect two or more DIGITAL computers. The DIGITAL Network Architecture (DNA) comprises products which support all DEC computers:

- DECnet-11M and DECnet-11M Plus
- DECnet-VAX
- DECnet-20

DECnet has a layered architecture which has a certain similarity with the OSI-7 layers model, as shown below:

	OSI layer	DNA layer
	end user	end user
7.	Application	User Network
6.	Presentation	Application
5.	Session	Session Control
4.	Transport	End Communication (Network Services)
3.	Network	Routing
2.	Data Link	Data Link
1.	Physical	Physical Link

DECnet offers the following user services (applications):

- file access (incl. transfer)
- remote command file submission and execution
- remote-login

- downline system loading
- task-to-task communication

DECnet does not offer electronic mail. This service is available as operating system utility (e.g. VMS mail) or as part of an office support package (ALL-IN-1).

DECnet supports all kinds of communication circuits, including packet switching (X.25) networks and offers adaptive routing, which means that within a complex network a user defined "least cost" path will be selected and the ability to detect and automatically route around line or node failures exist.

DECnet offers also BSC connections to IBM computers or even an SNA interconnection (IBM 3270 cluster controller emulation and 3720 controller emulation).

Currently the implementation of DECnet has reached phase IV of DIGITAL's implementation plan. DEC has announced that the phase V implementation, which is expected in about two years, will offer full OSI compatibility.

IBM System Network Architecture (SNA)

This architecture defines a uniform set of commands, procedures, message formats and protocols used by SNA products to connect and communicate with one another, but does not specify the internal design of these products. SNA can operate under MVS and VM operating systems. SNA has a very strong hierarchical control structure.

A SNA network consists of several components, as illustrated by Fig. 2. The major components are:

- host processors controlling all or part of the network (e.g. S/370, 30XX processors, 4300 processors)
- distributed processors, providing functions similar to the host processors, except for network management (e.g. system /36, system /38)
- communication controllers managing the physical network (e.g. 3720, 3725)
- custer controller, controlling workstations (e.g. 3274)
- workstations (e.g. 3278, PC)
- SNA access methods, logically controlling the flow of data through the network, providing an interface between the application subsystems and the network and protecting applications systems from unauthorised access (e.g. ACF/VTAM)
- applications subsystems e.g.

IMS (Information Management System)

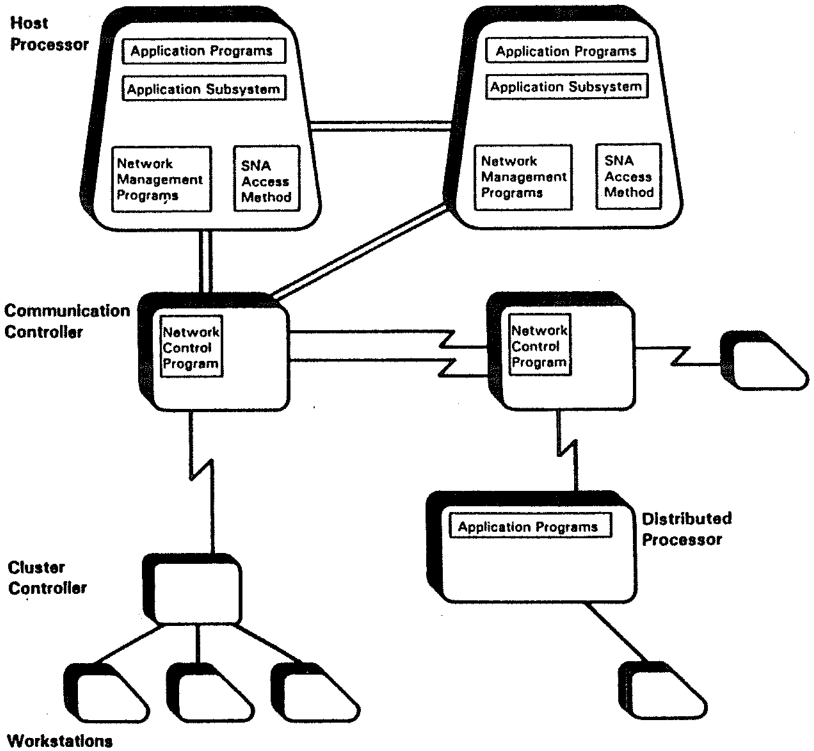


Figure 2 - SNA Network Components

CICS	(Customer Information Control System)
JES	(Job Entry System)
DISSOS	(Distributed Office Support System)
PROFS	(Professional Office System)
SNADS	(SNA Distribution Services)
DIA	(Document Interchange Architecture)
DCA	(Document Content Architecture)
Note:	The last three products are strictly speaking not application subsystems but architectures forming part of SNA.

- application programs which are user written and can be of any type
- network management programs, assisting network operations, detecting and reporting errors, maintaining network performance statistics (e.g. NetView)
- network control programs routing data and controlling its flow between communication controller and other network resources (e.g. ACF/NCP)

Network components are interconnected by links, which can be either of synchronous type (SDLC) or S/370 data channel. Asynchronous and BSC link are also still supported.

Conceptual elements of SNA are:

- end users, which are person at a workstation and the application program, the person is using
- nodes (distinguishing host nodes, communication controller nodes, peripheral nodes)
- transmission groups, which are links between adjacent subarea nodes
- subareas, consisting of a host or communications controller node and its peripheral node
- network addressable units (distinguishing Logical Unit (LU), Physical Unit (PU) and System Service Control Point (SSCP))
- domains, consisting of an SSCP and the network resources it can control
- sessions, which are logical connections that enable two network addressable units to communicate with each other
- routes, consisting of a service of nodes and the transmission groups that connect them.

It would go beyond the scope of this presentation to describe the concept and function of SNA in more detail. Although the concept of SNA is rather unique and therefore difficult to correlate with the OSI 7-layer-model, IBM does provide a description of the SNA in a layered organisation, which has similarities with the OSI model, and is shown by Fig. 3.

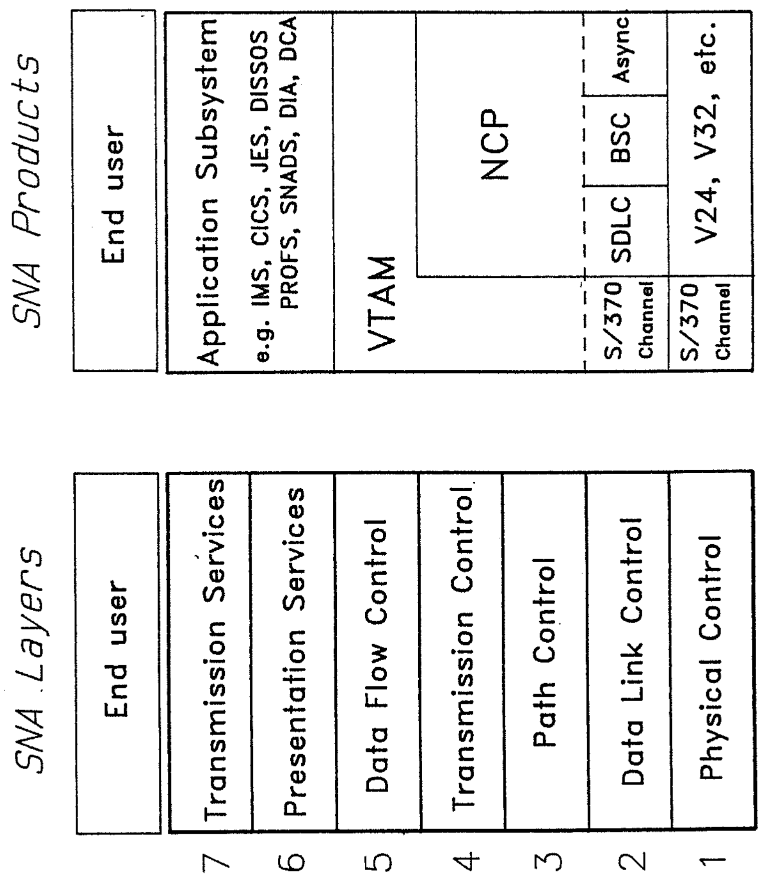


Figure 3 - SNA Layers and Products

It should however be noted that the functionality of the 7 SNA layers is, apart from layers 1 and 2, not really equivalent to those of the OSI layers. Fig. 3 also attempts to map SNA products to SNA layers.

Although IBM has started to release OSI products (e.g. the transport and session layer) and is involved in OSI related standardisation committees, the company has, at the current time, not made a firm commitment to OSI.

IBM-RSCS

RSCS is the abbreviation for "Remote Spooling Communications Subsystem Networking", a software forming the basis for communication within EARN (European Academic Research Network).

EARN consists of a set of independent host nodes, connected by means of public leased lines. A central computer in each country provides international connectivity and some central services.

EARN was sponsored by IBM until the end of 1987 and RSCS was originally developed for networking under IBM's VM/SP operating system (the JES2 and JES3 application subsystems now support RSCS also under MVS).

Emulation for this networking software is available for a variety of non-IBM systems, such as Siemens BS 2000, DEC VAX/VMS, CDC Cyber, UNIX. In fact, the majority of computers participating in EARN are non-IBM machines.

The RSCS software supports the following functions:

- remote printers
- remote workstation
- data and file transfer between remote systems

On the basis of RSCS protocols EARN provides the following services:

- file transfer (incl. program and document transfer)
- receive mail and send mail to one or more network users (mail and conference function)
- exchange immediate messages with people on the network (forward function)
- job submission and execution

EARN links some hundreds of computers in all major western European countries. Connection is also provided to USA (BITNET), Canada (NORTHNET), Japan, South East Asia and Australia.

The network is managed by a Board of Directors which is composed of academic representatives from each country. The network is, since the end of 1987, financed by the users.

A general policy for migration to ISO/OSI protocols has been adapted. A first step will be the transition from leased lines to public packet switched networks. The whole migration progress is expected to last several years.

Communications Protocols Used in Relevant ESA Networking Applications

Applications Supported

The European Space Agency (ESA) operates besides its ground station communication network ESTRACK, a second communications network, called ESANET, which supports ESA's general purpose applications, e.g. administrative communication, office communication, non-operational technical applications, etc.

ESANET connects the different ESA site locations, but also extends further to national space agency sites and to European aerospace industry. There exist also links to NASA (GSFC) and to IKI in Moscow.

The network currently provides data communication services, message (mail) services and facsimile services.

It futhermore supports IRS, the information retrieval service of the Agency operated at ESRIN in Frascati, Italy, and the European subnetwork of SPAN (Space Physics Analysis Network). The US part of SPAN is managed by NSSDC at the Goddard Space Flight Centre near Washington D.C., whereas the European branch is managed by ESOC. The European SPAN comprises 5 nodes which are permanently interconnected by means of leased lines and about 10 nodes directly addressable via public packet switched networks. More than 100 additional computers are reachable via a combination of permanent and switched circuits.

ESOC is using SPAN also as a tool to experiment with online dissemination of mission data products. Up to now, such products are generally distributed off-line, mainly on magnetic tape.

ESA has decided at the beginning of this year to establish new data services concerning cataloguing, archiving and retrieval of data from scientific satellites. This new services will be based on a distributed archive but centralised catalogue and directory services. The new distributed system, to be implemented and managed by ESRIN is named ESIS (European Space Information Service).

Protocols Applied

Most ESA general purpose applications are based on IBM compatible mainframes using MVS and VM operating systems. The network applications in this domain are consequently based on SNA.

For interconnection of office systems (PROFS-to-PROFS communication) RSCS is being used (intergrated into SNA). RSCS is also being used for the connection to EARN at ESOC, which is also accessible from the other ESA sites.

IRS access is still largely based on asynchronous protocols.

For remote VAX-to-VAX connections, DECnet is being used. DECnet is also the native SPAN protocol.

The majority of ESIS users are expected to be also SPAN users and will therefore access ESIS via SPAN, i.e. via DECnet. In order to increase ESIS access possibilities, gateways to EARN and to JANET (the UK academic network) are planned.

DECnet is furthermore planned to be used in the networks distributing data products from the first European Remote Sensing Satellite (ERS-1) and in the network providing remote access to EURECA experiment data.

Finally, the use of TCP/IP for interfacing to certain NASA applications is currently being studied.

ESA's networking policy is to migrate, whenever and wherever possible, to ISO/OSI protocols. It is hereby considered that this will probably not apply to the IBM computer subnet. Detailed migration strategies are currently not yet worked out, but one migration step will definitly involve the use of an internal X.25 network, currently being installed. It is furthermore expected that the use of DECnet phase V will constitute one migration path.

STANDARD FORMATTED DATA UNITS (SFDU)

A Technique for Interchange of Data

Dr C. Mazza

Division/EOS ESA

Robert-Bosch Str., 5 Darmstadt, FRG

Introduction

History has shown that scientific knowledge is acquired at a speed which is directly related to the amount of information that can be exchanged between scientists. Traditionally this was done through symposia, conferences, articles and specialised literature.

The age of space exploration combined with the enormous progress made in the domain of computers has opened unprecedented possibilities for the scientific communities in terms of availability of information.

However, the lack of standardisation whenever data are exchanged between heterogeneous environments is an obstacle to the full exploration of this huge potential. Moreover, data do not normally carry any information about their format and their representation. This information is usually available in Interface Control Documents normally not directly in a computer-interpretable representation. As the life of space borne data may be several decades the information concerning how to interpret the data needs to be available for the same period of time.

The Consultative Committee for Space Data Systems (CCSDS), an international committee founded by Space Agencies from various countries, has recognised these problems and formed a panel dealing with Standard Data Interchange Structures. This panel has defined a concept, called SFDU (Standard Formatted Data Unit) for the purpose of

interchanging data between heterogeneous computer and communication environments in a uniform and automated manner. The SFDU can be seen as a high-level service for the interchange of data at the level of the application layer (layer 7) of the ISO OSI model.

Standard Formatted Data Units (SFDU)

The SFDU concept is a data structuring technology and an operational approach proposed to solve the problem stated in the introduction.

It is to be noted that an unpleasant consequence of the fact that data are not self-describing is that each user ends up with re-writing in his own data handling software the description of the data.

The following definition is extracted from ref (1).

"The SFDU concept defines interprocess data objects whose formats are described in a standard way for ease of identification and interpretation. Each SFDU is an individual conceptual object, consisting of a standardised label and data content, which is sent from an "originator" to a "recipient"."

This concept is depicted in Figure 1.1.

To fully understand the SFDU concept, few definitions are needed and an explanation of the underlying principles.

When talking of a data object we need to distinguish between:

- its *format*, i.e. "the assignment of each data element of a data object to a field or sub-field and to a specific location or address on a given physical medium or in a device". Ref (2)
- the *semantic information* which is needed, in addition to the format, in order to be able to interpret the data object, i.e. the physical nature of the data, the engineering units in which they are expressed, the type of computer-representation chosen for the data (e.g. integer, real, floating-point, boolean)

In the SFDU concept both the format of a data unit and the semantic information needed to parse a data unit constitute a so-called *Data Description Record* (DDR).

A DDR is a set of statements expressed in a so-called *Data Descriptive Language* (DDL).

A DDL is "a formal notation for specifying the conceptual structure of data objects", i.e. their format and the related semantic information.

Instance of a data object, is a particular set of values of each of the data elements constituting a data object.

A necessary condition for being able to parse instances of a data object is to possess its DDR. If a DDR is expressed in computer-inter-

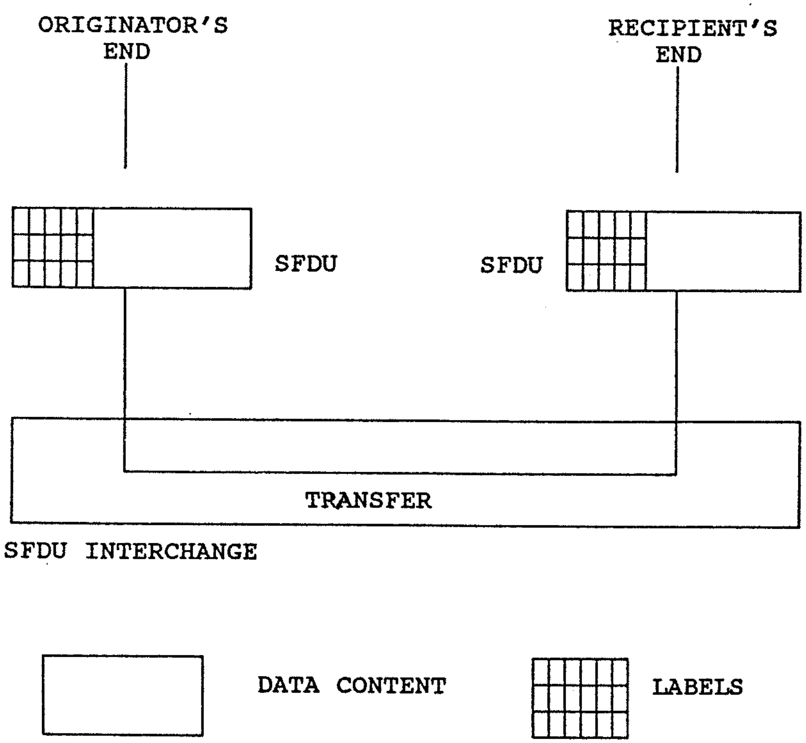


Figure 1.1

pretable DDL, then it becomes possible to (a) exchange between computers both the data instances and the related DDRs; (b) use the DDRs in order to automatically parse the data in an application program.

The SFDU concept involves registering with a *Control Authority* (CA) each DDR with a unique identifier. A CA is an organisational unit capable of registering, archiving and distributing DDRs upon request. Each instance of a data object in an SFDU environment should carry a reference to the corresponding DDR.

SFDU Structure

The basic structure chosen for the SFDU is called type-length-value or TLV encoding (Fig. 1.2). It is a technique recommended by ISO (International Standard Organisation) for data exchange (see e.g. ISO x 409).

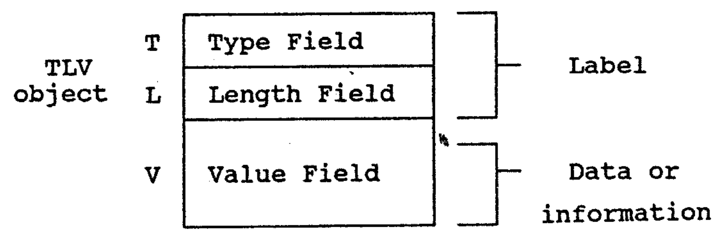


Figure 1.2

SFDU structures are therefore formed of TLV objects, the T and L fields of which constitute a label and the V field of which contains the data or information to be exchanged.

The following two sections are reproduced from section 2 and 3 of ref (2). They describe the field specifications of an SFDU and the construction rules to be applied when using SFDUs.

Structure and Field Specifications

"The basic SFDU structure is called Type-Length-Value or TLV encoding. This structure, comprising a TYPE, a LENGTH, and a VALUE field, is referred to as a TYPE-LENGTH-VALUE Object (TLVO) and is the fundamental structural element used to build the recommended SFDUs.

In this approach, data exchanged between open (independent) data systems are tagged with a TYPE identifier and a LENGTH indication, as shown in Figure 2-1. In SFDU usage, the specification of the TYPE (T) and LENGTH (L) field combination (TL field or label) is identified by an embedded version identifier at a fixed location in the TYPE field. A restricted character set of ASCII, denoted by RA, is the set of characters, allowed in the TYPE field. This set, totalling 36 characters, is comprised of the numeric characters 0-9 and the upper case Roman letters. Two versions are recommended and both have fixed lengths for both the TYPE and LENGTH fields. The TYPE field contains an identifier (ID) of the data descriptive record (DDR). The DDR contains the definition of the format and the parsing rules of the VALUE field. The TYPE field contains a global identifier referred to as ADI, which is comprised of the Control Authority ID and the DDR ID. The LENGTH field is interpreted as a numeric value that represents the length of the VALUE field in units of octets. While the TL field structure and representation are highly restricted by the Recommendation to only two versions, the VALUE field can be quite varied in terms of its internal structure and representation. A more detailed breakdown of the three fields is shown in Figure 2-2. ...

Type Field

Control Authority ID

The first sub-field (Octet 0-3) of the TYPE field is the Control Authority (CA) ID, which shows the organisational entity that has the registration responsibility for the DDR. The registration process es-

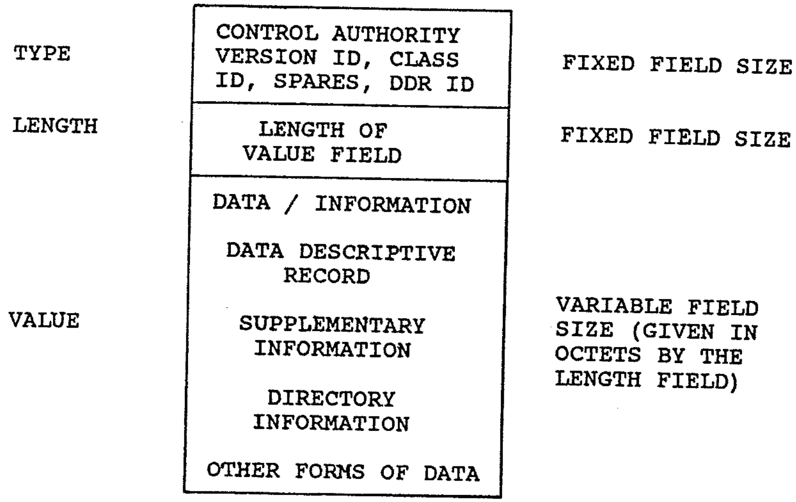


Figure 2-1 TLV Encoding Structure

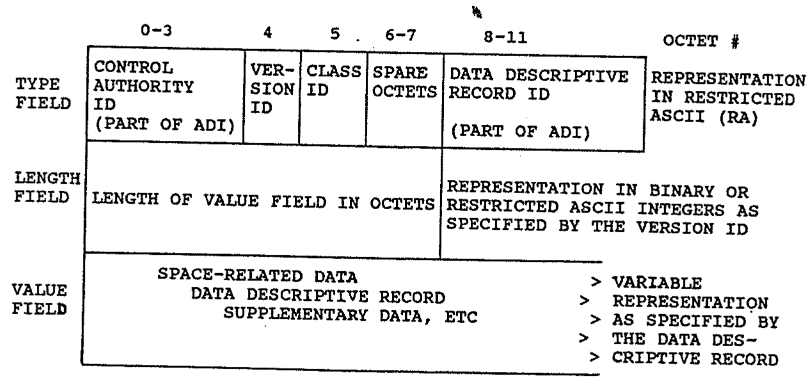


Figure 2-2 Basic Structure of the TLV Object.

tablishes a method by which each specified data object specification/definition (i.e. DDR) can be uniquely referenced... The only allowed instances of this sub-field are characters of the RA set.

Version ID

The second sub-field (Octet 4) identifies the structure (given in Figure 2-2) and coding of the label. There are currently two recommended versions. Version 1 (ID = RA character 1) defines the LENGTH field as an RA numeric character string. Version 2 (ID = RA character 2) defines the LENGTH field as unsigned binary.

Class ID

The third sub-field (Octet 5) is used to classify the VALUE field for the purpose of SFDU interpretation. The recommended values for the class ID are shown in Table 2.1.

Table 2.1. Class ID Instances

Class ID	Classification of VALUE Field
D	Data Descriptive Record
I	Data
S	Supplementary Information
Z	Comprised of one or more TLV Objects and/or SFDUs

Spare Octets

The fourth sub-field (Octets 6-7) is set to RA numeric characters 00 for Versions 1 and 2.

DDR ID

The last sub-field of the TYPE field (Octets 8-11) is used along with the CA ID to identify the DDR. The DDR ID sub-field consists of four RA characters.

Length Field

This field is used to specify the length of the VALUE field in octets and is represented in either the numeric character subset of RA (Version ID =1) or binary (Version ID =2). The eight octets specified for this field provide for lengths of 1 x 10^8 octets (version ID = 1) and 1 x 10^{19} octets (Version ID = 2). This field should always be completely filled, with leading zeros as necessary.

Value Field

This field contains various forms of data and information as shown in Figures 2.1 and 2.2. This variable length field can be in any desired code or representation that can be expressed with a DDR.

Construction Rules

The SFDU defined in this Recommendation is an aggregation of two or more TLV Objects. The first TLV Object shall always have the following unique TYPE field instance: **CA ID = CCSD; Version ID = 1 or 2; Class ID = Z; DDR ID = 0001** and shall be denoted hereinafter as a T_Z TYPE field. This TLV Object indicates that the SFDU follows the recommended construction rules given below.

Rule 1

The SFDU is composed of: (i) a T_Z TYPE field, (ii) a LENGTH field, and (iii) a VALUE field comprised either of a sequence of one or more TLV Objects with Class ID not equal to Z, denoted by T_{NZ}, or of a sequence of T_{NZ} Objects and SFDUs in any order.

This rule is expressed formally in Annex B and is illustrated in Figure 3.1.

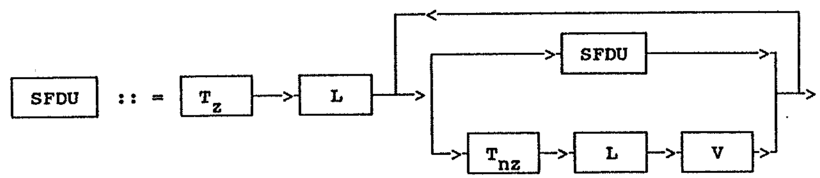


Figure 3-1 Diagrammatic Definition of the Recommended SFDU

T_Z is a Class Z TYPE field. $T_{\neg Z}$ is a Class non-Z TYPE field;
V is a VALUE field for a $T_{\neg Z}$ TYPE field and L is a LENGTH field.

Rule 2

If an SFDU contains a TLV Object with Class ID = D (denoted by TLVO(D)), this TLVO (D) must precede any TLV Object that requires this TLVO (D) for its interpretation.

Rule 3

The portion of the VALUE field of each TLV Object with Class ID =D which is used to specify the rules for an interpretation of a data object must be expressed in only one DDL. These interpretation rules for the same data object may also be expressed in other Class ID = D TLVOs with the same restrictions.

Processing SFDUs

Processing SFDUs involves the following five steps: data description, data registration, instance generation, transfer of data and interpretation of data.

These steps are described in some detail hereafter in a scenario typical of a space mission.

It is to be noted that a driver of the SFDU concept was the possibility of working in multi-national environments. Therefore, distinction is made in the SFDU literature between "closed" and "open" systems.

A **closed system** is considered to be "a system with its own private data formats and protocols which it uses internally and does not share on a broader basis".

An **open system** is one which uses publicly available formats and protocols so that... anyone can communicate with the "open" system by following the "open" system standards.

A space data system can be "opened" in various different points.The CCSDS has formulated a recommendation to encapsulate data in Standard Formatted Data Units to be made available by the system for the end users at the access points.

We assume now that someone, referred to as the "originator", has to define a space data system. Using the SFDU concept he would (Step 1) define the structure and the description of the data that the

system will handle. This is a necessary activity traditionally done with Interface Control Documents. Here, the originator would generate DDRs (Data Description Records) using a DDL (Data Description Language) to describe formats and semantic information of each data object.

The originator would then (Step 2) take the generated DDRs to an appropriate unit in his organisation called the Control Authority (CA) to register them. The CA, distinguished from other CAs by a unique identification number (CA ID) would then assign a unique identification number, the DDR ID, to each DDR. Both the CA ID and DDR ID become part of an SFDU header in T field (the ADI). Once the DDRs are registered and archived by the CA organisation they can be made available on request on an electronic medium (e.g. a communication line, or a magnetic tape or a diskette).

It is to be noted that DDRs can be transmitted either separately from or together with the data instances which they describe. Both possibilities exist in the SFDU concept.

Once the system is implemented and the data encoded according to the DDRs registered with the designated CA, the originator is ready to generate and deliver data instances (Step 3).

The data instances have to be packaged in the value fields of SFDUs according to the registered format and the necessary headers attached (type- and field-length). The instances are then ready for local use (within the system) or for transmission to a recipient user.

The transfer step (Step 4) is considered to be transparent to the SFDU process, insofar as this process is independent of the SFDU concept. Nevertheless, it is an essential step in order to be able to exchange SFDUs and appropriate protocols need to be established between the originator and the recipient according to the supporting media (storage and communication).

When recipient receives the SFDU via established protocols, he is then ready (Step 5) to interpret the data, i.e. to unpack them and to parse them. As he holds already or alternatively he receives with the data instances, the corresponding DDRs, he can incorporate these DDRs into his own data reduction software, thus achieving the primary goal of the SFDU concept, i.e. the automatic interpretation of the data.

Benefits and Applicability of the SFDU Concept

The SFDU concept was arrived at by starting from the statement of a problem, deriving from this a list of requirements and finally expressing a concept which fulfils these requirements.

It offers several benefits some of which are indicated hereafter.

- It enables an automatic exchange of space data in heterogeneous data processing environments.
- It reduces time and costs associated with having to accomodate several data formats only manually documented and in some cases not even properly documented.
- It allows the extension, practically indefinitely, of the useful life of data, as their description is stored electronically.
- It permits the automatic parsing of exchanged data.
- It simplifies the access to and the correlation of data from multiple sources and from different disciplines.
- It facilitates the management of large volumes of data.
- It permits the efficient establishment and preservation of audit trail information.

In space data systems the concept can be applied in various areas, e.g. typically in:

- ground stations (telemetry and telecommands)
- control centres
- payload operation centres
- data archives
- mission support systems
- remote user facilities
- telescience operations

Earth Observation Data. How to Apply the SFDU Concept

The Committee on Earth Observation Satellites (COES) has formed the so-called Working Group on Data (WGD), the mandate of which is to co-ordinate Data Management Issues related to earth observations with particular emphasis in the area of standardisation of user product formats, standardisation of new media formats, standardisation of catalogue and directory related information, archival practices and network communications.

Standards for data from various types of earth observation instruments (e.g. radiometers, scatterometers, radar altimeters) are being defined by this group.

As the work of the CCSDS on SFDUs is relevant to them an effort is being conducted at International level from one side in order to introduce and promote within GEOS the SFDU concept and from the other side to investigate within the CCSDS further data structures and tech-

niques which would allow without major difficulties the encapsulation of existing standardised data formats into Standard Formatted Data Units.

References

1. *CCSDS 610.0-G-5 – Space Data Systems Operations with Standard Formatted Data Units*: System and Implementation Aspects. "Green Book", Issue-5, February 1988

2. *CCSDS 620.0-B-1 – Standard Formatted Data Units: Structure and Construction Rules*. "Blue Book", Issue-1, February 1988.

THE netCDF SELF-DESCRIBING, PORTABLE FILES – A BASIS FOR "PLUG-COMPATIBLE" SOFTWARE MODULES CONNECTABLE BY NETWORKS

David W. Fulker

Unidata Program Center
University Corporation for Atmospheric Research
Boulder, Colorado, USA

Motivation

Accessibility and Usability of Data

Central to many scientific endeavors is the need to access data from a variety of sources. The Unidata Program was established, among several purposes, to improve the accessibility and usability of data for university research and education in the atmospheric and related sciences. As one part of this effort, a set of data structuring standards, associated access software, and usage conventions, known collectively as the Network Common Data Format (netCDF), has been developed, drawing upon a related scheme employed at the National Aeronautics and Space Administration (NASA). The National Space Science Data Center at NASA has demonstrated how such a scheme enhances the usability of data, especially across the traditional discipline boundaries.

This paper focuses on the Unidata design for a netCDF-based library of software suitable for the modern environment of networks and computers, in which desk-top workstations are becoming commonplace, the appetite for scientific data is growing at a rapid rate, and technologies for transporting, replicating, and storing data are improving radically.

Reusability of Scientific Data Processing Software

Another objective of the Unidata Program is to facilitate interactive analysis and display of atmosphere-related data, especially on small computers, practical for use in the classroom or the office. Because of the unpredictable nature of research-oriented computing, Unidata provides more than a monolithic, turnkey, software system; instead, modules are offered that may be combined with software developed by the universities, and even by their studends. The Unidata modules are intended to serve as templates for university-developed modules, and the design strategy emphasizes reusability of software.

Reusable software modules are especially valuable because they can reduce the repetitive programming of intrinsic, common tasks required for new computer applications. This principle has been employed with great success for a variety of purposes, but the management of scientific data has not generally been among them. An important exception is a system developed at the New Mexico Institute of Technology, on which certain Unidata software modules are based.

Self-Describing Files and Interdisciplinary Research

Integrating Data from Diverse Sources

As important geophysical research questions become increasingly interdisciplinary, a number of difficult data management issues must be carefully considered. Prominent among these is the handling of ancillary data, in which we include descriptions of measured parameters, units of measure, time of recording, coordinate systems, structure and organization, data types, and all other information required to make effective scientific use of a data collection.

Traditionally, data are recorded separately from their related ancillary data. In fact, ancillary data may be available only in printed form or, in the worst case, ancillary data are buried deep inside computer programs used to analyze the measured data. The difficulties in obtaining and using ancillary data across discipline boundaries are greatly exaggerated: analysis programs for one discipline may be ill-suited for another, conventions may differ with regard to coordinate systems, units of measure, etc.; searching of discipline-specific printed literature for information on data management conventions is likely to be frustrating.

Data Structures with Dimensionality

Common to all areas of scientific computing is the use of data structures having dimensionality greater than zero. Typically, these take the form of vectors or matrices (grids) where the elements along a given dimension correspond to discrete values of a well defined coordinate system. This characteristic is, perhaps, the single most important reason why the large body of commercial database management software is poorly suited for use with scientific data.

Any scheme that purports to improve the usability of scientific data must permit effective use of multidimensional data structures, including their characterization, in quantitative fashion, among ancillary data that can be interpreted easily by computers.

NASA's Common Data Format

An important and obvious (though difficult) improvement to the usability of data across discipline boundaries can be achieved through the use of self-describing files. If a large proportion of the ancillary information accompanies the associated measured data in a single file, then interpretation is substantially simplified for the researcher who has had no prior experience with the measurements. In principal, obtaining the file is all a researcher need do.

An important application of this principle has been demonstrated (ref) by the National Space Science Data Center at the Goddard Space Flight Center, a NASA organization in Greenbelt, Maryland. By storing a large collection of data, from very diverse sources, in a scheme designated the Common Data Format (CDF), NASA has proven that a collection of highly generic analysis and display functions can be applied effectively to data that are not ordinarily compared or integrated. [ref]

The CDF facilitates the storage of data as files that are self-describing and that contain collections of related, multidimensional variables and parameters. Sufficient ancillary data are encoded within a CDF file to create a variety of well annotated displays, as well as to perform a variety of "data-driven" processing functions. Among these are remapping and algebraic functions that can be used to correlate data sets measured on different coordinate systems or that have other distinguishing characteristics.

CDF Data Access Functions

The effectiveness of NASA's CDF design arises from an abstraction that is sufficiently general to encompass a wide range of scientific data but that is sufficiently narrow to permit practical realization. The abstraction is manifest in a rather small collection of data access functions (13 subroutines) that provide all the necessary services to build a CDF file and retrieve elements therefrom. In the NASA implementation, these data access functions are generally used indirectly via a separate software layer designated the Virtual Data Table, which is beyond the scope of this paper.

Chained Software Modules for Processing Text

Much Software Is Developed Repeatedly

Every software programmer has experienced coding and recoding certain segments of software that are common to many different programs. A common need is to search for a given string of text and replace it with another. More generally, one needs to search for a pattern of text and replace it with a text string that may contain substrings from the pattern being replaced.

The world of text processing, ranging from compilers to word processors, is replete with examples of intrinsic tasks that arise repeatedly during software development. Throughout much of the history of programming, many intrinsic needs have been met by developing software repeatedly, or almost repeatedly, with minor variations.

Subroutines/Functions Have Confining Interfaces and Uses

The most common antidote (and a very effective one) for repetitious software development is the use of function and subroutine "libraries." Such libraries perform generic tasks and may be employed by subroutine or function "calls" from the common programming languages, using argument lists or "bindings" defined by the author of the library.

In general, however, the interfaces to and the uses of subroutine and function libraries are somewhat confining, for two reasons. First, it is customary for relatively small amounts of information to pass between a program and the subroutines or functions it employs; specifically, arguments usually consist of a few program variables and constant values; whole files are seldom used as arguments, though they could be. Second, subroutines and functions can be used only within a conventional, compiled programming language, such as FORTRAN or C; they cannot be used directly by the user from the command language.

A major difficulty in using subroutines and functions for managing scientific data is the requirement for effective use of the computer's file system. In general, the best tools for working with the file system are interactive and are used from the command language; they are often difficult or impossible to use from compiled programming languages. This is the main reason that subroutine and function arguments seldom reference whole files.

UNIX "Pipes and Filters" Utilize the File System

In contrast to subroutines and functions, procedures that are invoked from the command language almost always include whole files among the arguments. Furthermore, the command language typically supports effective use of the file system, permitting, for example, the use of "wild card symbols" to generate argument lists that include all files whose names match a certain pattern.

Of particular interest are the UNIX "shell" command languages. Under UNIX, the typical procedure employs two implicit files, designated the standard input and the standard output. The UNIX shell permits the chaining together of procedures, using a "pipe" that connects the standard output of one procedure to the standard input of the next. Each procedure in the chain is called a "filter" that typically has a small number of arguments to control its behaviour. The pipe and filter paradigm is a highly intuitive one for users who wish to create a sequence of processing tasks, say from a raw data source to an output display.

Pipes and Filters Have Proven Highly Reusable

The UNIX pipe and filter paradigm has proved especially effective for a variety of text processing activities, including the development and interpretation of programming languages. Crucial to this success is that the basic UNIX filters have proved to be abundantly reusable. There are filters that merge, sort, edit, select, translate, and perform a large variety of intrinsic text processing functions. These may be chained together in many sequences and with various arguments, creating a very large collection of capabilities without repetitious software development.

Underlying the reusability of UNIX software is an implicit set of rules for the text files that are piped from one filter to the next. Some filters (a "C" compiler or a "Tex" formatter, e.g.) require highly specific input forms; others (the "sed" stream editor, the "grep" line selector, and "sort," e.g.) can except any file containing strings of characters terminated by end-of-line markers.

"Plug-Compatible" Software for Scientific Data

New Mexico Tech-Scientific Data Through Pipes and Filters

Often, scientific data processing may be described concisely as a sequence of processing stages, starting, for example, with the merging and/or subsetting of raw data, proceeding with transformations, statistical calculations, and algebraic computations, and concluding with the generation of graphical displays. Thus, the UNIX pipeline and filter paradigm is as intuitive for scientific data management and processing as it is for text processing. This similarity has been exploited in a successful data processing system developed at the New Mexico Institute of Technology.

The New Mexico Tech system is titled Candis (for Common Data Analysis and Display System) and provides scientific software modules that may be chained together to create an impressive collection of data management, processing, and graphic display capabilities. Candis runs under a UNIX shell and employs the same "pipe" features used for text processing and other UNIX programs.

Unlike the implicit and rather loosely defined rules for text files that are piped from one filter to the next under UNIX, Candis employs a specific file structure (also called the common data format, though it differs from the NASA version) as the standard input and output form.

Unidata Variation: "Plug-Compatibility" Based on the netCDF

Unidata is developing a collection of "plug-compatible" software modules that draw heavily upon the New Mexico Tech concept. However, the Unidata design differs in three primary ways from the Candis system: 1) the Unidata standard input/output form, called the Network Common Data Format (netCDF), is based on the NASA CDF, though it differs in ways that are discussed in a subsequent section; 2) because Unidata software is required to operate under DECs VMS operating system (in addition to UNIX) the Unidata system will not rely upon a UNIX "shell" and the associated "pipes" to achieve chaining; 3) the Unidata modules will be connectable by networks as discussed below.

Similarities to the UNIX Paradigm

Both the Unidata and the Candis modules may be invoked directly from the command language. Just as in the UNIX paradigm, most modules expect their inputs and create their outputs as whole files, similar to the UNIX standard input and standard output files. Some modules, especially those (such as merge utilities) supporting several input files, accept multiple file names as arguments. Other arguments are used as flags and parameters to control processing, i.e. filtering, functionality.

Distinctions from the UNIX Paradigm

As mentioned previously Unidata has adopted the netCDF file structure as the standard input and output form for all modules except those that take raw data for input or that produce print or graphic output. This standardization is essential because of the complexity and variety of data that must be passed between modules in a full-blown scientific processing model. In essence, the netCDF serves as a "mechanical and electrical standard" for "plug-compatible" software modules.

Such an interface standard provides ready access to ancillary data, because the common data format defines self-describing files. This is necessary because many of the Candis and planned Unidata modules place limitations on the character of their input files. The access to ancillary data permits immediate conformance checking of the input data.

Another distinction from the UNIX paradigm concerns parallel processing. Under UNIX, the "shell" realizes a sequence of pipes and filters by creating a number of parallel tasks, one for each filter in the pipeline. (Typically, these tasks are not strictly parallel—they usually operate in time-shared mode on most UNIX computers.) Because lines of text are inherently sequential, this form of parallelism is practical: in essence, each filter passes along one line of text to the next filter as soon as the processing of that line is complete. The netCDF supports data structures that are not inherently sequential; therefore, it may be that the Unidata "plug-compatible" software modules will operate

strictly in sequential fashion. (This design decision has yet to be finalized. The Candis system does achieve a measure of parallelism, relying heavily on UNIX pipes for doing so.)

Elemental Data Management and Processing Modules

Working in collaboration with New Mexico Tech, Unidata plans to release a collection of plug-compatible, elemental, scientific data management and processing modules in the fall of 1988. Individual descriptions are beyond the scope of this paper, but the elemental modules to be offered generally fall into six categories:

Selectors
Selectors choose items from the input netCDF file and use them to create a new netCDF file as ouput. Depending on the module and the method of selection, the output file may simply be smaller, or it may have lower dimension than the input file. For example, one module reduces dimensionality by creating a cross section.

Constructors
Constructors merge and otherwise build netCDF files from existing ones. Depending on the module and the character of the input, the output file may simply be larger or it may be of higher dimension than the input file. For example, one module can be used to create a three dimensional cube from several input files, each containing two dimensional slices.

General Utilities
There are modules to perform a variety of format translations on and to print information about netCDF files. Eventually, this category may include a variety of operators for indexing, sorting, and transposing the contents of netCDF files.

Graphics Generators
Graphics generators create displays, including graphs, contour plots, vector plots, etc.,, that represent the contents of netCDF files.

Mathematical Operators
Mathematical operators perform statistical, algebraic, and other calculations on variables contained in the input netCDF file. In some cases (such as the smoothing operator) the output file has the same variables as the input and, in other cases (such as the partial derivative operator), new variables are added to the output file.

Specialized Data Convertors
This is a catch-all category that encompasses a variety of discipline- and data-specific functions. Among the capabilities will be those for converting a variety of data from their archive formats to netCDF files.

It is intended that users of the Unidata system will use these elemental modules as templates for developing their own "plug-compatible" software. Furthermore, Unidata expects to adopt much of this user-developed software for inclusion in the central software library.

Networking of Plug-Compatible, Scientific Software

"Distributed" Modules and Files

An key aspect of the Unidata plan for plug-compatible software is the network model in which such software is envisioned for use. In essence, the modules and files of a processing chain can be distributed among several computers interconnected by a suitable network. This means that the modules of a process may run on different computers, and the files of a process may reside at several points on the network. Furthermore, it is envisioned that files may be mounted from portable storage media, such as compact disks, conventional tapes, and other devices.

This model is feasible because Unidata has adopted a networking standard, which is based on the widely used TCP/IP protocols developed by the Defense Advanced Research Projects Agency (DARPA). These protocols are used both for long haul networking between Unidata sites and for local area networks supporting general-purpose scientific workstations and specialized "servers." As it becomes practical to do so, Unidata will make a transition to corresponding protocols adopted by the International Standards Organization (ISO).

Realizing the Model on a Network of Diverse Computers

This distributed model is difficult to realize because of the diversity of computers that Unidata intends to support and that may be interconnected via the DARPA and ISO protocols. In particular, the self-documenting (netCDF) files, by which modules are chained together,

must not be computer-specific; their structures and data values must survive network-file-transfer with integrity.

Furthermore, the netCDF must be sufficiently flexible to encompass the wide range of file sizes and numerical representations likely to be encountered among diverse computers. In general, it is practical only to convey relatively small data units via network while extremely large ones can be conveyed via tapes, compact disks, etc.

Sun Microsystems' External Data Representation (XDR)

Fortunately, Sun Microsystems has developed an effective means of representing complex data structures for network transfer between dissimilar computers, as required for the Unidata processing model. The Sun technique is designated the External Data Representation (XDR), it may be employed with a variety of protocols, including DARPA and ISO, and it is becoming well accepted by the international networking community. The XDR may also be used for data structures that are stored on portable media such as tapes and compact disks.

Even though the XDR serves as a standardized, computer-independent format, it is reasonably compact. If necessary, there are various ways that compaction schemes can be employed in conjunction with the XDR.

Unidata's Network Common Data Format (netCDF)

The XDR is actually a rather complete data description language. Therefore, it is well suited to describing data abstractions such as the NASA Common Data Format (CDF). In fact, the Unidata netCDF is nearly equivalent to the CDF, represented in the XDR form. The Unidata netCDF access software is similar to the 13 CDF access routines developed by NASA' though Unidata has bindings for C as well as FORTRAN.

Within Unidata, the principal purpose of the netCDF is to define a "plug-compatible" interface between elemental data management and processing modules. Because the netCDF employs XDR, the netCDF allows modules on diverse computer systems to connect with one another across a computer network and/or to employ data stored on portable media. Though agreement has yet to be reached, the netCDF is intended to be suitable for adoption by NASA – if it is, then Unidata and NASA could begin to establish an important standard for the transport and storage of geophysical and other scientific data.

Below are listed the principal characteristics of netCDF files:

Self-Describing
Unidata's netCDF files are self-describing with respect to dimensionality, parameters, units, data types, and so forth.
Portable
By virtue of the XDR standard, netCDF files are portable among a diverse collection of networked computers.
Interface Standard for Plug-Compatibility
The generality and versatility of the netCDF permits its use as an interface standard between the "plug-compatible" modules of a complete processing chain.
Access Via Succinct Subroutine Bindings
Creating and retrieving data from netCDF files is accomplished via a succinct set of subroutine bindings that are very similar those of NASA's 13 CDF subroutines.

In practice, employment of the netCDF will incorporate most of the New Mexico Tech conventions for the usage of Candis files. Of special note is the processing history generated by Candis: each software module appends a line of text to a "history" record (contained within the file, of course) that documents the processing step performed by the module, including the settings of the module's arguments. Thus, as planned for Unidata, every netCDF will contain a history of the processing steps that led to its creation

Conclusion

New Data Processing Models Are Being Demonstrated

Efforts at NASA and New Mexico Tech have led to new abstractions and models for managing, processing, and displaying scientific data. These techniques hold promise for improving the usability of scientific data holdings and for increasing the reusability of associated software. Unidata is developing a system that combines concepts from the NASA and New Mexico Tech efforts with a network processing paradigm.

The Unidata endeavor employs networking and data representation standards that are particularly relevant to the focus of this workshop session, "Compatibility Problems: Computer-to-Computer Communications and Software Transportability." In particular, the Network Common Data Format (netCDF) will serve as the basis for the chaining together of elemental software modules into complex processes. These modules, and the files on which they act, may be distributed among diverse computers that are linked by standard-conforming networks.

Data "Centers" May Play New Roles

Potentially, the Unidata network processing model could be extended to encompass data centers throughout the world. In essence, the first few modules of a processing sequence, and the files on which they act, could reside on computers at the data center, and subsequent modules and associated (typically, much smaller) files could reside on workstations and other computers proximate to the researcher.

Additionally, the Unidata model is compatible with centers performing "publication" (i.e., bulk duplication on compact disks or other media) of data collections for distribution to research organizations. In this case, we suggest that published data be accompanied by software modules suitable for use as the first few elements of a processing chain. For example, such software could perform subsetting and organize the output as a netCDF file.

In this way, common data processing functions could be performed on a large, diverse class of computing systems by using a relatively small collection of elemental software modules. This software could be optimized for transportability and would serve to facilitate the interchange and use of scientific data among researchers (in various disciplines on a global scale) whether by network or other medium.

COMPATIBILITY PROBLEMS: COMPUTER-TO-COMPUTER COMMUNICATIONS AND SOFTWARE TRANSPORTABILITY

Carl C. Abston

World Data Center A

Boulder, CO. USA

Computer-To-Computer Communications

Computer-to-computer communications is a very wide-ranging field. In this discussion, I will not try to do an in-depth analysis of all types of communications but rather limit this discussion to items of interest to the World Data Center System.

Asynchronous Communications

Terminal asynchronous communications have been used since the early 1970's. The present communications between data centers via the EDS time sharing network uses this type of communications. It is slow, error prone, and limited in its capabilities, however it works and will continue to be used for a long time to come.

RJE Synchronous Communications

Beginning in the late 1960's or early 1970's, RJE (Remote Job Entry) communications were created. This type of communications allowed an RJE work station to communicate with a mainframe computer. Batch jobs (card decks) could be submitted remotely (sometimes from different geographical areas) to a mainframe computer and have the resulting printouts transmitted to the RJE station for printing. There were a num-

ber of these synchronous protocols such as 2780, 3780, bi-synch, NTR, WS200, etc.

At WDC-A in Boulder, we are, in fact, still using a 3780 protocol in an RJE mode to communicate with a UNISYS mainframe computer in Asheville, NC, via a 9600 baud dedicated land line. The 3780 software and hardware board (made by AST) is installed on a PC XT. The XT is also connected to an LAN. A user submits a job from his local PC via TCP/IP to the VAX. The VAX transmits the file via LAN to the XT. The XT transmits the file to a UNISYS DCP-40 which goes to a modem, land line, DCP-40 in Asheville, and finally into the UNISYS mainframe computer job queue.

The return procedure is similar. A print file generated on the Unisys computer is moved to the DCP-40, then transmitted via a modem and DEDICATED land circuit to a DCP-40 in Boulder which intern moves the file to a PC XT. Every 5 minutes a demon running on the VAX-750 looks for print files on the XT and moves them into the VAX Print Queue and finally the file is printed on the VAX impact printer. This system works reasonably well and most of the complications are transparent to the user. However, it is a complicated and very limited method of communication.

Local Area Network

Concept

Local Area Networks (LAN) require a rather different concept in thinking. In fact, it may be useful when thinking about LAN's to forget most of what you presently know about computers and communications. In the past, we have thought of a computer system mostly as a central processing system, or more specifically as a CPU with peripherals and terminals added to the CPU and communications as point-to-point exchanges.

LAN's are a very different approach. You first begin with the heart of the system, a rather unimpressive looking coax cable. To this cable you add taps (connections) to each computer that has a resource to be shared. Each of these computers now becomes a server node on a network. Each server may specialize in something. For example, one server may have two tape drives, another lots of disk storage, printers, optical drives, another one may hold multiply accessed data bases, another may be a bridge connecting two LAN's, or a router or gateway between geographically separated LAN's, and so on. Now add all of the user workstation PC's to this network and you have created an operational LAN where every PC workstation has access to every server node of the network, including all peripherals. As requirements change, this modular system network can be reconfigured to setisfy those new requirements.

In the older computer systems, as new users are added, the present users loose performance because the computer systems have a fixed amount of computing power. Therefore, as more users are added, the fixed computing power must be divided by more tasks. In the LAN approach, more PC's are added to the network, each with their own CPU. Therefore the existing users are not necessarily affected. If the new users require a lot of tape drive time, then another node with tape drives may be added. If the networks become congested with data transfers, that network can be split into two or more networks and bridged together. The network and nodes on that network can be changed to match the frequently changing user requirements, whereas the more conventional system requires rather drastic action, such as replacing the entire system with a newer and bigger computer to adjust to a changing user climate.

Ethernet LAN

Several different types of LAN's are used. IBM's token ring is popular in the IBM mainframe computing centers. Other centers are using Arcnet, Gateway, G-net, Omninet, Plan2000, Starlan, Pronet, PCnet, Cluster, Ethernet and many others. However, Ethernet Local Area Network (LAN) type connections are perhaps the most widely used in multivendor computer-to-computer communication systems today. This is not to say that Ethernet LAN's are the only satisfactory method nor even the best method for all situations, rather a method which seems well suited to Data Center problems and their multivendor hardware systems.

To understand a little more about Ethernet LAN's, let's look at what is required to create an Ethernet LAN system. Each computer contains an Ethernet controller card. The card taps into a coax cable. Ethernet controller cards are available for a wide range of PCs ($400), minicomputers ($3000) and some mainframe computers. Each computer also needs a copy of a protocol with TCP/IP being the most widely used outside of large IBM facilities. IBM also supports the SNA communication method and third party vendors are now supporting SNA to TCP/IP bridges.

By using Ethernet and TCP/IP, it is possible to set up multiple LANs and link them together using a $2500 bridge. This approach is used when traffic begins to slow down transfer rates. Ethernet is a broadcast protocol. Every node on the network reads every message. If the message contains the correct address that node responds, otherwise the message is disregarded on that computer. When a bridge is present in a system, it reads all messages from both LANs, dynamically learns which addresses belong to which LAN and restricts messages not directed to the other LAN. This can greatly reduce the load on heavily used LANs without reducing capabilities. In other words, every user on either LAN still has access to every node and both LAN's.

One potential problem with LANs is that an Ethernet LAN segment is limited to a 1000 foot communications range. This can be somewhat solved by adding a $3000 multiport repeater to the net. It then permits up to 8 LAN 1000 foot legs to be connected creating an 8000 foot network. Up to two repeaters can be present between any two nodes in an Ethernet network.

To connect LANs at geographically separated data centers, a gateway or Router is added to the network in each location. This permits LAN messages to be transmitted via a modem and land or satellite communications circuit. Although the response is limited by the speed of the communications circuit, the capabilities are the same as using a local node. Routers are protocol sensitive, therefore an IP Router will only understand and process TCP/IP messages. Routers are available for Decnet and XNS (used by Novell). Gateways, however, are not protocol sensitive but they are slower and more expensive.

TCP/IP

The TCP/IP protocol, however, only allows file transfers not intelligent use of a node.

NFS

The Sun NFS system rides on top of the TCP/IP basic protocol and permits the disc of a Unix based VAX or a Unix based SUN computer to look and behave like a standard PC DOS local drive. FTP Systems is also developing an NFS system which permits a PC to be used by another PC in an intelligent mode.

Novell

Novell is a multi-user, multitasking operating system using XNS as a LAN Protocol. Novell supports most of the different types of LAN's, Protocols, and hardware controllers. For example, all of systems listed on the previous page are supported. It is a very popular and powerful system today.

After the Novell operating system has been installed on a server PC, all PC's in the network will need a piece of software to communicate with the server. This method allows any user PC, intelligent access to the server. For example, the server disc may be F:, therefore any PC can access the server disc by entering F:. A program run on a user PC can use the server disc to read or write data. Also the server permits many users simultaneous access to the exact same data base. This system is used at WDC-A in Boulder to support updates and searches for both the Data Request and Mail List systems.

The Novell system will responsively support a large number of users depending on the application and the hardware involved. When a user's PC accesses the Novell server, the server provides only disc access. The user's PC actually runs the application software which extensively uses the local PC memory and CPU and allows the server's CPU to be almost idle.

However, it is also necessary to point out some limitations. For example, not all peripheral devices are presently supported by this system. Which peripherals are and are not supported changes by the week.

Software Transportability

Software Transportability has long been a dream of software developers and applications people alike. However, for the most part, it has remained a dream.

During the last eight years we have seen an enormous amount of commercial software development which is widely used throughout the world. Why is this possible? Because the IBM type XT and AT personal computers have become truly world machines. Commercial programs designed for Word Processing, Accounting, Telecommunications and Graphics are very popular and widely used.

In the data centers, we have long discussed common software, but not too seriously. The problems have always been related to the fact that we have very different computer hardware in each of the World Data Centers.

Computer Hardware

Although it is possible to write transportable software for a variety of hardware systems, it is not very practical. For example, if a piece of

software is designed to use a floating point processor but is run on a machine without such a processor, the program either doesn't run at all, or runs very slowly. On the other hand, if you actually need a floating point processor but write the software to avoid using one, it will run very slowly on a computer which has the processor. The same thing is true with input-output, and so on. In other words, to handle various hardware systems, you must design the software to run on the system with the lowest common denominator and, therefore, the software is overly complicated, difficult to maintain, runs slowly and really does not satisfy anyone. Therefore, it is an impractical approach.

The IBM AT-type personal computer improves the situation. Software developed for the IBM PC will run on any IBM PC-type computer no matter where in the world it is located. Perhaps, for the first time, we have an opportunity to effectively share some type of software between centers.

Software

Different computer languages are used at different data centers. For example, Cobol was used extensively at some of the WDC-A installations but not used at the WDC-B and WDC-C centers. Fortran-77 is extensively used at the WDC-A's. The C language is extensively used at the WDC-A's in Boulder, but may not be generally used elsewhere. Even Basic is used at some centers. In the past no compatible hardware existed among all data centers.

To exchange software meant using the lowest common denominator for a software language, which was low level Fortran. This language is clearly limited in its capabilities and runs very slowly on many systems.

However, with the widespread use of IBM-type AT personal computers, a lot of problems with common software disappear. Every data center in the world probably has at least one IBM-type PC. A program written in any language can be compiled at the originating center and the executable module distributed to all other centers. Updates could be send via the mailbox and telecommunications system. A center does not need to have a compiler for that language or even understand the language of the distributed software. Therefore, a piece of software written in any data center could be effectively used in any other data center which has a similar requirement. This really is the first opportunity that software exchanges may become practical.

The following is an example of a possible practical application in common software. In 1984, WDC-A for STP introduced a joint geomagnetic catalog identifying the geomagnetic collections at all the World

Data Centers including WDC-A Boulder, WDC-B Moscow, WDC-C1 Copenhagen and Edinburgh, WDC-C2 Kyoto and Bombay. The program which created this joint catalog is in daily use at WDC-A in Boulder running on an IBM type PC AT. This same software could be used by any center to update their geomagnetic collections. The data bases are exchanged every few months. The total size of the data base is 2 million characters, however, the monthly updates are small enough to be exchanged via telecommunications. Every data center handling geomagnetic data is using some method to keep track of these data. Why not use the same software everywhere running on similar PC hardware?

Each center, of course, has variations of the collected elements to be stored and retrieved, format, etc. However, a short conference among representatives from Moscow, Boulder, Edinburgh, Copenhagen, and Kyoto could probably arrive at a common format and procedures in a fairly short time.

Currently, each center is developing all of its own software for data and information processing and handling and several centers handle exactly the same data. From a systems point of view, this method is not very efficient. The number of staff hours available for program development is very limited in each center. It would be much more effective from a theoretical point of view to have one center develop an inventory system for geomagnetics, another center develop a system for ionospheric data, and so on. Then the executable modules could be exchanged among all centers. This approach could more effectively utilize the available labor and expertise, however, it requires that each center one piece of common hardware (an IBM AT compatible) and the desire to use this approach.

Summary

A lot of options are available for computer-to-computer communications. Before seriously discussing one method versus another, the more basic question is a definition of the problem.

What do the World Data Centers wish to do with computer-to-computer communications?

The same time is true for software exchanges. What do the centers wish to do?

COMMUNICATIONS AMONG DATA AND SCIENCE CENTERS

James L. Green

National Space Science Data Center

Greenbelt, MD 20771

Abstract

Computer-to-computer communications networks are being used extensively by World Data Centers and other national and international scientific data and information systems. These networks are providing the means to bring users at science centers and operating staff at data centers closer together. This paper is not meant to be an exhaustive study on network usages, but will provide examples of how various data and science centers are interacting with each other over the Space Physics Analysis Network, or SPAN. It is important to note that there is extensive use by data centers of many other networks besides SPAN, but the following is believed to be fairly representative of those type of uses. Network access by remote scientists allows, for example, the rapid request of archived data, the rapid and timely transfer of data and information, and the ability for remote users to directly interact with a large volume of information as never before possible.

Introduction

Computer-to-computer networks have existed for over 15 years, but it has been only within the last few years that the World Data Centers (WDCs) and other national and international data and information systems have made use of this type of communication. In addition to data centers, "science centers" are found at universities, government, and many industry laboratories. The users at these science centers have begun to use the existing computer networks to provide services to other colleagues and to data centers.

The National Space Science Data Center (NSSDC) manages the Space Physics Analysis Network (SPAN), which is NASA's largest science computer network. Space and Earth science researchers using SPAN are located in universities, industries, and government institutions all across the United States and Europe. These researchers are in fields such as magnetospheric physics, astrophysics, ionospheric physics, atmospheric physics, climatology, meteorology, oceanography, planetary physics, and solar physics. SPAN users have access to space and Earth science data bases, mission planning systems, information systems, and computational facilities for the purposes of facilitating correlative space data exchange, data analysis, and space research.

This paper is not meant to be an exhaustive study on network usages, but will discuss how personnel at science centers and data centers use the communications infrastructure that SPAN provides to greatly enhance space and Earth science research the world over. It is important to note that there is intensive use by data centers of many other networks besides SPAN, but the following is believed to be fairly representative of those types of uses. It is clear that rapid changes are occurring in communication technology and information systems design and implementation, making this report only an outdated snapshot by the time it is published.

The Space Physics Analysis Network

As indicated by the original name, SPAN initially linked together space plasma physicists working on NASA solar-terrestrial research programs. The solar-terrestrial programs that are supported by SPAN involve the Dynamics Explorer, International Sun-Earth Explorer, and the International Solar-Terrestrial Physics spacecraft, just to name a few.

Within the last two years, many Earth science and astrophysics institutions have also connected to SPAN. The astrophysicists are analyzing observations from spacecraft such as the International Ultraviolet Explorer and doing multi-wavelength data comparisons with spacecraft and ground-based instruments. In addition, over 15 institutions and research laboratories from the ocean community are now connected to SPAN. These oceanographers are using instruments aboard spacecraft to study all the Earth's oceans on a global scale. For a more detailed description of SPAN, the reader is referred to Green, 1988a.

A computer network system similar to SPAN has been developed in Europe for use by the European Space Agency (ESA) and its associated research community. This computer network is called the European SPAN and is fully internetworked with the corresponding NASA SPAN system. For more information on European SPAN, the reader is referred to Sanderson, 1988.

Data Centers and User Facilities on Span

The following is an overview of the services and types of interactions that are occurring with the use of computer networks. As discussed earlier, the emphasis is on the use of SPAN.

National Oceans Data Center (WDC-A for Oceanography) – The NODC has developed a system called NOSIE (NODC SPAN Information Exchange), which provides information about oceanographic data and services available at the NODC (see Hamilton and Ward, 1988). NOSIE is an interactive, menu-driven system that has four categories: General Information Catalogs and Inventory, User Services, and Data Submission Guidelines. Some of the most interesting aspects of NOSIE are a list of the most recent data acquisitions and a feature by which users can request data in specific geographic regions or time periods and the system returns a count of the amount of data meeting their criteria. At this time there are no online data sets, but users can order data at the end of an interactive computer session.

NODC also transfers ocean data to the Scripps Institute of Oceanography (SIO) in California for verification.. Once the SIO scientists have verified the data, they are again returned over the network to NODC for archiving and final distribution. Without the network to facilitate the verification procedure, the distribution of the data to users would have been delayed several weeks.

NSSDC (WDC-A for Rockets & Satellites) – The NSSDC is responsible for archiving and distributing processed space and Earth science data from NASA spacecraft. New thrusts at the NSSDC include putting all information about the data in its archive online and providing electronic access to the most requested data sets. This allows SPAN users access to NSSDC-held information and selected data archives 24 hours per day. The NSSDC has made extensive use of SPAN and has over 15 online information and data systems available for remote uses. The reader is referred to Green, 1988b, and Green, 1988c, for a discussion of these data systems and their capabilities.

Although the NSSDC has used SPAN primarily as a network that brings users electronically to its facilities, SPAN has been used on occasion by NASA as a quick reaction capability to move near realtime data and information all over the world (see, for example, Green and King, 1986, and Thomas et al., 1987).

WDC-C1 for Solar-Terrestrial Physics (STP) – Within the last few months, WDC-C1 for STP (at the Rutherford Appleton Laboratory in Chilton, England) was connected to SPAN by means of a direct connection to the European High Energy Physics Network (HEPNET). HEPNET and SPAN are internetworked.

WDC-C1 for STP has had a connection to the Joint Academic Network (JANET) and BITNET for many years and provides access to its many services for the university, government and industry science community in the United Kingdom. It is expected that a gateway will be established between JANET and SPAN, using the newly connected computer. At present, WDC-C1 capabilities on JANET include interactive access to ever-growing data bases of solar-geomagnetic indices, ionosonde data, AMPTE-UKS data, middle atmosphere temperature and composition data, and computer modeling software such as the MSIS model. These online services, which started in 1984, are now handling 150 to 200 queries per month. The data volumes handled by file transfers from WDC-C1 to requesters are in the tens of kilobytes per file range, with 20 to 40 files per month.

WDC-C2 for Geomagnetism – WDC-C2 at the University of Kyoto uses SPAN not only for receiving requests (several per month) but also for obtaining data and software. WDC-C2 does not currently support an interactive system for remote user access, but one for managing their geomagnetic indices archive is being planned. Recently, the NSSDC networked software used for plotting and analyzing orbit data to WDC-C2. This software was successfully installed and has been used with the received orbital data from the AMPTE CCE spacecraft (sent by magnetic tape because of the large volume).

WDC-C1 for Solar Activity – This organization did not use SPAN until recently, when it participated in the exchange of spatial data with scientists at the Solar Maximum Mission at Goddard Space Flight Center and solar physicists at Marshall Space Flight Center. However, future plans for use of the network include transmission of solar spectroheliograms observed daily at Meudon to requesters all over the world. Since the volume of the spectroheliogram data is considerable, WDC-C1 SA is currently working on compression codes to limit the volume of the data networked and then decompress it after it has arrived at its destination.

NOAA/Space Environment Laboratory – SEL, in Boulder, Colorado, deals primarily with data from near real-time sources (spacecraft and ground-based) up to 30 days. It is currently developing the capability to routinely use SPAN to network ground-based full disk solar observations to the Solar Maximum Mission (SMM) Data Analysis Center at Goddard Space Flight Center (GSFC). These images will be used by the SMM project for daily planning for observations of key features of the sun by the spaceborne telescope. It is expected that up to 26 images (2 megabits each) per day will be networked to GSFC. Once transferred, these images will be stored at GSFC and used in SMM correlative research activities.

SEL has experimented with using SPAN to receive data from the Big Bear Solar Observatory (through Caltech), the Stanford Solar Observatory, and the Solar Magnetograph at Marshall Space Flight Center. The received data goes into their Forecast Center operations at SEL. At this time, the method appears to be viable and many of those sites in addition to universities in Hawaii, Alaska, and Alabama, are interested in receiving some of these data.

Future plans are to link the existing Space Environment Laboratory Data Acquisition and Display System (SELDADS) into SPAN to support remote user access to real-time solar geophysical data. It is estimated that half of the existing 60 users will access this system over SPAN when it becomes operational.

Lunar and Planetary Institute (LPI) – LPI routinely receives about 90 requests per month of which 75 percent are from SPAN users. Some of the services accessible at LPI include Geophysical Data Facility (GDF), the Bibliographic Search Service (a bibliographic data base of about 25,000 references to lunar and planetary literature), the online Lunar and Planetary Science Conference Program, and the Planetary Image Center. Most of these services have been accessible since October 1986. In the past year, there have been 10 planetary scientists doing active research using the GDF. These users log onto the facility, display regional or global data, model these data, and transmit the results back to the remote institution (as much as 0.2 megabytes) for further analysis. It is expected that general usage of the facility will grow by at least 25 percent over this next year, with at least a 50 percent growth in the library services.

EROS Data Center – EROS, in Sioux Falls, South Dakota, uses SPAN and the U.S. Geological Survey's GEONET for the transfer of data and for access to supercomputer facilities. Although there are no online interactive request services, EROS receives three to five requests per month over SPAN. EROS transfers mostly image data (from NOAA and NASA spacecraft) on request. The image data transferred are approximately 2-3 megabytes in size. EROS provides a gateway system between SPAN and GEONET. It is expected that EROS will dramatically increase its use of SPAN when it takes on major EOS polar platform data handling and distribution responsibilities.

Space Telescope Science Institute (STScI) – The STScI in Baltimore, Maryland, will be the active repository for the data from the Hubble Space Telescope (HST) when it is launched in approximately a year. Currently, a system called the Data Archive and Distribution Service (DADS) is under construction, which will have the capability to manage over 15,000 gigabytes (15 terabytes). This system is expected to be fully operational about one year after launch of the space telescope. DADS is a multimillion dollar capability, and is the most ambitious system of its kind ever developed. DADS will have extensive information systems (catalogs, directories, inventories, etc.), online data of all HST data products, and ancillary information serving the extensive remote and local astrophysics and planetary user community. The European archive of HST data will be located at the European Coordinating Facility (ECF) at Max Planck Institute, West Germany. In addition to providing access and data transmission to users over SPAN, STScI plans to use SPAN to update the data information archive of HST data that ECF will also be managing.

University of Miami – Many universities gain access to the oceanographic data being collected at the University of Miami by using SPAN. The University of Miami routinely networks compressed data from the Advanced Very High Resolution Radiometer (AVHRR) instrument on board a polar-orbiting NOAA spacecraft. The AVHRR data are received in real time at the university and are quickly processed (stripping out the infrared portions), compressed, and transmitted to the University of Rhode Island via SPAN. There the data are decompressed and remapped into a standard set of projections used for several real-time ship activities, such as cruise support and chart generation. These images are also networked to Harvard University for their Gulf Stream prediction models. In this example, a Lempel-Ziv compression algorithm is used.

European Space Research Institute (ESRIN) – ESRIN, in Frascati, Italy, has just been given the responsibility for the development of a distributed information system for the European Space Agency (ESA). This system will essentially consist of a uniform interactive user interface at several insitutions in Europe that will provide for an enhanced query facility into the services that each institution offers. The selected sites for the first phase of this project include the IUE facility at Villafranca, Spain; the HST facility at Garching, West Germany; the

EXOSAT facility in Noordwijk, Netherlands; the SIMBAD data base in Strasbourg, France; and the WDC-C1 for STP, United Kingdom. The existing European SPAN system provides the needed connectivity for this effort.

In addition, ESRIN maintains a system that is used for the routine transfer of quick-look data from the San Marco spacecraft to the investigators in the United States. San Marco is a joint U.S., West German, and Italian spacecraft that was launched in late March 1988. The low-altitude equatorial spacecraft is tracked from a station in Kenya, where the data are transmitted to computers in Rome. There, quick-look data are processed and then networked to ESRIN, where they are temporarily stored. From the SPAN node at ESRIN, the data are accessible by users in the United States – at Goddard Space Flight Center, the University of Maryland, and the University of Michigan.

Conclusion

The use of computer-to-computer networks such as SPAN and the European SPAN is providing a new dimension for quick access to key data in the space and Earth science international community of scientists. Networks like SPAN are being used so extensively that it is hard to keep up with the many and varied uses. In addition, it is clear that networks have established themselves, in just a few short years, as another of the essential tools needed for conducting effective space and Earth science research. It is obvious that in the future communications networks will play an ever-increasing role in movement of data, results, information, and ideas.

Acknowledgment

I would like to gratefully acknowledge contributions from Kin Leung, Stephen Tellier, Guy Simon, Allen Watkins, Mike Hapgood, David Speich, Doug Hamilton, and Dr. M. Sugiura, which are much appreciated.

References

Green, J.L. and J.H. King, *Behind the scenes during a comet encounter, EOS (feature article)*, 67, 105, March 1986.

Green, J.L., *The Space Physics Analysis Network, Computer Physics Communications*, North-Holland, Amsterdam, 49, 205-213, 1988a.

Green, J.L., *NSSDC information systems: Status and plans,* Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b pp. 24–32.

Green, J.L., *NSSDC applications for computer networking,* Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988c. pp. 168–175.

Thomas, V., J.L. Green and L. Porter, *SPAN links astronomers around the globe in support of supernova, Information Systems Newsletter*, 10, p. 4, 1987.

Sanderson, T., *The European Space Physics Analysis Network, Proceedings of the 21st ESLAB Symposium*, Bolkesjo, Norway, ESA SP-275, 1988.

Hamilton, D. and J. Ward, *Online access to NODC information services, to be published in the proceedings from the Workshop on Geophysical Informatics*, Moscow, 1988.

COMMUNICATIONS AMONG CENTERS

Carl C. Abston

World Data Center A

Boulder, Co. USA

Present Telecommunications Among Centers

During the past five years several of the World Data centers including Boulder, Moscow, Kyoto, Edinburgh, Chilton, and Copenhagen have been using telecommunications to exchange messages. The system basically involves the EDS time sharing system located in Detroit, Michigan USA. It is a node on the Telenet commercial network. Access to telenet or Tymnet depends on the individual country. In some cases, the user can dial directly into telenet, in other countries, it is necessary to go through a local network which connects into telenet. No matter how the connections are made, EDS acts like a mail box where messages are left or picked up. Occasionally we have used this system to connect to two centers together in real time to discuss some technical problems. The mailbox approach has proved to be both useful and inexpensive for most applications. For example, when two centers are cooperating in a join project requiring considerable discussion about various details, the mailbox has greatly facilitated the successful completion of these projects.

Transferring data however, has not proven to be an overwhelming success but not a failure either. We have tried data transfers using Kermit (a hand shacking protocol) to transfer data between Boulder and the EDS computer. This method sometimes works and sometimes doesn't, but at best it is hard to use. Some of the PC based software, such as PC-Talk, works as well as anything. The data is usually transferred correctly, although the throughput is slow, about 1000 characters per minute and retransmissions at somethings necessary. Small amounts of data less than 50,000 characters are routinely moved to the mailbox and quickly read by WDC-C in Chilton, England.

In the Soviet Union, other problems exist. The Interface into Telenet is through a central communications hub. This hub is handling all telecommunications from Moscow and is therefore very busy. A fixed time slot for telenet is not always satisfactory. At times we have a great need to communicate with each other but at other times we have very little need to talk. There also appears to be more problems transmitting and receiving data files accurately between Moscow and the EDS mailbox than with the other centers.

Data Transmission Improvements

Although data transmission via telecommunications has not been altogether successful, some improvements appear to be possible.

To successfully transmit data via telecommunications two items are necessary. First, the data absolutely must be sent and received correctly. Second, it should require a reasonable amount of time and effort.

Correct data transmissions can only be assured using protocols. Fortunately, there are a large number of these in common usage today. Unfortunately, the larger machines, such as the minicomputers (VAX-20's) used in time-sharing do not support most handshaking protocols. I am again reminded that the best software runs on the smallest machines. For example, the EDS time sharing system is very limited in its protocol capabilities whereas the PC systems are rich in capabilities. Hundreds if not thousands of "bulletin boards" exist in the US today. These systems consist of a telephone line, a 300/1200 baud modem, a PC XT or AT-type computer and an operating system such as FIDO. These systems only support one user at a time, however they support a dozen different protocols and allow a user to send or receive data files with a minimum of knowledge and key strokes. In other words, the PC systems are years ahead of both minicomputer and mainframe computers in asynchronous telecommunication transmissions.

To better illustrate the point, let me take a recent example. WDC-A in Boulder and USGS in Washington D.C. have been jointly cooperating in the creation of a CD-ROM holding the Gloria data base from the Gulf of Mexico. The accession software was written at WDC-A Boulder, USGS in Washington and JPL/NASA in Pasadena, Ca. Integrating these three packages into a total software system required many alterations in each package. The software was joined and tested at USGS. Both Boulder and Pasadena, transmitted compressed binary object software via standard dial-up telephone lines into a XT based bulletin broad in Washington. 1200 baud modems without MNP were used. The Boulder object code was 300 kilobytes but compressed into 104 kilobytes and was transferred in 18 minutes. In all of the transmissions, no errors occurred and all of the received software worked as designed. From this example, I know telecommunications can be used to exchange small (up to half a megabyte) data sets without errors and in a reasonable time period.

Improvements in transmissions, both in speed and in accuracy, can be achieved in the following ways:

1. Modems can be upgraded from 1200 to 2400 baud. Infact, the newest systems coming on the market allow telecommunications speeds as high as 9600 baud over standard dialup lines in the US, Europe and Japan.

2. Wherever possible, the newer varieties of modems with error correction protocols such as MNP 1, 2, 3, 4, or 5, should be used.

3. Data being transmitted should be compacted. For example, using PKARC on a PC, it is possible to achieve a 3-10 to 1 compaction improvement. In other words, reducing the data by 3 times, usually improves the speed by 3 times. There are some exceptions.

4. Use the fastests protocols. Some protocols are faster than others. Choosing a faster protocol can greatly improve throughput speed. For example Kermit is usually slow, XMODEM/CRC is faster, and YMODEM is even better. Also, protocol implementation is as important as the protocol technique itself. For example, we experimented with Kermit of several systems using 9600 baud lines and communicating within our center. Transmissions using Kermit between a PC AT and a Charles Rivers multi-user Unix based system, achieved an actual throughput rate of 7200 baud. Transmission between a PC AT and a Data General minicomputer system, achieved a rate of 4500 baud. And the same test between a PC and our VAX 750, achieved a 5800 baud rate. We were using the same PC, the same communications lines, single user minicomputers, the same protocol, and different implications. When we did the same experiment using a mainframe with 1200 baud communications, we only achieved a transfer rate of 400 baud. Efficiently written code, is efficiently written code, which always works better, no matter how fast the machine you are using.

World Data Center International Data Transfer Improvements

The present EDS DEC-20 time-sharing system, currently being used by many of the world data centers is old, getting older, with no new software being introduced. Improving data transfers with this machine appear limited at best. The National Oceanic and Atmospheric Administration has been using the telemail system to exchange messages and data files between its various centers. The system is setup to support PC's, protocols and binary data transmissions. Therefore, I would recommend that several centers join in testing the feasibility of using the TELEMAIL system for data transfers between countries.

This approach may not be economically feasible for WDC-C in Japan.

WDC-A (Boulder) Telecommunication Improvements

WDC-A in Boulder is currently in the process of connecting the present Ethernet TCP/IP LAN into the University of Colorado's fiber Optic Ethernet TCP/IP LAN. An IP-Router will be added to the WDC-A LAN. When this is complete and all LAN addresses standardized, WDC-A will have access to several additional networks including ARPRA Net, Bitnet, NASA's SPAN net as well as several local Colorado and University networks. The interconnection between SPAN and the University of Colorado is via a GATEWAY. WDC-A and CU's LAN are Ethernet TCP/IC systems whereas the NASA Span network is DEC-NET. The Gateway not only provides interconnectivity but protocol conversion as well.

Telecommunications between WDC-A's (Boulder, Washington, Asheville)

Several in the WDC-A's including the four located at Boulder, WDC-A in Oceanography in Washington and WDC-A for Climate in Asheville have begun the process to integrate these centers via telecommunications. Each Center either has or will establish an Ethernet TCP/IP LAN. Each center LAN serves as a backbone interconnecting all computers in each respective center.

These three LAN backbones will be connected via IP-Routers, modems, and 56 kiloBaud satellite circuits. Therefore, any PC workstation will have access to any TCP/IP server node anywhere in the system. For example, every center will have a VAX minicomputer running TCP/IP. A PC in Boulder, will therefore have the ability of transfer a file to and from any VAX in the system as well as from any NODE on the network. The complications of IP-Routers, modems, and telecommunications will be transparent to the user. The user will continue to issue the same command as is presently used. The only real difference for the user, will be the speed of transmission.

LAN's have proven to be a very reliable method of transmitting and receiving data. At WDC-A in Boulder, all data, read or written to and from tape drives, has been moved across a TCP/IP LAN at least once. There have been no errors.

This project should be completed within the next 12 months and we will be closely evaluating the performance and flexibility of the system.

We are also working to add a central catalog system permitting outside terminal access to information regarding all data holdings in these centers. The first phase of the project, is to provide the capability for a user to dial into a central computer and get information about the data held in each centers. This computer will be an IBM AT-80386-type machine running the Unix operating system.

The second phase will allow the user access to individual center data inventory systems. There are of course, a lot of problems to be solved before the total communication system becomes an effective reality.

Summary

Many types of telecommunications are possible. The above few pages highlight the current work and communications philosophy now taking place between several of the WDC-A's.

Similar work is probably taking place among many of the other world centers. After detailing all ongoing projects and their interconnectivity, it may be possible to better define the direction and philosophy to improve communications between all WDC-s.

NSSDC APPLICATIONS FOR COMPUTER NETWORKING

James L. Green

National Space Science Data Center
Greenbelt, MD 20771

Abstract

The National Space Science Data Center (NSSDC) supports many computer network connections in order to provide remote access to the NSSDC by all its diverse users from all over the world. The NSSDC supports connections to Bitnet, the National Science Foundation network (NSFnet) and its internet, the X.25 international packet, and the Space Physics Analysis Network (SPAN), which it manages.

With the ease of electronic access dramatically increasing over the last few years, the NSSDC has created a major new thrust in "online" computer information systems and services accessible to remote users 24 hours per day. The NSSDC online information systems include the IUE Interactive Request System, the NASA Climate Data System (NCDS), and the Master Directory (MD), just to name a few.

The IUE Interactive Request System allows scientists to order observations taken by the International Ultraviolet Explorer spacecraft. The computer networks connected at the NSSDC allow remote users to access the IUE request system and are used to deliver, on the average, about 50% of the data. Tests are currently under way in which some IUE data are compressed before being networked to the remote node and are then decompressed, reducing the communication load on the wide area networks used.

This paper will provide an overview of all the worldwide network connections utilized by the NSSDC and will discuss some of the data center applications of these networks. In particular, the IUE Interactive Request System will be discussed in detail.

Introduction

There has been an explosion in the use of available communication technology for the movement and access of information and data. Many large universities and nearly every NASA center and other government institutions that work with NASA data have relatively high-speed local area networks and many wide area network connections. With the ease of electronic access dramatically increasing over the last few years, the National Space Science Data Center (NSSDC) has created a major new thrust in "online" computer information systems accessible to remote users 24 hours per day. Currently, not all the information about the NSSDC archive is remotely accessible to users, and less than 2% of the NSSDC's total digital data archive is on line (see Green, 1988b), but these systems are already a major achievement in providing rapid access to NASA-acquired science data that is unprecedented in archive data management.

This paper will focus on one example of the major NSSDC online interactive systems that are used extensively over international computer networks. These two systems are typical examples of how the NSSDC is responding to user demands for rapid access to NASA-acquired data. The reader is referred to Green, 1988c, for more details of all the NSSDC systems and capabilities.

Worldwide Network Connections

Many computer network connections have been made to provide remote access to the NSSDC by all its diverse users. Figure 1 is a breakdown of the major network connections by communication protocol. The Bitnet connection only supports mail communication between many universities in the United States and the NSSDC. Selected science nodes throughout the world 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 the NSSDC: SPAN (Green, 1988a) and the NASA Science Network (NSN). SPAN is a DECnet network that contains over 2050 nodes in the United States and is internetworked with over 10,000 nodes in the United States, Europe, Canada, and Japan (through other networks such as HEPNET). SPAN is used exclusively by space and Earth scientists working primarily on NASA-related missions and projects. SPAN is managed by the NSSDC. NSN (which uses the TCP/IP protocol) is internetworked with other wide-area networks, such as

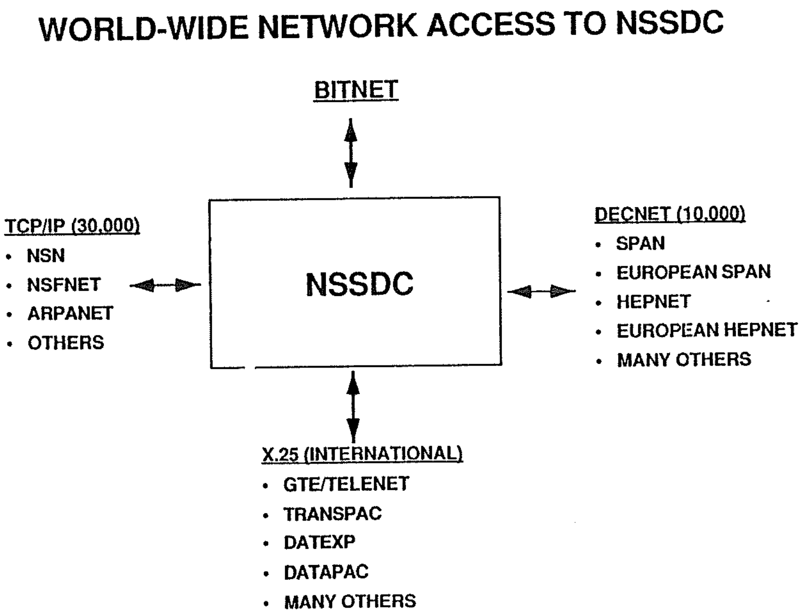


Figure 1 - This figure illustrates the wide-area network access to the NSSDC. The NSSDC manages the SPAN computer network, which supports many connections to other wide-area DECNET networks such as HEPNET. Another major NASA network is the NSN which provides TCP/IP connectivity to many other computer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

ARPANET and the NSFnet, that can reach many thousands of computers. In general, these wide-area networks are of relatively low speed but are serving a tremendously valuable service for remote users to gain access to NASA computer resources and to communicate with fellow researchers all across the country. The bulk of the wide-area network traffic is for informational purposes such as remote logon and mail.

The wide-area networks provide the pathways for remote users to access the NSSDC facilities at any time day or night (excluding times for system maintenance). The online interactive information systems at the NSSDC that are accessible to the remote user have been briefly discussed by Green, 1988c. The IUE Request System will be discussed in the next section, since it illustrates the heavy use of the new wide-area network communication technologies combined with advances in mass storage and database management.

The Online IUE Request System

The IUE Interactive Request System became operational in November 1987; by January 1988 requests were routinely serviced with this system. The request system consists of large online mass storage, menu driven interactive information access software (running on two VAX front end computers), a high-speed local-area network connecting the online storage with the interactive front ends, and the wide-area networks SPAN and NSN (see Perry, 1988).

In general, 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 International Ultraviolet Explorer (IUE) data into the NASA Space and Earth Science Computer Center's IBM 3850 Mass Store in order to better accomodate the large user demand through faster request response. It is important to note that the NSSDC typically manages its archive off line. Storing all the IUE data on line was done with full project cooperation and to gain valuable expérience with highly requested online data sets. The IUE data that are currently on line consist of over 61,000 unique star images and spectra.

The IBM 3850 Mass Store device is controlled by an IBM 3081 computer and is connected to the NSSDC interactive VAX front end computers by a high-speed local-area network (called SESnet) as shown in Figure 2. The 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

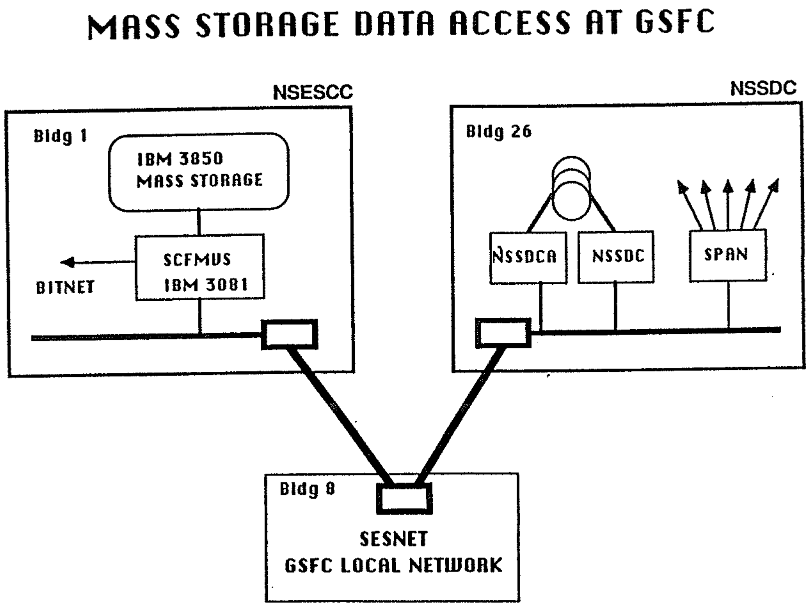


Figure 2 - All the available IUE data is stored in the NSESCC mass storage system. The high-speed local-area network at Goddard Space Flight Center, SESnet, provides connectivity between the IUE Interactive Request System (which resides on the NSSDC interactive front end computers) and the IBM mass storage system.

request coordinator networks the IUE data from the Mass Store device through SESnet. 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 easily handled by this system, saving the manual locating and loading of the data. Request for IUE data also come to the NSSDC through letters and phone calls (not all users are on computer networks).

Figure 3 shows the monthly averages of IUE images requested by individuals from 1979 to 1988. The NSSDC also sends large amounts of IUE data to other archives; these requests are not included in this figure. From 1979 to 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 montly number of IUE images requested in 1988 (an average computed from the first six months of that year). As clearly seen, there has been a dramatic increase in the amount of IUE data requested in 1988, reaching approximately 385 images/spectra per month. The 1988 requests have come from many scientists at 15 institutions in the United States, Europe, and Canada (locations services by SPAN).

It is important to note that the IUE Interactive Request System is used by scientists for over 75% of all requests. The computer networks are used to deliver, on the average, about 50% of the data (in July 1988, all images requested were networked over the SPAN), while the remainder, usually the larger data volumes requested, are satisfied by sending magnetic tapes. 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 remote node and are then decompressed, therefore reducing the communication load on the wide-area network.

Since there has been no new funding from NASA Headquarters for increased IUE data analysis, the request results for 1988 clearly show in Figure 3 that the tremendous increase in requested data is believed to be due to the convenience this system provides to the user. The following factors are a major part of the user convenience provided by the IUE 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

- No need to send replacement tape to the NSSDC (currently the network is a "free" service to the users)

To be able to use low-rate communication networks such as SPAN requires that the size of data requested is relatively small. The IUE example is a good one, since the data are managed by the stellar object observed, which in itself forms a small enough data subset to be easily networked. The amount of time required to satisfy a request is 2 to 15 minutes, depending on the communication line rate between the NSSDC and the requester and the load on the network at the time.

Future NSSDC Network Services and Activities

It is clear, from examples like the IUE 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. These types of services are in great demand. Worldwide communications technologies are extremely important in providing the pathways necessary for remote users to come to data centers all over the world.

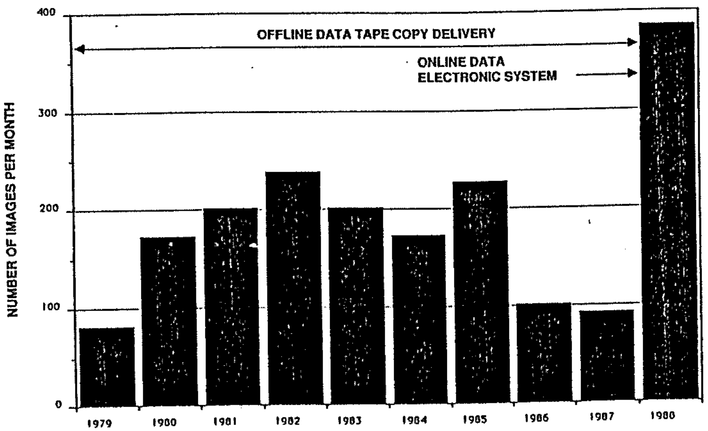


Figure 3 - This bar chart shows the average number of IUE images requested (per month) by individuals for each year since the archive was opened in 1979. Although the IUE Interactive Request System became operational in November 1987, it was not until January 1988 that remote users routinely accessed the system. The huge increase in the average number of images distributed in 1988 can easily be attributed to the better service that is now provided electronically.

ESA DATA AND INFORMATION DISSEMINATION SERVICES AND TECHNIQUES

An Overview

E. Jabs

ESA/ESOC

Robert-Bosch str. 5

D-6100 Darmstadt FRG

Introduction

The role of the European Space Agency (ESA) as concerns space science and earth science (remote sensing) is not to undertake research but to enable and support research, which other words: ESA is not a research organisation but a service organisation. This restriction is demonstrated by the fact that for typical missions in the solar-terrestrial discipline only a few experiments or instruments are financed and developed by ESA. The great majority of experiments is developed and financed by national European research institutes. As a consequence acquired experiment data has to be delivered to the institutes, which are responsible for the development of the experiment and will then become responsible for the scientific reduction and analysis of the data.

ESA's responsibility during the operational phase of a mission is more concerned with mission execution in accordance with defined overall objectives and with platform monitoring and control. Routine experiment monitoring and control will be executed as an institutional service according to experimenter specifications and requests, as far as these are compatible with the mission constraints.

The NSSDC will continue to aggressively pursue the "electronification" of its information about data and, to the extent reasonable, its archived data (see Green, 1988c, and King, 1988a). Much of the new data coming into the NSSDC will be managed by the online interactive systems, but much work needs to be done.

The NSSDC will continue to improve its connectivity. Higher speed communications lines will be necessary as more users log onto NSSDC systems and more and more data are networked to remote users. Over the next few years, SPAN will continue to grow, adding new nodes worldwide and supporting such activities as the International Solar-Terrestrial Physics program.

Acknowledgement

I would like to gratefully acknowledge Karen Satin who has done an expert job in technically editing this article, and my secretary, Marjorie Layfield, for her accurate and rapid typing of the manuscript.

References

Green, J.L., *The Space Physics Analysis Network, Computer Physics Communications*, North-Holland, Amsterdam, 49, 205-213, 1988a.

Green, J.L., NSSDC data holdings: scope, complexity, and future requirements. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b 205–215.

Green, J.L., *The NSSDC Information Systems: Status and Plans*, Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988c. pp. 24–32.

King, J., *Mass storage technologies for NASA's future data environment, to be published in the proceedings from the Workshop on Geophysical Informatics*, Moscow, 1988a.

King, J., Exchange of data and information between WDC-A-R&S and WDC-B2. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b pp. 281–284.

Perry, C.M., *The National Space Science Data Center and International Ultraviolet Explorer: 1978 – Present, to be published in the 10th Anniversary IUE Conference Proceedings*, 1988.

Service Characteristics of ESA Space Science Missions

The datailed services, which ESA provides for the support of unmanned scientific missions, depend on the operational requirements and characteristics of these missions.

They can, when looked at from an operational point of view, be divided into three categories:

- Principal Investigator (PI) type
- observatory type
- survey type

PI type missions fly experiments/instruments developed by Principal Investigators, who are also defining the experiment operations and are responsible for scientific analysis of acquired data. PIs require during certain periods interactive (on-line) access to their spaceborne experiment. ESA PI type missions belong mainly to the solar-terrestrial, planetary and microgravity disciplines. Typical examples are GEOS and GIOTTO, and to some extent EURECA.

Observatory type missions are characterised by platforms carrying a few major instruments, usually funded by ESA. While the Agency has the instrument operations expertise and responsibility, instrument use is shared by a wide user community. The operations require therefore an observatory environment, i.e. interactive access to the instrument from a central place, which is for practical reasons co-located with the platform control centre and where the user community is represented by means of guest observers. Observatory mission fall mainly into the astronomy discipline. Examples are IUE, EXOSAT.

Survey type missions carry one big instrument or a few major instruments, funded by ESA, which are usually scanning the domain of interest in a predetermined manner. The data users do not have expertise with instrument operation and do not require interactive access to it. Survey type missions generally belong to the astronomy discipline and earth science (remote sensing) discipline. Examples are GOS-B, Hipparcos, ERS-1.

The services of an ESA control centre in support of the above classified missions fall into two groups. The first group, comprising those services which are of an institutional nature and do not contain a direct interface with the scientific customer (payload data user) e.g.:

- acquisition of data from orbiting spacecraft
- sorting and temporary storage of raw data
- platform monitoring and control
- navigation

This group of services is in this context of no further interest. The services of the second group are directly customer oriented and comprise:

- instrument/experiment monitoring and control
- generation of mission data products and dissemination to "Primary Data Users"
- data archiving services

In addition ESA supplies a general, mission independent, on-line bibliographic service to European industry, national space agencies and research institutes, called IRS (Information Retrieval Service).

Data and Information Dissemination Services

Instrument/Experiment Monitoring and Control

The monitoring and control of PI **type experiments** is, in principle, the responsibility of PI's. They do perform this task during in-orbit commissioning of their experiments, and whenever they are resident at the control centre (usually located at ESOC, Darmstadt, Germany), e.g. during campaigns. During the remaining time PIs delegate responsibility for those routine activities and critical interventions, which cannot be automated, to control centre operations staff.

Nevertheless, most PIs require in addition quick-lock (summary) data from their experiments within a few days. This service enables them, during their absence from the control centre, to generate command requests aiming at optimization of experiment operations.

Quick-lock data products are usually mailed in form of print-out or graphs, on paper or microfiche, sometimes in computer readable form on magnetic tape. Command requests are usually sent by telex, sometimes by voice or by mail.

In a few cases PIs have requested and obtained a data link to a terminal or computer at their host institute. However, in all those cases a variety of procedural problems had to be overcome and remotely generated command requests were never automatically fed into the control system.

As for **observatory type missions**, instrument control is the responsibility of the control centre and may, to some extent, be delegated to the resident guest observer, performing his observation ideally in an on-line (interactive) mode. Dissemination of quick-lock data therefore does not apply.

Guest observes have, however, already expressed the desire, to conduct on observatory session from their home institute in the same manner as they would do at the observatory.

In case of **survey type missions** instrument monitoring and control is performed by control centre staff and does not require any involvement of, or interaction with, data users.

Dissemination of Mission Data Products

This service is for ESA space science missions provided by ESOC (or the dependent control centre VILSPA, located near Madrid in Spain), for Meteosat also provided by ESOC and for other remote sensing missions provided by ESRIN. It differs largely for the three mission categories:

Investigators on **PI type missions** receive so called "experimenter tapes" containing raw instrument telemetry, as well as corrected time and refined orbit/attitude information. Sometimes additional auxiliary data, e.g. selected platform telemetry and, rarely, summary data are also provided on magnetic tape. Delivery delay is usually in the order of several weeks.

Investigators on **observatory type missions** receive "observation data sets" comprising pre-processed instrument data from their observation sessions, together with calibration parameters and relevant auxiliary data. In some cases selections of finally processed (calibrated and corrected) data are also provided or even data resulting from a preliminary scientific analysis. The medium used in all this cases is magnetic tape, sometimes supplemented by display hardcopies or plotter output.

In the category of **survey missions,** the requirements for space science (usually astronomy) missions differ significantly from those of remote sensing missions: data users of **astronomical survey** mission require only raw or pre-processed instrument data with calibration parameters and relevant auxiliary data. The delivery time can be in the order of months and magnetic tape as medium is fully adequate.

Data from **remote sensing missions,** which are typically available in form of images, interest a great variety of users with different requirements:

High resolution Meteosat images, as an example, have to be processed on ground and redisseminated via satellite (again Meteosat) within about one hour after first acquisition for use by European weather offices. Low resolution images are also redisseminated via satellite with a similar delay for use by a much wider community. Meteorological information extracted at the processing centre at ESOC is disseminated via the world wide meteorological ground communications network.

Certain data products from ESA's first remote sensing satellite, ERS-1, to be launched in 1990/91, have to be delivered to the end user within a few hours after acquisition of instrument telemetry. Generation of these "Fast Delivery Products" is therefore performed at the receiving ground station and dissemination will involve a medium speed (64 kbit) on-line communications network or alternatively a special dissemination facility using satellite circuits and small terminals (APOLLO system).

Dissemination of Archived Space Data

The archival service for ESA scientific missions is currently rather rudimentary. It involves storage of new instrument/experiment data and in some cases of auxiliary data on magnetic tape for a period of 10 years. For observatory type missions archiving may also include the storage of additional data products (preprocessed and calibrated data, selected sets of finally processed data) at the observatory until the end of mission.

The basic service does not offer any cataloguing or retrieval facilities. It can cope with request for regeneration and delivery of experimenter tapes with a response time of several weeks. But it does not provide adequate facilities for secondary data users, i.e. persons which are not part of the original experiment team. As a consequence, the request rate for delivery of archived mission data is very low.

ESA has decided at the end of last year to offer in future improved services in this area. The system to be implemented is termed ESIS (European Space Information Service).

As concerns archiving, retrieval and dissemination of Meteosat images, and products, a more advanced service, which is generally accepted by users is provided.

Offered are images of various formats and resolutions either in form of hardcopies (photograph produced by a laser beam recorder) or in computer readable form on magnetic tape. A product catalogue is

available, but not on-line accessible. Both ordering and delivery procedures are based on ordinary mail. Resulting delays are fully acceptable to users, which are generally researchers at universities or other institutes.

In the area of remote sensing for land and sea applications ESA provides a preoperational archiving and dissemination service with data acquired from non ESA satellites. The service is provided by the EARTHNET Program Office (EPO) in ESRIN. There are an on-line, remotely accessible product catalogue available, browsing facilities at the ESRIN site, a variety of products (images in hardcopy form or computer readable form on magnetic tapes) and an on-line product ordering facility.

All products are disseminated to customers as ordinary mail. In the early 80's successful experiments were undertaken, in delivering products (images) to users via small earth terminals and a satellite link, applying a tape-to-tape transfer procedure. The low budget of the current applications cannot afford the still relatively high cost of such on-line data dissemination method.

Bibliographic Information Retrieval Service

ESA operates at ESRIN a bibliographic information retrieval service, IRS (also called QUEST), as a supplementary service to European industry and research institutes. IRS contains space science bibliographic information, but also chemical abstracts and many other disciplines. In total about 150 files, containing more than 50 million items, are on-line accessible requiring a total direct access storage of currently 80 Gigabytes.

Bibliographic searches may be executed from interactive display terminals and PCs from all over Europe, but also in many non-European countries. Searches can also be executed as batch jobs. Results of interactive searches are on-line available on the screen and in the local memory of the terminal, from which a hard copy can be taken. Printout of searches can also be delivered off-line by mail. Full-text documents (in hardcopy or microfiche form) can be ordered on-line but will be delivered as ordinary mail from the document archive, which is not located at ESRIN.

Dissemination Techniques Currently Applied

On-line Catalogue and Information Access and Data Product Ordering

On-line catalogue access and on-line data product ordering facilities are available to users of the EARTHNET archive. The access network to be used is identical to that of IRS. In addition major users have access also via public X.25 networks. The catalogue and order facility is implemented as special to purpose software on a Micro-VAX. The product catalogue contains also a schedule of planned acquisition of new products within a given reference period. The ordering facility is largely based on ORACLE, a standard relational data base management system.

The IRS system offers interactive access to a file index, file description and user manual for searches. The product which can be ordered on-line in this case is a document which is stored on microfiche rather than in computer readable form, in one of several European document archives.

IRS itself is a conventional bibliographic information storage and retrieval system, which was conceived more than 20 years ago. It uses the techniques of "index files", "inverted files" and "Thesaurus files" in addition to the "linear file".

The access is either via the ESA internal network (ESANET), via a set of dedicated lease lines, or via public packet switched networks.

On-line Data Dissemination

This data dissemination technique is, apart from Meteosat image dissemination and distribution of extracted meteorological information, currently not applied on a regular basis in the context of the above reviewed ESA data and information services. The Meteosat system, however, is rather specific and not really representative for applications which are the subject of this symposium. It is therefore here not described in any further detail.

On-line data transmission techniques were, in the context of some previous ESA missions, used on an ad-hoc basis (remote terminal access and file transfer). The specific techniques applied at that time are meanwhile technically obsolete and therefore not further discussed.

During the ICE mission, the encounter of the ISEE-B spacecraft with the comet Giacobini-Zinner in September 1985, SPAN was used to send experiment data in near-realtime from NASA's Goddard Space Flight Centre to ESTEC in the Netherlands. During the Giotto mission, and in particular during the months proceeding the encounter with comet Halley, Span was extensively used for world-wide mail exchange, but also for a variety of data and file transfers from and to "exotic locations" which were in most cases not preplanned, but arranged on a very short notice.

All this applications had experimental character but were very successful. It should in this context also be mentioned that the currently still existing data link from ESOC to IKI was initially set up to facilitate exchange of navigation data from the European Halley probe and from the two USSR spacecrafts.

Transportable Media Usage

Magnetic Tape

The by far most frequently medium used within ESA for archiving and dissemination purposes is magnetic tape, in particular standard industry compatible, 9-Track, half inch tape with a density of 6250 bpi. The 6250 bpi tape standards is in all respects advantageous to the 1600 bpi standard, which is still being used, if absolutely required for compatibility reasons. The long term data reliability of 6250 bpi tapes is, of course, still problematic.

IBM now offers a new magnetic tape technology, which uses as a medium tape cartridges, rather than tape reels. A cartridge contains a half-inch wide tape, on which data is recorded in 18 tracks with a recording density of approximately 38.000 bytes per inch. The capacity of one cartridge is 200 megabytes, when a block size of 24 kbytes is selected.

At a block size of 4 kbytes it has about the same capacity as a conventional 2400 feet reel written with a density of 6250 bpi. The cost of a cartridge corresponds about to that of a 2400 feet reel. An obvious advantage of the new cartridge system is the considerably smaller size and volume of the medium.

Load, rewind and unload times are much smaller than for a conventional tape system. Data transfer rates are higher (up to 3 megabytes/sec). The system provides automatic tape threading and with optional equipment also premounting and automatic feed of tape cartridges on tape drives. IBM claims increased subsystem date reliability because of the new media (chromium-dioxide), greater tape protection by cartridge, thinfilm head, and increased error correction.

ESA plans to install such tape cartridge system at ESOC still this year and to use it for archiving purposes as a replacement of a 6250 bpi magnetic tape system.

In those cases, where ESA has to acquire and record data with very high speed, e.g. about 100 Mbits per sec., high density digital tapes (HDDT) are being used. These tapes are usually also maintained as archive tapes, although this technique is rather problematic (unreliable and very maintenance intensive, i.e. costy). HDDTs are not suitable as dissemination medium.

Optical Disks

This medium is promising a long stable life time of the record (about 30 years) and seems therefore to be very attractive as archiving medium. The small volume and low weight make it, in principle, aslo attractive as dissemination medium. Two types of technologies have to be distinguished:

a) **WORM** (Write Once, Read Many times)
This technology, which is usually applying a 12 inch disc and provides storage of 2 Gigabytes, allows the user to create a record himself – drives allow to write and read. ESA is evaluating this technology since many years and has devices of this type in quasi-operational use since about two years for archiving of EARTHNET products.
Reliability seems to be satisfactory, problematic are the very high medium cost (approx. $ 300 per disc). Also problematic is, that there exists still no standard for the recording technique and the medium. This is a further reason why the use of a WORM disk as dissemination medium is currently unattractive.

b) **CD-ROM** (Compact Disc – Read Only Memory)
This medium, based on 5 1/2 inch disc with a storage capacity of 600 Mbytes, is attractive when many copies of the same record are required. The generation of a master copy, usually performed by specialised companies, is rather expensive (approx. $ 3.000), each additional copy costs in the order of $ 10. CD-ROMs are very suitable as data dissemination medium, provided the number of required copies justifies the high generation cost.
ESA has so far undertaken only a few experiments with CD-ROMs, executed by the EARTHNET office at ESRIN.

c) **Floppy-Discs**
The rather low capacity of this medium (currently 1.2 Mbytes as a standard) make it not really suitable for the majority of space data dissemination applications. ESA is currently not using floppy discs in this domain. As the capacity is expected to rapidly grow, this situation might change. Floppies could become an interesting, but only supplementary medium.

New Projects, Applying Advanced Data Dissemination Techniques

EURECA Data Disposition System

EURECA 1, the first mission of the "European Retrieval Carrier", to be launched in 1991 carries a variety of microgravity experiments, or more specifically, material science experiments. Experimenters require access to certain experiment data within a few hours after acquisition, in order to optimise the scientific return. The data have to be made available at a remote "Microgravity User Support Centre" (MUSC) or at the users host institute. ESA has decided to implement at ESOC a remotely accessible data base, which provides the required preprocessed data, shortly after acquisition. Old data are cyclically overwritten. It is thus the responsibility of the experimenter to collect his data within a certain time.

This data disposition system will be based on a VAX computer and be accessible via DECnet protocols. Users at the MUSC will access it via a leased line, other users via the public packet switched network. The implementation of this system will be presented in greater detail at this workshop by Dr. M. Jones from ESOC.

On-line Browsing Facility for Quick-Look Images

ESA's EARTHNET program at ESRIN is in the process of implementing a remotely accessible, browsing facility which will allow to inspect on-line certain types of ERS-1 quick-look images and select quick-look products from non-ESA remote sensing satellites. The facility could be considered as an extension of the already existing remotely accessible catalogue.

Images with reduced resolution will be stored in digitised form on magnetic disc, possibly later on optical disc memory of the computer hosting the catalogue/browsing facility. During a catalogue search the image of interest can be transmitted to the remote user terminal and displayed there. This procedure requires of course that the remote user has in his computer facility a matching software package installed and that his terminal has a defined minimum set of graphical capabilities.

In order to offer acceptable response times (image transmission delays), medium to high speed data links, that means satellite links, will have to be used in the communications network. It is planned to use the concept and communications technology, developed for ESA's APOLLO program mentioned, already earlier on. In this concept several (or even many) sporadic users share the use of a high speed satellite circuit by means of a time division multiple access procedure, in order to get the cost per user down to an acceptable level.

There remains some uncertainty whether all potential users will be prepared to pay the still considerable communications cost, once the system enters routine phase.

ESIS Information Dissemination Services

During the requirement definition phase of the ESIS (European Space Information System) project, introduced already above, particular effort was undertaken to ensure, that the envisaged ESIS services really meet the needs of potential future users. These were therefore in various ways involved in the discussion of requirements. The main result was naturally the specification of user friendly and state-of-the services in the area of data archiving, data query, data retrieval and dissemination. Fig. 1 shows the conceptual design of the system.

A more detailed description of the agreed requirements can, in the context of this presentation, not be undertaken.

Another result of this intensive discussions with potential future users was the agreement on advanced "User Interconnection and Information Services", covering the functions of

- electronic mail
- directory services
- file transfer
- remote interactive session

Whereas the requirements for the last two services are rather standard and fully covered by the defined relevant ISO-application services, specific requirements emerged for the first two above services.

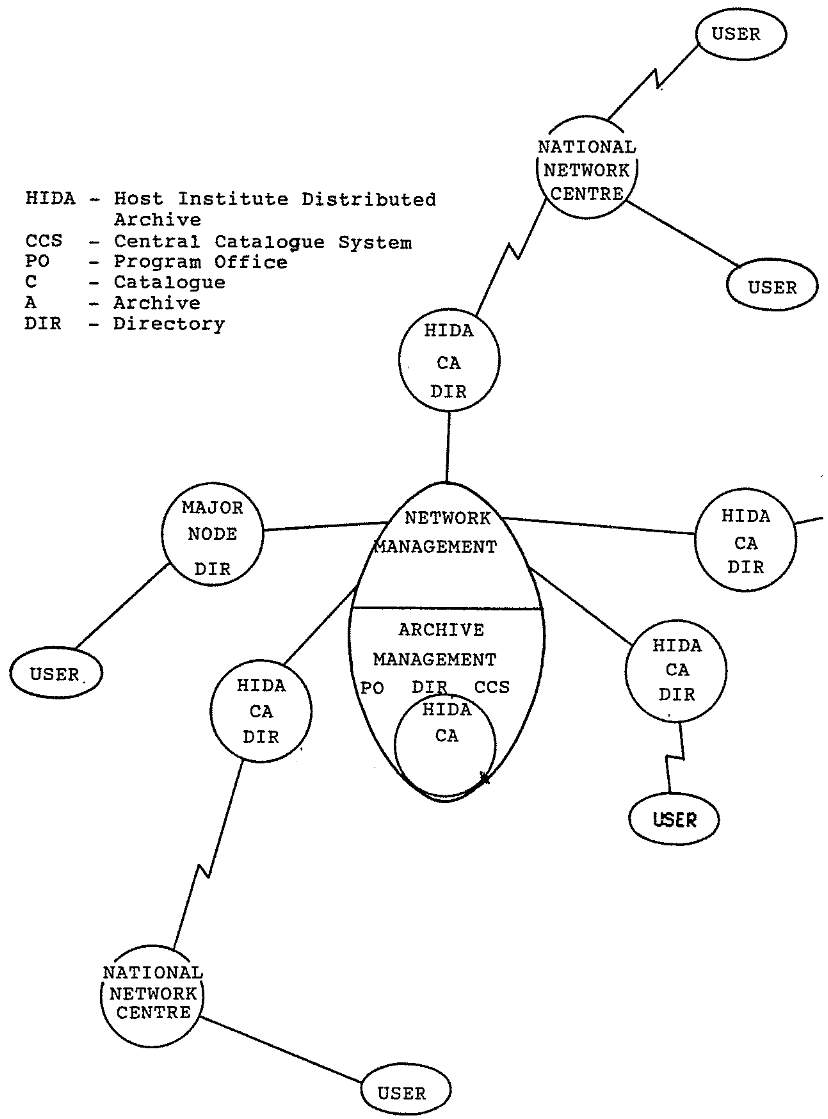


Figure 1 - ESIS Conceptual Design

Electronic Mail

Covering all the services which consist of a message exchange between two or more network users.

The following modes are required:

a) **Person-to-Person Mail**

Requirements for this standard function are generally in line with the ISO/CCITT recommendation. Particular emphasis is placed on elaborate delivery notification.

b) **Bulletin-Board Service**

Collecting information which is of interest to a group of persons (individuals) and offers them the possibility to

- access message directory
- browse through messages
- search through messages for specified arguments/keywords
- retrieve messages (e.g. for local printing)
- deposit messages

All functions should be available in interactive mode and batch mode.

c) **Newsletter Services**

providing the following features:

- to display the list of available newsletters with a short description of each
- to make subscription to a specific newsletter and to cancel subscription
- notification to user that a new issue of a newsletter is available by means of a message providing the abstract or the table of contents.

The newsletter system should be connected with the user directory, introduced below.

Directory Services

providing in general information about other users and available services

a) **User Directories**

maintaining the following information on users:

- user name
- address
- telex, fax and phone number
- electronic address

and providing related interactive capabilities.

b| **Application Directories**

providing information about applications and/or facilities offered by data centres to external users.

c| **Yellow Page Service**

providing classical cross reference service between user directory and application directory (and other directories, if available).

Architectural design studies for implementation of the above facilities are currently ongoing.

COMMUNICATIONS BETWEEN DATA CENTRES: OFF-LINE AND ON-LINE TECHNIQUES

Michael Jones and Julian Sternberg

European Space Operations Centre
Robert Bosch Strasse 5
6100 Darmstadt
West Germany

Introduction

This paper surveys the subject of communications between data centres from the point of view of space data systems. The survey particularly concentrates on the transfer of data to end users of the payloads mounted upon spacecraft.

The discussion is divided into two parts:

- distribution of data using "hard" media is off-line distribution using magnetic tapes and other "physical" media
- electronic data distribution and interchange, i.e. on-line data distribution, which can, depending on system design and capacity provide real-time or near real-time data transfer.

The paper shows that both of these means of data communication will continue to be used in the future, since they provide different, and in many ways complementary services. Thus, the present strong trend towards use of on-line data communications is unlikely to render "off-line" exchange of data obsolete. The discussion covers the various services which have to be provided in practical data distribution systems, even though certain of these are often considered to be part of the "application".

Examples mainly from past missions, are given by way of illustration.

Use of Hard Media To Transfer Data

Conventional Magnetic Tapes

For storage and dissemination of satellite data, magnetic tapes (9-track) have become a de-facto standard medium for almost every past mission. Provided some simple rules are followed regarding the physical blocking of data on tape, such tapes can be made acceptable to all manufacturers' hardware and software, an important point where one centre is supplying data to many different institutes having limited resources. With the trend to high densities (6250 bpi) the reliability of this medium is also improving.

The rules for mapping data onto tape have, in many cases, been extended beyond what is needed to ensure readability, and aim also to achieve compatibility at the *applications software* level. The most successful example of this is perhaps the FITS (Flexible Image Transport System) format for astronomical data arrays; its objective is to define and restrict the ways in which one may write commonly occurring data structures to tape, in order to make the tapes acceptable to many software packages and installations (Ref 1).

The distribution of telemetry data on tape is illustrated by a case study for the recent ESA missions EXOSAT and HIPPARCOS. EXOSAT (in orbit from 1983 to 1986) was an X-ray astronomical "observatory" mission with about 300 participating principal investigators, some of whom had not been closely involved with the project prior to the acceptance of their observation proposals. A greater burden was therefore placed on ESA to provide support to data recipients in the form of background documentation, tape format descriptions, and some scientific analysis software. The EXOSAT detectors and the on-board software to support them were in some ways testing a new technology, and this inevitably led to improvements being made during the mission--changes in operational procedures, on-board software, ground software and data formats.

HIPPARCOS (the European astrometry mission to be launched in 1989) is, by contrast, a "non-observatory" mission, with a very small number of already expert participating institutes, which will not receive direct assistance from ESA in the main scientific data reduction. Its on-board software is designed specifically for the unique purpose of this mission--thus there should be no reason to modify or exchange it – unlike for EXOSAT, where the choice of on-board software modes (and resulting data formats) could only be made with reference to the purpose of each specific observation.

Although these two missions are so different as concerns the exploitation of their payloads and data, it is quite remarkable that the data products for both of them are generated with a common methodology. Its main elements are:

- high degree of automation and security
- tape production is *off-line*, i.e. not tightly synchronised to the real-time process of telemetry acquisition/recovery
- avoidance of irreversible conversion of raw data as far as possible
- output tapes used for two distinct purposes: archive and despatch to users
- telemetry data are sorted and catalogued on behalf of the user

The "sorting and cataloguing" is performed in order to cope with the increasing flexibility of data structures for more recent missions. An example of flexibility is the generation of different structures according to the selection of on-board software modes or parameters, or indeed the use of packet telemetry itself, which (without some assistance) would be rather more difficult for an end user to handle. The objective is to convert the data into a set of fixed-record-structure files, together with a single catalogue or directory to describe the files and their relationship to the payload status.

In future missions, particularly those of the observatory type, it is likely that the standard data products will include the results of some scientific analysis. This was (intentionally) not the case for EXOSAT. *Raw* products (or with limited processing) can be *safely* generated, but may be unwelcome to a non-specialised guest observer because they require higher level of expertise in the particular mission. *Processed results*, on the other hand, are desirable in principle, but their generation is unsafe in the sense that it may not at first be performed optimally and may have to be repeated or corrected later, with a consequent cost impact.

Optical Discs

Although conventional magnetic tapes are well established and continue in regular use, the trend towards satellite missions with higher data rates and/or longer durations has led to a search for more compact data storage media. The Hubble Space Telescope, with a projected lifetime of 15 years is an example here; to store all its data products on normal tapes, while still in principle possible, would be extremely inconvenient.

Optical disc storage was the first promising candidate and has been used experimentally for about a decade. The discs can store typically 1-2 Gbytes, i.e. the equivalent of in the order of 10 6250 bpi magnetic tapes. The discs themselves are cheap, compact and durable.

Unfortunately, agreement on hardware and software standardisation does not seem likely to arrive in the near future, and the idea of optical discs as a freely exchangeable medium is premature. The "compact disc" formats (CD-ROM) *are* standard, but they must be written by the disc manufacturer and so are not appropriate for dissemination of satellite data.

The "write once" nature of the discs puts constraints on the software to handle them and the ways in which data may be encoded. Clearly new approaches to software will be necessary, possibly using more conventional media (magnetic disk, magnetic tape) as intermediate storage.

In view of the above, it is concluded that optical discs *can* be used to advantage within one centre, for example as an archiving medium, but for dissemination to other centres magnetic tapes will continue to play the principal role for the near future.

Other Magnetic Tape Media

The high storage capacity of optical discs is matched and even exceeded by newer formats of magnetic tape--cartridges, cassettes, digital audio tape. A capacity of about 50 Gbytes is available in one case (BTS "DRS 100"). In the long term, tape media have a better potential than disc media for developing common standards and for maintaining compatibility with existing applications software (e.g. through the FITS standard) and at first sight they hold great promise.

However, the present variety of tape formats is sufficient to discourage most scientific institutes from investing in these media at a time when evolution is still rapid.

Moreover, they tend to suffer from the following disadvantages with respect to conventional magnetic tapes:

- slow data transfer rates
- unreliability when used for *long-term* storage

The last-mentioned point is crucial for a data archiving centre. Even for conventional tapes it is necessary to rewrite the contents periodically, perhaps every 2 years, in order to ensure continued readability; a shortening of this time or an increased risk of data deterioration is highly undesirable.

Nevertheless, the development of these newer media will be followed with great interest, and it is expected that they will eventually play an important role.

On-line Data Exchange and Distribution

Discussion

For space data two main trends are apparent viz:

- increasing demand from users for on-line, near real-time, availability of their data or a portion of it.
- telescience: this is a 'two-way' process involving the user receiving data from his on-board experiment and sending commands to it, the latter being in general conditioned by examination of the experiment data received. Clearly for those experiments where telescience is genuinely necessary, the response times will need to be adequate.

In fact these trends reflect the burgeoning use of data communications, which has resulted from increasing availability of communications services (together with supporting hardware and software) coupled with falling transmission costs.

However, use of on-line communications brings in its wake certain problems which may have to be taken into account. The three main problems are:

- Security: i.e. prevention of unauthorised access to the communications system and via this to computers and data bases connected to it.
- Data Back-up and Archiving: electronic transfer or broadcast of data does not remove the need for archiving it.
- Cost: despite falling communication costs, electronic transfer of large amounts of data can still be very expensive and therefore may not always be the correct solution, particularly since it may not be possible to process the data at a commensurate rate to that at which it is received.

In space data systems, all of these problems must be trackled thus:

- Security: a variety of methods exists to icnrease security e.g. subscriber authorisation techniques, data encryption, separation of critical (e.g. operational systems) from data distribution systems either (a) by physical separation or (b) by using separate physical connections and protocols to connect, (i) data distribution system and users and, (ii) operational systems (to control spacecraft/payload) and data distribution system.
- Data Back-up and Archiving: a data back-up or archive has to be provided so that users can retrieve data which they fail to acquire by electronic means. This could be the control centre, a specialised archiving facility or designated users with the appropriate resources.
- Cost: data should not be transmitted on-line unnecessarily, particularly where large data volumes are involved. A solution used for scientific data is to send "key parameters", which would be sufficient to indicate the presence of interesting "events" as a means to allow selection of full resolution data. As remarked earlier large volumes of raw data should not normally be distributed electronically. This is better done on magnetic tapes, optical disks or some such suitable medium.

Examples of On-line Data Distribution

Three examples of on-line data distribution are discussed in the following subsections:

- METEOSAT, which has been in operation since 1977 offering a number of relatively sophisticated data distribution services, even by today's standards.
- the EURECA data disposition system: this is currently in the early stages of its development and will be put into operation during the first EURECA mission, currently foreseen for 1991.
- ESA's In-Orbit Infrastructure, which is foreseen to start its operations in the late 1990s

These examples cover a sufficiently broad scope to be able to bring out a number of main issues of on line communications.

Example 1: METEOSAT

The METEOSAT Operational Programme1 is, in principle, based upon a geostationary spacecraft stationed at approximately 0 degrees N, 0 degrees E. This spacecraft is equipped with:

- a multi-spectral radiometer taking images of the Earth in the visible and infrared bands of the electro magentic spectrum
- transponders to support retransmission of image data and other data from the ground to users with suitable earth terminals
- a transponder to collect data from so-called "data collection platforms" (DCPs) and retransmit it to the control centre for subsequent distribution.

The control centre is at the European Space Operations Centre (ESOC) at Darmstadt (FRG) together with the Meteosat Ground Computer System (MGCS) and the Meteorological Information Extraction Centre (MIEC).

Leaving aside the DCPs, METEOSAT has two forms of on-line data dissemination:

- Images are acquired in the computer system at the control centre (the MGCS), preprocessed, split into "formats", i.e. image segments, and then retransmitted from the ground station to a transponder on the spacecraft. The formats may be received by users with suitable earth terminals.
- "Products" such as wind vectors and sea surface temperatures and cloud top heights (7 product types in all, see Ref 2), are extracted from the images associated with certain "synoptic" hours. These are transmitted over a meteorological network of data lines

METEOSAT thus has the following characteristics:

- it has a sort of "pipeline" processing system on ground i.e. images are received each half hour, pre-processed, split into formats and then disseminated via METEOSAT according to a schedule. This is a continuous process; similarly, data products have to be produced by certain regularly occuring deadlines (the principal ones occuring twice every 24 hours).
- it primarily serves *operational* users (national weather forecast services): these will normally have the necessary infrastructure to acquire and exploit the data sent to them (i.e. equipment, 24-hour support etc.)
- an archiving service is provided at the control centre (Ref 2), in which all image data, auxiliary data and image products are stored. This service is frequently used by non-operational users (e.g. research institutes).

1 A further description is given in Ref 2 and in the papers referenced therein.

Example 2: EURECA Data Disposition System

EURECA (European Retrievable Carrier) is a low earth orbiter, due to the launched by the US Space Shuttle in 1991. It is in the nature of a space laboratory containing 15 experimental facilities over 50 experiments can be operated.

.Its main mission characteristics1 are as follow:

downlink data rate: 512 Kbits/s
prime ground station for control and data acquisition at Maspalomas, Canary Islands, Spain
No of passes/day over the ground station: 5-6
average pass duration: c.a. 6 mins
average data collection rate: c.a. 20 Mbytes/day
packet telemetry: data from individual instruments and the "platform" is put into identifiable packets according to the Packet Telemetry Recommendation (Ref 2) of Consultative Committee for Space Data Standards (CCSDS)

The EURECA mission is quite different to METEOSAT in that it is a so-called "PI mission", with a set of Principal Investigators (PIs) for whom the various experiments will be operated. Indeed the PIs will provide some of the experiments, ESA providing certain "core facilities" of the payload (Ref 4). Payload operations will be carried out according to a complex schedule which takes account of the wishes of the PIs and their associates as well as the available on-board resources (e.g. power).

1 See Ref 4 for more details.

Unlike METEOSAT dissemination, which is an active process, appropriate for the cyclical nature of a meteorological observing system, the EURECA Data Disposition System (DDS), is as its name suggests., passive. It makes the data available, but it is the responsibility of the user to "take" the data.

The configuration concept is shown in Figure 1. This figure shows the separation of the EURECA spacecraft control computer, (called the EURECA Dedicated Control System, EDCS) and the computer which interfaces to the users (called the Data Distribution Computer System, DDCS). Security is provided mainly by using different protocols and data bus systems for operational communications (e.g. EDCS to DDCS) and communications with users (DDCS to users). The DECNET protocol is used between the DDCS (Data Distribution Computer System) and the end users, whereas the native protocol of the Hyperbus is used between the Control Computer (EDCS) and the DDCS. The scope of usage of the relatively insecure DECNET is thus restricted.

The EURECA DDS may be summarised as follows:

- telemetry data is acquired on EDCS from the ground station;
- these data are kept on-line in computer files on the EDCS separated according to the packet identifier (packet ID) which appears in the CCSDS primary header of each packet. This provides a convenient separation of the data of the various users. These files reside on the EURECA spacecraft control computer;
- the user *requests* data transfer (i.e. it is not automatically disseminated), his requests being directed to the DDCS;
- requests and subsequent data transfers use DECNET (of Space Physics Analysis Network (SPAN), Ref 5), both EDCS and DDCS are DEC/VAX computers;
- the main security measure is the separation of EURECA operations and data distribution (i.e. DDS) onto separate computers (see below); requested files are transferred from the EDCS to the DDCS using the native protocol of the Hyperbus (not ETHERNET), which interconnects EDCS and DDCS. Thus the user has no direct access to the operational telemetry files.
- the DDS is used to accept requests from users for payload operations: it thus constitutes the biginnings of a telescience system.

Example 3: The ESA In-orbit Infrastructure and its Ground Facilities

The ESA In-orbit Infrastructure (IOI) essentially comprises:

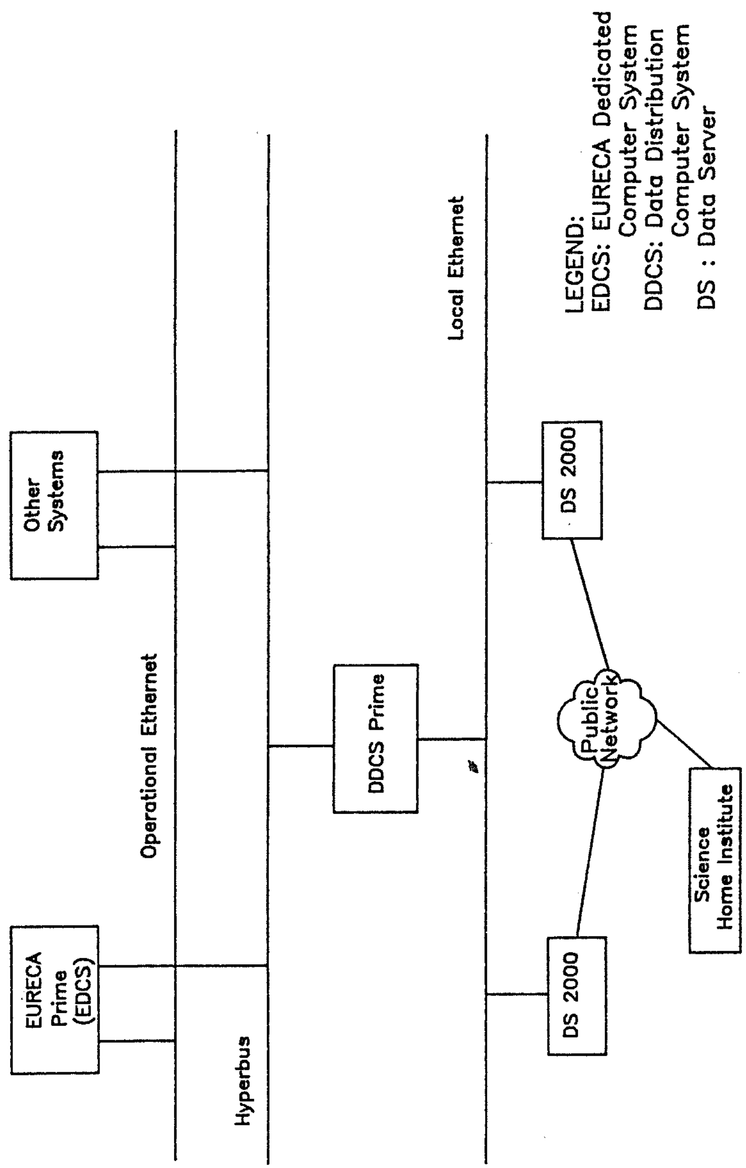


Figure 1 - Configuration Concept: EURECA Data Disposition System.

- the European contribution (COLUMBUS) to the International Space Station, made up of three flight elements, the Attached Pressurised Module (APM), the Man-Tended Free Flyer (MTFF) and the Polar Platform (PPF)
- the Hermes Space Plane
- the European Data Relay System (DRS), supported by two satellites (DRSS-W and DRSS-E) providing in-orbit links to flight elements

The In-Orbit Infrastructure is intended to become operational in the late 1990s and it will have a lifetime of some 30 years.

It has been decided to support the IOI with a *distributed* ground infrastructure comprising so-called Element Control Centres and other facilities (such as ground stations, earth terminals for DRS, Engineering Support Facilities,...). In essence the distribution of control centres is as follows:

HERMES Flight Control Centre (HFCC) at Toulouse France
COLUMBUS Control Facilities:
- Manned Space Laboratories Control Centre (MSCC) at Oberpfaffenhofen, near Munich, West Germany
- Polar Platform Control Centre (PPF CC) in the United Kingdom (to be confirmed)
- Central Mission Control Centre (CMCC) in Darmstadt, West Germany

Data Relay Satellite Operational Control Centre (DRS OCC) in Fucino, Italy

In addition, US facilities will also be involved e.g. the Space Station Control Centre will control the APM, which will be permanently attached to the Space Station.

One important characteristic of this ground segment is that the control centres are not completely independent. They may have to support combined operations involving more than one flight element. Moreover, because of the safety aspects associated with manned space flight some form of back-up of centres may be necessary, which also implies a coupling (e.g. mutual back-up).

It is in the first place the communications system which will make the IOI into a unified system. The communications system must handle digital, voice and video data, the two last-named being particularly important for astronaut communications and operations.

The main problems of communications of such a distributed set of facilities:

- to ensure availability of a certain set of communications services to subscribers at all major centres
- to provide required performance for transporting data needed for control of the flight elements (c.a. 2 Mb/s)
- to give a telescience service to users (in particular owners or operators of payloads mounted on flight elements)

On particular problem is to achieve an appropriate balance of centralised and decentralised functions. What must be avoided is the interconnection of a set of dissimilar networks each of which offers non-equivalent services, since this will be inefficient and unreliable.

Figure 2 gives an impression of the overall structure of the communications system. It is based upon:

Mesh network

High performance Interconnection Subnet (a 'backbone' network) linking the major centres

Subnets at major centres (e.g. to connect users and other facilities associated with each centre)

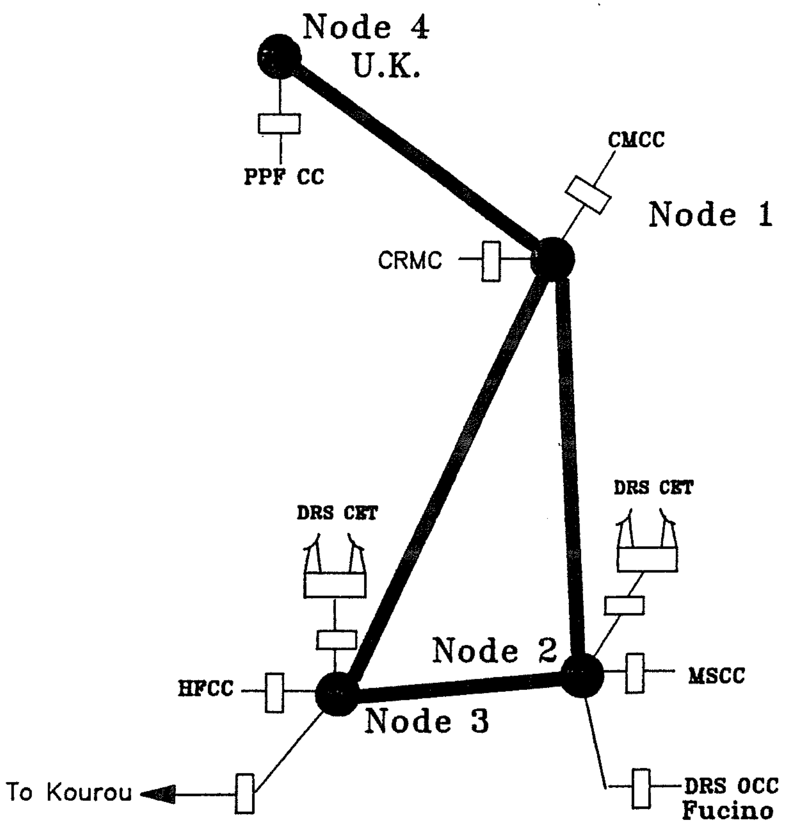
Hierarchical control structure in which
- each subnet has its own Network Control Centre
- the Network Control Centres lower-level subnets are subservient to those of higher level subnets

Common equipment/protocols on main Interconnection Subnet

High-rate (tens or hundreds of Mb/s) acquired directly from Flight Elements normally via DRS and a terminal local to the point of acquisition of the data

Facilities providing services to more than one centre (e.g. DRS Central Earth terminals) are connected directly to the main Interconnection Subnet to ensure availability to standard network services subscribers communicating with these facilities.

The concept outlined above is necessarily sketchy and incomplete. It is what emerged from a concept study plan carried out by ESA staff and staff from National Agencies of the Member States. This phase precedes the so-called Detailed Definition Phase (in effect the "Phase B" due to start in late 1988 or early 1989) and it remains to be seen if



CMCC: Central Mission Control Centre
HFCC: Hermes Flight Control Centre
MSCC: Manned Space Laboratories Control Centre
DRS: Data Relay System
DRS OCC: DRS Operations Control Centre
CET: Central Earth Terminal
CRMC: Communications Resources Management Centre
—[]— Gateway

Figure 2 - Schematic of Ground Facilities of In-Orbit Infrastructure (Simplified)

this approach will continue to be followed in this and the ensuring phases of the ground segment development.

It is noted that:

- the security problem has not yet been addressed for the IOI, although it will be important in such a distributed system involving safety-critical operations
- archiving of payload data will be the responsibility of users (e.g. owners or operators of payloads)
- large volumes of payload data should not be transferred over the ground communications network: they should either be acquired directly by the user, (via his own DRS terminal) or at a Central Earth Terminal and then transmitted onwards on transportable media.

Conclusions

This paper has briefly surveyed both off-line and on-line systems for communication between data centres. The conclusions may be summarised thus:

General

Despite the strong increase in the use of telecommunications systems, the use of 'hard media' such as magnetic tape and its successors is likely to remain important for bulk data transfer, particularly when that data must be subjected to complex and lengthy processing after transfer (i.e. when response time is not a significant factor).

Off-line Techniques

Despite interesting and significant developments in new media (optical disks, cartridge tapes and digital audio tape) conventional magnetic tapes are likely to remain in use for some time yet.

A number of reasons for this have been identified, but the most important ones are possibly:

- lack of mature and/or predominant industry standards
- user commitment to developed infrastructure (e.g. software for supporting tape-oriented standards such as FITS)

On-line Techniques

It was noted that on-line provision of data, probably in near real-time, is being demanded more and more by users. A number of issues were identified:

- security
- need to archive and back up data
- cost

The way in which these issues arose in a set of three case studies was examined.

References

1. *FITS: A Flexible Image Transport System, Wells, DC., Greisen, E.W. and Harten, R.H. Astronomy and Astrophysics Supplement, 44, 363-370 (1981)*

2. *The METEOSAT Archiving and Retrieval System*, Mazza, C., (July 1988) Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988 pp. 231–238.

3. *Packet telemtry recommendation*: "Blue Book", Issue 1, (May 1984) (Consultative Committee for Space Data Systems (CCSDS))

4. *EURECA in the COLUMBUS scenario*, Andresen, R.D., and Nellessen, W., Space Technology, 8 No 1/2, 45-57 (1988)

5. *Cooperation between Data Centres: Interoperability*, Jones, M. Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988 pp. 290–304.

NSSDC DATA HOLDINGS: SCOPE, COMPLEXITY, AND FUTURE REQUIREMENTS

James L. Green

National Space Science Data Center

Greenbelt, MD 20771, USA

Abstract

The National Space Science Data Center (NSSDC) has acquired over 6,000 gigabytes of digital data and several million feet of film products from NASA space and Earth science missions since it was established in 1967. In order 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 on-line database management machines; and optical mass storage devices; and it manages NASA's largest computer-to-computer wide area network.

The NSSDC data holdings cover the major science disciplines of Earth (land, oceans, atmosphere), space plasma, planetary physics, and astrophysics. In addition to handling NASA-acquired astronomy data, the NSSDC cooperatively manages (with the Space Data and Computing Division) the Astronomical Data Center (ADC) and distributes over 500 astronomy catalogs of celestial objects.

Based on current agreements with upcoming NASA missions, the future NSSDC data holdings will increase dramatically, approximately doubling every 2 years, reaching nearly 30,000 gigabytes by 1995. Innovative ways of extracting the most pertinent information about such large volumes of data and implementation of large mass storage systems will be discussed. This paper will provide an overview of the scope and complexity of the NSSDC holdings and how it will be addressing the ever-increasing volumes of data that it will be archiving.

Introduction

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

The NSSDC manages the World Data Center A for Rockets & Satellites (WDC-A-R&S). Under the World Data Center agreements, the NSSDC archive has generally been open to the international community of scientists and its personnel are participated in the international exchange of scientific data.

To carry out this charter, the present role of the NSSDC includes data archiving, preservation, maintenance, data life cycle planning, and systems development. In addition, the NSSDC serves science users through finding, retrieving, and distributing data. This paper will describe the scope and complexity of the NSSDC holdings and will discuss the future requirements for handling the large amounts of data that are expected to be archived.

Data Acquisition

The NSSDC usually receives data from NASA principal investigators or directly from NASA projects where facility instruments are being flown. The establishment of the NSSDC as a long-term archive is done through the NASA Management Instruction (NMI) 8030.3A. This NMI requires that all NASA projects write a Project Data Management Plan (PDMP) that details, for instance, what data products will be archived. The NSSDC also obtains data from other government and international agencies usually involved in space research. International agreements with NASA and other governments allow the NSSDC to archive and distribute data from upcoming missions such as ROSAT (German x-ray spacecraft) and San Marco (Italian spacecraft).

Scope and Complexity of NSSDC Holdings

The NSSDC, which was established in 1967, is the largest archive for processed data from NASA's space and Earth science missions. The NSSDC manages over 4,000 unique data sets. The size of the digital archive is approximately 6,000 gigabytes (6 terabytes) with all of these data in their original, uncompressed form. Over the last three years, the NSSDC has taken in approximately 250 new data sets per year. In addition to NASA-acquired data, the NSSDC cooperatively manages (with the Space Data and Computing Division) the Astronomical Data Center (ADC) and distributes over 500 astronomical catalogs.

The bulk of the NSSDC archived digital data is stored on 120,000 magnetic tapes, forming an "offline" archive. Out of this total archive, the 7-track low-density tapes number approximately 37,000. The 7-track data continued to come to the data center for archiving from principal investigators until 1982.

Currently, the NSSDC is migrating many of its older data sets, stored on the low-density data tapes, to higher density tape and tape cartridges (see Data restoration Program section). The tape cartridge system will allow the rapid migration of data onto other forms of online and offline mass storage.

Nearly 100 gigabytes of archived data have been loaded onto an online mass storage system. These data comprise all the International Ultraviolet Explorer (IUE) data and astronomical catalogs (Green, 1988b). The online data are used extensively and provide extremely rapid response to requests.

Two types of optical disks, Write Once Read Many (WORM) and Compact Disk Read Only Memory (CD-ROM), are also used at the NSSDC. A number of data sets are coming into the NSSDC on WORM disks with data from several instruments on the Nimbus-7 and Dynamics Explorer spacecraft. The total volume of these data is tens of gigabytes, but it is expected to reach well over 500 gigabytes by the end of this year (King, 1988).

The NSSDC also manages a very large photo and film archive and many geophysical computer models. The NSSDC photo/film archive consists of over 2 million feet to film, 41,000 sheets of microfiche, atmosphere, ionosphere, magnetosphere, and trapped radiation environment, just to name a few.

Requests and Data Distribution

From the time it was established in 1967, until 1985, all requests for NSSDC-held data were in the form of letters, phone messages, or personal visits to the data center. During this period, "offline" management of the archived data at the NSSDC was employed. Currently, requests for offline archived data typically take 2 weeks (if the data are held locally at the NSSDC) and, in some cases, up to 1 month (if the data come from our offsite storage facility). Satisfying these offline requests is a labor-intensive process. The charge for obtaining data from the NSSDC is usually for replacement costs in materials and supplies (example: a blank tape or equivalent is needed for one tape worth of archived data).

Figure 1 is a bar graph of the number of NSSDC requests actions per year from 1983 through the present time. The data actions that are displayed in this figure are the number of offline requests, tapes sent, interactive sessions, network data transfers, and optical disks sent. As discussed above, prior to 1985 only offline requests were accommodated by the NSSDC. Since 1985, the number of interactive sessions and network data transfers has rapidly increased. For the year 1988, the bar chart shows twice the 1/23 year statistics so that it can be compared with the previous years. For the year 1989, the bar graph shows only a projection of what data actions are expected, based on extrapolating the recent request rate. Figure 1 shows the steady increase of the interactive use of NSSDC online information systems and the key role that networking is starting to play in the transmission of data to the remote investigator. The use of any of the major networks for transmission of data is limited to only small data segments (usually less than 1 megabyte per transfer). The steady increase in the network access to the NSSDC is a combination of a steady increase in the number and capabilities of the information systems that are made operational, the increase in the number of scientists using networks, and the increase in the number of the network connections (Space Physics Analysis Network or SPAN in 1985, Public Packet Switched Network in mid-1986, Bitnet in late 1986, and the TCP/IP Internet in early 1987). What is not shown in Figure 1 is over 20,000 copies of NSSDC catalogs, NSSDC special reports, WDC-A-R&S SPACEWARN Bulletins, and NSSDC news documents that are sent out on request each year.

The distribution of data from the NSSDC includes the conventional tape, hard copy, photo, and floppy disk, but within the last few years an ever-increasing amount of data is moved to the remote requester over electronic networks such as SPAN and the National Science Foundation Network (NSFnet). The NSSDC now has several CD-ROMs with a variety of geophysical and planetary data that will be distributed. It is expected that the optical disks, WORM and CD-ROM, will be a popular distribution media. In addition, the WORM disks will also be appropriate for archive storage of data at the NSSDC and will play an ever-increasing role. As shown in Figure 1, the NSSDC expects to distribute over 700 optical disks in 1989.

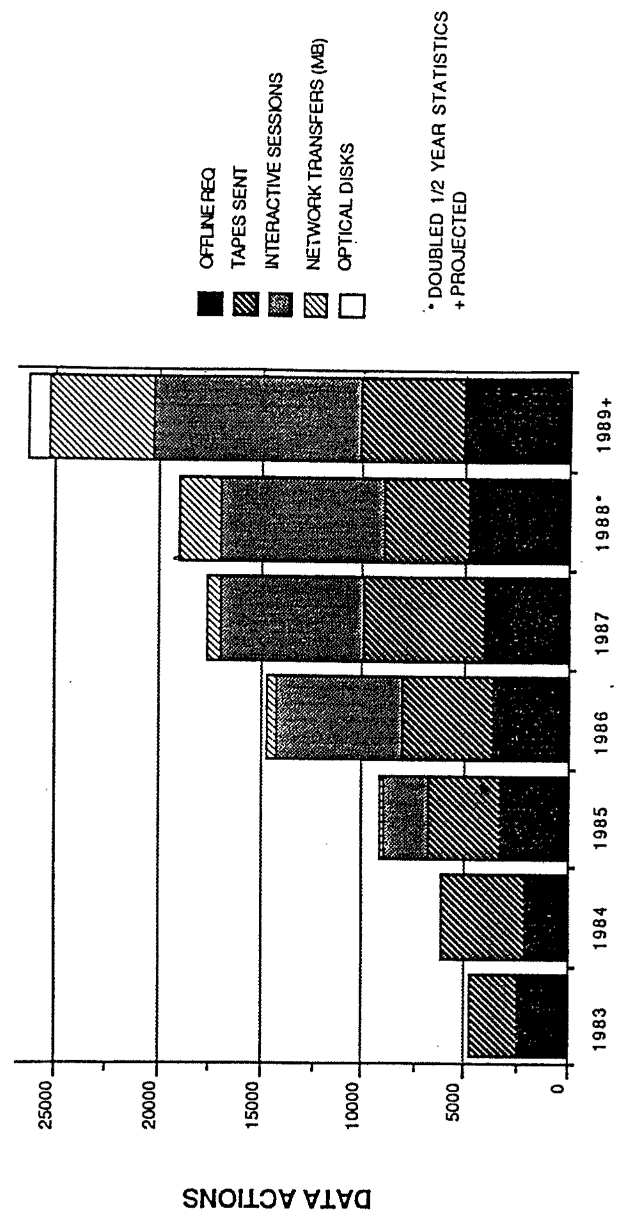


Figure 1 - This bar graph shows the number of NSSDC request actions per year from 1983 through the present time. For the year 1988, the bar chart shows twice the 1/2 year statistics so that it can be compared with the previous years. For the year 1989, the bar graph shows only a projection of what data actions are expected based on extrapolating the recent request rate. The data actions that are displayed in this figure are the number of offline requests, tapes sent, interactive sessions, network data transfer, and optical disks sent.

Figure 2 is a pie chart illustrating the breakdown in the type of requesters of data from the NSSDC. The majority of requests for data and information are from the international and U.S. university science community, followed by the general public and U.S. industry.

Data Restoration Program

Over the last year, the NSSDC initiated a program to review the holdings of its archive in all effort to identify what older data sets need to be migrated to new storage technologies for long-term retention. We have referred to this effort as the "data restoration program", since these efforts have breathed new life into many of the older data sets that were threatened from loss from older media.

The current data restoration program began as a pilot effort and was envisioned to be a pathfinder for developing procedures necessary for the identification and long-term data management of key NSSDC holdings. The atmosphere/meteorology field was chosen as our first discipline effort, since there were a limited number (about 100) of data sets to evaluate that encompassed a large fraction of our older archive (about 40,000 digital and analog data tapes). With advice from NASA Headquarters, the NSSDC selected five atmospheric and meteorological scientists that had an extensive background in the use of diverse data sets for scientific analysis to serve on an advisory panel.

To aid in the evaluation, a questionnaire was sent out to the investigators and science team contacts who originally supplied the data to the NSSDC. We obtained approximately 70% replies to the questionnaire. After forming the final panel in January 1988, the results of the survey and descriptive material on the data sets were provided and evaluated. The initial face-to-face meeting occurred in April.

The advisory panel prioritized the restoration effort such that the key older data sets would be rewritten and backed up first. Since May 1988, the NSSDC has restored approximately 1,200 tapes, including 10 complete data sets. The recovery effort has been time consuming because of such factors as logistics of retrieving the data from the NSSDC auxiliary storage location and hand-cleaning many of the tapes before reading them. It is interesting to note that there has been a reasonably low error rate encountered in reading the older data sets.

Recently, the restoration activity has dramatically increased in scope with the use of facilities at the NASA Space and Earth Science Computing Center (NSESCC). Within the next month, the copying of the data will be onto the new IBM 3480 tape cartridge systems that have recently been installed at the NSESCC. It is our goal that over 15,000

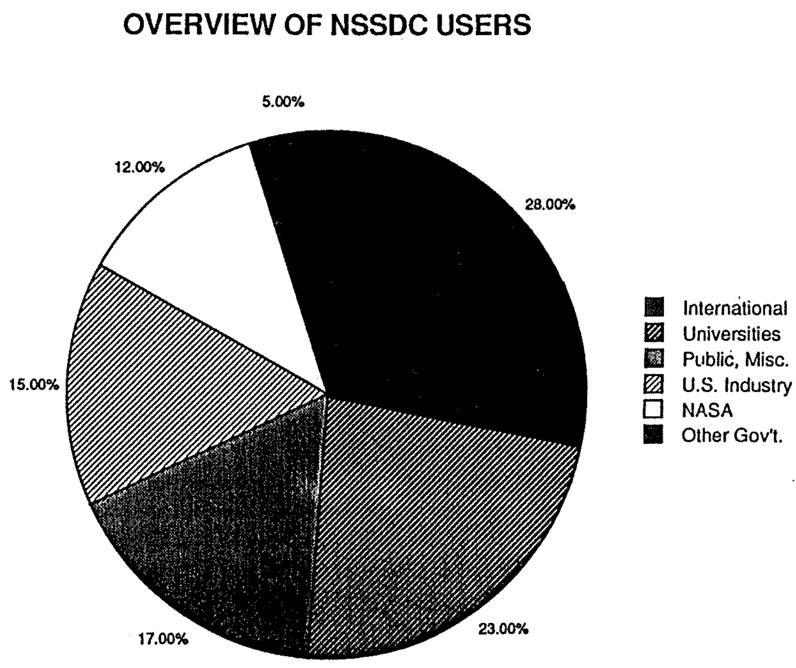


Figure 2 - This pie chart shows the type of requesters for data from the NSSDC. The majority of requests for data and information are from the International and U.S. university science community, followed by the general public and U.S. industry.

additional tapes will be restored (copied) within the next year onto the tape cartridge system. Plans are under way for the inventory of the restored data to be loaded into a new, advanced, interactive data management system (which is just now under development) and placed online for use by the general scientific community.

Future NSSDC Archive

Future NASA missions will take full advantage of new instrument and communication technologies. The combined effect of these technologies will increase the amount of data taken by 1995 by over 2,000 times its present rate. In 1985 NASA was acquiring approximately 360 gigabytes of data per year. Assuming both currently approved and most likely approved NASA missions, the acquired data volume by 1995 will reach well over 2,400 gigabytes per day (Carper, 1987). This a staggering rate. Correspondingly to the dramatic increase in the amount of data received from spacecraft, the size of the NASA data archives must also dramatically increase.

Figure 3 shows only the digital data volumes per scientific discipline that have been archived and are expected to be archived at the NSSDC. As discussed above the NSSDC has currently about 6,000 gigabytes of data. The size of the archive past the year 1988 is a projection and considers the current arrangements with the NSSDC and the missions that are currently approved through the PDMP agreements. If this projection holds true, by 1995 the NSSDC will have over 28,000 gigabytes in its archive, more than quadrupling its current size. In planning for the upcoming volumes of data to be archived, the NSSDC is developing innovative ways of dealing with the storage and long-term management of digital data (King, 1988, and Green, 1988b).

STANDARDS – One of the most important aspects of long-term data management that must be addressed is the issue of standards. These standards must encompass media, data, and information formats. Not only will standards facilitate the exchange and analysis of data, but they will be of tremendous benefit to an archive by allowing a cost-effective software framework to be constructed that would enable the rapid ingest of incoming data. The NSSDC has recently established a standards office to begin to promote the use of current standards and participate in the development of new ones.

ONLINE INFORMATION SYSTEMS – 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. The NSSDC has developed several online information systems

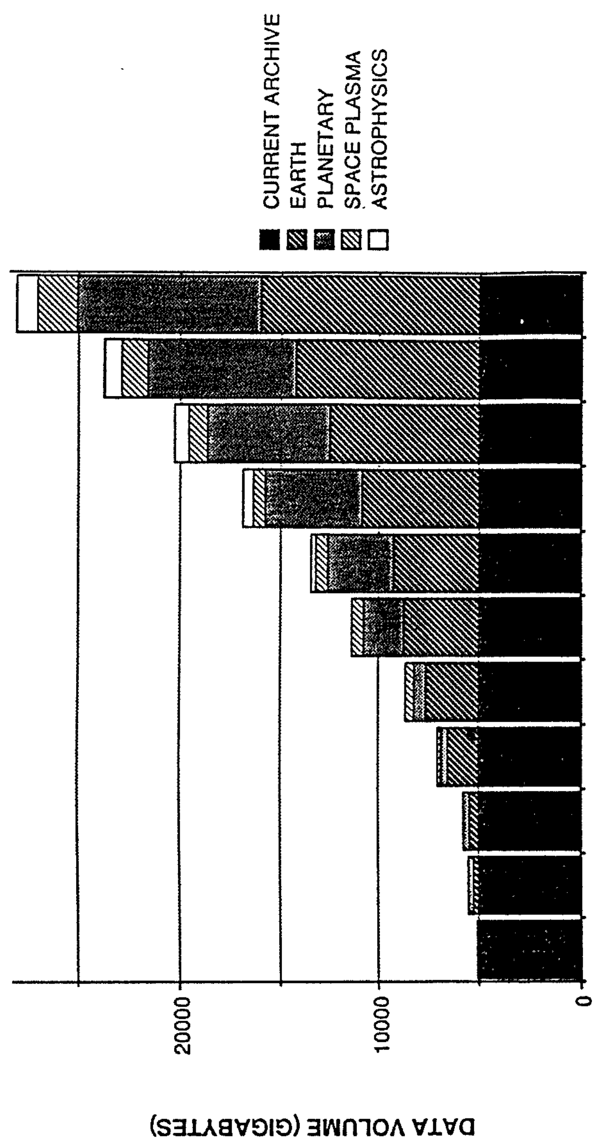


Figure 3 - Data volumes per scientific di cipline that have been archived and are expected to be archived at the NSSDC are shown in this figure. Today the NSSDC has about 6,000 gigabytes of data in its digital archive. The size of the archive past the year 1988 is a projection based on approved NASA projects. By 1995, the NSSDC will have over 28,000 gigabytes in its archive, more than quadrupling its current size.

describing its archive (Green, 1988b). These online information systems continue to grow in capability and complexity, forming a vital link with users worldwide. In addition, some of these systems should be able to automatically ingest data (in certain formats) that have just arrived in the archive. The auto-data ingest capability will be essential if the NSSDC is to have a chance at keeping up with the ever-increasing volumes of data.

ONLINE/NEARLINE DATA STORAGE—More and more of the new data sets coming into the NSSDC will be stored on optical disks, tape cartridges, and other nearline mass storage systems (King, 1988). This capability allows for the rapid promotion of the data to satisfy a request either directly from a user or by one of support personnel at the NSSDC. Within the next few months, the NSSDC will be distributing its first set of CD-ROMs (planetary data generated through the Planetary Data System Program). With a back order of over 80 requests for these CD-ROMs and the outlook of several more key data sets being committed to CD-ROM (King, 1988), this type of storage technology will be used extensively.

ARCHIVE BACKUP—As put of new NSSDC mass storage initiative, plans are now being devised at the NSSDC that, once in place, will provide for a complete "Safe storage" or backup of the NSSDC digital archive. The concept behind safe storage is ability to copy the NSSDC digital archive (or the most important part of the archive) and physically locate that backup in another location. With such large volumes of data to back up, a cost-effective solution requires an extremely dense media with a very small cost per megabyte of storage, a high data transfer rate, and adequate data compression schemes to further reduce the volume.

DATA COMPRESSION—Data compression techniques may be a very important tool for managing a large fraction (perhaps all) of the NSSDC archive in the future. Some applications would consist of compressing online data in order to have more data readily available to remote users, compressing requested data that must be moved over low rate ground networks (decompressing the data at the destination), and compressing all the digital data in the NSSDC archive for a cost effective backup that would be used only in the event of a disaster. Of these, currently, the NSSDC is distributing compressed data (however, uncompressed is still the most requested form) from the International Ultraviolet Explorer over the SPAN network (Green, 1988a). In the case of the archive backup, even though scientists are reluctant to provide NSSDC with compressed data for distribution, there is little argument against data compression techniques being applied in order to provide for a const-effective backup of the entire digital archive.

It is easy to predict that the archive of the future will change significantly, in the way it manages, locates, and distributes its data. The NSSDC will have to effectively manage huge data volumes and provide response to user requests for data in a timely manner. In addition to the conventional types of distribution, it is anticipated that much more data will be distributed over networks and CD-ROMs will dramatically increase in popularity.

What today's archives must be able to do is to continue to appraise and apply new technologies, or the amount of data they will be able to maintain will be severely limited in comparison to the rates at which new measurements are being made.

Acknowledgement

I would like to gratefully acknowledge Karen Satin who has done an expert job in technically editing this article and to my secretary, Margie Layfield for her accurate and rapid typing of the manuscript.

References

Carper, R., J. Dalton, M. Healey, L. Kempster, J. Martin, F. McCaleb, S. Sobeiski, and J. Sos (editors), *Mass storage systems for data transport in the early Space Station era 1992--1998*, NASA TM 87826, July 1987.

Green, J.L., *The Space Physics Analysis Network, Computer Physics Communications*, North-Holland, Amsterdam, 49, 205-213, 1988a.

Green, J.L. *NSSDC Information Systems: Status and Plans,* Proceedings of the Workshop on Geophysical Informatics, Moscow, 1988b pp. 24–32.

King, J., *Mass storage technologies for NASA's future data environment, to be published in the proceedings from the Workshop on Geophysical Informatics*, Moscow, 1988.

MAINTENANCE OF DIGITAL AND ANALOGUE DATABASES

M.A. Hapgood

WDC C1 for STP
Rutherford Appleton Laboratory
Chilton, DIDCOT
Oxfordshire, OX11 OQX
United Kingdom

Introduction

This paper presents some ideas arising from the experience of building and operating databases at the WDC-C1 for Solar-Terrestrial Physics.

Digital databases

Quality control

It is important to develop methods of operation that minimise data errors. At our Centre we have developed a number of procedures to assists this work:

The most important point is to ensure that no stigma attaches to staff when they make occasional errors loading data into databases. Staff should be encouraged to report such mistakes so that the errors can be corrected.

It is important to build checks into the software used to load and update databases. Tests can include checks on data type and on the range of validity of data. The latter can be divided into absolute checks

(e.g. rejecting impossible values such as negative numbers for quantitites that must be positive) and conditional checks (warning of unusual values outside the normal range).

Another useful check is the regular generation of standard products from the databases, e.g. graphical and tabular summaries. If operated by scientific staff, these can provide an independent and regular review of data quality.

Scientific staff must be responsible for quality control and must be given adequate time to do this work.

Storage Media

It is vital that digital data are copied to new storage media in a timely fashion. It is important to get this timing right. Clearly one must not wait for current media to become obsolescent (our centre, in common with many other centres, has had recent experience of rescuing data from punch cards and 7-track tapes just before loosing the ability to read those media). At the other extreme, we must be careful when using state-of-the-art systems. It is essential to pick systems that will have a reasonably long operational lifetime (at least 5 years). Perhaps the best criterion for managing archives is to copy data to new media once firm standards emerge and are implemented by a number of commercial suppliers.

Analogue data

Why do we retain the use of analogue data? Clearly digital data are better from several points of view: (a) it is easy to use large amounts of digital data; (b) it is easy to copy them; and (c) it is easy to generate analogue products (Tables, graphs, etc.). In addition, analogue data are seen to be old-fashioned and therefore there may be some prejudice against such data.

However, much relevant old data are only available in analogue form, especially data required for the study of long-term changes. Also, some data are still produced only in analogue form. Therefore, we need to keep useful analogue data.

In our centre we adopted a number of rules for retaining analogue data:

- they must be capable of interpretation. That is, they must be readable, identifiable (time, location, calibration,...) and understandable by staff. The latter point is often the most stringent test since the skills needed to use the data may have been lost by staff retirements.

- it is important to give priority to the retention of database of high scientific value, such as data from observatory networks with large spatial and temporal extent (ionosondes, magnetometers,...).
- data centre staff must be able to justify the retention of old data when challenged.

It is useful to consider measures to improve the storage of analogue data--especially by reducing their volume. There is often pressure to reduce the space occupied by data centre archives. One important measure is the use of microfiche: (a) to produce new analogue data on microfiche; and (b) to copy old data to microfiche – preferably using silver films, which have good archival properties (unlike dye-based films). In the long-term we would like to digitise the old data. However, this is expensive and will require collaboration among centres.

General remarks

Modernization

The modernization of data centres is a continuous process – perhaps better termed the evolution of data centres. This change is required:

- because science changes. There is often a need for new datasets to meet new scientific requirements, e.g. the increasing importance of solar wind plasma and interplanetary magnetic field data in Solar-Terrestrial Physics. It may also be necessary to abandon old datasets as the old techniques are phased out, e.g. it may be necessary to abandon the Kp index as the hand-scalling of magnetograms is being phased out; a similar but not identical index could be introduced by computer-scaling of digital magnetograms.
- because technology changes, especially storage media. See section 2.

Thus data centres must assign some of their resources to ensure their continuing evolution.

Maintenance

The maintenance of databases and their access software is an important issue. Proper maintenance is vital to the continuing success of any data centre but is a major use of resources (perhaps 30%). This must be recognised by data centre managers.

BANK OF GEOPHYSICAL MODELS

A.Ya. Feldshtein, K.L. Latyshev**,*
*I.N. Nesterov**, Yu.S. Tyupkin**

**Soviet Geophysical Committee, Moscow, USSR*
***Kaliningrad University, Kaliningrad, USSR*

Computation experiment (CE) as a method of geophysical studies is now widely used. This method has the following stages: choice of a mathematical model for the process examined; choice of a numerical method for the task solution; compilation of a program to realize this method; computer calculations; analysis of results. Above all, the last stage may lead to correction of the mathematical model, to choosing another computation scheme or to modification of it; to implementation of different programs or their parts (subroutines). So, with the CE implementation in the practice of geophysical studies approximately the same situation emerges, which existed about two decades ago, when it became obvious that the utilization of data bank conception was essential and necessary for geophysics. Today, perhaps, it is practical to discuss the creation of a bank of programs for modelling geophysical processes. But what is implicated by this? Continuing the analogy by the concept of data bases, it becomes clear that it is necessary to have software, which would provide the user, who is not a professional programmist, with means to control without difficulties the run of the CE and to introduce some changes both by an interactive mode (which is, as a rule, a change of some or any parameters of the process) and by substituting certain blocks (subroutines) of the model for the alternative. The concrete realisation of this approach repeat the stage which were experienced by the DB concept in the course of its development.

First of all, it is necessary to describe the CE process as a set of universal components and their interaction. The following classification is suggested in /1/:

Model – a program corresponding to a certain mathematical model of a process,
Module – a program inherent to a composition of a model,

Class – an entity of functionally-equivalent modules,
Element of the class – a program module included into the class,
Structural elements of the model – elements of available classes of the model used as a standard.

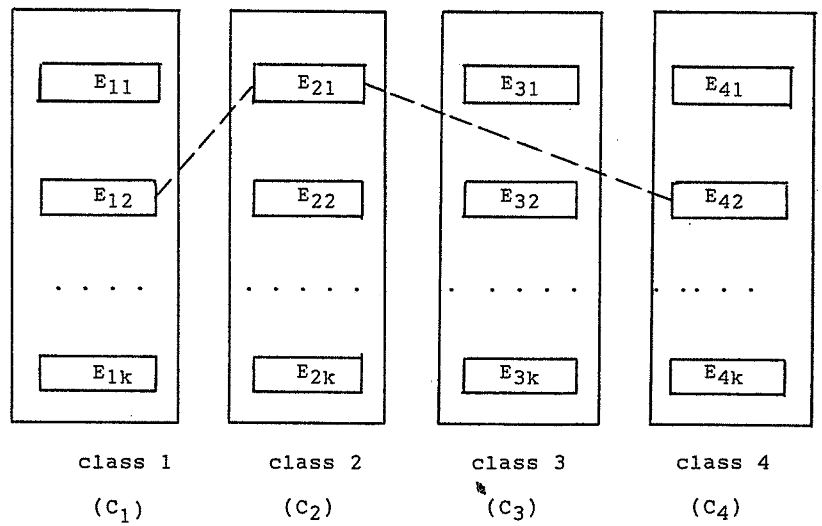


Figure 1 - Diagram of the model calculation process using the elements E_{12} C_1, E_{21} C_2, E_{41} C_4

Secondly, a models bank management system (MBMS) should be developed providing:

1. Introduction of new models, classes, elements. The essential integration of new components in the contents already existing in the Models Bank (MB) should be realized.

2. Providing the user with the following services:

- to form requests on modelling in a form convenient for the user in the problem-oriented type request language;
- to keep the values of experimental data and of frequently used parameters on a personal data base;
- to use different sources of the entry information: displays, personal data bases, data bases of general purpose, etc.;
- to model values of some entry parameters on the basis of other physical-mathematical models automatically linked;
- to keep the modelling results on a personal data base and to output them onto different outside devices;
- to specify the devices and the formats of its output in requests for modelling;
- to indicate in requests for modelling the necessity to establish the control points via the set up intervals of modelling time;
- to ask for reference information about the MBMS and its request language.

Natural the MBMS (as well as DBMS) is oriented to the concrete operation media.

As the first step in that direction upon the order of the WDC B for STP the package of applied programs ARMIZ, in which many of the functions listed above were realized /1,2/, was developed by the scientists of Kaliningrad University. The existing version of ARMIZ is realized for OS ES 6.0, and TSO is also used. The functional contents of the ARMIZ available at the Center are oriented on the studies in the field of ionosphere-magnetosphere physics. The physical-mathematical models included in the package are built up on the basis of equations of quasihydrodynamics, and they enables us to calculate the major characteristics of the ionosphere-magnetosphere plasma: spatial-time distributions of concentrations and temperatures of charged and neutral particles, speed of their movement at an interval of altitudes from 50 km to several of the Earth's radii. To be more concrete, the current configuration provides for the solution of the following tasks:

- calculation of parameters of the midlatitude ionospheric heights of layers E and F;

- calculation of parameters of ionosphere along the fixed circular tubes of the geomagnetic field;
- computer modelling of the Polar wind;
- calculation of parameters of mesosphere and lower thermosphere in the region of 50-150 km heights;
- calculation of the electricity fields and the speed of convection.

References

1. I.N. Nesterov, K.S. Latyshev. *Applied programs package ARMIZ. The system filling. In: Software for geophysical study*, issue 3, Moscow, 1986.

2. S.A. Grygoriev, K.S. Latyshev. *Applied programs package ARMIZ. Functional filling. In: Software for geophysical study*, issue 4, Moscow 1987.

CONSIDERATIONS REGARDING CHOICE OF DATA FORMATS FOR ARCHIVING AND USE OF GEOMAGNETIC DATA

Eigil Friis-Christensen

WDC-C1 for Geomagnetism
Danish Meteorological Institute
Copenhagen
Denmark

Introduction

From the point of view of a data manager at a data center, life would be so much easier if every data supplier used a common standardized data format for scientific data. In spite of this it still seems that we are far from this situation.

It seems therefore appropriate to consider and analyze this situation a little further, maybe with more emphasis on the needs and requirements of the data collector and the ultimate data user.

Founded on my own experience as a data collector, data manager and data user I will describe the considerations which has led to our choices of data formats used in the Division of Geophysics at DMI. Since the WDC-C1 for geomagnetism is an integral part of this, the choice of format has influenced also the data formats selected for the archives of digital data at the datacentre. Although our choice has been made based upon local considerations, it is my feeling that many of the considerations are more general, and may explain some of the reasons why a common standard format has never really gained universal acceptance. In spite of this choice, the intention of this report is by no means to discourage the use of common formats, because as soon as effective networks between centres have been established, the evolution will be in favour of using sophisticated scientific data base system for all the general data manipulation and display tasks. Until this is the

case, and until sufficient powerful and easy-to-use scientific data base management systems have been developed, there is still a need to make efficient use of local data, facilities, and resources.

During the considerations of choice of data formats one of things which turns out is that the exact choice of format may not be the most important issue. It is more important that the process of selecting a format in itself forces people to consider the data and the applications to be used so that consistency in the archiving of data is secured.

Data Formats and Data Base Formats

For some people the data format just means the way the individual numbers are packed into records which may be read sequentially by means of a program dedicated to a particular data set. For others it involves a complete description of the data, its history and its actual transformation into physical units, which may be read by a general program which is independent of the particular data set. The latter definition is normally referred to as the data base format, because a mere series of data records does not necessarily mean a data base.

When dealing with digital data in a number of different contexts, it soon turns out that the ideal data base format does not exist. The choice of the way to organize the data will always depend on the particular application and whether the application is a routine application or not. It is very often also a trade-off between simple but fast applications on one side, and flexible but complicated manipulations on the other side. It may also depend upon the local hardware and operating system, although during recent years it seems that at least on this front there is a great interest in developing operating systems towards a more common standard or,– and this is required to establish networking between different computer systems – provide the user with an interface which is independent of the actual computer hardware. Thus, on the other hand, demands a large overhead in the computer which requires excessive computer power which today may not always be available, but seems to be less of a problem in the future considering the current speed of technical developments.

In particular there seems to be different, and sometimes contradictory, requirements regarding the data-formats seen from the data collector's and from the data user's point of view. The following table demonstrates some of the main differences:

FF870313	Name of VFF-file
88.05.06	Date of creation
6560	Recordlength in datafile
26	Number of parameters
56	Number of rows in datafile
0.327670E+05	Value indicating missing data
T	Parameters stored in columns (T/F)
F	Datafile blocked (T/F)
F	Notes shown at opening of file (T/F)

Param	Name	Unit	Type	Dim.	Byte	Print-format
1	STARTDATE	YYMMDD	R	1	0	f7.0
2	STARTTIME	HHMMSS	R	1	4	f7.0
3	STARTDAYNR		R	1	8	f7.0
4	STOPDATE	YYMMDD	R	1	12	f7.0
5	STOPTIME	HHMMSS	R	1	16	f7.0
6	STOPDAYNR		R	1	20	f7.0
7	IAGACODE		C	1	24	a4
8	GENUMBER		R	1	28	f7.0
9	GEOG.LAT.	deg	R	1	32	f8.2
10	GEOG.LON.	deg	R	1	36	f8.2
11	INV.LAT.	deg	R	1	40	f8.2
12	MLT-UT	hours	R	1	44	f8.2
13	DECL.	deg	R	1	48	f8.2
14	QLTYPE		R	1	52	f7.0
15	OFFSET_H1	nT	R	1	56	f8.2
16	OFFSET_H2	nT	R	1	60	f8.2
17	OFFSET_Z	nT	R	1	64	f8.2
18	QL_H1	nT	R	1	68	f8.2
19	QL_H2	nT	R	1	72	f8.2
20	QL_Z	nT	R	1	76	f8.2
21	H1	nT	H	540	80	f8.2
22	H2	nT	H	540	1160	f8.2
23	Z	nT	H	540	2240	f8.2
24	H1_F	nT	H	540	3320	f8.2
25	H2_F	nT	H	540	4400	f8.2
26	Z_F	nT	H	540	5480	f8.2

END

	Col 1.	Col 2.	Value
BASELINE	21	23	.10000000E+01
SAMPLEINTERVAL	21	23	.20000000E+02
SAMPLEINTERVAL	24	26	.20000000E+02
FILTER_LOW	24	26	.88000000E-03
FILTER_HIGH	24	26	.25000000E-01
NOTES			
END.			

Type R is real data, C is character data, and H is integer*2 (16 bit) data.

Data Collector's Requirements:	**Data User's Requirements:**
1) Few modifications of data structure throughout the data processing history (facilitates software development).	1) Uniqueness in interpretation of data
2) Information about status and history of data (within record or in an associated file).	2) Status of data preferably final.
3) Self-describing structure (same software for different data-sets)	3) Self-describing structure including necessary ancillary information.
4) Packed data (reduces space and processing time requirements)	4) Fast input/output operations, unformatted-binary data.
5) Primary sorting key is station (data are originally collected station by station)	5) Data organized in an appropriated way for the application. (Ordered by time, random access).
6) Ancillary data kept in separate file (reduces need for updating).	6) Ancillary data kept together with the data (facilitates data usage).

The General Archival Data Format (GADF)

The choice of the data collector will in many cases be a simple sequential ASCII data format. Although simple, such a format does not satisfy some of the requirements mentioned above. For this reason the "General Archival Data Format" (GADF) has been defined and used by WDC-C1 as its primary archival format. The main principles of this format is a simple sequence of records each comprising a time series of data. The structure of the record is imbedded as information in the record itself, which means that the same basic format may be used for a variety of data with different time resolution and record length.

A binary (16 bit) representation has been used for the data part. This is a "de facto" standard on most computers, and is an ideal storage unit for geomagnetic field data, and, in fact, any experimental data. In addition to the physical data the record consists of a binary header which contains all the housekeeping information which an external user would not normally require, but which is essential for processing the data from the raw form into the more calibrated and manipulated data. Finally the record consists of an ASCII header containing information about time and coordinates of the station. A simply scanning program can then just read the ASCII parts of the header and in this way provide an overview of the contents of a particular data file.

The format is in a way close to the existing 1-minute data format which has been used for a long time as a standard at WDC-A, Boulder. It is therefore very easy to convert data from the GADF format to the WDC-A format. But there are three important differences. First the self-descriptive structure of the record which means that the same format (and software may be used for data with different time resolution. Secondly the binary format, which means a factor of 3 in packing density. This means a lot when dealing with high resolution data on small disk systems like those used on normal Personal Computers. Finally we have incorporated additional information about the data which otherwise had to be stored and provided explicitly. This additional information contains the current status and history of the data, which is essential to keep track of throughout the different processing steps, which convert the raw data into useful information.

If the data are to be used on an integrated system only, it might have been advantageous to make a distinction between the data itself and the information about the origin, quality and transactions of the data. When, however, multiple transfer of data from one system to another or to magnetic tape is often the case, then it seemed appropriate to incorporate this information in the data record itself.

The Vectorized Flat File System (VFF)

Although the above mentioned format is appropriate when dealing with geomagnetic time series, making stacked plots etc., the format is not very suitable for data analysis where the geomagnetic information is regarded as a distribution in space of 3-component magnetic field vectors. And in particular if this information is to be analyzed together with, say, radar measurements of ionospheric electric fields. For this purpose a more general data file format has to be used.

A number of such systems have been described and used in the scientific community; see for example a description of a FLATDBMS system by Smith and Clauer (1986). This system is essentially a two-dimensional array of data elements which are defined and described in an associated header-file. In applications, however, where the main part of the data consists of time series, and are manipulated as such, the structure of the FLATDBMS file slows down the input/output process, since a large number of read and write statements have to be executed to load a time series into memory.

A different, and more sophisticated approach to the problem of defining a suitable data format for scientific data has been undertaken at the National Space Science Data Center (NSSDC) by Treinish and Gough (1987), who defined the common data format (CDF), which is intended to be used on the SPAN network in connection with large spacecraft data bases. The CDF file system consists of nested multi-dimensional data-arrays, which makes it possible to describe nearly any type of scientific data-set.

While the CDF-system is a large and by definition a rather complex and flexible system, the present version does not run for example on a small system like a PC, although a PC version could probably be developed. The first mentioned FLATDBMS, although it is probably more limited in its range of applications.

In an attempt to optimize applications of this nature, which is vital on small systems as well as on heavily loaded main-frame computers, the FLATDBMS concept has been modified to incorporate the possibility of 1-dimensional arrays (vectors) as elements in the basic two-dimensional array. Normally a vector would consist of a time series of data for one station-component for a specified time interval. Although this is a principally simple extension of the original FLATDBMS concept, it does mean that additional demands have to be met by the supporting software. The extension presented here could be considered as a small and simple step in the direction of the CDF-system, but without the multidimensional structure characteristic of the CDF system. The important aspects of the FLATDBMS system is conserved. Namely an associated header file which contains all the information necessary to interpret and manipulate the data. An example of a header file is given in the table.

Not all the desired elements of the VFF-system have been finished yet, but a preliminary version is presently being used in our research work. The basic system consists of a system of FORTRAN subroutines which are used to access and manipulate the data. Further a graphical display system working with this concept has been developed as one of the first general application systems. Of course programs to convert data from the archival GADF format into the VFF format has been developed as well.

Our experiences have shown that although quite some work had to be put into a development of a system like this, the benefits of not having to write a new program for every application are sufficiently large to justify the efforts. It is our objective to incorporate the system in future WDC-C1 activities like for example center generated plots of limited and specifically selected data.

It is not within the scope of this report to go into a detailed evaluation of different formats. The objective of the present report is to demonstrate that some principal ideas may be used with advantage also in the development of systems in smaller institutions using less powerful hardware.

Conclusions

Even in a small organization it is difficult, and at least inefficient to use only one format for all different applications. But selecting a few fundamental data formats for use for specific ranges of applications is certainly a time saving choice. If considerable effort has been put into the design of these formats it also turns out that the programs to convert data from one form to another are fairly straight forward.

References

Smith, A.Q., And C.R. Clauer, FLATDBMS: *A flexible, source-independent data management system for scientific data.* STAR Lab Rep. D106-1984-1, Stanford Univ., Stanford Calif., 1984

Treinish, L.A., and M.L. Gought, *A software package for the data-independent management of multidimensional data*, EOS Trans. AGU, 633, 1097

THE METEOSAT ARCHIVING AND RETRIEVAL SYSTEM

Dr C. Mazza

European Space Operations Centre

Robert-Bosch str. 5

6100 Darmstadt FRG

Introduction to METEOSAT

Meteosat is a family of satellites, the first of which was put into orbit by the European Space Agency (ESA) in 1977. For six years METEOSAT was considered a pre-operational programme, i.e. it was operated by ESA as an experimental project.

In November 1983 the so-called Meteosat Operational Programme (MOP) began, which is approved until end 1995. The MOP is controlled by Eumetsat, an international body representing European interests in space meteorology.

The responsibility of the satellite operations and the related ground segment rest with the Directorate of Operations at the European Space Operations Centre (ESOC) in Darmstadt, Federal Republic of Germany. The ground segment comprises the equipment necessary to operate the satellites in orbit and provides for the dissemination of processed data, the extraction and quality control of meteorological products and the archiving and retrieval of image and image-related data.

The approved Meteosat Operational programme contains three new meteorological geostationary satellites to be launched in late 1988, 1990 and 1992. Until these satellites are placed in orbit the programme will be supported by the pre-operational satellite, Meteosat-2 and a third pre-operational satellite, P2, launched in June 1988.

The Meteosat System

The Meteosat System consists of two major components, the Space Segment and the Ground Segment.

The Space Segment consists of one or more spin-stabilised satellites in geostationary orbit. The primary satellite is located over the Gulf of Guinea, at the crossing between the Equator and the Greenwich meridian (0 degree N, 0 degree E); reserve satellites will be located nearby in a hibernated condition.

The main components of the Ground Segment are the Data Acquisition, Telecommand and Tracking Station (DATTS), the Meteosat Ground Computer System (MGCS), the Meteosat Operations Control Centre (MOCC) and the Meteorological Information Extraction Centre (MIEC). MGCS, MOCC and MIEC are located at ESOC and the DATTS is situated in open country about 40 km from Darmstadt.

The primary functions of the Meteosat Operational System can be identified in three basic missions:

- Earth imaging
- Dissemination of image and other meteorological data
- Data collection and distribution

and two additional main functions:

- Meteorological processing
- Data archiving and retrieval

Earth Imaging

The principal payload of the satellite is a multispectral radiometer. This provides the basic data of the Meteosat system as visible and infared radiances producing images of the full earth's disk from geostationary orbit. The radiometer operates in three special bands:

- 0.5. – 0.9 micrometer – visible band
- 5.7 – 7.1. micrometer – infrared water vapour absorption band
- 10.5 – 12.5 micrometer – thermal infrared (window) band

The infrared and water vapour images are composed of 2500 lines of 2500 picture elements whilst the visible image has 5000 lines of 5000 picture elements. The spatial resolution at the subsatellite point is approximately 5km for infrared and water vapour images and 2.5 km for visible images.

Earth images are generated at half-hourly intervals. Each image is transmitted to the DATTS continuously, in real-time, on a line by line basis, and from there to the MGCS for further processing, distribution and archiving.

Image Dissemination

Meteosat is equipped with high-power amplifiers which allow the relay of earth images and other meteorological information from the ground via the satellite to small user reception stations.

The data transmitted is mainly processed Meteosat image data but also includes conventional meteorological charts received at ESOC from the Deutsche Wetterdienst at Offenbach and images of the Western Atlantic and the Americas generated by an American GOES satellite.

The data are relayed via the two Meteosat dissemination channels to user stations,which must lie within about 75 degrees of the subsatellite point. Two forms of image transmission are used, high-resolution digital data, for reception by Primary Data User Stations (PDUS) and analogue (WEFAX) data, received by Secondary Data User Stations (SDUS).

Data Collection and Distribution

Meteosat has a total of 66 telecommunications channels designed for the collection of environmental data from automatic or semi-automatic Data Collection Platform (DCP), which may be located at any point within the Meteosat coverage area. Regional DCP, that is DCP which operate exclusively with the Meteosat system, and international DCP, mobile DCP likely to move through the coverage areas of all geostationary meteorological satellites, are supported by Meteosat.

Meteosat acts simply as a relay so that environmental data transmitted by the DCP are received in ESOC, and then redistributed in a variety of different ways. The DCP, installed at a fixed site on land or on mobile supports such as ships, buoys, ballons or aircraft, may be of two types:

- 'self-timed', transmitting messages automatically, according to a predetermined time and frequency schedule, or
- 'alert', transmitting every time a parameter value exceeds a preselected threshold.

Meteorological Data Distribution

The MOP satellites will also incorporate a Metorological Data Distribution (MDD) mission. It is envisaged that two telecommunication channels at 2400 baud will be used for the relay of fascimile weather charts, DCP messages and meteorological bulletins to purpose-built user stations.

Meteorological Processing

Another objective of the Meteosat system is the extraction and distribution of meteorological parameters from the basic image data. The Meteorological Information Extraction Centre produces routinely seven meteorological products:

- cloud motion winds
- sea surface temperatures
- cloud top height maps
- cloud coverage data
- upper tropospheric humidity values
- a basic climatological data set
- precipitation index

These products are extracted by a fully-automated set of software within MGCS. The first five products listed are quality controlled by a meteorologist before the results are coded and distributed; the remaining two are archived on magnetic tape. The resolution of the cloud top height maps is about 20 km whilst the remaining products are derived on a grid having approximately 200 km resolution. In addition, ISCCP (International Satellite Cloud Climatology Project) data extracted on a regular basis.

Archiving and Retrieval

The Meteosat objectives are completed by the requirement to archive all images, image related data and meteorological products. The raw image itself comprises some 300x106 bits of information each half hour, thus the quantity of data to be archived is extremely large.

The primary archive medium utilised for Meteosat data is 9-track, 6250 bpi computer compatible magnetic tape and twelve tapes are

needed to store data from one day of Meteosat operation. Individual images can be retrieved from this archive and copied onto computer compatible magnetic tapes or used as input to a laser beam recorder which produces high quality photographic images.

More details on the Meteosat Archiving and Retrieval System are given later in this article.

The Meteosat Ground Computer System

The MGCS is a large integrated computer system used for the processing of Meteosat data and for control of the spacecraft. As well as its primary data link with the DATTS, the MGCS is also connected by a computer to computer link with the Global Telecommunication System (GTS) of the WMO.

This interface is made with the Regional Telecommunications Hub (RTH) in Offenbach, FRG, and allows parameters extracted by the Meteosat Processing System to be transmitted to the user community. The link is also used in the reverse direction, since conventional meteorological data are needed for many of the meteorological computations made within the MGCS.

The main component of the MGCS is the operational mainframe (MF) computer. This is supported by a number of smaller sub-systems, consisting of mini-computers and ancillary devices, dedicated to specific tasks. Another mainframe computer is operated in MGCS as a general purpose batch-computer, however, it is used as a back-up facility for the operational machine. All the subsystems are switchable between the two MF providing a flexible configuration which can be adapted to operational contingencies.

The image data flowing from the DATTS via the high speed data link enters the computer system at one of two minicomputers known as Front End Processors (FEP). The FEP performs the image acceptance tasks with each image line being processed in real-time, at a rate of one line every 0.6 seconds. During this period the raw data are also stored on a temporary magnetic tape archive. From the FEP the processed image data are passed to the mainframe and stored on magnetic disks ready for this purpose, each containing two generations of the processed raw image thus providing a rolling store of six images at any one time.

A further system of minicomputers, Back End Processors (BEP) communicate with the spacecraft via the high-speed data link and DATTS. They are used for spacecraft control functions, for the Data

Collection System, for the dissemination of processed image data to the spacecraft and for linking with the RTH at Offenbach.

Two further minicomputers support the meteorological interactive display system used for the quality control of the derived meteorological products and image display.

The Meteosat Archiving and Retrieval System

The Meteosat system provides for the archiving of all image data and meteorological products in digital format; in addition some image data are also archived in photographic form. The flow of data on the MGCS is indicated in the following diagram.

The incoming raw data from the ground station are passed to a minicomputer known as a front-end processor (FEP) where the data for each image are accepted and processed in real time. Every image line of the raw data is composed of 2560 32-bit words each of which contains data for all three activated radiometer channels.

The Digital Archive

The digital archive is maintained on HDCCT, recorded on 6250 bpi tape decks. This enables all the data for one day to be stored on approximately 12 tapes. The digital archive includes all image data received by the MGCS.

In addition to the image data, the system also archives meteorological products. All of the products, with the exception of the Cloud Top Height, are archived on 6250 bpi tape.

It is intended in the future to archive the images on optical discs, but for the time being the reduction of the cost per byte archived achived with optical discs does not yet justify the investment.

The Photographic Archive

Three slots per day, normally at 00, 12 and 15 UTC, are archived onto photographic negative film using a VIZIR laser beam recorder. The film size is 425 x 460 mm, and a basic spot diameter of only 40 micrometer gives some 11,000 separate points for each of 11,000 lines. 64 grey levels are available.

Although the film size allows pictures of up to 40 cm square to be produced, in normal use four 20 x 20 cm pictures are obtained from each section of film. Thus the normal archive standard is for each image to be recorded as a 20 x 20 cm negative, one for each radiometer channel. The film are called negative because in all cases the clouds appear as dark areas. This means that when contact prints of the negative films are made, the paper print shows clouds as white, i.e. in the conventional representation. This representation can be reversed on the laser beam recorder if required.

Archive Retrievals

Digital data, either images or meteorological products, may be retrieved onto CCT recorded at 1600 bpi in a readily usable format. Facilities are available which enable the retrieval of sectors or data windows from complete images. On the request of the user retrieved image data may also be rectified using the nearest neighbour rectification scheme.

Photographic products (visible, infrared, or water vapour) can be produced for either the full disk or specific windows. These products are normally supplied as either 40 x 40 cm or 20 x 20 cm contact prints. Thus prints of windows automatically enlarge the specified area. Because only three skiots of full disk images are produced each day for the photographic archive, specific requests may first have to be retrieved from the digital archive.

Archive Catalogues

Several types of archive catalogue are currently available.

- The Meteosat Catalogue of Digital Data
- The Meteosat Catalogue of Image Negatives
- The Meteosat Image Bulletin

The first two are conventional catalogue listing all available information in the two archives.

The Meteosat Image Bulletin is a monthly publication which contains, for each day of the month, one image from each of the three channels plus an abbreviated version of the full catalogue for that day. This bulletin is available on subscription or as individual issues from the adress at the end of the publication.

A new facility introduced during 1986 provides access to the catalogue of Meteosat images through the ESA-Information Retrieval Service (ESA-IRS). On-line access to ESA-IRS can be made from a user terminal, e.g. word processor or microcomputer work station, through a variety of public telecommunication networks. A small charge is made by acessing the database and the user requires his own personal password which is supplied free of charge by ESA-IRS or a National Centre. The Meteosat catalogue is updated on a monthly basis.

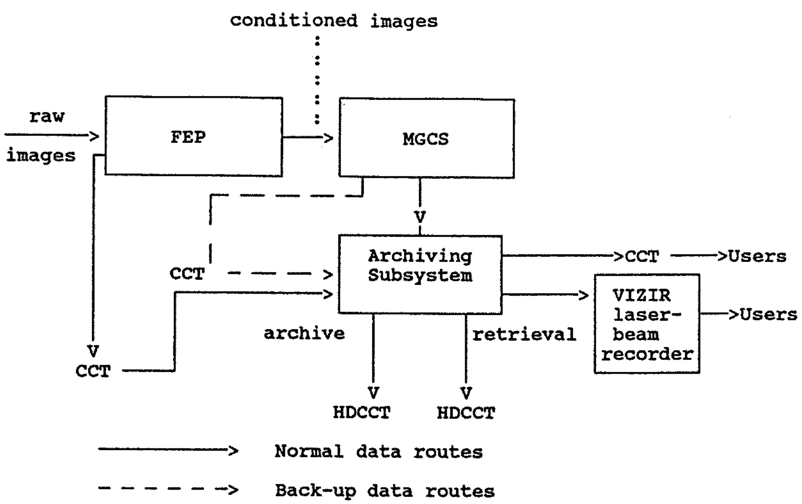


Figure 1 - The METEOSAT Archiving and Retrieval System.

The information presented in this article is directly reported, with only minor modifications from the following ESA publications:

ESA BR-32-ISSN 250-1589 of September 1987
Meteosat System Guide-- Vol 11--Meteosat Data Services, September 1980
Meteosat, the Operational Programme

A PROTOTYPE DATA BASE MANAGEMENT SYSTEM

B.V. Danilchev

Soviet Geophysical Committee, Moscow, USSR

Introduction

A prototype data base management system--System/S--is intended to automatize the procedures of storing, processing and manipulating the data and is developed to be used by non-programming user (1). System/S uses the relational data base model (2), which is recognized to be the most suitable for all kinds of users. The system is oriented mostly toward time series processing, and uses the magnetic tape as the main media for storing the data base (DB).

System/S allows to automatize different routines of data handling, input from a sequential data set into the DB; output from the DB into sequential data set; manipulating the DB with a help of query language; data editing and making inquires about the data, stored in the DB.

System/S automatize also the scientific data processing with the help of special set of application programmes. A special User Interaction Language (UIL) allows any user to manipulate the DB easily.

Data Model

The relational model of a DB provides a non-programming user with the most usual and natural way to present his data--a two-dimensional table. A user sees the DB as a set of tables, each of them having the following properties.

- the name of a table is unique in the DB;
- the name of a column is unique in the table;
- data elements in a column belong to the same type;
- duplication of lines is not allowed;
- the order of lines in a table is not sufficient.

An example of a table is presented on Figure 1.

Each table in the DB has a key – a subset of columns, which keeps up the property 4. A key exists in any table, the set of all columns will always be the key, and for many tables the key is much smaller. For example, the table INDICES (figure 1) has the key, including three columns, YEAR, MONTH, DAY. Key reflects the semantics of the data and is used to improve the effectiveness of DB operations.

The type of data is fixed for a column, it can be created simultaneously with the table. Examples of type descriptions are shown in figure 2.

UIL has much in common with well known query language SEQUEL2 (3), it includes commands to create a table, delete a table or part of it; input and output data in and from the table; select data, entering queries, from one or more tables; and to join two tables. An example of a command to input data from a magnetic tape, containing dayly sunspot numbers into INDICES table is shown on figure 3, and a command to select from the same table a date and flux value of all those days in the first part of January 1987, when sunspot number was more than 100, is shown in figure 4.

It can be easily seen, that a request for a selection will also be a table. It is possible to name this new one, and store it in the DB for future use. However, a new table may not suit some semantic features of the DB as a whole, that is why the DB, managed by System/S, is devided into common part and a number of private parts – one for every user. Tables, created by a user are stored in his private part.

Scientific Data Processing

The main difficulty a non-programmer, solving a scientific problem, meets is in the fixed link between a program and data; that is why additional programming is needed for switching to other data. Using a standard model it is possible to create an interface between DBMS and application programs management system, so that it will adjust the data automatically. Additional expenses will appear to adapt new application programs, but it has to be done only once, and those expenses will be covered by easy usage.

An example of a table, containing dayly radio flux values and a command of UIL, calling a program, computing correlation between, fluxes with frequencies 8800 MHz and 2695 MHz are shown in figure 5 and 6. To apply the same program to another table, INDICES for example, a user need only to change names, as shown in figure 7.

INDICES

YEAR	MONTH	DAY	WOLF_NUMBER	MHZ2800
87	01	01	30	120.1
87	01	02	120	168.2
87	01	03	110	133.6
87	01	04	96	170.1
.	.	.	.	.
.	.	.	.	.

Figure 1.

```
MONTH : ( 1, ... , 12 )

HEMISFHERE : ( S, N, E, W )
```

Figure 2.

```
INPUT

 PORT :      TAPE : SA102/SL/WOLF
 TABLE :     INDICES
 WINDOWS :   [ 1,2] -> YEAR,
             [ 4,2] -> MONTH,
             [ 7,2] -> DAY,
             [10,3] -> WOLF_NUMBER
```

Figure 3.

```
SELECT

 TABLES :    INDICES
 CONDITION :  YEAR = 78  &  MONTH = 1
       &  DAY < 16  &  WOLF_NUMBER > 100
```

Figure 4.

RADIO_FLUX

YEAR	MONTH	DAY	MHZ8800	MHZ2695
86	02	10	286	170.2
86	02	11	290	177.4
.	.	.	.	.
.	.	.	.	.

Figure 5.

```
PROGRAM :   CORRELATION

  TABLE :    RADIO_FLUX

  COLUMN1 :  MHZ8800

  COLUMN2 :  MHZ2695
```

Figure 6.

```
PROGRAM :   CORRELATION

  TABLE :    INDICES

  COLUMN1 :  MHZ2800

  COLUMN2 :  WOLF_NUMBER
```

Figure 7.

Conclusion

An experimental use of System/S is going on, and the experience obtained allows to plan its further development, adding a subsystem to manage meta-data, or descriptions or tables and programmes, accessible to a user, together with services, concerned with quantitative maintenance of tables, raising the level of friendliness; increasing the number of application programmes together with automatizing their adaptation.

References

1. B.V. Danilchev. *Solar activity data management system. World Data* Center B. Moscow, 1985.

J. Martin. *Computer Data-Base Organization*. 2-nd ed. Prentice-Hall Inc. Englewood Cliffs, N.J., 1977.

3. D.E. Chamberlin, et al. *SEQUEL 2, a unified approach to data definition, manipulation and control.*—IBM Journal of Research and development, 1976, v. 20, N11.

THE PROSPECTS OF CREATION AND USAGE OF COSMIC RAYS DATA BASE

A.V. Belov, E.A. Eroshenko, Yu. S. Tyupkin

IZMIRAN , Troitsk, Moscow Red., USSR

Soviet Geophysical Committee

Molodezhnaya Str. 3, Moscow, USSR

Introduction

Secondary Space particles which are recorded on the Earth's surface are generated in the atmosphere by primary charged space particles of rigidness in the 10-1000 GV range. These particles respond to changes of the electromagnetic properties in the interplanetary space (scale ranging from 10E9 to 10E15 cm) which are the results of various Solar active phenomena. Due to high speed and relatively free propagation, they provide information about the current phenomena well in advance before the other features reveal these disturbances. This property can be taken as a basis for using cosmic ray (CR) for prediction of geoactive phenomena. However, cosmic rays carry distorted information on the interplanetary space because they are affected by the Earth's atmosphere and magnetosphere processes. Methods which take into account all these effects are sufficiently well developed but they are labour-consuming. And these complexities with processing make users avoid the CR data when considering interplanetary space problems. We believe, however, that if the WDCs could provide the users with not only the intial CR data but also with the results of secondary processing, this would be of great assitance to researchers.

The WDC-B CR Data Base.

The worldwide CR network consists of about 100 cosmic ray stations including about 12 Soviet stations. Till recently, the worldwide network

CR data were stored at WDCs in tabulated form, but as a result of a great effort of the WDC-C2 we have most of the ground-based CR data on magnetic tape now. These data form the foundation to the WDC-B CR data base.

At present, the WDC-B in cooperation with IZMIRAN are prepairing CR data of the Soviet network in the standard format for international exchange. CR data are often used together with data on other disciplines of Solar-terrestrial physics. This is the reason why CR data base includes:

- mean hourly values of CR intensity at the worldwide network of stations;
- mean hourly values of the interplanetary space and solar wind parameters;
- mean hourly indexes of geomagnetic and solar activity;

the special software which is being created in WDC-B.

Processing programs

As mentioned above, the users are interested in the results of secondary processing rather than in the initial data. Therefore, to use the CR data base more effectively a software for processing of these data should be developed. We believe that the following programs have to be included in this software:

1) Calculation of the coupling coefficients for each local CR detector.

This program is based on calculations of asymptotic directions for particles of different energies (3,4). Different versions of this program have been developed in Japan (1), and in USSR (2), and the coupling coefficients have been calculated by the authors for minimum and maximum periods of solar activity for isotropic and different aniotropic components of CR (5,6).

2) Calculation of the barometrical coefficients and their variations for the network of stations.

It is no secret that any inaccuracy in the barometerical coefficient leads to spurious effects in any type of variations of CR examined. The program uses the atmospherical pressure data and uncorrected CR data at each station. IZMIRAN and WDC-B are testing different methods and adopting one for recalculation of these coefficients for each station.

3) Determination of long-term CR variations rigidity spectrum and CR modulation parameters in the CR rigidity region 1-30 GV using the annual data of neutron monitors, ionozation chambers, muon telescopes and satellite and stratosphere CR observations (7);

4) Determination of daily mean amplitude, phase and rigidity spectrum parameters for three harmonics of CR longitudinal distribution for data of each station and worldwide network as a whole. A version of this program is realized in WDC-B. The program uses the worldwide network data and coupling coefficients and calculates the listed above parameters for any time interval;

5) Determination of the isotropic CR variations rigidity spectrum and the geomagnetic cutoff rigidities variation for the different points of the Earth's globe by a spectrographic method (8,9).

This program is developed at IZMIRAN and SIBIZMIR, and it uses the CR data base of hourly and daily mean values of neutron monitors network and meson telescopes.

6) Discrimination of isotropic and anisotropic CR variations and definition of their parameters by the method of global surver.

Different versions of this program were developed and realised in USSR (9) and in Japan (10).

7) Discrimination of the zonal components of CR anisotropy and definition of their rigidity spectrum for the daily mean data of high-latitude monitors and muon telescopes (11);

This program was developed by IKFIA and IZMIRAN groups.

8) Determination of daily mean values of all components of the CR gradient in the vicinity of the Earth for the quiet and moderately disturbed period (12, 13);

9) Determination of solar CR energetic spectrum and anisotropy for GLE intervals. The program is realised in IZMIRAN, and it operates with the five- or one-minutes neutron monitors data.

10) Estimation of probability of directly coming solar neutron registration.

We believe that the combination of the software listed above and the CR data base will be very useful for researcher, because there is no doubt that a more extensive class of solar-terrestrial physics problems can be solved in this case. For example:

1. Continuous diagnostics of electromagnetic conditions of the near-Earth space:

a) definition of the IMF intensity and of the direction of the magnetic field line in the plane of the ecliptic, the IMF irregularity degree, and the power spectrum of IMF fluctuations;

b) definition of the daily mean values of CR gradient in the vicinity of the Earth associated closely with high velocity solar wind fluxes, the position of current sheet in the heliomagnetosphere and sporadic disturbances of the solar wind;

2. Forecasting of the shock wave front arrivals causing a SC geomagnetic storm.

3. Estimation of radioactivity during powerful solar flares.

4. Definition of the environment temperature in the upper atmosphere.

5. Diagnos of the magnetospheric current system during strong geomagnetic storms.

The examples of results which can be obtained by means of this software (12) are demonstrated in Fig. 1. The day to day CR gradient variations (Gx,Gy,Gz) for 6 revolutions of the Sun in 1974 are presented in Fig. 1a.

Fig. 1 illustrates also the average position of the neutral sheet (1b), and behaviour of the solar wind velocity (V), isotropic CR variations (A_0), zonal components of CR anisotropy (Az), and latitudinal (1b) and azimuthal (1c) components of CR gradient in the interplanetary space during one revolution of the Sun in 1974. Fig. 1d shows the Gx distribution separately for the rise and the fall of the solar wind velocity.

The process of transformation of the preliminary CR data into "active data" (to our mind) is presented in Fig. 2.

References

1. Nagashima K. Rep. Ion. Space Res. Japan, 26, 189 p., 1971.
2. Krymsky et al. Geomagn. i Aeron., v.6, N6, p. 991, 1966; v.7 N1, p. 11, 1967; v.7, N4, p. 620, 1967.
3. Mc Cracken K.G. et al. IQSY Instruction Manual No. 10, 1965.
4. Shea M.A. et al, Suppl. to IQSY Instruction Manual, No. 10, Cosmic Ray Tables, 1968.
5. Belov A.V., Eroshenko E.A. The reception coefficients of the neutron monitors, 17-th ICRC, Paris, v.4, p. 97-100, 1981.
6. Yasue S., Mori S., Sakakibara S., Nagashima K. Coupling coefficients of CR Daily Variations for Neutron Monitor Stations, Rep. of CR research laboratory, Nagoya University, No. 7, Japan, 1982.
7. Belov A.V., Gushchina R.T., Dorman L.I., Sirotina I.V. Rigidity spectrum of CR variations in solar magnetic cycle. 20-th ICRC, v.3, 1987, p. 319.
8. Baisultanova L.M., Belov A.V., Dorman L.I., Yanke V.G. Magnetospheric effects in CR during Forbush-decreases. 20-th ICRC, v.4, 1987, p. 231.

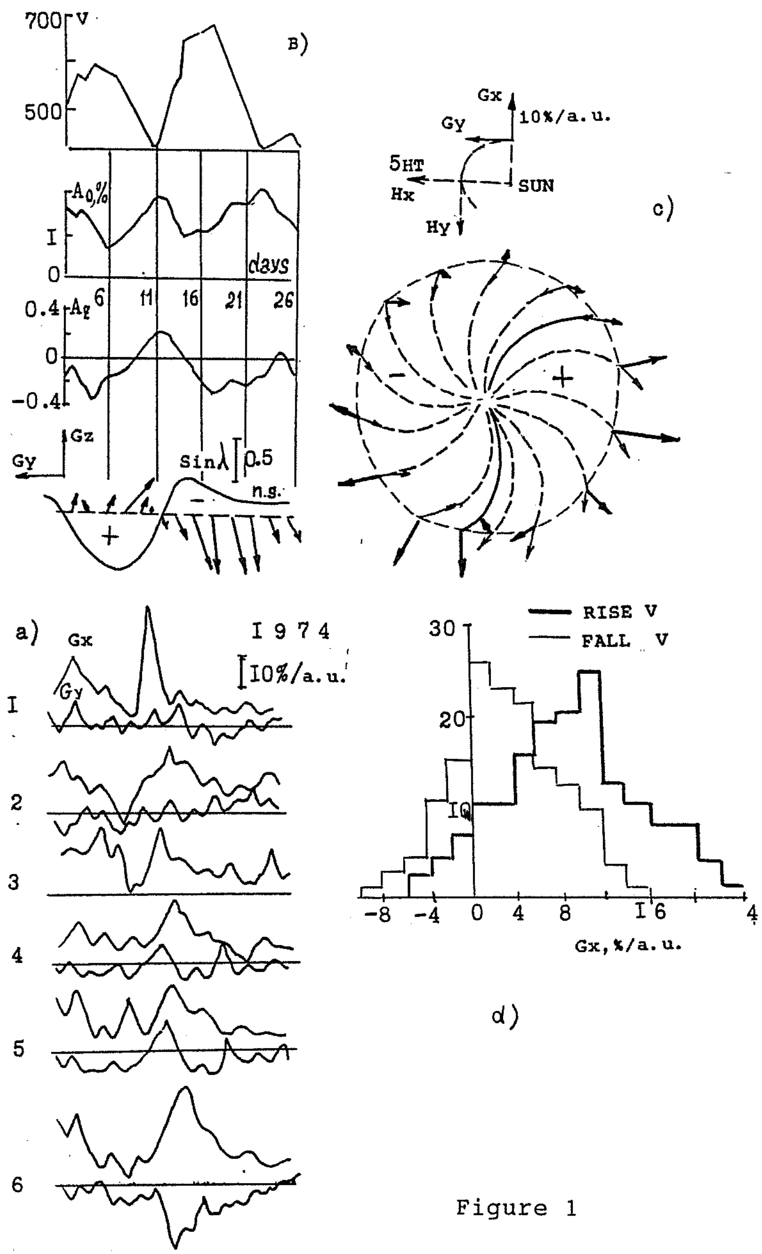


Figure 1

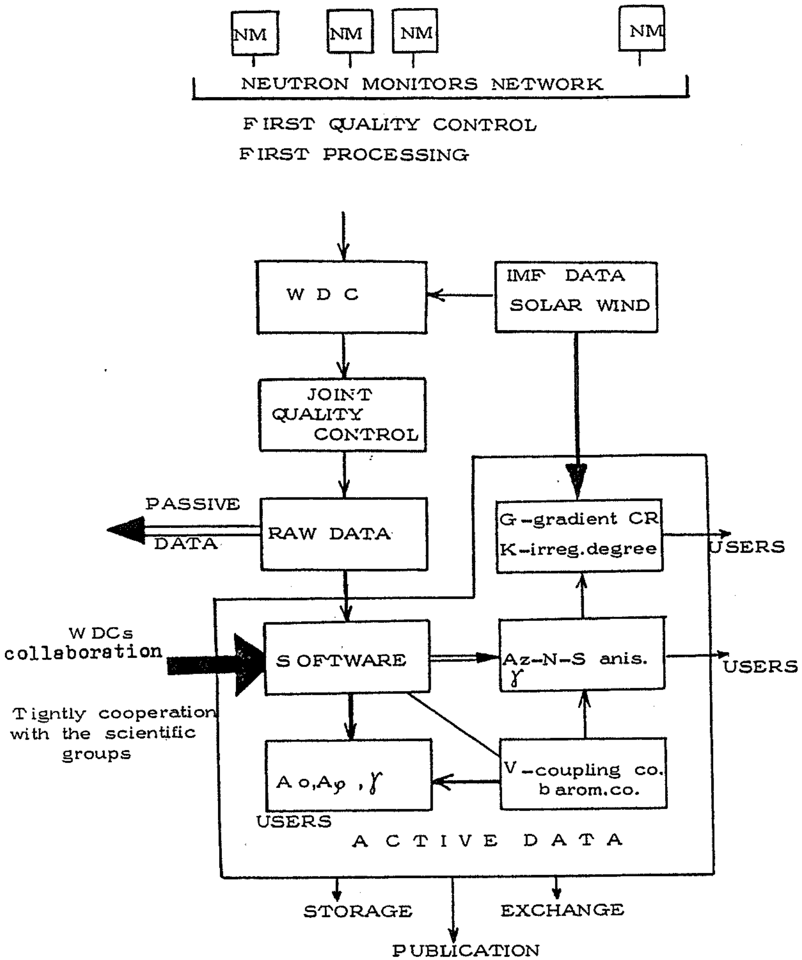


Figure 2

9. Dvornikov V.M., Sdobnov V.E., Sergeev A.V. Determining the parameters of magnetospheric current systems from ground based CR and geomagnetic field observations. 20-th ICRC, v.4, 1987, p. 220.

10. Kudo S., Mori S., Swinson D.B., Yasue S., Rue Y. Solar cycle dependence of diurnal anispotropy. 20-th ICRC, v.4, 1987, p. 140.

11. Belov A.V., Dorman L.I., Oleneva V.A. The north-south CR anisotropy during two last solar minimum. 20-th ICRC, v.4, 1987, p. 186.

12. Belov A.V., Dorman L.I., Eroshenko E.A., Oleneva V.A. Cosmic ray gradient variations in 1974. 20-th ICRC, v.4, 1987, p. 137.

13. Ichinose M., Morishita I., Nagashia K. IMF sense-dependent solar diurnal variation of CR. 20-th ICRC, v.4, 1987, p. 148.

TOWARD AN IGBP DATA AND INFORMATION SYSTEM

Juan G. Roederer

Geophysical Institute
University of Alaska
Fairbanks, USA

Introduction

Any major international program of scientific cooperation must meet many different, often conflicting, requirements (Roederer, 1985): (1) It must maintain an appropriate balance between centralized coordination of activities and the freedom of research of the participating scientists; (2) It must offer meaningful opportunities for participation of scientists from all nations, regardless of their economic development; (3) It must address a number of well-defined scientific problems, but these must be of sufficiently diverse merits and interest so as to motivate wide participation of scientists and secure the support of the public and government officials.

An international program needs an overal scientific theme and an overal strategy to develop the theme. The main theme of the International Geosphere-Biosphere Program (IGBP) is to describe and understand that regulate the total earth system, the changes that are occurring in this system, and the manner in which they are influenced by human actions (ICSU, 1988).

Organizing a "megaprogram" such as the IGBP requires the formulation of a strategic plan for the conduct of the projects, the anticipation of contingencies and the formulation of alternative approaches. This planning process must be built upon the following premises: (1) One cannot dictate to the participating scientists what they should do, hence, the plan must emanate from the participating scientists themselves; (2) the plan must be realistic from the financial and human resources point of view; (3) the plan must be a logical extension and/or

complement of existing and other already planned scientific enterprises.

While the main frame of an international research program should consist of not more than a set of research guidelines on what *should* be done where, when and by whom, the success of an international program will be determined by the central services that *will* be set up specifically to aid the participating scientists. Such services mainly involve data and information exchange and analysis, and the coordination of research projects, missions and campaigns.

Special services that could be established for the IGBP include: specialized data centers; rapid information exchange offices; specialized forums (workshops); computer-interactive data analysis centers; central computer facilities for modeling and simulation; specialized geochemical analysis and geophysical instrumentation centers; and theoretical research centers.

This article deals with a series of basic issues and questions that should be addressed and answered before any specific recommendations are made regarding the establishment of an IGBP Data and Information System. The expression "data and information system" used throughout this article is defined as "the conglomerate of technical and human resources available to locate and obtain IGBP-related (i) data; (ii) information on past, current and planned research activities; and (iii) published materials." It is not our intention to describe a proposed structure and operation of such system.

The Role of Information Systems in Human Society

Since the end of World War II, human society has undergone a profound transition from an "industrial society" to an "information society", in which industrial, economic and military power is conditioned to information-processing power, and societal well-being, social organization, and government are conditioned to the information transfer capacity among elements of the population (Bell, 1973). Today, elements of the information society pervade even the economically least developed countries. In the early 1800's, a more gradual transition occurred when the industrial society emerged from a predominantly agrarian society.

It is important to realize that these transitions were driven almost exclusively by *science*, when the results of scientific research were applied to technological development: it was science that made the transition into the industrial society possible, and it is science that is making the high-paced development of the information society possible.

For the purpose of this article it is useful to review a few key facts concerning the development of science and the scientific method. Science emerged when it became apparent that the images of the world and environmental events, acquired through the senses and registered by the human brain in natural day-to-day experience, contained inaccuracies and subjectivce biases that interfered with the development of an increasingly complex society. It became apparent that in order to establish a repertoire of reliable information on cause-and-effect relationships, environmental exploration and documentation would have to be expanded from subjectively "relevant" phenomena to others that bore no direct relation to, or had no effect on, the human organism. It was also realized that a merely passive, qualitative, random observation of environmental events did not yield sufficient information. Active, quantitative probing and systematically planned experimentation became a necessity: the empirical method was born. Our sensory systems needed extension to achieve higher resolution and accuracy in the acquisition of environmental information, and scientific instruments were developed to make the measurements required for a quantitative description of processes occurring in a wide range of spatial and temporal domains, extending way beyond the domains of everyday experience.

A most crucial fact in this development was the realization that the organized documentation of facts in books, reports, and data repositories was absolutely essential for the recording and preservation of scientific results, their statistical interpretation and, in general, for the development of an "objective truth" about environmental events.

The importance of this aparently trivial statement, and the fact that its significance reaches far beyond the realm of pure science, can be dramatized with the following observation. Human brains today are belived to be no different, anatomically, from the brains of humans who lived, say, five or even ten thousand years ago (no genetically significant change could have taken place in such a "short" time-span). What is it, then, that allows us today to do all these incredible things which our distant ancestors were unable to do, such as integrating differential equations, building airplanes or planning an IGBP? It is the retrievable information that humans have stored physically in the surrounding environment--information that does not deteriorate and become subjectively distorted as all that is stored in human memory and passed on orally to others (Roederer, 1979). In other words, what has changed

over these millenia is not the structure of our brains, but the information and data systems built *outside* our bodies!

It is instructive to describe the need for data and information systems in the context of the evolution of research and development (R&D). A report by Arthur D. Little, Inc (1978) identifies three "Eras" of R&D, leading to the transition into the information society. Each Era persists into the next; all three coexist today.

In the *"discipline-oriented" Era I,* starting some time before the turn of the century, basic research and discipline-centered R&D are the main sources of new knowledge. This is the era of the great discoveries in science and the inventions in technology--the theory of relativity, molecular biology, the construction of automobiles and airplanes, for example.

The *"mission-oriented" Era II,* starting during the 1950's, has as its basic ethic that of "organizing to do a job". It involves the great R&D enterprises and projects such as space exploration, particle accelerators, biotechnology and the nuclear industry.

The *"problem-oriented" Era III* has as its basic ethic the solution of society's problems. It is the era of the information society *par excellence*. The Arthur D. Little, Inc. report has identified ten problem categories as fundamental targets in Era III: environment; energy; economic well-being; safety; public health; transportation; crime preventions; and the administration of justice, housing and welfare.

Era I information systems mainly handle "end products" of research, such as articles in scientific journals, books, etc.; producers and users of data normally belong to the same research group. The principal output of Era I is scientific and technological *knowledge*. Era II involves data-intensive research efforts, but the main output, *information*, still remains within limited groups of the scientific and technological community.

In contrast, the problem-oriented Era III information and data systems mainly handle cross-disciplinary data flow (often intensive raw data flow). Data producers and users usually belong to different groups and even to different disciplines, but they must be able to communicate with each other and work cooperatively in data analysis and interpretation. This presents a vast spectrum of organizational and even psychological problems (Roederer, 1988). The data needed are often of synoptic type, acquired in large monitoring networks, observatories, national laboratories, or based on large-scale statistics or surveys that cannot be operated or conducted by isolated groups or institutes. Intensive *data* flow is the touchstone of Era III. It is clear that the IGBP and its scientific projects will involve mainly Era III-type data and information requirements.

It is important to note that as one advances from one Era to the next, government agencies are found to carry increasing responsibilities in the development and implementation of required R&D and related data and information systems, and international cooperation is required to provide the necessary geographic "coverage" for the environmental database.

Data vs. Information

A few definitions are in order (Roederer, 1981). We usually think of the concept of "*data*" as embodying sets of numbers given in some digital or analog representation, encoding the values of some physical magnitude measured by a certain device under certain circumstances. We usually think of the concept of "*information*" as embodying statements that represent answers to preformulated questions or that describe the outcome of expected alternatives (in information theory, a precise mathematical definition of information is used). And we may give an operational definition of scientific "*knowledge*" as any comprehensive information about a given system that allows making predictions about the system's future or postdictions about its past.

Data are meaningless without the information on what physical magnitude they represent; on the instruments used in the measurement; on units and formats, etc.; and on the particular circumstances in which the data were taken. Information, in turn, is meaningless without knowledge of the questions or alternatives that it is supposed to answer. Knowledge is meaningless without specification of the system to which it refers.

Information is extracted from data whenever the data are subjected to some mathematical treatment that leads to the answer of preformulated questions. A remote sensing satellite picture is nothing but a collection of data representing light emission intensities in a two-dimensional array of solid-angle pixels. Information is extracted from that data only when a given pattern is searched for by an automatic device or by a human being looking at the picture and letting the brain recognize the pattern in question. A tape recording of magnetospheric VLF waves is nothing but a collection of data representing electromagnetic wave intensities in a given frequency band as a function of time. Information is extracted when the recording is, say, Fourier-analyzed, or when it is played through an audio amplifier into the human ear and the pattern of perceived tones is recognized by the brain.

In the two above examples, data have been converted into the form of sensorially detectable signals, with the human cognitive apparatus – the brain – effecting the information extraction. It is of fundamental importance to realize that in each of these data/infromation transformation processes the *human factor* intervenes in a most crucial--and unavoidable – way. Indeed, information-extraction from any kind of data always *must* engage the human brain at some stage. If not in the actual process of information extraction--pattern recognition in the above examples – the brain is engaged in the formulation of the alternatives or questions to which the information to be extracted refers, or in the design of the apparatuses, algorithms or programs used in data and information handling.

It is also of fundamental importance to realize that information extraction *always* will require a process of pattern recognition at some stage, because questions and/or alternatives translate into patterns of parameter values – the data – that need to be searched for and recognized in order to obtain the answers the information conveys. For instance, the questions "is there a forest fire?", "is there a drought?" translate into a set of patterns that need to be searched for and recognized and measured in a LANDSAT image; the question "is there a whistler?" translates into a certain pattern that needs to be recognized by listening to, or Fourier-analyzing, a VLF record. All this of course also applies to information extraction from data that have no relation to sensorially detectable magnitudes. The answer to the question "was there a magnetic storm?" requires pattern recognition in plots of geomagnetic data. It is in this particular area, the human-machine interface, where the development and application of new technology has the greatest potential of success for IGBP data and information systems.

Quite generally, we may assert that information only becomes information when it is recognized as such by the brain. Data will remain data, whether we use them or not. Yet what is one person's information may well be another person's data. In science, information itself is almost always expressible in quantitative form and can become data out of which information of a higher level can be extracted. One thus obtains the hierarchical chains of information-extraction processes common to practically all research endeavors. An example is the conversion of raw or "*level I*" data, such as the length of the column of mercury in a thermometer, to "*level II*" data, which usually represent the actual values of physical magnitude as determined by some algorithm applied to level I data. A thermographic record or a Landsat image are examples of level II data. Similarly, "*level III*" data are obtained by processing level II data (mostly from multiple data suites) with the use of mathematical models so that information can be extracted on the global behavior of the system under observation. A weather map is a typical example of level III data. Finally, we may add a category of "*level O*" data, in which we include samples (such as ice cores, butterfly collections and moon rocks) and other material objects (such as archeological finds, scrolls, art objects); while these are not always expressible in numerical form, they do bear the characteristic features of "data".

Data can be transduced (i.e., converted from one form or medium into another); transmitted; compressed or integrated (for instance, time-averaged, or converted from multiple data suites into single-parameter values, respectively); stored; and retrieved. In each process, there is a loss of information through the introduction of noise and the involuntary or deliberate destruction of data. Information theory provides a framework with which involuntary random perturbations can be treated quantitatively. Data compression or integration processes are in themselves information-extraction processes in which the resulting information (e.g., the average values, or the values of a given function of the original data) automatically becomes data.

Issues Concerning IGBP Data and Information Systems

Data and information are closely knit concepts and should not be separated artificially in a discussion of management policies, even if the persons handling data systems and information systems usually come from different professional communities. In this chapter we shall discuss some major issues that must be addressed and resolved before a policy for the establishment and management of an IGBP Data and Information System can be formulated.

Interlinking Data

An important part of data of interest to the IGBP belong to the existing world data base. Not always will it be possible to identify of "flag" these data as appropriately tailored for global change studies. There is a need to design strategies for easy and speedy retrieval of the IGBP "component" of data currently existing in national and international repositories, and for earmarking data to be deposited there in the future.

A central IGBP data directory will be needed, as well as an international centralized organization that could refer a scientist or engineer speedily to the appropriate data repositories (with the exception of

some limited data sets in a few "traditional" disciplines). Furthermore, as more and more "Era III" (Chapter 2) data become available and the scientific research becomes more and more interdisciplinary, the establishment of "interactive data centers" where scientists congregate to work with a common data base cooperatively and in computer-interactive form will become a necessity.

Interlinking Information

There is an increasing need in all Erta III research ventures to find out in real-time who is doing what, where, when and why (the "five W's"). We don't mean here catalogs of projects published in book form 1-2 years after the projects have started: we mean *continuously updated* and electronically accessible directories. In some branches of science this already exists; for instance, the Satellite Situation Center at NASA's Goddard Space Flight Center. What is needed is an "IGBP Research Situation Center", operating on the basis of an electronic communications network linking it with the principal foci of IGBP activity in all participating countries and pertinent governmental, academic and industrial establishments.

International electronic mail networking between individual scientists, data centers, libraries, major research institutions, government agencies, and world data centers must be expanded far beyond the present state. Of significant promise is the recent progress in facsimile transmission systems. While these will not displace electronic mail, they have capabilities such as the transmission of pictures, graphs, tables and hand-written information, that an electronic mail system does not have.

The Human Factor

A directory is only as good as the persons collecting and sorting the information. A literature search is only as good as the list of keywords chosen by the person doing the search. A data base is only as good as the persons handling the data. Random errors and deliberate or involuntary destruction of information occur whenever data or information is transformed, transferred or transcribed; even if much of this can now be done by machines, such machines must be programed by humans.

It is important to exploit the capabilities offered by modern technology to maximize automation and minimize the need for human intervention in data and information systems. For instance, the development of new parallel processing and content-addressable memory systems with context-dependent information retrieval may lead to the real possibility of "intelligent searches", which are less dependent on key-wording. Quite generally, we have already stated that new technology offers its greatest promise in the area of "warm body/cold machine" interface at the input and output levels. An example of this are computer visualization techniques for the graphic display of level III data. Such techniques combine the best of two worlds: computer power with the power of the human visual apparatus, which still is the best (though not always the fastest) pattern-recognition system available.

Standardization

One of the greatest difficulties in the organization of *data* systems, especially in the environmental sciences, is the question of standardization of data. By this we mean not only the standardization of data formats, but also the standardization of the measurement instruments themselves reproducibility and intercomparability of data constitute the very basis of the scientific method, and this requires uniform, standardized measurement processes. On the other hand, the continuity of the measurement process *per se* is essential to the acquisition of a long-term data base for the study of environmental change. In particular, protocols for the standardization of ecosystems data are badly needed.

Some aspects of *information* systems per se also need standardization. For instance, programs and pertinent user instructions for electronic communication systems, particularly the log-in and retrieval procedures, differ greatly among available systems, and switching from one to another can be a nightmare even for the most experienced scientists. This problem is particularly acute when it comes to interdisciplinary work, such as will be required in IGBP projects.

Environmental Monitoring and Long-Term Data

Scientific research is based on the measurement of observable quantities and the establishment of functional relationships between those that are linked through interactive processes. There are three basic types of measurements. First, there is the class of "pioneering" measurements that may lead to the discovery of previously unknown or unsuspected relationships or phenomena. Second, there are the meas-

urements made to verify, confirm, reproduce or statistically consolidate a newly found relationship or behavior. Third, there is the class of systematic, continuous, carefully calibrated, absolute measurements necessary to obtain a comprehensive understanding of long-term trends of a given natural system.

Scientists involved in the study of natural phenomena often prefer to deal only with the first and second classes of measurements, because these may lead more readily to publishable results and topics for dissertations, leaving the less glamorous long-term measurements to the operational agencies. However, such long-term measurements, especially the monitoring of environmental parameters, are extremely important in providing insights into the global behaviour of a system. A close collaboration between scientists in research institutions and operational agencies in all participating countries is essential to develop and carry out the core projects of the IGBP.

The "Gray" Literature

The exponential proliferation of the "gray literature" – unreviewed preprints and reports with limited, usually author-controlled distribution – is presenting increasing difficulties to librarians and scientists alike. Rather than focusing on curative medicine, one should practice preventive medicine to help arrest this trend. This will require doing something about the main causes for this proliferation. They are: 1. The speed of progress in many scientific disciplines is so fast that scientists cannot afford to wait until their articles are published to promulgate their results. 2. Scientists are increasingly disenchanted with the peer review processes of many reputable journals, which in their opinion diminish their chances of publishing unusual results or bold, innovative ideas. 3. In many countries there is a notable lack of incentive for the scientists of their governmental agencies to publish results in the scientififc literature. 4. The often long delays in the public release of proprietary or classified information discourages publication of results that may be many years old.

Policy Considerations

A policy for the development of an IGBP Data and Information System and for its management should accomplish three objectives. 1. It should set general guidelines for the development and the operation of a system that is able to satisfy the needs of all participating scientists; 2. It should recommend procedures or methods for the management of data and information supplied by national agencies; 3. It should establish the basis for an international coordination of the access to those relevant data and information operations that are not part of the IGBP Data and Information System per se.

There are a number of policy issues that must be resolved. Some are discussed below, in no particular order of priority.

User-Driven Data and Information System

Any formally established system should be driven, and to a certain degree controlled, by the community of users. Data and information organizations have a tendency of developing a bureaucracy of their own unless an oversight body is established. On the other hand, however, a good measure of initiative must be left to the technical staff of a data and information system so that they can expand their user market with innovative ideas about improvements of their service, particularly, the incorporation of new technologies.

A mechanism involving the community of users is also needed to make consensus decisions about what data or information is to be stored in a dedicated IGBP repository; what activities are to be listed in a "W-5" directory (Section 4.2); how certain data should be standardized; which sources of data or information are considered unreliable. etc.

Quality Control

One of the most difficult issues in data and information management is quality control. There is "good" and there is "bad" data and information. In some cases, quality can be expressed in terms of the statistical significance of the data. But often this is not possible, either because of the incompleteness of the observations, the lack of some needed complementary information, some politicaly motivated bias in reporting, or, simply, because of a lack of adequate scientific training of the originator(s) of the data. While in general quality control is the responsibility of the scientists generating the data or information, the agency funding the research shares in this responsibility. Mechanisms to ensure the most regorous adherence to scientific principles in data acquisition should be set up; this is especially critical for ecological information and data.

Protection of Data and Information

Some data are invaluable because of the uniqueness of the event that they describe. Other data have been obtained at a great cost to their originator--often the government. All data need adequate protection from natural hazards, vandalism, terrorist action, computer "viruses" and accidental erasure. This could be a costly endeavor, but it must be taken into consideration in the formulation of data and information policy.

The Cost of Data and Information

It is virtually impossible to establish the monetary value of a given set of scientific data or a given type of information, to the point that it would be highly unrealistic to attempt setting up a "self-supporting" IGBP Data and Information System. The most one could achieve is an operation in which the nominal cost of each information or data request is recovered (processing of the request, materials used and mailing).

References

Bell, D., *The Coming of the Post-Industrial Society*, Basic Books, Inc., New York, 1973.

Arthur D. Little, Inc. *Into the Information Age,*, report, V. Ciuliano, Project Director, Am. Libr. Assoc., Chicago, 1978.

ICSU, International Council of Scientific Unions, Constitution of the Special Committee for the International Geosphere-Biosphere Program: A Study of Global Change, 1988.

Roederer, J.G., Human brain functions and the foundations of science, Endeavour, Vol. 3, No. 3, pp. 99-103, 1979.

Roederer, J.G., Considerations in the development of a national geophysical data policy, EOS, Vol. 62, No. 27, pp. 569-570, 1981.

Roederer, J.G., The proposed International Geophere-Biosphere Program: some special requirements for disciplinary coverage and program design, in *Global Change*, T.F. Malone and J.G. Roederer, editors, ICSU Press, pp. 1-19, 1985.

Roederer, J.G., Tearing down disciplinary barriers, Astrophys. and Space Sci., 144, 659-667, 1988.

A DATA MANAGEMENT PLAN FOR THE SOLAR-TERRESTRIAL ENERGY PROGRAMME (1990--1995)

David Rees

Department of Physics and Astronomy,
University College London
Gower Street, London WC1E 6BT, UK

Abstract

This document describes, in outline form, the data base, data management and communication requirements for the Solar Terrestrial Energy Program (STEP) which has recently been initiated and sponsored by the Scientific Council for Solar Terrestrial Physics (SCOSTEP). STEP will run in the period 1990 to 1995, and will involve a number of space missions, and many ground-based, ballons, aircraft and rocket facilities. In many instances, these facilities will be used in innovative ways and in large-scale coordinated programmes. The experimental studies will be supported by strong initiatives in appropriate theory and numerical modelling work. It is expected that these programmes will generate much larger data bases than previously common in the Solar Terrestrial Physics (STP) community. The success of STEP will depend to a large degree on the ability of the STP community to store, retrieve and inter-communicate the new data obtained during STEP. The data and related requirements outlined here are based on the current plans of the International Steering Group for STEP. These plans are currently in the process of development through the wider involvement of the STP Community. A special Panel is being set up by the STEP Steering Group to deal with aspects of 'Informatics', as these become more defined in detail by the community involved in STEP.

Introduction

The following statement is taken from the pre-amble to the Solar Terrestrial Physics Program, as presented to, and approved by, the Scientific Committee on Solar-Terrestrial Physics (SCOSTEP).

"The Solar-Terrestrial Energy Program (STEP) is a program of the Scientific Committee on Solar-Terrestrial Physics (SCOSTEP) of the International Council of Scientific Unions (ICSU) scheduled to run over the period 1990 – 1995. The essence of the program lies in the study of energy transfer processes across boundaries between regions which, until now, have been mainly studies in isolation. During the 1990s, STEP will help to coordinate selected large-scale international projects whose results will help to further our understanding of the coupling processes at work in the interfaces between adjacent regions of space (e.g. solar wind – magnetosphere, ionosphere/thermosphere – middle atmosphere). The studies will involve both theoretical and experimental investigations... It is expected that these studies will require widespread use of advanced electronic communication systems. Such communication systems will facilitate planning and execution of STEP projects, and will also allow the maximum scientific participation and exploitation of the large amounts of data which will be acquired, deposited into joint data bases for the cooperative analysis of the various projects which will fall within the overall STEP program."

Overall organisation of STEP

A large series of programmes, such as those envisaged within STEP, require a high level of coordination and exchange of information. This coordination and information exchange affect every facet of the programmes from outline and initital planning, to the coordination of joint observing intervals, and the subsequent exchange of data and ideas, leading to the joint intepretation of results, and the presentation of new findings to the wider community.

"SCOSTEP has approved the formation, in 1987, of a STEP Steering Group, reporting directly to the SCOSTEP Executive. The STEP Steering Group consists of 10 scientists, with a broad geographic distribution, and several Ex-Officio members, representing other International Scientific Organisations which have considerable interests either in the objectives of STEP, or in the facilities which STEP projects will set up, utilise or exploit.

The executive summary of the rationale and objectives of STEP are described in Appendix. To provide a forum for discussion, formulation and approval of STEP Sceintific Projects, and to arrange coordination of essential common facilities, the STEP Steering Group has set up a number of Working Groups and Panels. The Working Groups are expected to be the focus of the experimental and theoretical studies which will be carried out under the auspices of STEP, while the Panels provide for the coordination of essential supporting facilities, including correspondence with National and International Agencies whose facilities may be requested or required in order to provide adequate substructure for widespread STEP-related activities.

The strategies and objectives of these Working Groups and Panels are presently being formulated, and it is clearly necessary to co-opt additional members to these Working Groups and Panels.

In the cases of the Working Groups, the co-opted members will be primarily the leaders of major projects which fall within the STEP Scientific Program, selected International Scientists, and the leaders of certain essential scientific facilities.

In the case of the Panels, the co-opted members will be invited from the major facilities and National/International organisations which are essential for the basic common substructure and support of STEP-related activities and projects.

In both cases, it is likely that additional, Ex-Officio, members will also be sought from other International Scientific Organisations which have a strong interest in STEP activities."

The 'Informatics' Panel, set up to coordinate information gathering, storage, communication and exchange, as well as aid the coordination of major projects, has a particularly difficult task. Its provisional membership, includes scientists from senior staff at national facilities and World Data Centers.

Data and Data Management

Instrumentation

At the heart of a new initiative such as STEP, novel instrumentation, and the opportunity to place instruments in key regions not previously explored, have a particularly important role. In the analysis of physical processes in environments ranging from the solar surfaces to the terrestrial surface, many diverse types of sensors are required. These sensors have to provide a combination of in situ and remotely sensed measurements and parameters for given spatial regions. For some investigations, localised measurements are required, however, form many studies, data for brief to extended periods, from a number of instruments located on separate, suitably located spacecraft, or from several ground-based locations will be required. Often, investigations will require incorporating multi-parameter data from several spacecraft and from several ground-based locations. Most of the data is initially recorded as time series, over variable periods of time and with widely diverse sampling intervals. In some cases, the original format of the data does not represent a time series, but time series of important parameters can be formed through analysis of the data in its original format (e.g., image processing). In the sections below, some selected examples of the types of data and their formats are outlined for the various regions of space relevant to STEP.

Some typical instruments used in each of the major discipline areas appropriate to STEP are listed below. The list is not exhaustive. New instruments and platforms, and innovative methods of storing and displaying data, particularly from combined data sets, will inevitably (and hopefully) modify any attempt to make such a list exhaustive at this early stage of planning.

In previous international programmes, theory and numerical modelling work have often been treated as separate entities, with little provision for their support and data requirements within the major projects and programmes. This has often led to a low level of support for essential modelling activities, and frequently, it has been very difficult to coordinate and communicate between experimental data holdings and numerical model data sets, retarding the exploitation and development of both aspects. For STEP, it is crucial to appreciate that theory and modelling activities are essential and that the appropriate model data sets are large and diverse. Adequate data storage and good inter-communication between the experimental and model data sets is required, as between the respective investigators. No particular examples of formats will be given: These days, with diverse and global models, the storage and display formats tend to be similar to those discused above for experimental data sets, since these are widely used for inter-comparison. Many models, for example of the solar wind--magnetosphere interaction, processes within the magnetosphere, and of the magnetosphere /ionosphere or ionnosphere/ thermosphere, are truly global, and often time-dependent. Rather special displays, corresponding to the global empirical models perhaps derived by combining the experimental results from many satellites or facilities over a number of years, are in common use, and will be quite demanding of data centres, as well as of great value. Naturally, comparison with traditional and new empirical and semi-empirical models will also continue to be extremely important: the appropriate formats and displays will be similar to those required by the numerical models.

Data Collection Modes

Studies of the solar-terrestrial environment involve three distinctive types of data acquisition:

- Long-term measurements at a specified sample rate (decades).
- Medium-term measurements over the lifetimes of the experimental apparatus at variable sampling rates (years).
- Short-term campaigns with variable sampling rates (days- weeks)

The data recorded during the modes of operation outlined above range from sequences of images, through strip charts containing traces of varying parameters, to magnetic tapes containing time streams of data recorded in digital form. Digital form is the natural format for data measured in-situ by instruments onboard satellites, rockets and ballons. Digital formats are also being used increasingly by ground-based investigators. For a number of investigations, data streams are now routinely transferred from the recording stations to the host laboratory using modern techniques of electronic telecommunications. However, it is important to note many data sets used in the STEP program will not be in digital form, because of the lack of resources available ot participating scientists in some of the participating countries. While conversion of some of these data to digital form for selected study intervals may be achieved, means must be found for the storage an dissemination of a significant amount of data which will be made available only in the form of "hard copy".

Data storage:

It has been customary, in the past, for individual experiments to process and store their own data sets. However, in cases of data sets which have widespread use, it is not unusual for those data sets to be provided to the World Data Centers after some reasonable (but often lengthy) time delay. Some typical data sets and formats are listed below to give some indication of the diversity of recording techniques and storage formats.

Ground Magnetograms

Strip charts of length 0.5 m and height 0.25 on which appear traces of the time variations of three orthogonal components of the terrestrial magnetic field at the recording site.

OR

the same information reduced in "image size" and stored on microfilm (rolls) or microfiche.

OR

the same information encoded on magnetic tape at sample intervals as short as 10 s and as long as one minute. The sampling time is at the discretion of the researcher or agency providing the data and can be non-standard on occasion (special study intervals).

Ground Ionosondes

Images showing the various layers of the inonosphere through the use of a grey scale to identify the location of refracted waves at specified frequencies. The ordinate axis on the rectilinear image often indicates the virtual height of the layer.

OR

the same information as a digital time stream from which the user, employing a standard computer routine, can reconstruct the ionograms.

Ionograms are usually recorded once every 15 minutes at a given location, although higher sampling rates may be used for specific scientific (as opposed to operational) reasons.

Satellite particle detectors

Data are in digital form and may be decommutated to allow a single time series of data to be extracted. Sampling times of 10^{-3} s are not uncommon for some experiments, so that a satellite experiment can produce data rates and volumes which are orders of magnitude higher than those required by typical ground-based instrumentation for a given sample interval. The storage medium is normally magnetic tape although, in some cases, the data are averaged to produce values at intervals more commensurate with microfilm or microfiche as a storage medium. For example, solar wind magnetic field data from the IMP 8 satellite are reduced to 15.36 s averages and are presented as time series on microfilm. Subsequently, hourly values of the interplanetary parameters are computed and made available to the scientific community in tabular and graphical forms.

It is common, in the solar-terrestrial community, for the individual responsible for acquiring data to make these data available on request for collaborative projects. Some reasonable amounts of data are also provided on request by many experimenters for colleagues to use for thier own research. Inclusion of an agreed acknowledgement of the source of the data in any subsequent publication is often the only precondition for its use. There is every reason to belive that this traditional means of data dissemination will continue over the years of the STEP program. However, the World Data Centers should be prepared to receive data sets in highly variable formats for some time to come.

Frequently, it is advantageous for the supplier of data to even provide the initial cost of the medium for transmitting the data to the user (microfiche, magnetic tape etc), it being cheaper or less onerous than having to cope with diverse media and quality control of items supplied by users. In the case of electronic data bases also, there is often no cost to the user for access to essential data. Similarly, it may be impractical to arrange reimbursement to the central data source for cost of data processing or acquisition. With the anticipated data volumes, and increased number of users, however, we might expect that the question of "user pays" may have to be considered in some detail. This is particularly critical in regard to access they require for their research and contribution to STEP. High costs of data are likely to severely deter wide participation.

Data Dissemination by Electronic Means

Between 1984 and the present, there has been a rapid and fundamental change in the way that scientists are able to communicate with one another. The development of the facsimile reproduction machine (FAX) allows data to be sent electronically, reducing communication times from days to seconds when warranted. There has been a rapid development of national and international electronic data networks, for example the commercial Bitnet and Telemail services, and the academic /research networks such as the Space Physics Analysis Network (SPAN) within the USA and to Europe, the Joint Academic Network (JANET) within the UK and the European Academic Network (EARN), and European Space Information System (ESIS), now being set up as a pilot scheme. The rapid development of such services, and the increasing access to such service by much of the research community within Europe and the USA and efficient Gateways between these services, has profound imlications for the way STEP scientists will be able to interact with each other and with data during the 1990s. It will become commonplace for a scientist to access a data set, held the other side of the world, and to be able to manipulate data within that data base in order to gain additional knowledge for the research in hand. One vital question is whether such modern communication links will be available to members of the STEP community for the entire duration of the STEP interval. It is also recognised that such networks require special equipment and personnel which are not always readily available to smaller research groups in Europe and North America, let alone in underdeveloped and developing countries. Even network access and data transmission costs may consume significicant resources, which should not be allowed to detract from research resources. A major effort will be required to provide information to the community

concerning optimum and appropriate cost-effective access to networks. In each case, solutions will have to be tailored toward available consistent with adequate access to STEP data. In the majority of cases, such remote access will make new and additional data available to the STEP community, while also enriching the intellectual effort available to work on and analyse new and complex research projects.

Use of the SPAN network, which is now linked to many sites in the USA, Canada, Europe and recently to Japan, is presently available to many participating institutes and scentists at a notional charge. This is not strictly a 'zero cost', since a significant front-end investment has to be made to provide a suitable host computer for the local SPAN node. Similarly, suitable system level software and licences have to be obtained, and experienced computer personnel have to be engaged to commission and maintain the networks as part of an enhanced overall capability. For Institutes within countries outside North America, the costs are often increased by the considerable additional costs of international network traffic. It is important that the most economic methods of installation and running the networks, in term of their impact on the local host computer node utilisation, and particularly for international electronic data traffic are well understood. It is quite possible that other 'mailbox' or network facilities will become more widely available during the STEP interval. It is certainly essential that good advice, based on sound research into the available facilities, is widely available to possible participants, to provide enhanced data communication at the most appropriate rate, by the most convenient means, and at a budget level which, as with the scientific or technical contribution, is likely to be highly variable, and dependent on location (country). While expansion of SPAN may work excellently for North America, Europe and Japan, other methods of linking to international data routers may be more appropriate for large countries such as India: extension of the uplink/downlink capabilities of satellites such as METEOSAT or NOAA/TIROS could perhaps provide and economical methods of linking numerous scientific groups dispersed over wide areas, where SPAN-like connections would be uneconomic.

In addition to advision on optimum solutions to the problems discussed above, it is inconceivable that a true unification of all electronic networks and mailbox facilities will happen in the near future. Inevitably, this means that numerous Gateways between distinct computers and network systems have to be established and maintained. The information about optimum use of equipment and protocols needs to be accumulated, organised and disseminated to interested parties. Similarly, the protection of networks and data bases, and the management of authorised accounts, and passwords represents a significant, complex and expensive burden, but one which has to be accepted if the quality of service is to be maintained.

Just as a true unification of all network protocols is unlikely, a series of diversified back-up methods of data storage and dissemination will be required. Magnetic tape is almost certain to remain a cost-effective method of shipping large data volumes, particularly when intercontinental distances are involved. Optical disk storage technology is still maturing, and is already being used to complement or replace magnetic tape back-up in many facilities. It is likely that optical disks will ultimately become as widely used as magnetic tape or floppy disk for the storage and shipment of intermediate to large data holdings.

In all of the above areas, and probably in new areas not yet defined, the STEP organisation has a complex task to perform: The goals are to involve as many STEP researchers as possible in programmes to which they can contribute, via the most powerful, yet affordable modern tools of communication. At the same time, it has to be recognised that the communication is a means to an end, not an end in itself. Practical problems, such as persuading diverse computer managers to maintain current directories of users and network nodes and addresses, are added to the normal duties to maintain accurate directories of experimental data, and that obtained from increasing sophisticated numerical models and simulations. National and International organisations and multi-national corporations involved in Computer equipment and telecommunications must be persuaded that there is a great long-term scientific, technical educational and economic benefit to be obtained from the wide-spread availability of low-cost network access and data traffic, with special consideration of developing and underdeveloped countries.

Special Considerations, particularly contrasts with previous major programmes:

1. In USA/Europe a larger proportion of participants (than for previous programmes) will store and maintain data in machine-readable form.

Initial formats will mainly be of *local* convenience as will the medium!

2. Given available protocols/software much reduced and raw data can be converted and verified and networked to data centres.

However; not all, and much very important and urgent data may still rely on charts, odd-format tapes, film etc.

Electronic Mail services have been identified as having great value to STEP. The following are particularly important:

- Bulletin board

- 'Situation' centre
- Information system

3. Data Analysis – Initial Science (Days--1-2 Years)

This is strictly the province of local institute /host computer/ data base.

It is essential that efficient network links facilitate data exchange direct between participating groups, on an international basis.

Data centres provide: Geophysical Indices; selected space data; models. (also catalogues of available data and repository)

4. Data analysis – mature phase 1 year – 10 years

Encourage deposition of working, multi-parameter/instrument/spacecraft data sets in D-C's, soon after the event. Too often, it is very hard to re-assemble sets 'later on'. Encourage use of simple and 'self-explanatory' formats, with adequate documentation.

Instrument knowledge/calibration is critical, i.e. quality control.

Science teams for Missions and facilities have to be determined to help!

5. Situation in East Europe, USSR, developing countries.

A very significant fraction of the global resource of knowledge, potential effort, and data, all potentially, and in many instances, actively, involved in STEP and other major international programmes, is still very hard to contact by electronic means.

Cost of data base and network facilities:

The Scientific, Technical and Programmer effort to set up/maintain protocols and standards is large. Issues such as the security and quality control of networks and data bases is another expensive item. The cost and management of running central facilities and leased lines (local PTT's) are, or can be, major problems.

It is essential to match local facilities to local resources and requirements. In this respect, considerable advice should be available from European and USA networks and D.C's to potential participants within developing countries. However, some of the optimum and most cost-effective means of data exchange, particularly with participants in developing countries, may require different approaches to those currently in use in theUSA and Europe.

Acknowledgements:

This presentation is made on behalf of the International Steering Group for STEP. The work undertaken in preparing this overview of the initial planning of the STEP Steering Group, and its likely impact

in the area of Geophysical Informatics rests very heavily on contributions from the entire Steering Group, and particularly from a detailed series of notes, on the topic of STEP data requirements, specially prepared by Dr. Gordon Rostoker, its Co-Chairman.

Dr. Don Williams, Chairman of the SCOSTEP Working Group which drafted the framework for STEP, and Prof Juan G. Roederer, President of SCOSTEP Bureau, and many others in the STP community, have played major roles in formulating the program, presenting it to the scientific community, and guiding it through the stages of discussion and approval by ICSU. Appendixes 1, 2 and 4 are taken directly from the booklet "STEP--A Framework for Action", published by SCOSTEP in 1988.

Appendix: Step Executive Summary

Introduction

During the past decade, SCOSTEP has aided the scientific community by organising many proposed studies of the solar-terrestrial environment into coherent programs of international scientific cooperation and by establishing special data and information services for these programs. These programs were focused largely on individual regions of the solar-terrestrial system: the sun, the solar wind, the magnetosphere, the ionosphere-thermosphere, and the middle atmosphere.

Solar-terrestrial research has attained a point in its evolution where it is desirable to put more emphasis on the comprehensive study of the mutual linkages between the various regions of space from the sun to the earth, in addition to the traditional study of the individual regions themselves. STEP will focus on the solar-terrestrial environment as a complex interactive system whose overall behaviour often drastically departs from the simple superposition of its parts.

The main goal of STEP will be to advance the quantitative understanding of the coupling mechanisms that are responsible for the transfer of energy and mass from one region of the solar-terrestrial system to another.

STEP will involve ground-based, aircraft, balloon, rocket and satellite experiments: theory and simulation studies; and dedicated data and information systems. Integral to the success of STEP is the set of

solar-terrestrial spacecraft missions approved by the Inter-Agency Consultative Group as the next cooperative project of NASA, ESA, ISAS, and INTERCOSMOS. The program will also take advantage of results obtained by other relevant spacecraft missions. STEP is expected to begin in 1990 and terminate in 1995; however, an extension of the STEP interval may be envisaged if some key spacecraft missions are delayed.

Priority Areas and Their Principal Goals

The basic framework of STEP will consist of Priority Areas, each one with a comprehensive goal of scientific understanding of the interaction mechanisms controlling energy and mass transfer between specific regions of the solar-terrestrial system.

The Sun as a Source of Energy and Disturbance

To achieve an understanding of the principal source mechanisms for electromagnetic and corpuscular emissions on the sun and in the solar environment, and to formulate physical models for improving the predictability of short-term perturbations (minutes to days) and long-term variability (years to decades).

Energy and Mass Transfer through the Interplanetary Medium and the Magnetosphere-Ionosphere System:

To achieve an understanding of the energy, momentum and mass transfer mechanisms across shocks and the boundaries that separate the distinct plasma regions of the solar-terrestrial system, and to study the acceleration, diffusion and convection processes and large-scale instabilities that distribute and modify the complex corpuscular flows and fields in that system.

Ionosphere-Thermosphere Coupling and Response to Energy and Momentum Inputs:

To achieve an understanding of the global processes which determine the coupling and interactions among the neutral and ionized species in the ionosphere-thermosphere system, and to study the response of the system to changes in solar input, and to energy and momentum transfer by particles, fields and waves from adjacent regions.

Middle Atmosphere Response to Forcing from Above and Below:

To achieve an understanding of the response of the middle atmosphere to changes in solar and near-space inputs and to volcanic, tectonic meteorological, biospheric and anthropogenic activity, and to study the extent to which this response feeds back to the regions of the geosphere below and above.

Solar Variability Effects in Regions Adjacent to the Earth's Surface:
To determine the influences of solar variability on the physical and chemical properties and the large-scale behaviour of the lower atmosphere, on man-made technological systems, on earth currents and on biota, and to formulate, test and study mechanisms responsible for these effects

Working Groups and Panels

The actual scientific research will be organized into a limited number of well-defined cooperative projects proposed by the scientific communities of the participating countries. These projects may deal with subjects entirely within a Priority Area, or cut across two or more of them. To define, plan and implement the research coordination, Working Groups and Panels have been established.

The activities in the above Priority Areas will be coordinated by Science Working Groups, one for each Area. In addition to these Working Groups, the science support and service activities will be defined and coordinated by several Programmatic Panels.

Two major panels, a *Panel on Informatics* and a *Panel on Long-Term Measurements*, will have the responsibility of formulating recommendations regarding the establishment and operation of dedicated research support centers and monitoring networks, as well as coordination the activities of these centers and networks with individual STEP projects.

Two other panels will become active mainly when the specific STEP research projects have been defined. They will serve as forums for the scientific discussion of topics of common interest to several of these projects: a *Panel on Common Mechanisms in the Solar-Terrestrial System*, and a *Panel on Experimental Techniques*.

Other panels may be established as the need arises. For instance, a Panel on Modeling and Simulation may be set up to test quantitative models of the solar-terrestrial environment and methods of numerical simulation, and to formulate recommendations on the use of standard models and simulation techniques.

IONOSPHERIC INFORMATICS WORKING GROUP OF URSI COMMISSION G

Bodo W. Reinisch

University of Lowell
Center for Atmospheric Research
450 Aiken Street
Lowell, Massachusetts 01854
U.S.A.

Abstract

Objectives and activities of the URSI Working Group G. 4 on Ionospheric Informations are described in this paper. The IIWG was formed during the XXIInd General Assembly of the International Union of Radio Science in Tel Aviv, 1987, to promote the application of information technology to acquisition, archiving and distribution of ionospheric data.

Objectives of the URSI Working Group on Ionospheric Informatics

In May 1987, the first International Workshop on Ionospheric Informatics was held at Novgorod, USSR, under sponsorship by URSI, COSPAR, and the USSR Academy of Sciences. Computer processing of ionospheric information, data driven ionospheric models, ground-based and satellite borne digital ionosondes, incoherent scatter radar, and ionospheric data archiving and distribution were the topics of this workshop. Commission G of URSI, recognizing the importance of this work, formed the Working Group G. 4 on Ionospheric Informatics /1/ during the XXIInd General Assembly in Tel Aviv, Israel in 1987. The terms of reference for the IIWG were defined as follows: "To promote the application of information technology to the acquisition, processing, archiving and distribution of ionospheric data." The IIWG held its first Business Meeting in Tel Aviv on 2 September 1987.

It was clear from the beginning that the group members cannot productively work on too large a number of topics simultaneously. It was agreed to concentrate on four immediate tasks, while other topics presented at the Novgorod Workshop /2/ will be focussed on later. The items requiring immediate attention were the following: (1) Availability of electron density profiles in the World Data Centers, (2) The validity of the electron density profiles in the presence of an E-F valley, (3) Handling of the data from the digital ionosonde network, and (4) Establishment of scaling conventions for obliquepath ionograms. It is clear that the somewhat arbitrary limitation to these four items leaves out other urgent concerns, such as the data formats for incoherent scatter radar data, the optimal media for archiving of ionospheric data, review of data base software, access to other geophysical data (geomagnetic indices, interplanetary magnetic field), ionospheric mapping techniques, and so on. The IIWG will address some of these subjects in due time, others are handled by other working or task groups. For instance, Working Group G.3 of URSI deals with ionospheric modeling and mapping, and the ISR stations have already organized their data in a report issued by NCAR /3/ (1988).

IIWG Activities

In July 1988, a Workshop on Ionospheric Informatics and Empirical Modeling was conducted at Helsinki during the COSPAR General Assembly, cosponsored by COSPAR, URSI and IAGA. The proceedings of this workshop will be published in "Advances of Space Research."

Following the workshop, the IIWG conducted a Business Meeting where reports were presented on each of the four priority topics. D. Bilitza (USA) presented the report on N(h) data availability at the World Data Centers, prepared by J. Allen (USA) with inputs from R. Conkright (USA), A. Feldstein (USSR), D. Willis (UK), and D. Bilitza. The WDC-C2 in Japan did not send any information. From the report it became clear that the existing data base on N(h) profiles is inadequate for accurate modeling of the global ionosphere, e.g., for the International Reference Ionosphere (IRI). Bilitza announced the publication of a report on the "Worldwide Ionospheric Data Base" /4/ which he is preparing at the National Space Science Data Center. This report, which will be issued later this year, contains a wealth of information on available ionospheric data, data centers, observation networks, etc.

K. Rawer (FRG) summarized the inputs from T. Gulyaeva (USSR) and J. Titheridge (N.Z) on N(h) reduction from ionograms. If all ionosonde stations would provide accurate N(h) profiles, the data base would be quite adequate for the modeling of (at least the bottom side) ionosphere. The two biggest problems are that firstly, ionosondes do not routinely calculate the electron density profiles, with the exception of the Digisonde 256 network, and secondly, the accuracy of the calculated F-region profiles is very poor when an ionization valley exists between the E and F layers. Titheridge's new physical model of the valley prescribes boundaries for the valley shape which limits the errors in the F-profiles. The N(h) report emphasizes the need for calibration of the ionogram inversion techniques. Only comparison with ISR profiles can provide this calibration for different times of day and different seasons. Reinisch (USA) noted that University of Lowell and MIT groups are conducting comparative studies using the Digisonde 256 at the Millstone Hill Incoherent Radar Site in an attempt to resolve the valley ambiguity. The Lowell group uses the changes of group delay, phase and amplitude with frequency for both the ordinary and extraordinary traces.

A topside sounding system with at least three polar orbit satellites and an extended ground-based digital ionosonde network would provide the data base required for accurate global modeling of the ionosphere.

Reinisch, using inputs from M. Abdu (BRA), Sizun (FRA) and A. Feldstein (USSR), reported on the available data from the digital ionosonde network. A typical output from the Digisonde 256 at Millstone Hill is shown in Figure 1. The autoscaled parameters foF2, foF1, h'F, h'F2, M3000, Fmin, foEs, MUFF2(3000), fminF, fx1, fminE, foE, h'E, h'Es, the average range and frequency spread for F and E layers, QF, QE, FF, FE, and the coefficients for the vertical electron density profile are listed on top of the printout. For archiving purposes, the Digisonde Users Group recommends a simple ASCII text file format /5/. The database file is shown in Table 1, with each block corresponding to data from one ionogram. This database format is flexible, expandable, and easy to read at most computers.

P. Bradley's (UK) report on the scaling of oblique-path ionograms emphasized the need for standardization of the scaling rules so that a consistent database can be established using data from different propagation links.

```
+---------------------------------------------------------------------+
+      ULCAR - MILLSTONE HILL, WESTFORD, MASSACHUSETTS                +
+                LAT 42.6, LONG 71.5W      DIP 72.9  fH 1.4           +
+      DIGISONDE 256 - VS.02.B      UNIVERSITY OF LOWELL. USA         +
+---------------------------------------------------------------------+

STATION YEAR DAY H  M      OUT OPT B  E  Q CAB XLZT NRW HEIG PROGRAM
  043   1987 293 17:30 UT-0483100 01-11 1 32E 4101 534 123A     2
                 --------

FOF2  FOF1  H'F   H'F2  M3000 FMIN  FOES  MUF   FMINF
 8.2  4.1   208.  260.  3.07   1.4   4.7  25.2   3.0

FXI   FMINE FOE   H'E   H'ES  QF    QE    FF    FE
 9.1  1.4   2.9   105.  110.   5.   10.    .2   1.2

AUTOSCALED TRACES (KM):
-----------------------
 3.  210.  190.  203.  203.  203.  208.  213.  218.  223.  233.
 4.  250.  253.  255.  260.  258.  258.  258.  258.  258.  258.
 5.  258.  263.  263.  263.  263.  263.  263.  268.  268.  268.
 6.  268.  273.  273.  278.  278.  283.  283.  288.  288.  293.
 7.  303.  308.  313.  318.  328.  338.  348.  373.  383.  413.
 8.  443.  488.  598.
 1. ***** ***** ***** *****  105.  105.  105.  105.  110.  110.
 2.  115.  115.  120.  120.  125.  130.  135.  140.  155.  185.

NORMALIZED AMPLITUDE AS AT REFLECTION HEIGHT 100KM IN [DB]
        TOPF    2.    3.    4.    5.    6.    7.    8.
 F       22.    0.   43.   52.   61.   61.   68.
 E       19.   58.
ES        0.    0.   36.   44.    0.

PROFILE - ULCAR
---------------
   W =    .0  KM
   FSTART PEAK HT   AU        A1       A2       A3       A4      DEV    ROOTS
   [MHZ]  [KM]     [KM]      [KM]     [KM]     [KM]     [KM]    [KM]/PT
E   .199 119.481 -38.098   8.694   -5.084                         1.7    -
F1 2.999 166.031 -41.266  -2.016   -3.301    1.319   -1.295     10.1    -
F2 4.199 271.786-101.118    .465   -4.132   -1.149     .176      4.9    -
```

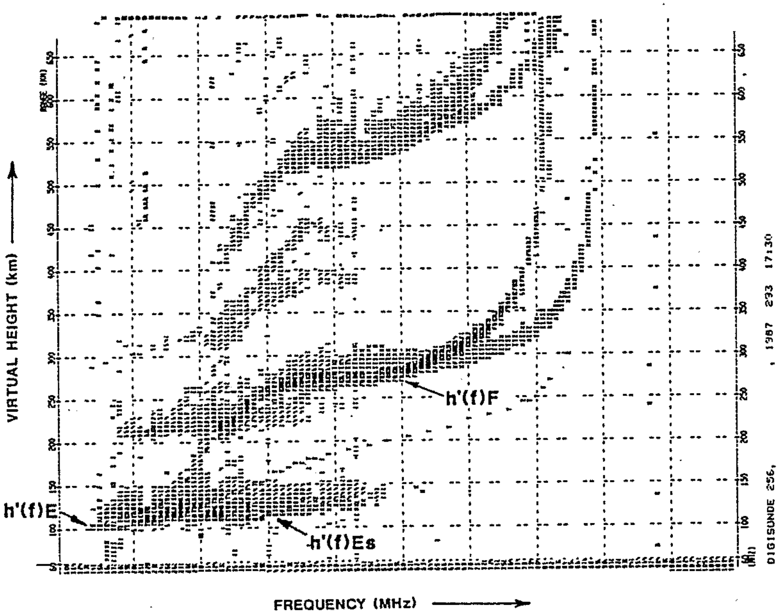


Figure 1 - Automatically scalled digisonde ionogram

Code	Format	Description	
	40I3	DATA FILE INDEX	
1	60Z1	IONOGRAM SOUNDING SETTINGS (PREFACE)	
2	50F8.3	SCALLED IONOSPHERIC PARAMETERS	
3	20I2	ARTIST ANALYSIS FLAGS	
4	10F7.3	GEOPHYSICAL CONSTANTS	
5	16F7.3	DOPPLER TRANSLATION TABLE	
6	400I3	ARTIST O-TRACE POINTS - F2 LAYER	VIRTUAL HEIGHTS
7	400I2		AMPLITUDES
8	400I1		DOPPLER NUMBER
9	400F6.3		FREQUENCY TABLE
10	150I3	ARTIST O-TRACE POINTS - F1 LAYER	VIRTUAL HEIGHTS
11	150I2		AMPLITUDES
12	150I1		DOPPLER NUMBER
13	150F6.3		FREQUENCY TABLE
14	150I3	ARTIST O-TRACE POINTS - E LAYER	VIRTUAL HEIGHTS
15	150I2		AMPLITUDES
16	150I1		DOPPLER NUMBER
17	150F6.3		FREQUENCY TABLE
18	400I3	ARTIST X-TRACE POINTS - F2 LAYER	VIRTUAL HEIGHTS
19	400I2		AMPLITUDES
20	400I1		DOPPLER NUMBER
21	400F6.3		FREQUENCY TABLE
22	150I3	ARTIST X-TRACE POINTS - F1 LAYER	VIRTUAL HEIGHTS
23	150I2		AMPLITUDES
24	150I1		DOPPLER NUMBER
25	150F6.3		FREQUENCY TABLE
26	150I3	ARTIST X-TRACE POINTS - E LAYER	VIRTUAL HEIGHTS
27	150I2		AMPLITUDES
28	150I1		DOPPLER NUMBER
29	150F6.3		FREQUENCY TABLE
30	20I2	MEDIAN AMPLITUDE OF F ECHO	
31	20I2	MEDIAN AMPLITUDE OF E ECHO	
32	20I2	MEDIAN AMPLITUDE OF ES ECHO	
33	20E9.4E1	TRUE HEIGHT F2 LAYER COEFFICIENTS	
34	20E9.4E1	TRUE HEIGHT F1 LAYER COEFFICIENTS	
35	20E9.4E1	TRUE HEIGHT E LAYER COEFFICIENTS	
36	20E9.4E1	TRUE HEIGHT MONOTONIC SOLUTION	
37	20E9.4E1	VALLEY COEFFICIENTS	

N O T E S

Nomenclature:

- Block - All data for one ionogram.
- Group - All lines of data for a single Code.
- Line - A sequence of Elements, CR/LF terminated.
- Element - A single datum in the specified format.

Table 1 - ARTIST Data Editing Program (ADEP) Block Format

Summary

The Ionospheric Informatics Working Group G. 4 of URSI has made a first step in collecting and distributing information on special topics of interest: N(h) profiles, digital ionosonde data, and oblique-path ionogram scaling. Reports on these topics will be published in 1989 in "Advances in Space Research" together with the proceedings of the Ionospheric Informatics Workshop at the COSPAR meeting in Helsinki.

References

1. INAG No. 51, Ionosonde Network Advisory Group, URSI WG G.1, "Ionospheric Station Information Bulletin No. 51", February 1988, WDCA for STP, NOAA, Boulder, Colorado 80303, USA.

2. K. Rawer, T.L. Gulyaeva and B.W. Reinisch (Editors) "Ionospheric Informatics," Adv. in Space Research, 8, 4, 1988.

3. "The NCAR Incoherent-Scatter Radar Data Base," Catalogue, May 1988, National Center for Atmospheric Research, Boulder, Colorado 80307, USA.

4. D. Bilitza, "The Worldwide Ionospheric Data Base."

5. R. Gamache, Z. Zhang, T. Bullett and B. Reinisch, "ADEP Database Format," Technical Note, University of Lowell Center for Atmospheric Research, Lowell, MA 01854, USA, April 1988.

DATA EXCHANGE BETWEEN WDC-A-R&S AND WDC-B2-STP

Joseph H. King

World Data Center-A for Rockets and Satellites
National Space Science Data Center
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

Abstract

Selected data exchanges between the U.S. and the USSR, both within the context of the World Data Center structure and outside that context, are described. (Note that the original presentation at the Moscow Workshop incorporated both this topic and information exchange, directories, and international coordination thereof; this latter topic is logically distinct, and is reported in a separate paper in these Proceedings.)

Introduction

World Data Center-A for Rockets and Satellites (WDC-A-R&S), located at NASA's National Space Science Data Center (NSSDC), has a charter to capture and exchange information about spacecraft and rocket launches and orbit and payload characteristics. As such, WDC-A-R&S has no archive of sensor data. However, owing to its collocation with NSSDC, it is able to respond to requests for data from the international scientific community by drawing on the data archived at, and the services of the staff of, NSSDC. In fact, about 30 percent of the 2800 requests for data and information handled by NSSDC in 1987 were from the non-US science community.

TABLE 1

SOVIET DATA PROVIDED BY WDC-B2-STP TO NSSDC/WDC-A

AUREOL 2	MASS SPECTROMETER	(H+ AND O+; 3 DAYS EARLY 1974)
PROGNOZ 6	9/77 - 1/78	5-MIN. AVERAGED MAGNETICFIELD VECTORS; 1-HR. PLASMA PARAMETERS
PROGNOZ 7	11/78 - 6/79	5-MIN. AVERAGED MAGNETIC FIELD VECTORS; 1-HR. PLASMA PARAMETERS
PROGNOZ 9	1983	5-MIN. AVERAGED MAGNETIC FIELD VECTORS
PROGNOZ 10*	4/85 - 11/85	10-MIN. AVERAGED MAGNETIC FIELD VECTORS; HOURLY ELECTRON, PROTON,ALPHA FLUXES
IZMIRAN MODEL		HIGH LATITUDE MAGNETOSPHERIC ELECTRODYNAMICS PARAMETERS

*THESE DATA HAVE BEEN ADDED TO NSSDC "OMNITAPE," AND ADDITIONAL PROGNOZ DATA WILL ALSO BE ADDED

TABLE 2

U.S. DATA PROVIDED BY NSSDC/WDC-A TO WDC-B2-STP

OMNITAPE	HOURLY SOLAR WIND FIELD/PLASMA COMPILATION	1963-87
IMP-8	5-MIN. MERGED SOLAR WIND FIELD/PLASMA	1977-80
IMP-8	SAMPLE 15 SEC. MAGNETIC FIELD DATA	
INTERNATIONAL REFERENCE IONOSPHERE MODEL		
MSIS-86 NEUTRAL ATMOSPHERE MODEL		

WDC-B2 for Solar Terrestrial Physics (STP) in Moscow is responsible for the archiving and dissemination of ionospheric, geomagnetic, solar wind, and cosmic ray data, and is the principal archive in the USSR for satellite data. For various reasons, its holdings of scientific data from USSR satellites are more limited than the corresponding NSSDC holdings of US spacecraft data.

Data Exchange Between NSSDC/WDC-A-R&S and WDC-B2-STP

There has developed over the past three years or so an exchange of data between NSSDC and WDC-B2-STP. Table 1 shows the principal data sets provided by WDC-B2-STP to NSSDC, and Table 2 shows the data sets which have flowed in the other direction. Note that the majority of these data sets are in the area of solar-terrestrial physics, and in fact solar wind field and plasma data make up the largest single component. This is not unexpected, given that solar wind data are relatively simple to understand and are needed for a wide range of studies of the ways the Earth's magnetosphere responds to solar wind variations.

Over the years, NSSDC has created, maintained, and updated the multispacecraft, 24-year set of hourly solar wind field and plasma parameters. This data set, on both magnetic tape and in the form of paper data books, has been widely disseminated in the USSR and internationally. It is particularly significant that within the past year it has been possible to fill some of the gaps in coverage provided by US spacecraft with data from the Soviet Prognoz 10 spacecraft. These Prognoz data have been normalized relative to the US data from IMP-8 to maximize consistency. Additional data from Soviet spacecraft will be used to fill additional gaps in order to create as continuous a record of the solar wind as possible.

Note from Tables 1 and 2 that not only are individual spacecraft data sets being exchanged, but also geophysical models. It will be important in the future as the level of international cooperation grows in the coordinated acquisition and utilization of data from many nations' spacecraft, that common models be used for determining magnetic conjugacies. This is but one of many advantages to the international exchange of geophysical models.

Other US – Soviet Data Exchange

US scientific data has been provided to individual Soviet scientists over the years. Typically such data has been provided in limited amounts, and has come from missions such as ATS, OGO, IMP, and ISEE through NSSDC. WDC-B2-STP recently requested permission from NSSDC/WDC-A-R&S that these data, already in the Soviet Union, be transferred to WDC-B2-STP for archiving there an further dissemination to the Soviet science community. The requested permission was granted.

This author does not know of all cases wherein Soviet scientific data may have been sent by individual Soviet scientists to members of the US science community. An example of Soviet data flowing into the NASA community has been ground magnetometer data being contributed to the NSSDC-sponsored Coordinated Data Analysis Workshops.

In 1987, a high level US – USSR agreement established Joint Working Groups in each of five discipline areas (astrophysics, planetary science, space physics, Earth science, and life science) to discuss areas of mutual benefit. Data exchange was a prominent element of the initial round of discussions. The following data flows have been stimulated by these agreements and may be just the beginning: Viking imagery and IMP-8 solar wind data from NSSDC to support Phobos data; Prognoz 7 and Vega plasma data from the Institute for Space Research in Moscow to NSSDC.

It should also be mentioned that major multinational thrusts, now in their early phases, are likely to yield significant data flows. These thrusts include the IGBP (International Geosphere Biosphere Program), STEP (Solar Terrestrial Energy Program), and the activities of the Interagency Consultative Group (IACG) in the solar terrestrial domain. The WDC's may play a yet to be determined role in the data exchange which will be vital to the success of these programs. (Associated information exchange is addressed in a companion article in these proceedings).

DATA FINDING IN A DISTRIBUTED INTERNATIONAL ENVIRONMENT

Joseph H. King

World Data Center-A for Rockets and Satellites
National Space Science Data Center
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

Abstract

This paper identifies the need for data directories, describes the NASA Master Directory, describes the multi-agency "Catalog Interoperability" effort being pursued in the United States, and ends with a call for increased international coordination in working towards the easy flow of information about available data sets which are distributed world wide.

Introduction

In order to maximize the potential scientific results from planned multinational scientific programs such as the International Geosphere Biosphere Program and the Solar Terrestrial Energy Program, and indeed to fully exploit all past, present and future data, it is vital that scientists with research needs be able to readily find relevant and available data sets, and to ascertain various characteristics of these data sets, including their locations, access procedures, etc.

Information about sensor data and data derived therefrom, or "metadata", may be thought of as being distributed across layers of a distributed metadata environment. The top layer of this environment characteristically captures very limited amounts of information about a very wide range of data sets; such a layer will be referred to herein as the directory layer. The bottom layer typically captures information about the individual granules, or pieces, of a data set (i.e. individual files, images, etc.); this is called the inventory layer. Frequently, inventories are specific to individual data sets.

There may be intermediate layers of the metadata environment, sometime specific to a given discipline or project, describing entire data sets much more fully than directories do.

The more knowledgable a scientist is about the data set he wants to work with, the lower in the metadata environment he will start a query. On the other hand, it is expected that many data sets exist worldwide, having various levels of accessibility, which researchers may not be aware of but which may be relevant to their studies. Thus the directory level functionality is viewed as very important in today's increasingly international and interdisciplinary environment. That only high level data set information is captured, and virtually all data sets require that same kinds of information for their high level characterization, makes directories good candidates for standardization.

The NASA Master Directory

The National Space Science Data Center (NSSDC) is developing a Master Directory intended to describe at a high level all data (whether held at NSSDC or elsewhere, whether originating from NASA missions or others, whether acquired on the ground or in space, etc.) relevant to NASA researchers in Earth science, space physics, planetary science, and astrophysics. The first "build" of this Master Directory (MD1) was released to the scientific community in October, 1988. MD1 was developed under the general guidance of a multidisciplinary group of NASA scientists mainly from the university community. It is presently implemented as a central database, with menubased query software, running on a network-accessible NSSDC VAX computer and associated backend Britton Lee database machine.

MD1 describes whole data sets (and even aggregates of data sets as defined at individual data facilities). For each data set described, values of relevant discipline(s) and of parameter(s) contained in the data are specified from controlled word lists. The data source (e.g., spacecraft), instrument, and investigator are named. Temporal and, when relevant, spatial coverages are given. Data sets associated with specific campaigns or projects may be so flagged. The place where the data are held, and a person at that place, are named. Words not on

controlled lists, but judged by the person supplying the information to be useful, may be entered as "general" keywords. All the foregoing information is managed as a relational database. Finally, descriptive text provides the querying scientist with some of the most important characteristics of the data set.

Information about a given data set is typically provided by a person at the site where the data are held. The vehicle for information transfer is the Directory Interchange Format (DIF), which is further described below.

In addition to information about data sets, MD1 also provides information about archives or other data facilities, about sensors and sources, and about campaigns and projects.

Querying of MD1 for data sets is by specification of any word(s)/phrase(s) which occurs to the querying scientist or by specification of values for any desired combination of the information fields available. In the first case, several fields of MD1 are searched for the value(s) specified. The fact that there is no control over these words means relevant data sets may not be found because the word(s) used by the queriers may be different than the words used by the person initially keywording the data set. In the latter case, the user first specifies for which fields he wants to specify value(s) or ranges (in the case of time/space coverages); then he specifies the desired values for each field of interest, selecting the possible values from controlled lists. The development of the controlled word sets is not easy. It is desired that each set be complete and orthogonal. It is expected that the present word lists, developed with much science community involvement, may change.

After the data set select criteria are specified, the database is searched and a list of data sets satisfying the criteria is returned. The user is then invited to specify which of the data sets he wants to read more information about. For a growing number of data sets, it is possible for the user to select an option by which he is transferred electronically to the site holding the data or more detailed inventory.

At the time of this writing, there were 180 distinct data sets described in MD1, and that number was growing. An evaluation period is now in progress. During the spring of 1989, user critiques will be assessed, and a follow on version of the NASA Master Directory will be developed and made available. Among the activities being pursued during the evaluation period is the assessment of alternative directory architectures. For instance, does it make sense to distribute the directory on floppy disk or CD-ROM with query software?

Catalog Interoperability

It is recognized that various organizations and nations will have overlapping but not identical discipline spans of interest, and hence will have need for their own directories. However, because persons associated with one organization need to know about some other organizations' data sets, it is important to have the ability to exchange information between directories of various organizations. Further, because information will frequently need to pass up to a directory from lower layers of the metadata environment, again a need for the convenient transfer of information bundles is realized.

To address this issue, there has been an ongoing effort in the US to promote "catalog interoperability." A group of data systems experts associated with various NASA discipline specific data systems, as well as persons representing the metadata systems of the National Oceanic and Atmospheric Administration and the United States Geological Survey have been working together to define the previously mentioned Directory Interchange Format (DIF). Thus, although these three agencies which share a common interest in Earth science data from various perspectives do not have identically structured top level directories, they are able to import and export DIF's and to make the needed format changes between the DIF and their databases.

The DIF is basically an ASCII file with a keyword-value structure following specified syntax rules. More information about DIFs is found in the DIF manual available from the author of this paper.

International Catalog Interoperability

It is important to facilitate the flow of information across national boundaries about data which can be made available across such boundaries. For years, paper catalogs of the World Data Center sites and of other data facilities have been the primary vehicle for such information flow. Increasingly it is becoming possible to electronically communicate across national boundaries, as is evident from the networking oriented papers in this Proceedings.

There is a direct analog to the multi-agency catalog interoperability effort described above in the international arena. It is important to define now what information about scientific data and their sources ought to be readily accessible to the international scientific community. It is important to define the commonalities in requirements and functionalities in the actual and intended top level directories of

various nations and multi-nation groups (e.g., ESA). It is important to define appropriate standards, technologies, user interfaces, etc., to be used in managing directory contents and transferring information across directories, and in providing individual scientists and other users access to these directories and the information therein. Perhaps the recently developed DIF can serve as a starting point for consideration.

Note that it is useful for members of the scientific community to be aware of data already existing, but also of data that will become available in the future.Therefore it is important that as plans for future scientific data gathering missions firm up, information about those missions (e.g., scientific spacecraft and their instruments) should be made accessible via these high level directories.

There is no group presently in place to guide the evolution of international "catalog interoperability," but it is highly desirable that such a group should be established under appropriate sponsorship.

Acknowledgement

I wish to acknowledge the role of Dr. James Thieman of NASA/NSSDC in leading both the NASA Master Directory and Catalog Interoperability development efforts.

COOPERATION BETWEEN DATA CENTRES: INTEROPERABILITY

Michael Jones

European Space Operations Centre,
Robert-Bosch Str, 5,
6100 Darmstadt,
W. Germany

Introduction

This paper examines the concept of 'interoperability' in the context of cooperation between data centres, illustrating the subject with a number of real examples drawn from experience within the European Space Agency (ESA).

The paper first defines interoperability (Chapter 2) and then in Chaper 3 goes on to discuss the notion of layers of interoperability. Chapters 4 examines three practical examples of interoperability and the paper concludes with a discussion of lessons learned from them.

Interoperability: Meaning and Scope

Interoperability is characterised by:

1. The capability to build end-to-end systems from elements or subsystems provided from different sources (e.g. Agencies)

2. The capability to replace one system/element in such a system by another from a different source, the substituted element providing the same services and products or a known subset thereof as that of the element which it replaced

Interoperability can reduce the total effort and cost involved in developing and operating an end-to-end system in a multi-agency

environment. However achievement of interoperability requires 'up-front' investment of effort for:

- agreement and common understanding of products and services provided by elements of a given type
- agreement and definition of interfaces between elements
- test/checkout of those interfaces; this involves checking that the elements work together correctly and in effect that each element provides its agreed services and products in the correct formats.

It is obvious therefore that interoperability is desirable from a point of view of cost effectiveness and in certain cases may be essential since without 'interoperability' certain projects may be financially or technically unfeasible.

The meaning and usefulness of interoperability is thus clear. The principle problem is how to achieve it. By means of looking at certain appropriate examples an attempt is made to derive some of the features of successful interoperability.

Levels of Interoperability

Interoperability can be considered to be in levels of layers which build on one another. These layers can be associated with services (and corresponding interfaces). This concept, which is similar in approach (and related) to the OSI 7-layer communications model is illustrated in Fig. 1.

Broadly speaking, one distinguishes between

- communications interoperability
- applications interoperability

Communications interoperability can be conveniently layered according to the OSI model, as indicated in the inset of Fig. 1. An important point is that communications interoperability can have an enormous range varying from:

- interoperability at layer 1
- full end-t-end interoperability involving all 6 levels and possibly part of the application layer

However in the more restricted cases (like (i)), the services of the intervening layers must either be dispensed with or emulated in the application. If the emulated services are not available on the other side of the interface, the consequence could be a significant loss of reliability, particularly if end-to-end services for error handling and routing, for example, are missing. On the other hand, keeping the interoperability interface at a lower level (e.g. 3 or lower) has a number of distinct practical advantages, for example:

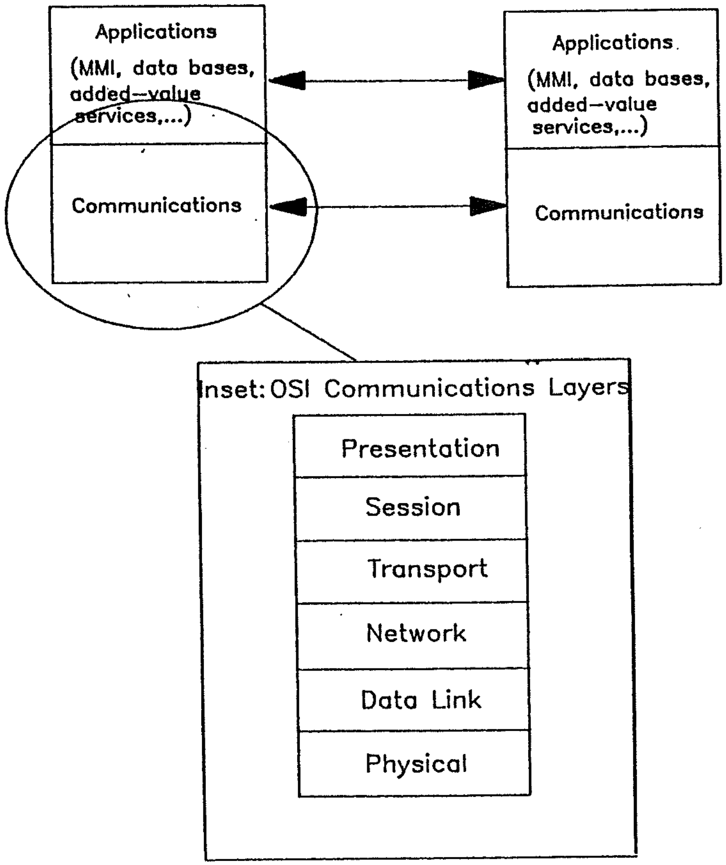


Figure 1 - Layers of Interoperability

- less complexity, i.e. easier to find common or compatible software/hardware from vendors
- more efficient, i.e. performance better, power requirements lower

At the applications level, various types of interoperability may be considered. These do not necessarily strictly correspond to a layering, but are dependent on the application. Examples are:

- man-machine interfaces; thus similar or identical MMIs may be provided for a given application at different centres
- software interoperability, allowing 'porting' of applications software (e.g. source code) to a number of different systems (e.g. different computer hardware or configurations)

Examples of Interoperability

Introduction

A surprising number of examples of interoperability can be found, involving cooperation between Agencies, data centres, institutes or independent units within a single Agency. Three examples were selected for discussion in this paper, namely:

- Giotto Operations: this is an specific mission from which a number of examples of interoperability can be drawn.
- Ulysses (operations of an ESA satellite using NASA facilities)
- Space Data Communications Network (SPAN), which gives an excellent example of communications interoperability.

Example 1: Giotto Operations

The ESA spacecraft Giotto was launched in July 1985. After a 'cruise phase', a sucessful flyby of the comet Halley took place on the night of 13-14th March 1986. Ref. 2 gives a good introduction to the Giotto Mission Operations systems, describing in some detail the complex ground system set up to support the mission. The discussion here will be restricted to highlighting the various areas of interoperability within this mission, principally:

(1) TT&C and X-band acquisition support, both for backup and during the encounter phase to maximise data recovery as well as providing a sort of 'hot-standby' security.

(2) Pathfinder, involving interoperability between Giotto and the USSR spacecraft Vega 1 and Vega2

The types of interoperability involved in the two cases are quite different, the former involving real-time operational usage of equipment from another Agency, the latter being more in the nature of a cooperation, with extensive data exchange.

TT&C and Science Telemetry Acquisition

During the Giotto Mission, the NASA Deep-Space Network (DSN) was used to provide

- Backup TT&C (Telemetry, Tracking and Command) support during the first days of the mission.
- additional tracking
- high-speed telemetry backup at encounter: parallel 24-hour support was provided so that maximum data coverage was ensured
- Telecommand capability for spacecraft emergency

DSN stations at Canberra, Madrid and Goldstone were used. It sould also be noted that the backup so provided was based on 64m. antennas and had in some respects a greater capability (performance) than the prime TT&C facility at Carnarvon which had a 15m. antenna. This meant that the backup could handle contingencies in which the telemtry signal level from the spacecraft was too weak to be acquired by the 15m antenna at Carnarvon.

(2) During the Cruise Phase use of the German National Facility at Weilheim to backup the prime TT&C station at Carnarvon, W. Australia, Weilheim was modified by ESA to provide compatible ranging support and reed-Solomon decoders forelemtry reception becoming in effect part of the ESA S-band Network for the duration of the Giotto Mission.

The operation with NASA DSN deserves further study here, since it contains some practical examples of the interoperability defined in (2). Turning first to telemtry, it is noted that

- it was received as a bit stream over NASA's NASCOM network
- ESA provided equipment to transform this stream into formats compatible with ESA facilities i.e. spacecraft formats with Reed-solomon decoding performed
- similarly data tapes provided by DSN stations for off-line processing were bit-stream structured, which had to be taken into account in the software which provided final experimenter tapes.

Although considerable effort had to be put into the checkout of this interface, this aspect of operation with NASA worked rather well, possibly because the interface on the NASA side was a rather simple one, so the protocal conversions were relatively straightforward. One problematic area was the organisational one of scheduling of the usage of DSN, particularly during the test phases where the absence of direct ESA participation on the DSN scheduling meetings (for geographical and logistical reasons) made the process somewhat cumbersome.

Regarding telecommanding, setting up the facilities for backup commanding involved finding a solution to the problem that the ESA and NASA commanding systems are based on radically different philosophies, the former being far more orientated to automation, in particular as concerns the spacecraft controller's man-machine interface. Broadly two choices were open to ESA in using NASA facilities for backup commanding:

(i) install a remote terminal to the NASA Mission Command and Control Center (MCCC)

(ii) integrate the interface into the ESA commanding system so that commanding via NASA 'looks like' (as far as possible) manual commanding on the ESA system. The spacecraft controller would then be offered the usual standard ESA facilities of automatic pre-transmission command validation, command execution verification, and command logging via the standard man/machine interface, although performance (e.g. command rate) might be different.

Solution (ii), which would have involved developing NASCOM interfacing software 'behind' the command MMI software on the ESA spacecraft control system was rejected on grounds of cost. Solution (i) was therefore adopted, but turned out not to be totally without problems, the first of which was that the MCCC terminals of the type used for commanding were no longer in production. The solution to this was to implement an emulation of the MCCC terminal on an IBM PC, which was then connected to GSFC via a 9.6. kbaud link to GSFC via INTELSAT. The resulting man machine interface was workable but uncomfortable since the command files were set up on the terminal and then had to be manually subjected to a series of operations (passage to ground station, configuration of ground station for uplink etc.) Verification if needed would have involved feeding the commands into the ESA control system as dummy commands (i.e. with uplink suppressed) so that verification against Giotto telemetry received on the ESA Control System could be done. In practice this was not necessary since it was mainly used for commands not requiring verification (ranging commands, on-off sequences, commands without verification)

The conclusion on commanding is that

- a backup was successfully implemented
- in practice it only had to be used a limited number of times
- from an operational viewpoint it was unsatisfactory because (a) it looked radically different from the usual ESA system (b) it was heavily manually biased (c) it had an exceedingly low throughput (c. one command every 2 min.). (d) it would create increased mission risk when used under contingency conditions, since un-

tried sequences and files had to be created. It would therefore have been unacceptable for heavy operational use

- the successful implementation of this backup commanding was in part made possible by the familiarity with the NASA commanding system gained as a consequence of the Ulysses Project; this shows the obvious but important fact that one attempt at achieving interoperability may make further attempts easier to accomplish

In fact ESA has learnt from this experience: for the Shuttle-launched European Retrievable Carrier (EURECA), which will use NASCOM interfaces in the deployment and retrieval phases of the mission, approach (ii) has been adopted.

Pathfinder Activities

Pathfinder involved the reevaluation of comet Halley's trajectory based upon cometary ephemeris data from the encounters of the USSR spacecraft Vega1 and Vega2 with the Comet 3 and 5 days respectively prior to the encounter. This involved three Agencies--NASA, ESA and Intercosmos:

- NASA performed very precise tracking of the Vega spacecraft in order to achieve the accuracy of the orbit determinations for the Vega spacecraft necessary for improvement in cometary orbit determination
- ESA and IKI used Vega data for the refinement of the comet's orbit. This involved establishing a communications 'hot-line' between ESOC in Darmstadt and the Institute for Space Research (IKI) in Moscow. This supported not only exchange of data but also remote access to ESA computers from Moscow, interactive dialogue (via terminals) and voice communications between the teams concerned

It should be noted that these activities were the culmination of a international campaign of analysis of astrometric observations of the comet involving twelve observatories and associated observers (see for example ref. 3).

Pathfinder is an example of what a well prepared and coordinated cooperation activity can achieve supported in the case of the comet orbit refinement by relatively simple digital data communication facilities.

Background

Ulysses is a joint ESA-NASA mission. It was planned for launch in May 1986 using the shuttle and a Centaur upper stage. All system integration and chechout was performed before the Launch was delayed by the Challenger failure. Launch is now planned for October 1990 using the shuttle and Boeing IUS upper stage. In their meantime the system is being converted to run on more modern hardware (DEC/VAX instead of MODCOMP Classic). The system structure and interfaces remain essentially the same, so the system as originally implemented is described here.

For the ULYSSES mission, ESA provides the spacecraft while NASA provides all launch facilities plus use of necessary ground stations, communications networks and the mission control centre infrastructure at JPL Pasadena. The ULYSSES-dedicated control and monitoring system is provided by ESA and integrated into the existing NASA facilities at JPL. It is this integration of an ESA-provided control and monitor system (H/W, S/W and staff) into a NASA infrastructure which is the subject of this section.

Split of Operations

Operations were split between NASA and ESA as follows:

NASA (JPL)

- Telemetry data Acquisition and frame synchronisation
- Long term Telemetry data Filing and Processing
- Command transmission
- Ranging, Doppler processing and Navigation

ESA (ESOC)

- Command Schedule Generation
- Command event checking
- Real time TLM processing (limit checks, calibration, display)
- Near Real time TLM processing (performance analysis etc.)
- Science Printouts
- Flight Dynamics

Interfaces

Human interfaces are not considered here. The prime data interfaces are":

- Telemetry Frames (real-time or recorded and replayed from spacecraft or from ground)
- Command Schedule
- Navigation/Flight Dynamics data.

Constraints

To minimise cost, the existing ESA Spacecraft control concept (as implemented in the Multi-Satellite Spacecraft Control System (MSSS) at ESOC) was used as far as possible.

The Data acquisition system provided by LJP was a version of their 'standard' Data Acquisition and Control system (DACS) (Based on VAX hardware).

The existing JPL commanding system was to be used, with minor extensions to accommodate ULYSSES specific features such as block commanding. This system was based on a standard format magnetic tape input to the command system (of the MSSS real time command, transmission and verification concept).

To minimise cost (particularly for hardware maintenance during the 5 year mission life) the ESA-supplied spacecraft control system was to be based on hardware already used in JPL, for which there was on-site support and an existing machine which could be used as backup (i.e. MODCOMP hardware).

Approach

The split of functional responsibilities between NASA and ESA is shown in Fig. 2. The main interfaces between NASA and ESA are as follows:

The *TLM Interface* between DACS (VAX) and RTTP (MODCOMP) was agreed as RS422 at the electrical level, x25 level 2 at the communications protocol level, and an application-specific protocol based on SFDUS (Standard Format Data Units) at the transfer level. This interface was fully defined down to the bit level in a mutually agreed and approved Interface Control Document (ICD).

The *CMD Interface* is based on the existing magnetic-tape interface to the JPL multi-mission command system. The necessary extensions to accomodate ULYSSES-specific command structures did not affect the existing tape/record structure, only the record contents. A mutually-approved ICD was generated for this interface.

For *Navigation/Flight Dynamics Data* a ULYSSES-specific tape format was agreed and controlled by ICD.

Development

The ESA-supplied spacecraft control and monitoring system (the RTTP) was developed at ESOC (Germany) and shipped to JPL for intergration into the NASA infrastructure some 6 months before the scheduled launch date. Some 4 months prior to this (i.e. 10 months before launch) the RTTP software system was taken to JPL on tape and installed on the (JPL provided) backup RTTP machine to enable exhaustive testing of the TLM interface between DACS and RTTP. The procedures for these tests being fully defined, documented and approved between JPL and ESOC during working sessions over the preceding months.

The command and Flight Dynamics tape interfaces were verified, using agreed test procedures, by exchange of mag tapes between ESOC and JPL.

Lessons

Although the Modcomp hardware was initially selected as compatible with NASA machines, during the time span of the project NASA 'standardised' on VAX hardware leaving the ULYSSES system as potentially 'non/standard' hardware with resultant maintenance problems.

The VAX-MODCOMP interface (X25 level 2) used hardware in both machines which was not part of the suppliers' normal range (because no 'standard' interface between the machines was available). This was the source of some integration problems.

The pre-testing of interfaces between ESA and NASA systems proved invaluable and led to a smooth subsequent integration of the full system.

The only significant transfer level interface problem arose because the agreed and signed-off interface control document was transcribed within JPL into a different document for their implementation department. This transcription contained a non-trivial error.

Example 3: Space Data Analysis Network (SPAN)

The Space Physics Analysis Network (SPAN) was originally designed in 1980 by NASA and put into initial operation in 1981. It is a multi-mission communications network to support cooperative space and earth science research and data analysis.

SPAN is a computer to computer network based on DECNET. It connects other DECNET-based networks such as the High Energy

Physics Network (HEPNET). In such cases no gateways as such are necessary. Gateways to ARPANET, BITNET EARN and ESANET have been implemented. Indeed, where possible, ESANET is used to carry SPAN traffic in Europe in order to keep down costs.

The network, consists of

- a US subnet
- a European subnet

The general structure of each subnet is in the form of a "backbone" network connecting main routing centres (or area nodes) with "tail circuits" connecting these routing centres to member institutes. The topology is illustrated in fig. 2, from which it can be seen that the US subnet has five area nodes, and the European subnet has one (viz. ESOC). The US subnet and the European subnet are connected by a fixed 9.6.kbit/s link between ESA's European Space Operations Centre (ESOC at Darmstadt, W. Germany and the Goddard Spaceflight Center (GSFC). In most cases, tail circuits are (in the USA) simple dedicated leased lines and (in Europe) the X.25 Public Packet Switch Network.

To maintain a reliable operational system requires a proper management structure. Overall SPAN management is by the National Space Science Data Center (NSSDC). Network managers are responsible for the day-to-day operations of each subnet (US or European), Routing Managers look after the operation of area nodes and Remote Node managers look after "end nodes" i.e. end points of tail; circuits in member institutes.

A last aspect considered in this overview is that of charging policy for which a common SPAN/HEPNET policy is followed. This is designed to make SPAN/HEPNET attractive to its subscribers while at the same time encouraging a correct usage of the Network. Details of the policy will not be given here, but it suffices to say that subscribers contribute mainly to the costs of transport to their local routing node. Costs of Network interconnection are charged at a relatively low-rate along the preferred interconnection (e.g. ESOC to GSFC) to force traffic along these routes.

A good example of the usefulness of SPAN is given in Ref 2, which describes its use to transmit data from the encounter of the ICE spacecraft with the comet Giacobini-Zinner. This provided support to Investigators involved in a European experiment on-board the ICE spacecraft and some 100 kbits encounter data were transferred from the computer of the NSSDC (National Space Science Data) at GSFC to ESOC via SPAN and from thence via ESANET to ESA's Technology Centre (ESTEC) at Noordwijk, Holland, where the data was processed. The first processed data appeared in the form of plots some 30min, after

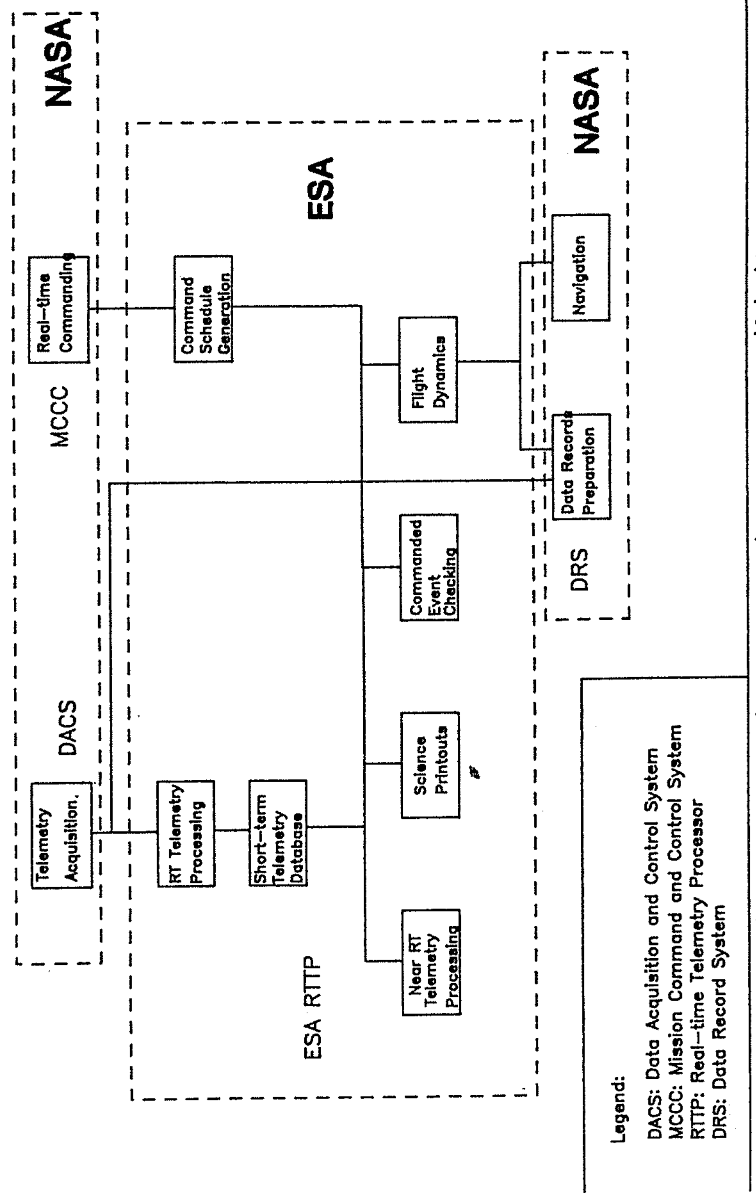


Figure 2 - Split of Functional Responsibility Between ESA & NASA

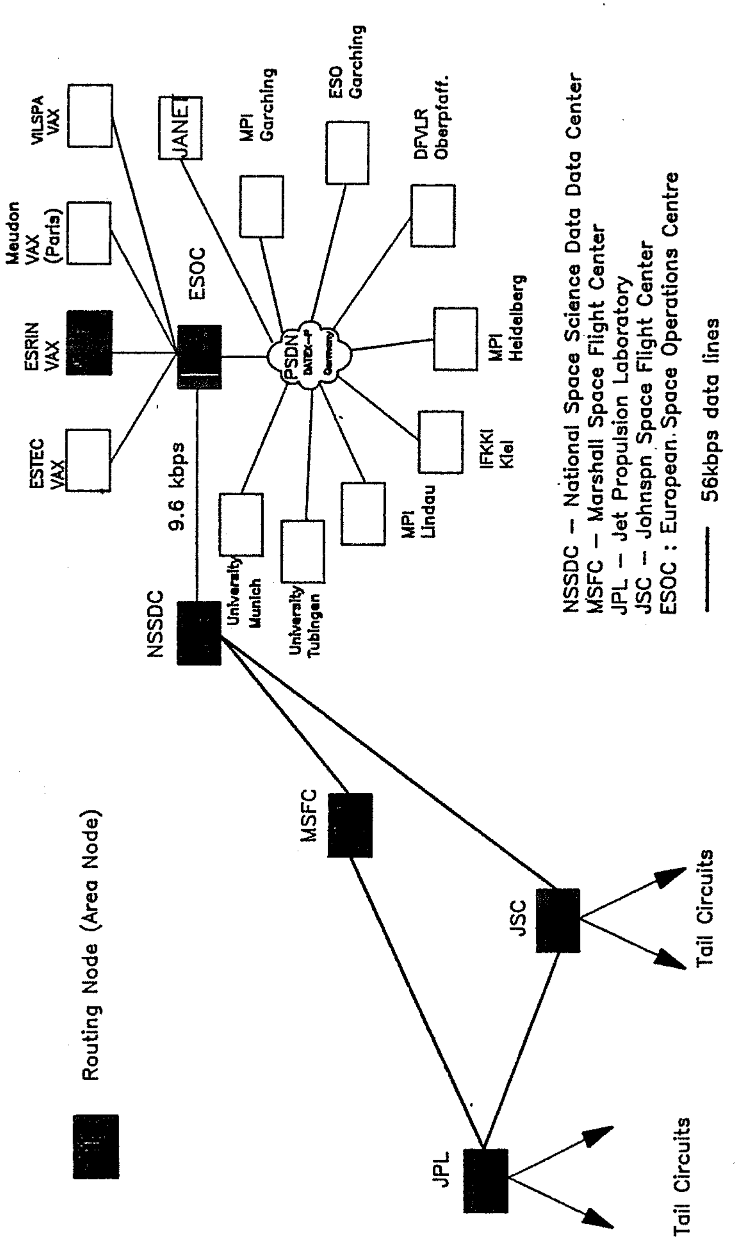


Figure 3 - SPAN Topology

the beginning of the transmission of the data from GSFC. These plots were transmitted back to the NSSDC via facsimile well in advance of the first Press Conference held five hours after the encounter.

In conclusion features which have led to SPAN's success are:

- computer hardware commonality (use of VAX/DECNET); this hardware commonality not only assists data transfer, but because of the availability of common software tools and office automation facilities on the VAX computers, related activities, such as joint preparation of papers by authors at different sites, are made easier.
- need for a proper management structure; this structure, covers overall Network policy on the one hand and day-to-day Network operations on the other, the latter involving a hierarchy of Network, Routing and Remote Node Managers, so that problems are dealt with at the appropriate level
- use of a charging policy which is attractive to users, but also promotes correct usage of the Network

Discussion and Conclusions

The paper has discussed a number of areas in which the degree of difficulty in achieving interoperability varies greatly. The following major conclusions may be drawn:

1. The idea of layers of interoperability is an important one. From the few examples given in this paper one can pick out cases, thus

Communications Interoperability:

- Ciotto DSN telemetry interface (NASCOM) – this is an interface at a low OSI layer (layer 1), which can be relatively straightforwardly implemented, but has drawbacks in services offered (and therefore in reliability).
- SPAN, which offers a comprehensive set of services via DECNET. It depends on VAX/DECNET commonality, and works well in a scientific environment. However its level of security is not sufficiently high which requires the SPAN environment to be isolated from any critical operational environment (e.g. for spacecraft operations).

Applications Interoperability: here Giotto backup commanding is a good example. In the solution taken, the spacecraft operations staff at the Control Centre (ESOC) had to deal with two different Man-Machine Interfaces, viz. that on the standard ESA system and the remote command interface to the NASA system. Although this was useable, an emulation of the ESA MMI on the NASA command interface would have been both safer and more convenient operationally. Better still would be a common approach to commanding services by ESA and NASA. Work on standardisation of commanding services is being done within the Consultative Committee for Space Data Standards (CCSDS), so in the long term this may well be achieved.

2. The layers mentioned in (1) are normally associated with standards (e.g. protocol specifications). Clearly standards are of prime importance for making interoperability practicable, and the layer at which interoperability takes place is chosen to be the highest at which standards are supported by the Agencies/Centres concerned.

3. Organisational measures are usually essential to achieving interoperability, e.g. in the choice of layer at which interoperability takes place, choice of 'dialect' of standard used, coordination on hardware/software commonality, formal agreements on cross support, etc.

A number of subsidiary conclusions can also be drawn thus:

- hardware, and basic software compatibility between different facilities can ease interoperability. However proper coordination is required, for example when equipment or software is upgraded. Careful configuration management of all the interoperating elements, at least as concerns their interfacing and common services is required.
- successful interoperability requires exhaustive verification of interfaces well in advance of that of the rest of the system. Again this is an area where good coordination, backed up by suitable agreements plays an important role.

References

1. Wilkins, D.E.B., 'The Giotto Mission Operations System', ESA Journal, 10 (3), 1986

2. Sanderson, T.R. et al., 'Near Real-Time Data Transmission During the ICE-Comet Giacobini-Zinner Encounter', ESA Bulletin, 45, pp. 21-23, 1986.

3. Morley, T.A., 'Astrometric Observations of Comet Halley', ESA Bulletin, 43, pp,. 19-23, 1985.

UAG SERIES OF REPORTS

Fewer than four UAG Reports are published at irregular intervals each year. Copies of these publications may be purchased through the NATIONAL GEOPHYSICAL DATA CENTER, Solar-Terrestrial Physics Division (E/GC2) 325 Broadway, Boulder, Colorado 80303, USA. Please note that some reports are available on microfiche only at $4.00 a copy and that a $10.00 handling charge will be added. In this case orders must include check or money order payable in U.S. currency to Commerce, NOAA/NGDC.

UAG- 1 IQSY NIGHT AIRGLOW DATA, by L.L. Smith, F.E. Roach, and J.M. McKennan, ESSA Aeronomy Laboratory, Boulder, CO, July 1968, 305 pp.

UAG- 2 A REEVALUATION OF SOLAR FLARES, 1964-1966, by Helen W. Dodson and E. Ruth Hedeman, McMath-Hulbert Observatory, University of Michigan, Pontaic, MI, August 1968, 28 pp.

UAG- 3 OBSERVATIONS OF JUPITER'S SPORADIC RADIO EMISSION IN THE RANGE 7.6-41 MHZ, 6 JULY 1966 THROUGH 8 SEPTEMBER 1968, by James W. Warwick and George A. Dulk, University of Colorado, Boulder, CO, October 1968, 35 pp.

UAG- 4 ABBREVIATED CALENDAR RECORD 1966-1967, by J. Virginia Lincoln, Hope I. Leighton and Dorothy K. Kropp, ESSA [now NOAA], Aeronomy and Space Data Center, Boulder, CO, January 1969, 170 pp.

UAG- 5 DATA ON SOLAR EVENT OF MAY 23, 1967, AND ITS GEOPHYSICAL EFFECTS, compiled by J. Virginia Lincoln, World Data Center A, Upper Atmosphere Geophysics, ESSA [now NOAA], Boulder, CO, February 1969, 120 pp.

UAG- 6 INTERNATIONAL GEOPHYSICAL CALENDARS 1957-1969, by A.H. Shapley and J. Virginia Lincoln, ESSA Research Laboratories [now NOAA], Boulder, CO, March 1969, 25 pp.

UAG- 7 OBSERVATIONS OF THE SOLAR ELECTRON CORONA: FEBRUARY 1964 - JANUARY 1968, by Richard T. Hansen, High Altitude Observatory, NCAR, Boulder, CO, and Kamuela, HI, October 1969, 12 pp.

UAG- 8 DATA ON SOLAR-GEOPHYSICAL ACTIVITY OCTOBER 24 - NOVEMBER 6, 1968, Parts 1 and 2, compiled by J. Virginia Lincoln, World Data Center A, Upper Atmosphere Geophysics, ESSA [now NOAA], Boulder, CO, March 1970, 312 pp, (includes Parts 1 and 2).

UAG- 9 DATA ON COSMIC RAY EVENT OF NOVEMBER 18, 1968, AND ASSOCIATED PHENOMENA, compiled by J. Virginia Lincoln, World Data Center A, Upper Atmosphere Geophysics, ESSA [now NOAA], Boulder, CO, April 1970, 109 pp.

UAG-10 ATLAS OF IONOGRAMS, edited by A.H. Shapley, ESSA Research Laboratories [now NOAA], Boulder, CO, May 1970, 243 pp.

UAG-11 [Superseded by UAG-30]

UAG-12 SOLAR-GEOPHYSICAL ACTIVITY ASSOCIATED WITH THE MAJOR GEOMAGNETIC STORM OF MARCH 8, 1970, Parts 1, 2 and 3, compiled by J. Virginia Lincoln and Dale B. Bucknam, World Data Center A, Upper Atmosphere Geophysics, ESSA [now NOAA], Boulder, CO, April 1971, 466 pp, (includes 3 parts).

UAG-13 DATA ON THE SOLAR PROTON EVENT OF NOVEMBER 2, 1969, THROUGH THE GEOMAGNETIC STORM OF NOVEMBER 8-10, 1969, compiled by Dale B. Bucknam and J. Virginia Lincoln, World Data Center A, Upper Atmosphere Geophysics, ESSA [now NOAA], Boulder, CO, May 1971, 76 pp.

UAG-14 AN EXPERIMENTAL, COMPREHENSIVE FLARE INDEX AND ITS DERIVATION FOR 'MAJOR' FLARES, 1955-1969, by Helen W. Dodson and E. Ruth Hedeman, McMath-Hulbert Observatory, University of Michigan, Pontiac, MI, July 1971, 25 pp.

UAG-15 [Superseded by UAG-30]

UAG-16 TEMPORAL DEVELOPMENT OF THE GEOPHYSICAL DISTRIBUTION OF AURORAL ABSORPTION FOR 30 SUBSTORM EVENTS IN EACH OF IQSY (1964-65) AND IASY (1960), by F.T. Berkey, University of Alaska, Fairbanks, AK; V.M. Driatskiy, Arctic and Antarctic Research Institute, Leningrad, USSR; K. Henriksen, Auroral Observatory, Tromso, Norway; D.H. Jelly, Communications Research Center, Ottawa, Canada; T.I. Shchuka, Arctic and Antarctic Research Institute, Leningrad, USSR; A. Theander, Kiruna Geophysical Observatory, Kiruna, Sweden; and J. Yliniemi, University of Oulu, Oulu, Finland, September 1971, 131 pp, $4.00 (microfiche only).

UAG-17 IONOSPHERIC DRIFT VELOCITY MEASUREMENTS AT JICAMARCA, PERU (JULY 1967 - MARCH 1970), by Ben B. Balsley, NOAA Aeronomy Laboratory, Boulder, CO, and Ronald F. Woodman, Jicamarca Radar Observatory, Instituto Geofisico del Peru, Lima, Peru, October 1971, 45 pp, $4.00 (microfiche only).

UAG-18 A STUDY OF POLAR CAP AND AURORAL ZONE MAGNETIC VARIATIONS, by K. Kawasaki and S.-I. Akasofu, University of Alaska, Fairbanks, AK, June 1972, 21 pp.

UAG-19 REEVALUATION OF SOLAR FLARES 1967, by Helen W. Dodson and E. Ruth Hedeman, McMath-Hulbert Observatory, University of Michigan, Pontiac, MI, and Marta Rovira de Miceli, San Miguel Observatory, Argentina, June 1972, 15 pp.

UAG-20 [Superseded by UAG-30]

UAG-21 PRELIMINARY COMPILATION OF DATA FOR RETROSPECTIVE WORLD INTERVAL JULY 26 - AUGUST 14, 1972, by J. Virginia Lincoln and Hope I. Leighton, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, November 1972, 128 pp.

UAG-22 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES (AE) FOR 1970, by Joe Haskell Allen, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, November 1972, 146 pp.

UAG-23 U.R.S.I. HANDBOOK OF IONOGRAM INTERPRETATION AND REDUCTION, Second Edition, November 1972, edited by W.R. Piggott, Radio and Space Research Station, Slough, UK, and K. Rawer, Arbeitsgruppe fur Physikalische Weltraumforschung, Freiburg, GFR, November 1972, 324 pp.

UAG-23A U.R.S.I. HANDBOOK OF IONOGRAM INTERPRETATION AND REDUCTION, Second Edition, Revision of Chapters 1-4, edited by W.R. Piggott, Radio and Space Research Station, Slough, UK, and K. Rawer, Arbeitsgruppe fur Physikalische Weltraumforschung, Freiburg, GFR, November 1972, 135 pp.

UAG-24 DATA ON SOLAR-GEOPHYSICAL ACTIVITY ASSOCIATED WITH THE MAJOR GROUND LEVEL COSMIC RAY EVENTS OF 24 JANUARY AND 1 SEPTEMBER 1971, Parts 1 and 2, compiled by Helen E. Coffey and J. Virginia Lincoln, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, December 1972, 462 pp, (includes Parts 1 and 2).

UAG-25 OBSERVATIONS OF JUPITER'S SPORADIC RADIO EMISSION IN THE RANGE 7.6-41 MHZ, 9 SEPTEMBER 1968 THROUGH 9 DECEMBER 1971, by James W. Warwick, George A. Dulk and David G. Swann, University of Colorado, Boulder, CO, February 1973, 35 pp.

UAG-26 DATA COMPILATION FOR THE MAGNETOSPHERICALLY QUIET PERIODS FEBRUARY 19-23 AND NOVEMBER 29 - DECEMBER 3, 1970, compiled by Helen E. Coffey and J. Virginia Lincoln, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, May 1973, 129 pp.

UAG-27 HIGH SPEED STREAMS IN THE SOLAR WIND, by D.S. Intriligator, University of Southern California, Los Angeles, CA, June 1973, 16 pp.

UAG-28 COLLECTED DATA REPORTS ON AUGUST 1972 SOLAR-TERRESTRIAL EVENTS, Parts 1, 2 and 3, edited by Helen E. Coffey, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, July 1973, 932 pp.

UAG-29 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1968, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, October 1973, 148 pp.

UAG-30 CATALOGUE OF DATA ON SOLAR-TERRESTRIAL PHYSICS, prepared by NOAA Environmental Data Service, Boulder, CO, October 1973, 317 pp. Supersedes catalogs UAG-11, 15 and 20.

UAG-31 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1969, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, February 1974, 142 pp.

UAG-32 SYNOPTIC RADIO MAPS OF THE SUN AT 3.3 MM FOR THE YEARS 1967-1969, by Earle B. Mayfield, Kennon P. White III, and Fred I. Shimabukuro, Aerospace Corp., El Segundo, CA, April 1974, 26 pp.

UAG-33 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(10) FOR 1967, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, May 1974, 142 pp.

UAG-34 ABSORPTION DATA FOR THE IGY/IGC AND IQSY, compiled and edited by A.H. Shapley, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO; W.R. Piggott, Appleton Laboratory, Slough, UK; and K. Rawer, Arbeitsgruppe fur Physikalische Weltraumforschung, Freiburg, GFR, June 1974, 381 pp.

UAG-35 [Superseded by UAG-92]

UAG-36 AN ATLAS OF EXTREME ULTRAVIOLET FLASHES OF SOLAR FLARES OBSERVED VIA SUDDEN FREQUENCY DEVIATIONS DURING THE ATM-SKYLAB MISSIONS, by R.F. Donnelly and E.L. Berger, NOAA Space Environment Laboratory; Lt. J.D. Busman, NOAA Commissioned Corps; B. Henson, NASA Marshall Space Flight Center; T.B. Jones, University of Leicester, UK; G.M. Lerfald, NOAA Wave Propagation Laboratory; K. Najita, University of Hawaii; W.M. Retallack, NOAA Space Environment Laboratory and W.J. Wagner, Sacramento Peak Observatory, October 1974, 95 pp.

UAG-37 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(10) FOR 1966, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, December 1974, 142 pp.

UAG-38 MASTER STATION LIST FOR SOLAR-TERRESTRIAL PHYSICS DATA AT WDC-A FOR SOLAR-TERRESTRIAL PHYSICS, by R.W. Buhmann, World Data Center A for Solar-Terrestrial Physics, Boulder, CO; Juan D. Roederer, University of Denver, Denver, CO; and M.A. Shea and D.F. Smart, Air Force Cambridge Research Laboratories, Hanscom AFB, MA, December 1974, 110 pp.

UAG-39 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1971, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, February 1975, 144 pp.

UAG-40 H-ALPHA SYNOPTIC CHARTS OF SOLAR ACTIVITY FOR THE PERIOD OF SKYLAB OBSERVATIONS, MAY 1973 - MARCH 1974, by Patrick S. McIntosh, NOAA Space Environment Laboratory, Boulder, CO, February 1975, 32 pp.

UAG-41 H-ALPHA SYNOPTIC CHARTS OF SOLAR ACTIVITY DURING THE FIRST YEAR OF SOLAR CYCLE 20 OCTOBER 1964 - AUGUST 1965, by Patrick S. McIntosh, NOAA Space Environment Laboratory, Boulder, CO, and Jerome T. Nolte, American Science and Engineering, Inc., Cambridge, MA, March 1975, 25 pp.

UAG-42 OBSERVATIONS OF JUPITER'S SPORADIC RADIO EMISSION IN THE RANGE 7.6-80 MHZ, 10 DECEMBER 1971 THROUGH 21 MARCH 1975, by James W. Warwick, George A. Dulk and Anthony C. Riddle, University of Colorado, Boulder, CO, April 1975, 49 pp.

UAG-43 CATALOG OF OBSERVATION TIMES OF GROUND-BASED SKYLAB-COORDINATED SOLAR OBSERVING PROGRAMS, compiled by Helen E. Coffey, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, May 1975, 159 pp.

UAG-44 SYNOPTIC MAPS OF SOLAR 9.1 CM MICROWAVE EMISSION FROM JUNE 1962 TO AUGUST 1973, by Werner Graf and Ronald N. Bracewell, Stanford University, Stanford, CA, May 1975, 183 pp.

UAG-45 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1972, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, May 1975, 144 pp, $4.00 (microfiche only).

UAG-46 INTERPLANETARY MAGNETIC FIELD DATA 1963-1964, by Joseph H. King, National Space Science Data Center, NASA Goddard Space Flight Center, Greenbelt, MD, June 1975, 382 pp.

UAG-47 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1973, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, June 1975, 144 pp, $4.00 (microfiche only).

UAG-48 [Superseded by UAG-48A]

UAG-48A SYNOPTIC OBSERVATIONS OF THE SOLAR CORONA DURING CARRINGTON ROTATIONS 1580-1596 (11 OCTOBER 1971 - 15 JANUARY 1973), [Re-issue of UAG-48 with quality images], by R.A. Howard, M.J. Koomen, D.J. Michels, R. Tousey, C.R. Detwiler, D.E. Roberts, R.T. Seal, and J.D. Whitney, U.S. Naval Research Laboratory, Washington, DC; and R.T. Hansen and S.F. Hansen, C.J. Garcia and E. Yasukawa, High Altitude Observatory, NCAR, Boulder, CO, February 1976, 200 pp. Supersedes UAG-48.

UAG-49 [Superseded by UAG-92]

UAG-50 HIGH-LATITUDE SUPPLEMENT TO THE URSI HANDBOOK ON IONOGRAM INTERPRETATION AND REDUCTION, edited by W.R. Piggott, British Antarctic Survey, c/o Appleton Laboratory, Slough, UK, October 1975, 294 pp.

UAG-51 SYNOPTIC MAPS OF SOLAR CORONAL HOLE BOUNDARIES DERIVED FROM HE II 304A SPECTROHELIOGRAMS FROM THE MANNED SKYLAB MISSIONS, by J.D. Bohlin and D.M. Rubenstein, U.S. Naval Research Laboratory, Washington, DC, November 1975, 30 pp.

UAG-52 EXPERIMENTAL COMPREHENSIVE SOLAR FLARE INDICES FOR CERTAIN FLARES, 1970-1974, by Helen W. Dodson and E. Ruth Hedeman, McMath-Hulbert Observatory, University of Michigan, Pontiac, MI, November 1975, 27 pp.

UAG-53 DESCRIPTION AND CATALOG OF IONOSPHERIC F-REGION DATA, JICAMARCA RADIO OBSERVATORY (NOVEMBER 1966 - APRIL 1969), by W.L. Clark and T.E. Van Zandt, NOAA Aeronomy Laboratory, Boulder, CO, and J.P. McClure, University of Texas at Dallas, Dallas, TX, April 1976, 10 pp.

UAG-54 [Superseded by UAG-85]

UAG-55 EQUIVALENT IONOSPHERIC CURRENT REPRESENTATIONS BY A NEW METHOD, ILLUSTRATED FOR 8-9 NOVEMBER 1969 MAGNETIC DISTURBANCES, by Y. Kamide, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO; H.W. Kroehl, Data Studies Division, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO; M. Kanamitsu, Advanced Study Program, National Center for Atmospheric Research, Boulder, CO; Joe Haskell Allen, Data Studies Division, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO; and S.-I. Akasofu, Geophysical Institute, University of Alaska, Fairbanks, AK, April 1976, 91 pp, $4.00 (microfiche only).

UAG-56 ISO-INTENSITY CONTOURS OF GROUND MAGNETIC H PERTURBATIONS FOR THE DECEMBER 16-18, 1971, GEOMAGNETIC STORM, Y. Kamide, Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, April 1976, 37 pp.

UAG-57 MANUAL ON IONOSPHERIC ABSORPTION MEASUREMENTS, edited by K. Rawer, Institut fur Physikalische Weltraumforschung, Freiburg, GFR, June 1976, 302 pp.

UAG-58 ATS-6 RADIO BEACON ELECTRON CONTENT MEASUREMENTS AT BOULDER, JULY 1974 - MAY 1975, by R.B. Fritz, NOAA Space Environment Laboratory, Boulder, CO, September 1976, 61 pp.

UAG-59 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11) FOR 1974, by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, December 1976, 144 pp.

UAG-60 GEOMAGNETIC DATA FOR JANUARY 1976 [AE(7) INDICES AND STACKED MAGNETOGRAMS], by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, July 1977, 57 pp.

UAG-61 COLLECTED DATA REPORTS FOR STIP INTERVAL II 20 MARCH - 5 MAY 1976, edited by Helen E. Coffey and John A. McKinnon, World Data Center A for Solar-Terrestrial Physics, Boulder, CO, August 1977, 313 pp.

UAG-62 GEOMAGNETIC DATA FOR FEBRUARY 1976 [AE(7) INDICES AND STACKED MAGNETOGRAMS], by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, September 1977, 55 pp.

UAG-63 GEOMAGNETIC DATA FOR MARCH 1976 [AE(7) INDICES AND STACKED MAGNETOGRAMS], by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, September 1977, 57 pp.

UAG-64 GEOMAGNETIC DATA FOR APRIL 1976 [AE(8) INDICES AND STACKED MAGNETOGRAMS], by Joe Haskell Allen, Carl C. Abston and Leslie D. Morris, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO, February 1978, 55 pp.

UAG-65 THE INFORMATION EXPLOSION AND ITS CONSEQUENCES FOR DATA ACQUISITION, DOCUMENTATION, PROCESSING, by G.K. Hartmann, Max-Planck-Institut fur Aeronomie, Lindau, GFR, May 1978, 36 pp.

UAG-66 SYNOPTIC RADIO MAPS OF THE SUN AT 3.3 MM 1970-1973, by Earle B. Mayfield and Fred I. Shimabukuro, Aerospace Corp., El Segundo, CA, May 1978, 30 pp.

UAG-67 IONOSPHERIC D-REGION PROFILE DATA BASE, A COLLECTION OF COMPUTER-ACCESSIBLE EXPERIMENTAL PROFILES OF THE D AND LOWER E REGIONS, by L.F. McNamara, Ionospheric Prediction Service, Sydney, Australia, August 1978, 30 pp, $4.00 (microfiche only).

UAG-68 A COMPARATIVE STUDY OF METHODS OF ELECTRON DENSITY PROFILE ANALYSIS, by L.F. McNamara, Ionospheric Prediction Service, Sydney, Australia, August 1978, 30 pp, $4.00 (microfiche only).

UAG-69 SELECTED DISTURBED D-REGION ELECTRON DENSITY PROFILES. THEIR RELATION TO THE UNDISTURBED D REGION, by L.F. McNamara, Ionospheric Prediction Service, Sydney, Australia, October 1978, 50 pp, $4.00 (microfice only).

UAG-70 ANNOTATED ATLAS OF H-ALPHA SYNOPTIC CHARTS FOR SOLAR CYCLE 20 (1964-1974) CARRINGTON SOLAR ROTATIONS 1487-1616, by Patrick S. McIntosh, NOAA Space Environment Laboratory, Boulder, CO, February 1979, 327 pp.

UAG-71 MAGNETIC POTENTIAL PLOTS OVER THE NORTHERN HEMISPHERE FOR 26-28 MARCH 1976, A.D. Richmond, NOAA Space Environment Laboratory, Boulder, CO; H.W. Kroehl, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO; M.A. Henning, Lockheed Missiles and Space Co., Aurora, CO; and Y. Kamide, Kyoto Sangyo University, Kyoto, Japan, April 1979, 118 pp.

UAG-72 ENERGY RELEASE IN SOLAR FLARES, PROCEEDINGS OF THE WORKSHOP ON ENERGY RELEASE IN FLARES, 26 FEBRUARY - 1 MARCH 1979, CAMBRIDGE, MASSACHUSETTS, U.S.A., edited by David M. Rust, American Science and Engineering, Inc., Cambridge, MA; and A. Gordon Emslie, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, July 1979, 68 pp, $4.00 (microfiche only).

UAG-73 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(11-12) FOR JANUARY - JUNE 1975, by Joe Haskell Allen, Carl C. Abston, J.E. Salazar and J.A. McKinnon, National Geophysical and Solar-Terrestrial Data Center, NOAA, Boulder, CO, August 1979, 114 pp, $4.00 (microfiche only).

UAG-74 ATS-6 RADIO BEACON ELECTRON CONTENT MEASUREMENTS AT OOTACAMUND, INDIA, OCTOBER - JULY 1976, by S.D. Bouwer, K. Davies, R.F. Donnelly, R.N. Grubb, J.E. Jones and J.H. Taylor, NOAA Space Environment Laboratory, Boulder, CO; and R.G. Rastogi, M.R. Deshpande, H. Chandra and G. Sethia, Physical Research Laboratory, Ahmedabad, India, March 1980, 58 pp.

UAG-75 THE ALASKA IMS MERIDIAN CHAIN: MAGNETIC VARIATIONS FOR 9 MARCH - 27 APRIL 1978, by H.W. Kroehl and G.P. Kosinski, National Geophysical and Solar-Terrestrial Data Center, Boulder, CO; S.-I. Akasofu, G.J. Romick, C.E. Campbell and G.K. Corrick, University of Alaska, Fairbanks, AK; and C.E. Hornback and A.M. Gray, NOAA Space Environment Laboratory, Boulder, CO, June 1980, 107 pp.

UAG-76 AURORAL ELECTROJET MAGNETIC ACTIVITY INDICES AE(12) FOR JULY - DECEMBER 1975, by Joe Haskell Allen, Carl C. Abston, J.E. Salazar and J.A. McKinnon, National Geophysical and Solar-Terrestrial Data Center, NOAA, Boulder, CO, August 1980, 116 pp.

UAG-77 SYNOPTIC SOLAR MAGNETIC FIELD MAPS FOR THE INTERVAL INCLUDING CARRINGTON ROTATIONS 1601-1680, MAY5, 1973 - APRIL 26, 1979, by J. Harvey, B. Gillespie, P. Miedaner and C. Slaughter, Kitt Peak National Observatory, Tucson, AZ, August 1980, 66 pp.

UAG-78 THE EQUATORIAL LATITUDE OF AURORAL ACTIVITY DURING 1972-1977, by N.R. Sheeley, Jr. and R.A. Howard, E.O. Hulbert Center for Space Research, U.S. Naval Research Laboratory, Washington, DC and B.S. Dandekar, Air Force Geophysics Laboratory, Hanscom AFB, MA, October 1980, 61 pp.

UAG-79 SOLAR OBSERVATIONS DURING SKYLAB, APRIL 1973 - FEBRUARY 1974, I. CORONAL X-RAY STRUCTURE, II. SOLAR FLARE ACTIVITY, by J.M. Hanson, University of Michigan, Ann Arbor, MI; and E.C. Roelof and and R.E. Gold, The Johns Hopkins University, Laurel, MD, December 1980, 43 pp.

UAG-80 EXPERIMENTAL COMPREHENSIVE SOLAR FLARE INDICES FOR 'MAJOR' AND CERTAIN LESSER FLARES, 1975-1979, compiled by Helen W. Dodson and E. Ruth Hedeman, The Johns Hopkins University, Laurel, MD, July 1981, 33 pp.

UAG-81 EVOLUTIONARY CHARTS OF SOLAR ACTIVITY (CALCIUM PLAGES) AS FUNCTIONS OF HELIOGRAPHIC LONGITUDE AND TIME, 1964-1979, by E. Ruth Hedeman, Helen W. Dodson and Edmond C. Roelof, The Johns Hopkins University, Laurel, MD, August 1981, 103 pp.

UAG-82 INTERNATIONAL REFERENCE IONOSPHERE - IRI 79, edited by J. Virginia Lincoln and Raymond O. Conkright, National Geophysical and Solar-Terrestrial Data Center, NOAA, Boulder, CO, November 1981, 243 pp.

UAG-83 SOLAR-GEOPHYSICAL ACTIVITY REPORTS FOR SEPTEMBER 7-24, 1977 AND NOVEMBER 22, 1977, Parts 1 and 2, compiled by John A. McKinnon and J. Virginia Lincoln, World Data Center A for Solar-Terrestrial Physics, NOAA, Boulder, CO, February 1982, 553 pp.

UAG-84 CATALOG OF AURORAL RADIO ABSORPTION DURING 1976-1979 AT ABISKO, SWEDEN, by J.K. Hargreaves, C.M. Taylor and J.M. Penman, Environmental Sciences Department, University of Lancaster, Lancaster, UK, July 1982, 69 pp.

UAG-85 [Superseded by UAG-91]

UAG-86 [Superseded by UAG-92]

UAG-87 CHANGES IN THE GLOBAL ELECTRIC FIELDS AND CURRENTS FOR MARCH 17-19, 1978, FROM SIX IMS MERIDIAN CHAINS OF MAGNETOMETERS, by Y. Kamide, Kyoto Sangyo University, Kyoto, Japan; H.W. Kroehl, National Geophysical Data Center, NOAA, Boulder, CO; and A.D. Richmond, NOAA Space Environment Laboratory, Boulder, CO, November 1982, 102 pp.

UAG-88 NUMERICAL MODELING OF IONOSPHERIC PARAMETERS FROM GLOBAL IMS MAGNETOMETER DATA FOR THE CDAW-6 INTERVALS, by Y. Kamide, Kyoto Sangyo University, Kyoto, Japan; H.W. Kroehl, National Geophysical Data Center, NOAA, Boulder, CO; and B.A. Hausman, National Geophysical Data Center, NOAA, Boulder, CO, November 1983, 197 pp.

UAG-89 ATMOSPHERIC HANDBOOK: ATMOSPHERIC DATA TABLES AVAILABLE ON COMPUTER TAPE, by V.E. Derr, NOAA Environmental Research Laboratories, Boulder, CO, July 1984, 56 pp.

UAG-90 EXPERIENCE WITH PROPOSED IMPROVEMENTS OF THE INTERNATIONAL REFERENCE IONOSPHERE (IRI): CONTRIBUTED PAPERS, MAINLY FROM THE URSI-COSPAR WORKSHOP HELD IN BUDAPEST IN 1980, edited by K. Rawer, University of Freiburg, Federal Republic of Germany, and C.M. Minnis, International Union of Radio Science (URSI), Brussels, Belgium, May 1984, 233 pp.

UAG-91 COMBINED CATALOG OF IONOSPHERE VERTICAL SOUNDINGS DATA, compiled by Raymond O. Conkright and Marcus O. Ertle, National Geophysical Data Center, NOAA, Boulder, CO, December 1984, 174 pp.

UAG-92 INTERNATIONAL CATALOG OF GEOMAGNETIC DATA, compiled by C.C. Abston, National Geophysical Data Center, NOAA, Boulder, CO; N.E. Papitashvili, Academy of Sciences of the USSR, World Data Center B2, Moscow, USSR; and V.O. Papitashvili, IZMIRAN, Moscow Region, USSR, August 1985, 291 pp. Supersedes UAG-35, 49, and 86.

UAG-93 IONOGRAM ANALYSIS WITH THE GENERALIZED PROGRAM POLAN, by J.E. Titheridge, University of Auckland, New Zealand, December 1985, 194 pp.

UAG-94 THE SOLAR MAGNETIC FIELD--1976 THROUGH 1985: AN ATLAS OF PHOTOSPHERIC MAGNETIC FIELD OBSERVATIONS AND COMPUTED CORONAL MAGNETIC FIELDS FROM THE JOHN M. WILCOX SOLAR OBSERVATORY AT STANFORD, by J. Todd Hoeksema and Philip H. Scherrer, Center for Space Science and Astrophysics, Stanford University, Stanford, CA, January 1986, 370 pp.

UAG-95 SUNSPOT NUMBERS: 1610-1985, (based on THE SUNSPOT-ACTIVITY IN THE YEARS 1610-1960, by Prof. M. Waldmeier, Copyright 1961, Swiss Federal Observatory, Zurich, Switzerland), revised by John A. McKinnon, National Geophysical Data Center, NOAA, Boulder, CO, January 1987, 112 pp.

UAG-96 SOLAR-GEOPHYSICAL ACTIVITY REPORTS FOR STIP INTERVAL XV, 12-21 February 1984 Ground Level Event, AND STIP INTERVAL XVI, 20 April - 4 May 1984 Forbush Decrease, compiled by Helen E. Coffey and Joe H. Allen, National Geophysical Data Center, NOAA, Boulder, CO, July 1987, 418 pp.

UAG-97 NUMERICAL MODELING OF POLAR IONOSPHERIC ELECTRODYNAMICS FOR JULY 23-24, 1983, UTILIZING IONOSPHERIC CONDUCTANCES DEDUCED FROM DMSP X-RAY IMAGES, by B.-H. Ahn, Kyungpook National University, Taegu, Korea; E. Friis-Christensen, Division of Geophysics, Danish Meteorological Institute, Copenhagen, Denmark; D.J. Gorney, Space Sciences Laboratory, The Aerospace Corporation, Los Angeles, CA; Y. Kamide, Kyoto Sangyo University, Kyoto, Japan; and H.W. Kroehl, National Geophysical Data Center, NOAA, Boulder, CO, April 1988, 133 pp.

UAG-98 TYPE II SOLAR RADIO BURSTS RECORDED AT WEISSENAU 1966-1987, by H.W. Urbarz, Astronomical Institute of Tubingen University, Weissenau Station, 7980 Rasthalde, Ravensburg, GFR, February 1990, 86 pp.

UAG-99 PROCEEDINGS OF THE WORKSHOP ON GEOPHYSICAL INFORMATICS, MOSCOW, AUGUST 14-18, 1988, edited by J.H. Allen, National Geophysical Data Center, NOAA, Boulder, CO; and V.A. Nechitailenko, Soviet Geophysical Committee, Academy of Sciences of the USSR, Moscow, USSR, January 1991, 304 pp.

☆U.S. GOVERNMENT PRINTING OFFICE: 1991-574-145/45038