NASA Technical Memorandum 100701

Report On The Second Catalog Interoperability Workshop January 12 - 14, 1988

James R. Thieman and Mary E. James

April 1988

NASA

Report On The Second Catalog Interoperability Workshop January 12-14, 1988

James R. Thieman
*National Space Science
Data Center
Goddard Space Flight Center*

Mary E. James
*Science Applications
Research
Lanham, Maryland*

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The topics for the Second Catalog Interoperability (CI) Workshop were divided into four main areas: the progress and future of the Directory Interchange Format (DIF); near-term plans for increasing Catalog Interoperability (CI); the longer-term goals of CI; and a demonstration, evaluation, and plans for the NASA Master Directory (Builds 0 and 1). Related topics also covered were: the DAVID-based plans for the astrophysics data system; the directory interchange demonstration requested among selected agencies by the Interagency Working Group on Data Management for Global Change; and the progress of the ESADS lexicon working group. This section summarizes the important points in the four main areas. Further details on these and the other topics may be found in the body of the report.

The Directory Interchange Format (DIF)

Significant progress has been made in the CI working group in the development of the DIF structure (an ASCII file in the form "label: value" allowing a standardized approach to the exchange of directory information among data systems). The manual defining the structure is available in a draft form and is maintained under a change control procedure. The version of Sept. 18, 1987 was "frozen" as the baseline for the creation of DIF structures describing data sets to be identified in Build 0 of the NASA Master Directory. The version is still a draft, however. Some important obstacles remain to be overcome before a complete manual can be put forth for the full scale creation of DIF structures for all of the important data sets. The DIF may be promoted as a standard for directory exchange in the future, but it was decided to delay such a process until after a testing period.

Discussion of the DIF in general concerned its flexibility in handling the highly volatile nature of data archiving in the future, specifically in the era of Eos. DIF maintenance and updating must be frequent to accurately characterize the existing data availability. This may require an automated process. It was recommended that procedures for updating and maintaining directory information be documented soon.

It was also recommended that clear guidelines be given on what constitutes a DIF directory entry. In most cases this is a data set, but closely related data sets which differ only in minor ways, such as time or spatial resolution, could and should be "aggregated" into a single DIF entry.

The general information content of the DIF is close to agreement within the CI working group (representatives of USGS, NOAA, and the existing NASA discipline data systems). The main obstacle is the creation of good sets of keywords to be used in the keyword fields. Both DIF creators and directory users will choose from these sets to characterize and search for data sets. Good multi-discipline sets are difficult to create. Several existing lists were evaluate and found to be too lengthy, too discipline-specific, and/or too detailed to be useful within a high-level directory.

Strawman lists of valid keyword sets were distributed and discussed for three categories of keywords: parameters, discipline, and location. The parameters category remains the most difficult. For the initial DIF structures, the creators were allowed to submit any parameter keywords they wished. Attempts were then made to form controlled lists based on the freely-submitted values. Three approaches were used: grouping based on "commonalities", grouping based on the units of the parameter, and grouping according to the American Geophysical Union manuscript index. Participants preferred both the first and last approaches about equally. The advisors recommended proceeding with the "commonalities" approach unless a better method can be found. This will require further work outside of the workshop.

The strawman lists for discipline and location keywords are small and quite general (e.g., Planetary Science for a discipline and North America as a location). It was felt this would be sufficient for a high-level directory which points to places where more detailed information may be obtained. Discussion of the strawman lists highlighted some shortcomings. In general, the advisors recommended that, although alternate lists might be suggested, the present lists were as good as any and could be used with minor alterations. It was also proposed that the numerical coverage fields, which define a searchable latitude/longitude rectangle for spatially confined data sets, should be used together with the location field or location-like mnemonics. Latitudes/longitudes might also be generalized to other coordinate systems for

other disciplines. Possibilities for spatial coverage and location will be further discussed by the working group.

Near-term Steps To Interoperability

Considerable progress on interoperability in general has been made since the last workshop in July, '87. All of the advisors' recommendations from that workshop are either accomplished or underway. The quest for data system interoperability is planned to proceed step-wise to gradually increasing levels of interoperability.

The first level, which is simple network interconnection of the data systems, is partially complete and allows a user to quickly transfer from one information system to another. The next step is to pass information through these connections: information on a user's interests as he transfers from one system to another and new or update information on data as it becomes available. Discussion centered on the types of information which should be transferred with a user in the course of a search. The potential utility of the various types of information is likely to vary among the existing systems. Testing should be done, perhaps with system prototypes, to determine the best approach for using transferred information to aid the user. Concerning the transfer of new or update information on data, automated DIF generation seems to be an important goal for future incorporation into data systems, but present data systems will probably require manual intervention in DIF creation for the near-term.

Existing and new data systems must have the necessary software to be able to create and use the information passed through the connections. Creation and implementation of such software requires development resources and must be included in development schedules quickly if these capabilities are to be available in the near future. For the NASA discipline data systems, this cannot happen before mid-'89 at the earliest, based on present schedules. At present there is little funding set aside for such purposes, and what would be required may differ from one system to the next. It was agreed to gather development timeline information for the participating data systems to determine the critical times for coordinating interoperability enhancements.

Long-term Interoperability Plans

Interoperability in the more distant future implies more uniformity in the overall data systems. This, in turn, requires clear guidelines on the nature of an interoperable data system. New data systems could be developed with these guidelines in mind and existing ones could evolve in the same direction.

The group discussed the boundaries within which the CI group should provide guidelines. "Guidelines" was felt to be the more appropriate term rather than standards, since different disciplines may have widely varying requirements for data systems and standardized systems may not fulfill the needs. Rigid standards are also likely to retard the use of new and better technology. Consequently, a future goal of the CI group would be to develop a document defining the concepts and capabilities of interoperable data information systems.

It seemed appropriate that "catalog" interoperability guidelines should concern metadata, the information about data, only. This would include directories, catalogs, and parts of inventories. Such areas as data browsing, graphics display, data access, etc. are important and should be handled by some working group, but they are outside of the domain of catalog interoperability.

Several specific areas for potential future guidelines were discussed. Direct catalog-to-catalog interconnections would be useful, but the cost of such and the method of permitting a user to connect to another catalog need to be determined. User interface uniformity is important, but this may be better implemented through uniformity in functionality rather than in software. Uniformity of directories throughout the interconnected systems may evolve because of the DIF, but it seems likely that each discipline will still have discipline-specific extensions to their directories. Uniformity of catalogs and inventories will be encouraged through adherence to the guidelines document. Use of a standard Database Management System (DBMS) would better enable cross-discipline and cross-system searches, but the costs will be much more difficult to assess.

2

The NASA Master Directory

The NASA Master Directory (MD) Build 0 was demonstrated using several approaches to searching for data sets of interest. This version of the MD has limited functionality since it was put together as a quick verification of the use of DIF structures as input to a directory. A number of potential enhancements were proposed for Build 1 of MD. The advisory group felt that all of the proposed enhancements would be worthwhile, but how many appear in Build 1 should depend on what can be developed without unduly delaying a proposed summer release date.

The DIF files created for Build 0 used the Sept. 18, 1987 draft version of the DIF manual mentioned previously. Nonetheless, it had over 100 DIF files loaded into the initial database, and it covers a large number of important data sets in the ocean and atmospheric sciences, the initial areas of emphasis. Expansion of the database will continue with the aim of having a sufficiently populated database by the time MD Build 1 is released so that the system may be tested by use in solving real problems. The entry of cross-discipline data sets would be especially useful for testing. Several of these types of data sets were suggested. Data sets from other agencies will also be described in the directory as a result of the interagency directory interchange project (see description in the report). Connections to other agency directories will be made available whenever possible. The advisors recommended a six month evaluation period of the directory by the general science community after release of Build 1.

INTRODUCTION

The second Catalog Interoperability (CI) workshop was held in Pasadena, California on January 12 through 14, 1988. The purpose of the workshop was to bring together members of the CI working group and the CI advisory group in order to assess the progress made since the last workshop in July of 1987, to work on some technical issues the working group was unable to resolve on its own, and to establish new goals and directions for the CI effort. The following sections are, for the most part, in chronological order according to the agenda (see Appendix A).

Attendees and Affiliations

Appendix B lists all of the workshop attendees, including working group members, advisors, and observers. The following is a list of just the advisors, their affiliations, and the area of representation.

ADVISORY GROUP

Raymond Walker, Chair	UCLA	Planetary
Vincent Abreu	U. of Mich.	Upper Atmosphere
Thomas Ciciarelli	USGS	Federal Agency
Peter Cornillon	URI	Oceans
Robert Gold	JHU/APL	Solar-Terrestrial
Helmut Jenkner	STScI	Astronomy
Dennis Joseph	NCAR	Federal Agency
James Pfaendtner	GSFC	Lower Atmosphere
Stephen Ungar	GISS	Land Science

Robert Gold was unable to attend due to illness.

Active working group members (those participating in the biweekly teleconferences) at the time of the workshop are listed below along with their affiliations and data system or organization they represent.

WORKING GROUP

Jim Thieman	GSFC	CI/MD
Joe King	GSFC	CI/MD
Blanche Meeson	GSFC	NCDS/PLDS
Ken McDonald	GSFC	PLDS
Erich Stocker	SAR	CI/MD
Mary James	SAR	CI/MD
Ted Johnson	SAR	CI/MD
Nagi Wakim	SAR	DAVID
George Saxton	NOAA	NESDIS
Jim Brown	JPL	NODS
Chuck Klose	JPL	NODS
Pat Hogan	JPL	NODS
Liz Smith	JPL	NODS
Brad Gaffney	JPL	NODS
Elaine Dobinson	JPL	PDS
Mike Martin	JPL	PDS

All except Ken McDonald of GSFC and Erich Stocker of SAR were able to attend.

Workshop Goals

After an explanation of the workshop logistics by Jim Brown, Jim Thieman began by stating the goals for the second CI workshop. Jim suggested that the following would be addressed:
- Resolve critical DIF issues
- Exchange directory interchange demonstration information
- Initial discussion of long range interoperability requirements
- Recommendations for near-term steps to interoperability
- Critique of Master Directory Build 0
- Recommendations for Build 1
- Plan for a DIF Standard

DIRECTORY INTERCHANGE FORMAT (DIF)

Review and Current Status

Jim Brown began the discussion on the Directory Interchange Format (DIF) with an overview of the current DIF structure.

The DIF provides a simple, logical, and readable way of exchanging directory level data set information. The DIF can and, with local extensions, is planned to be used for purposes other than the NASA Master Directory.

The CI working group had frozen the manual which describes the format (the DIF manual) on September 18, 1987 and had formalized a change control process for accepting changes to the DIF manual. As changes are suggested by CI working group participants, the changes are circulated and discussed. Once a consensus is reached, the changes are accumulated and incorporated into the next version of the DIF manual.

Jim presented an overview of the current DIF fields, and a discussion followed. Ralph Shapiro questioned how the data center field would be used noting that Eos (Earth observing system) will be using two facilities: a temporary archive and a permanent archive. It was felt that the DIF structure would be flexible enough to handle this, but noted that frequent updates would be necessary in the EosDIS (Eos Data Information System) environment to indicate which data are in temporary and which in permanent archives. Eni Njoku suggested that procedures for updating and maintaining directory information should be established. Jim Thieman agreed that these could be established as a procedural appendix to the DIF manual.

Ralph Shapiro noted that PDMP's should include an indication of methods for updating directory and catalog metadata. Ray Walker reminded the group that the Master Directory (MD) would be providing general information, and that the catalog level could provide information relating to the dynamic nature of data sets. Peter Cornillon suggested adding a time delta indication to the media field to indicate the rate at which data were being processed for that data set. Stephen Ungar suggested that supplementary information on processing centers could provide information on the standard rate of data acquisition. In addition, Ralph suggested that the existence of browse data for a data set should be indicated along with the data set description in the Master Directory.

It was agreed that a new version of the DIF manual baseline would be released once changes resulting from workshop discussions had been processed by the working group.

DIF Open Issues

Jim Thieman and Jim Brown gave an overview of the outstanding DIF issues. They noted that the standardization of keywords, and the publication of these keywords as appendices in the DIF Manual, was a critical issue that needed to be resolved.

Parameter Keywords

Jim Thieman stated that the problems with determination of a useful and usable set of parameter keywords was the number one obstacle to the DIF's progress. The prototype directory system (MD0) data providers were allowed to submit any "parameter" keywords they wanted to. From the diversity and inconsistency of the accumulated list of parameters, we learned that some type of parameter standardization would be necessary. A hierarchy was suggested for the final standardized list so that a user could use the hierarchy to make a logical choice of parameter from among a large number of choices. The difficult task still remains to come up with a reasonably small list of standardized parameter keywords (hierarchical or not), covering all disciplines, and sufficiently general for use at the directory level. George Saxton suggested that we look into the NASA Thesaurus, but it was noted that the NASA Thesaurus was not comprehensive enough in all the areas we needed. Mike Martin stated that we need to feed our efforts back into the NASA Thesaurus and Technical Information Group.

Peter Cornillon wondered why measured parameters and derived parameters were not listed together and was told that the parameter words were supposed to indicate what was actually contained in the data set(s) rather than what can be derived. It was also noted that we should keep in mind that different parameter classifications make things simpler for users, DIF submitters, and MD developers and that in general, users should be given the highest consideration. Peter felt that derived parameters should be linked to measured parameters so that a radiance data set would be a valid data set for SST, pigment, etc,. Victor Zlotnicki felt that the derived parameters could be listed as general keywords. Elaine Dobinson suggested that we keep in mind what each of the data systems are doing with keywords and benefit from their experience wherever possible. Jim Thieman continued the discussion by noting that the list of parameter keywords submitted for Build 0 of the Master Directory had already grown very large and that some sort of classification scheme was needed. Possible approaches that he and Mary James had considered were:

- Group by Commonalities
- Group by Units
- Group by AGU Index

A handout showed how the submitted parameters would be grouped within each of these schemes.

Jim Thieman then noted that input from the CI Advisory group and working group was needed to determine which would be the best approach and then to complete the lists. Jim suggested that participants try to produce a list of geophysical parameters from their area of specialty either using one of the classification approaches indicated, or an alternative approach.

Subsequent discussions on this subject by the advisory group resulted in a recommendation that the grouping by commonalities should be the preferred mode of approach. Several possible sources for obtaining the commonality groupings were offered, such as the NASA Thesaurus, and the NOAA NEDRES parameter lists. The CI staff at NSSDC will continue to work on this problem, investigating the various avenues and trying to formulate a solution which is feasible within the resources available.

Science Disciplines - Names and Hierarchy

Jim Thieman then showed the very general list of disciplines that had been generated and added to the DIF manual as an appendix. This list contains four major scientific disciplines corresponding to NASA code offices. Additionally, some subdisciplines were identified in this hierarchy. Jim requested that all should examine the list and any suggestions for enhancements to these current lists be brought up in subsequent discussions in the workshop or forwarded to the CI working group.

After private discussion the advisory group suggested that, although they could come up with alternative discipline and location (see below) keyword lists, they were not certain that they were better. Consequently, they recommended keeping the current list of keywords for use in the present environment.

Location Words

A very general list of location keywords had also been generated for each of the disciplines and listed as an appendix to the DIF manual. These locations were generally large areas or objects such as planets (planetary science), continents (earth science), magnetosphere (space physics), etc.,. Participants were also asked to look at this list and submit comments during or after the workshop.

Joe King stated that for all of the keywords, creating complete and orthogonal lists would reduce the chances of missing relevant data sets. Ralph Shapiro questioned whether these lists would be dynamic enough to keep up with requested modifications as new data sets are entered. Jim Thieman reminded everyone that, at the directory level, these types of lists would probably be quite general and that more detailed lists would be found in lower-level catalogs. From the discussion it was clear that good lists, in which the choice to take was obvious, are very difficult or impossible. to create and maintain. As noted at the end of the previous section, the advisors ultimately recommended keeping the present list of location keywords for the time being. There was, however, considerable later discussion on the methods of spatial searching within the directory. (See advisory group recommendations and discussion at the beginning of day 2.)

Data Set Aggregation

Another important issue was the question of what constitutes a directory entry. Normally, a data set is the basis for a directory entry, but data sets often differ from each other in only minor ways, such as ten different data sets which are essentially the same as each other except the data are averaged over longer and longer time intervals (seconds, minutes, hours, days). In subsequent discussions it was argued that the directory user should not have to wade through ten separate entries to find the particular data set of interest. The data sets should, instead, be aggregated into a single entry and the differences among them should be noted in the summary. Likewise, there may be reasons why a single data set should be broken or "projected" into two or more entries for situations such as large gaps in spatial or temporal coverage, where it is important to not span these gaps in the single-valued start and stop date or latitude/longitude fields currently required in the DIF. It was agreed that a clear statement of guidelines for aggregating data sets should be contained in the DIF manual.

INTERAGENCY DIRECTORY INTERCHANGE DEMONSTRATION

Introduction

Eni Njoku explained that NASA HQ is very interested in interoperability and, in particular, access and delivery of data to science users. This is a topic of interest in other federal agencies as well. Several interagency groups are interested in making the location of and access to data more efficient for coming programs such as the Eos "Mission to Planet Earth". The Interagency Working Group on Data Management for Global Change, which includes NASA, NOAA, NAVY, USGS, NSF, and DOE representatives, is a group particularly interested in this goal. This group hopes to establish a national virtual data and information system with a uniform approach to directories, catalogs, pricing, network protocols, quality control, data formats, etc.

They are working toward this goal by first requesting a demonstration of the ability of some of the agencies to exchange directory-level data set information and display it within each agency's directories. It is planned that this will be accomplished through the exchange of a few DIF files among the participants (NASA, NOAA, USGS, and NCAR) and the loading of the DIF information into the NASA, NOAA, and USGS directories. Network links are being established among these directories to facilitate the exchange of DIF files and to potentially allow users to be connected from one directory to another to efficiently pursue information about data of interest.

Eni pointed out that some of the outstanding issues facing this group are data system hierarchy, data redundancy, and networks. There is a need for to establish a network architecture for the virtual information system. Ray Walker suggested that the CI group should investigate the participating data systems to determine what compatibilities exist. Eni requested that someone from the CI group present the status of interoperability at the next interagency meeting. Ralph Shapiro pointed out that EosDIS is working within an international framework and that the need to efficiently locate and access data is an international problem and not just interagency.

Tom Ciciarelli wondered how the costs associated with a virtual information system would be handled. Eni explained that once electronic links were in place, there would be a standardized billing procedure. George Saxton said that NOAA feels that directory level access should be free. Distribution of the costs of access below the directory level is an issue which still needs to be resolved.

Representatives from the three non-NASA agencies (NOAA, USGS, and NCAR) participating in the directory interchange demonstration gave overviews of their systems and their planned involvement in the demonstration.

NOAA Involvement

George Saxton gave an overview of the NOAA organization and an introduction to the NOAA directory/catalog system which is currently being developed. The NOAA directory is very similar to the envisioned NASA MD structure, but some incompatibilities may exist. George indicated that the possible incompatibilities centered mostly on keywords and keyword hierarchies and that until the CI group agreed on keyword lists it would not be possible to fully assess the directory compatibilities.

NOAA is in the process of creating the DIF files for the data sets which will be used for the interchange. There will be four DIF files, representing one data set from each of the main NOAA data centers. In addition, the technical considerations of NOAA/NASA network links are being investigated. Plans call for a TCP/IP protocol link to be installed within a few months between GSFC and NOAA offices in Suitland, Maryland.

USGS Involvement

Tom Ciciarelli gave an overview of the Earth Science Data Directory (ESDD). The ESDD currently runs on an Amdahl system in their Reston, Maryland offices. It contains over 1400 entries describing natural resource, cartographic, and geologic data sets and data bases. These entries were primarily submitted from state geological surveys and natural resource departments, and the data set population continues to grow. The functions in ESDD (boolean operations, keyword search, lat/lon searches, substring searches,...) are similar to those envisioned in the NASA MD and access to the system is free, but an account must be set up to permit use.

USGS has already submitted two DIF files which have been loaded into the Master Directory, and has targeted three more entries for submission. Management and technical discussions for the installation of a network link between NASA and USGS are underway. The link has been ordered and should be installed in the March-April time period.

NCAR Involvement

Dennis Joseph described the NCAR data archive and distribution organization. Dennis noted that NCAR currently has about 300 data sets and is continually acquiring more. He explained that for cost and security reasons NCAR does not have an open catalog or directory but relies on NCAR staff to assist users with data searches. Consequently, their present participation in the interchange of directory information would be as contributors of several NCAR DIF files. Network links into NCAR are not planned at this time.

8

In addition to directory type metadata, NCAR maintains statistics about data sets (e.g. tables of the number of orbits over a given time, or a plot of ocean sample collection locations). It was suggested that such tabularized, summarized, and cross-referenced information is valuable and there should be a way of obtaining it within the directory/catalog systems. Such information may be stored within the summary text in the directory, but it may be more appropriate at the catalog level depending on the nature and size of the information.

NEXT STEPS TOWARD INTEROPERABILITY

Interoperability - Present and Future

Jim Thieman began the discussion of the near and long-term aspects of interoperability with the definitions of the increasing levels of interoperability which might be achieved. Level 1 interoperability involves providing simple network connections among data systems. In level 2 interoperability, information is sent over the network links either in the form of "context transfer" assisting the user in the search process, or in automated update information sent back and forth among the systems to keep metadata up-to-date. Level 3 involves limited standardization of data systems' functionality and user interface so that the user does not have a difficult time whenever confronted with a new data system environment. Level 4 is a completely uniform system in which all data systems appear as part of one uniform virtual data system.

CI Present Status

Jim brought the group up to date on efforts that have taken place to address the advisory group's recommendations from the previous workshop, noting that the recommendation for a document that addresses future interoperability requirements is very important to the CI effort. All of the recommendations are either accomplished or underway. The recommendations and specific responses which have taken place are listed below.

1. The Master Directory should be separated from CODD and a solar-terrestrial catalog should be developed.
   - The Interagency Consultative Group (IACG) is concentrating in the solar-terrestrial area first, recommending the creations of a solar-terrestrial data system.
   - NSSDC is currently working on a proposal for a solar-terrestrial data system with CODD as its initial catalog.

2. The Master Directory should include data from all NASA data systems and other agencies.
   - The initial emphasis placed on ocean and atmospheric data has resulted in the creation of many DIF structures in those areas.
   - NASA Headquarters is sponsoring a special effort to identify and assure entry of information on all NASA-supported as well as other earth science data of interest.
   - The Interagency Working Group on Data Management for Global Change has requested a directory interchange demonstration in which the various agencies will exchange sample DIF information their holdings.
   - A project being performed by NSSDC at the request of the CODMAC committee will result in information on important NSSDC and other outside data sets being put into the directory.

3. The NSSDC should prepare a simple document that defines the system requirements for the Master Directory and solar-terrestrial catalog.
   - The MD high level requirements document was prepared and distributed.
   - The NSSDC proposal for the solar-terrestrial data system is being formulated for NASA Headquarters Code ES.
4. Investigate ways to provide access from the Master Directory to catalogs from other data centers and agencies.
   - MD connects to all available NASA discipline catalogs and several data set catalogs. More are planned and are being added all of the time.
   - The directory interchange demonstration will result in establishment of network links to NOAA and USGS.

5. The NSSDC should begin studies to plan for future data systems with increased interoperability.
   - Initial discussions on the future of interoperability have begun in the the biweekly teleconferences.
   - The discussion of future interoperability is a part of the second workshop and will be continued in the post-workshop teleconferences, ultimately yielding a draft of a document defining data system interoperability.

Level 1 Status and Plans

Jim explained that with the Master Directory prototype which is currently available, we have achieved a useful portion of level 1 interoperability. There are currently 6-8 connections to other data systems or catalogs available in Master Directory and that number will continue to increase. There are usually not, however, defined paths for a user to go easily from a data system to the Master Directory or to another data system. Is it important for a catalog user to be able to connect to the MD or another data system? Discussion suggested that this question needs to be answered through actual use of the interconnected systems.

Level 2 Status and Plans

Level 2 interoperability involves passing information among systems. Peter Cornillon questioned why all catalog level information as well as directory information should not be in the MD. Ray Walker indicated that layering of information was found to be necessary for several reasons. Jim Thieman then asked if the level 2 "context transfer", the passing of search information along with the user from system to system to aid in efficiency, is a high priority. Should information be passed, and if so, how much should be transferred? Future implementations of most data systems (PDS, NODS, PCDS) are more than a year into the future, and any modifications to the software in those systems to accept transferred context information and use it will be a while off. Other options for more immediate implementation of level 2 capability are interface programs which could perhaps accept the transferred information in a format similar to the DIF (search interchange format?) and feed it into the data system.

The question was also raised as to whether it is important to have automated methods of keeping information up-to-date by methods such as automated DIF creation, transfer, and ingestion. Ralph Shapiro stated that in Eos automated information transfer to the MD is essential from sheer data volume considerations and the need for keeping up with data reception.

Level 3 Plans

The plans for achieving level 3 interoperability call for formulation of a document containing recommended standards or guidelines for data systems. This would be used by existing data systems as they came out with new versions of their systems, and also in the planning for new data systems. The document would be formulated out of discussions within the CI group. Jim questioned how deep within the directory/catalog/inventory systems this document should extend and whether it made sense to standardize systems down to the inventory level. Where should the domain of the CI group stop? There are other ESADS initiated groups which deal with subjects such as data browsing, graphics, etc. In discussion, it was suggested that catalog interoperability be limited to the metadata aspects of the data systems, but there should be a group which is concerned with data system interoperability as a whole. The final resolution of this issue was left for the advisory group.

DAVID

Nagi Wakim presented an introduction to the Distributed Access View Integrated Database (DAVID) software system as a possible solution to some of the needs of long term (levels 3-4) interoperability.

Nagi informed the group that during the Fall of 1987, it was demonstrated that the DAVID software was able to access several heterogeneous database management systems on several different computers with different operating systems. This demonstration was performed using databases synthesized from NODS, PDS, and PCDS information.

DAVID is currently being developed as a solution for the long range data system plans of the astrophysics community. Nagi introduced the DAVID library concept and cataloging structure, explaining that this would be used in the astrophysics data system currently being planned. It is possible that similar concepts could be used in other disciplines and achieve a uniform-appearing data system.

INTEROPERABILITY USER REQUIREMENTS - LONG TERM

Discussion then began of the questions raised by Jim Thieman's earlier presentation on interoperability. Jim started the discussion of long-term interoperability by listing several topics within the subject, including:
- Direct Interconnection Among Major Systems
- Free Access to All Metadata
- Uniform User Interfaces
- Uniform Directories
- Catalog Uniformity
- Inventory Uniformity
- Online Data Ordering
- Online Data Access
- Data Manipulation / Graphics Capability
- Common DBMS / System Software

Jim noted that ESADS had targeted CI as a three year effort and that, due to the phasing of data system development plans, it was not likely that level 3 could be reached in that time frame. Jim requested input as to the relative importance and cost of these aspects of interoperability.

First came the question of whether there should be direct interconnections among the data systems such that transfers could be made from one catalog to another and how "free" the access should be. There was much discussion regarding the difference between operational costs and development costs. Blanche Meeson noted that network costs may be increasing in the future. Eni Njoku noted that a certain level of baseline resource would already exist and not enter into the cost (e.g. networks that would be in place regardless of the CI efforts). Carroll Hood said that catalog-to-catalog interoperability would greatly increase the required efforts and cost. There was much discussion about what type of interconnections were needed among the catalogs.

As to a uniform interface, Blanche Meeson and Elaine Dobinson felt that this might be possible at directory levels, but less likely at catalog and inventory levels. Peter Cornillon felt that interface uniformity is critical, but Mike Martin felt that standards rather than absolute hardware and software uniformity were critical. Stephen Ungar felt that screen appearance was not as important as functional uniformity. Helmut Jenkner suggested that we agree on a basic set of functions - both screen layout and directory functional concepts - and use these as recommended standards.

Jim Thieman asked if we should consider a uniform, distributed MD. Jim Brown felt that non-uniform data types exist among the different disciplines, and consequently different data systems will not have uniform directories. Blanche Meeson stated that discipline directories may want to maintain relationships that are not in the MD. Jim Brown noted that the DIF imposes commonalities on data systems and that its use tends to make directories more uniform.

Concerning uniform catalogs and inventories, Jim Brown noted that a data systems concepts or guidelines document will encourage uniformity across disciplines and agencies. Pat Hogan reminded the group that uniformity and interoperability are not the same things. Jim Brown felt that data access, ordering, manipulation, graphics, etc. would be outside of our group's charter, but some group should handle the overall picture.

Common DBMS/System software among the data systems would allow queries to be made which can span across data systems to more efficiently search for data. Blanche Meeson stated that querying across multiple levels (e.g. directory and catalog level) will cause additional problems with assessing costs. Eni Njoku noted that it is more important to know when this will be available than to determine what it will cost.

There was a great deal of discussion regarding access to inventory (and browse) data and the costs of this access. George Saxton noted that NOAA makes data available at cost to individuals who are not associated with agencies. Eni Njoku suggested that we investigate to what level data access should be free of charge and present an issue paper regarding data access cost. Ray Walker suggested that in order

13

to do this we need more information regarding how many users we could expect and how much data they would have access to. Joe King noted that we will have to consider the cost per query and the number of queries we could expect. There was general agreement that there should be free access to metadata, but it needs to be determined to what level (e.g. inventory or catalog).

PRESENTATION AND DISCUSSION OF RECOMMENDATIONS FROM DAY 1

Advisory Group Recommendations

At the start of the second day Vincent Abreu presented the advisory group's recommendations from their private session held the previous afternoon. The recommendations included:
1. Retain high level discipline and location keywords presently used in MD0, with fine tuning.
2. The parameter classification approach based on commonalities should be used until a better method is found.
3. The CI group should rely on user suggestions to gradually optimize the keyword system.

The advisory group indicated that the coverage field should be expanded to include a keyword to identify a coordinate system. Stephen Ungar also suggested that we explore the use of location mnemonics. Peter Cornillon said that location and coverage should not be separated and that the location group should include coverage, height and depth locations, and location names. Joe King asked if the coordinate system would be an indication of how to interpret values such as whether values were right ascension and declination or geographic coordinates. Helmut Jenkner said yes, but argued that coordinates should have a third dimension.

Vincent Abreu also suggested a system in which data sets are keyed to "objects" (similar to a discipline, location), "coordinates" (areal coverage), and "parameter" (geophysical parameters).

Working Group Recommendations

Jim Thieman then presented the working group recommendations which included:
1. The CI group should generate a guidelines document for interoperable directories, catalogs, and inventories, giving the concepts or capabilities which should be common. The method of implementation is left to the developer. It should be understood that compliance with these guidelines will not yield a "virtual data system".
2. Use of common software for master directories in the government agencies may be feasible and worth pursuing.
3. A timeline showing planned development for each of the data systems would be useful in working toward coordinated CI development..
4. Catalog interoperability should not include all that is necessary for data system interoperability (i.e. only metadata access is addressed in the CI group).

A brief discussion followed in which Eni Njoku indicated that a virtual system is one in which a user should be able to query several catalogs in one session. Jim Brown suggested that functional guidelines are the first step to level 2 and 3 interoperability. It was agreed that the data system timeline will be very useful in proceeding with the promotion of interoperability.

INTEROPERABILITY USER REQUIREMENTS - NEAR TERM

Information Transfer from Directory to Catalog/Data System

The issue of context transfer, i.e., passing search information (such as the chosen data set identifier) along with the user when he is connected to another system, was discussed. Is it useful for transferred search information to be used to allow a user to bypass the top levels of the catalogs and go directly to more detailed information on the data set(s) of interest? Blanche Meeson thinks that discipline directories may have additional information not contained in the Master Directory which may be useful not to bypass. In addition, if the data set ID is one of the major pieces of information to be passed, PCDS is not set up to be able to use a data set ID. Elaine Dobinson noted that a data set ID could be used in PDS, but the user may gain some information by starting at a higher level in their catalog. Peter Cornillon pointed out that some users will want a narrowing of the chosen number of data sets at the catalog level, and some may want an expansion.

Type of Information Transferred

Jim Brown noted that it would be possible to pass other search information as well as the data set ID, and allow users to suggest whether they would like to see the query expanded or refined. Mike Martin suggested that we consider sending information regarding the user's preference for things such as screen type, editor, etc., to the catalog systems. Joe King suggested that an option may be to send information concerning independent variable ranges, keywords, and query operators to data systems.

Jim Thieman reiterated that, in the near term, how data systems use information passed will be data system dependent and that any use of context transfer is likely not to occur until at least mid 1989. It was decided that the question of whether to implement these with internal software modifications or interface programs should wait until we see a time-line of system development. Eni Njoku did indicate that recommendations as to when software modifications may be made available is critical.

There was also general agreement that data systems should adopt some sort of data set ID that would be used throughout their directories, catalogs, and inventories where appropriate. Jim Thieman asked if we would be able to at least explore some of the options for passing information before the 1989 time frame. Elaine Dobinson offered to provide the PDS demonstration software to anyone wishing to use it in a test of passing context.

Automated DIF Generation

Automated DIF file generation would involve passing directory-level information as it is currently stored in the data systems to the Master Directory. This would probably involve some field and keyword mapping, since correspondence will not always be one-to-one. It was suggested that if the process were entirely automated that there may be problems with undetected inconsistencies in data set descriptions. Victor Zlotnicki requested some kind of DIF template for the descriptions. Peter Cornillon wanted to know how many DIF files were going to be prepared for the Master Directory, since a relatively small number might obviate the need for automation.

Stephen Ungar suggested that the method of DIF preparation be left to the data system. Mike Martin thinks a standardized procedure for generating DIF files will be necessary (either automated or manual) in order to ensure accuracy and success. Jim Brown indicated that the initial capture from the data systems would likely be manual, but that, in the future, automated generation may be more likely. Jim Thieman noted that DIF files will be used for update as well as new entries. Stephen Ungar pointed out that small data sites are not likely to be able to automatically submit or update DIF files.

DIF CREATION EXPERIENCES, SUGGESTIONS, AND COMMENTS

A panel of experienced DIF creators (Mary James, Liz Smith, and Brad Gaffney) discussed the methods they used and the problems they encountered in the preparation of the initial Master Directory entries. Among the problems discussed were the outstanding DIF manual issues (e.g., keywords), updating procedures, and dataset aggregation and projection.

During the panel discussion, several topics were singled out, including the request for listing distribution media rather than storage media. The need for listing total volume (i.e. megabytes) of data was also proposed. It was agreed this would be useful. It was also suggested that multiple times and coverages be allowed rather than splitting a data set into several directory entries to prevent false hits on searches of data sets with wide gaps in time or spatial coverage. This proposal was more controversial and the group did not reach a consensus.

Several guidelines were suggested as additions to the DIF manual. After much discussion of the problems involved in data set directory entry representation, it was noted that aggregation guidelines need to be produced.

There was general agreement that several useful DIF tools, such as a prechecker and a method of extracting DIF files from the MD, should be provided by the Master Directory.

MASTER DIRECTORY BUILD 0 DEMONSTRATION

A brief presentation and demonstration of Build 0 of the Master Directory was given by Ted Johnson and Mary James. They went through several scenarios indicating possible ways of using the Master Directory. It was noted that Build 0 is limited in functionality in comparison to Build 1, which is planned for mid '88. Build 0 allows testing of some of the search paths and information displays. It also indicates some of the compromises which were made in the DIF definition in order to keep it simple (e.g., no relationship of which sensor goes with which source for a multi-sensor, multi-source data set). It is hoped that critiques of Build 0 will provide a more solid base from which to develop Build 1. A reference guide to using Build 0 was handed out. Anyone may access Build 0 through the SPAN network by typing:

$ SET HOST NSSDCA   Username: DIR_DEMO

and then following the menus.

MASTER DIRECTORY PLANS

After the demonstration of MD0, Jim Thieman presented several potential enhancements for the next version of the Master Directory. Among the functions that were suggested for a fully functional MD were:
- Additional Keyword Searching Capability (multiple values, lists of valid keywords)
- Other Search Methods
- Text Searching Capability
- Information Display Enhancements
- Contact Person Information Display
- Supplementary Information on Source, Sensor, Data Center, and Campaign

Discussion of the need for these enhancements was left to the advisory group private session.

PRESENTATION AND DISCUSSION OF RECOMMENDATIONS FROM DAY 2

Advisory Group Recommendations

At the beginning of the third day Vincent Abreu presented the advisory group's recommendations from day 2. These recommendations included:
1. Data bases (such as PDS, PCDS, PLDS, NODS...) have to beaccessible.
2. Users must be able to locate (those) data bases with ease.
3. Increase interaction between disciplines; unify approaches to discipline oriented data base cataloging.
4. Implement and fully populate (MD1) as soon as possible
    a. Make it work
    b. Make it accessible
    c. Populate with a sufficiently large data base
5. Encourage wide science use of Build 1 of the Master Directory and use the feedback to evaluate the system.
6. Based on a 6-month experience with MD1, the advisory group will carry out a final evaluation of the system and suggest high level requirements and future developments.

More detailed recommendations concerning the Master Directory included the following:
1. Do an assessment of contributors (who will be submitting DIF files).
2. Why is TAE being used? This could be a potential obstacle.
3. When possible, include PIs in the writing of DIF files.
4. Careful consideration should be given to aggregation of data.
5. A plan for maintenance of DIF files is needed.
6. Review requirements for all data systems.
7. Estimate number of accesses at the data centers and for MD1:
    a. Who will use it;
    b. What resources are required; and
    c. How are the estimates done.
8. Determine the functionality of NASA data centers as well as USGS, NOAA,...

In the discussion of these recommendations, it was stated that the advisors are primarily concerned with determining availability and access to the data. The review of data systems requirements and functionality should be used to determine what cataloging concepts have already been looked at. Eni Njoku expressed concern that each of the data systems should have a science advisory group and will be looking into this.

The advisory group requested a report on these efforts prior to the next workshop.

Working Group Recommendations

Jim Thieman then presented the working group's recommendations on the previous day's discussions which included the following.
1. Responsibility for correcting or updating DIF files lies with the originating data center unless agreement is made to pass the responsibility to another data center or NSSDC. SAIC may be the "originating data center" for many of the DIF files resulting from their earth science data set identification.
2. The next version of the DIF manual should occur after the Build 0 evaluation and testing period. Expansion of the database should proceed slowly during this period and emphasize DIF files useful for testing. Presently existing DIF files and new ones created before the next version are likely to require many changes.

3. As a part of the Build 0 evaluation and testing, we ask for a written critique from the advisors indicating consensus agreements on the needed changes to the directory and the DIF.
4. There should be agreement on a time when a "final" version of the DIF structure is to be specified. All changes to the DIF after that time must be "upwardly compatible" and not require changes to existing DIF files. Database expansion can proceed more rapidly at that point.

Peter Cornillon, speaking for the advisors, indicated that they wanted to see a system with greater functionality and population in order to evaluate the MD concept. Phil Zion suggested using an actual science problem as an evaluation, and gave the example of the many data sets being used for ozone research. Victor Zlotnicki suggested populating the MD with a comprehensive sample for a subdiscipline. Joe King suggested that we additionally populate with several data sets likely to be used as cross-discipline correlative data.

Blanche Meeson and Stephen Ungar offered examples of cross-discipline data sets that may possibly be found in MD. This brought up the problem of needed standardization and procedures for assigning keywords to entries, especially where they must be retrievable by scientists in several disciplines. The advisory group suggested the need for testing with a 'comprehensive' population of several disciplines.

DIF AS A STANDARD

Jim Thieman introduced the idea of promoting the DIF as a standard, noting that existing standards such as FITS and CDF had been examined and, although good for data packaging, they would not be the optimum method for the exchange of directory metadata. The DIF manual could be put into the form of a standards document and distributed to the affected communities for comment. Advice on the procedure for creating a standard would come from Don Sawyer at the NSSDC and Jim Brown at JPL, who have had experience in these matters. Potential communities for the standard were NASA, other agencies, the U.S. science and data system development community, and the international science and data system development community; each succeeding one representing a more arduous journey to approval.

It was agreed that the CI group should make efforts to maintain awareness of other international and interagency groups' efforts in the same area. Eni Njoku requested that the CI group prepare a statement of the current status of the DIF and send that along with the DIF manual to him for the upcoming CEOS meeting, an international group concerned with data management. It was also recommended that formal standardization of the DIF not be pursued at this time.

PROPOSAL FOR INTEROPERABILITY AGU SESSION

Vincent Abreu announced that there will be a special session on access to aeronomy and space physics data at the Spring AGU meeting, and encouraged CI participation. Ray Walker noted that this experience could be very helpful in getting user comments and requirements. Jim Green from NSSDC will chair this session, and interested participants should contact him for further information.

LEXICON WORKING GROUP PROGRESS REPORT

Joe King gave a status report of the Data System Lexicon Working Group (DSLWG) efforts noting that the DSLWG was addressing the problem of proposing a common set of terminology to be used among participating members in the data management field. The DSLWG is focusing on data and data system oriented words and, in the DSLWG meeting on 1/11/88, procedures for updating and adding words to the lexicon were established. Joe announced that the DSLWG will be having periodic teleconferences to further the progress of the lexicon.

SUMMARY SESSION

After final separate meetings for the working and advisory groups, Ray Walker summarized the discussion among the advisors indicating that:
- The advisors re-emphasized that MD evaluation should use real science applications;
- There needs to be adequate population in the MD to work with real science problems;
- There needs to be increased functionality to use the MD; and the functions should include
  a. Multiple keywords,
  b. Contact names and information,
  c. Supplemental information on data centers and sources,
  d. General keyword searching, and
  e. An improved interface (if time permits);
- Level 3 and 4 considerations should be held until MD1 can be evaluated;
- The study of capabilities and requirements of other data systems is critical; and
- The advisors would like 6 months to evaluate MD Build 1.

The advisors hope to submit their report to the CI Working Group around March 1. This report will detail their recommendations, expectations, and a brief critique of MD Build 0. Peter Cornillon noted that he would not be able to evaluate MD Build 0 form a scientist/user standpoint because there is insufficient population and capabilities.

The working group developed a general timeline for the near-term progress in CI which was roughly as follows:

Mar.-Apr. '88
Build 0 evaluation , Advisors submit critique , Major DIF changes completed , Build 1 requirements finalized , New DIF manual version frozen

Jul.-Aug. '88
Useful population of database for evaluation , Build 1 released

Jan.-Feb. '89
Evaluation by broad science community (6 months), Alter CI plans in accordance with evaluation

A great deal of discussion centered on how best to target data sets for additional MD population. For a good evaluation period it seems necessary that there be comprehensive coverage of all important data sets in at least two or three related disciplines.

As a guideline to evaluating Build 0, it was suggested that the science advisors review the questions in the Master Directory Requirements Document. This was sent in the same package with the previous workshop report. The questions indicated what could and could not be answered by a directory built on the DIF structure. The advisors need to determine if some of the questions which cannot be answered in the present directory database are important and need to be answerable in the future.

Jim Thieman asked the advisors to provide a list of requirements they consider to be important for Build 1. It was agreed that the CI group would send the advisors an indication of what enhancements could be implemented in a given time period, and the advisors will then respond. Within a few months, the MD1 functionality ought to have been determined through interactions between the MD development team and the CI advisors.

In closing Jim Thieman noted that the next workshop is scheduled for July 1988, but that may be changed based on the timing for MD Build 1 development, for solution of other CI issues, and how far CI participants were able to progress with their various action items.

APPENDIX A - WORKSHOP AGENDA

CATALOG INTEROPERABILITY WORKSHOP II AGENDA
January 12-14, 1987
NASA/JPL
Pasadena, California

Tuesday, Jan. 12th,   Building 264   Room 739

08:00	COFFEE	
08:30	Logistics/Introductions	Brown/Thieman
08:45	Directory Interchange Format (DIF) Review and Developments	Brown
09:15	DIF open issues Parameter keywords Science Disciplines - names and hierarchy Location words Data set aggregation	
10:30	BREAK	
10:45	Plans for Interagency Directory Interchange Demonstration Description, Participants, number of DIF files involved, schedule	Njoku/Thieman
	NOAA Involvement	Saxton
	USGS Involvement	Ciciarelli
	NCAR Involvement	Joseph
12:00	LUNCH	
13:00	Next Steps Toward Interoperability	
13:00	Presentation of potential interoperability scenarios	Thieman
13:30	DAVID Interoperability Presentation	Wakim
13:45	·Interoperability Requirements Discussion	

Interoperability user requirements - long term

High Level Interoperability Requirements
Discussion (Need \and Cost)
    Direct interconnection among major systems
    Free access to all metadata
    Uniform user interfaces

Uniform directories
Catalog uniformity
Inventory uniformity
Online data ordering
Online data access
Data manipulation / graphics capability
Common DBMS / system software

15:00   BREAK

Interoperability Requirements Discussion (continued)

16:00   Advisory Group private session Building 264 Room 114B
          Working Group session   Building 264 Room 739

18:00   Resolution of Cosmic Issues (Happy Hour)

19:00   Dinner

Wednesday, Jan. 13th, Building 198 Room 102

08:00   COFFEE

08:30   Advisory/Working Group Summation/Recommendations

08:45   Interoperability Requirements Discussion (continued)

Interoperability user requirements - long term
    (continued as needed)

Interoperability user requirements - near term

  Information transfer from directory during search - method

Automated DIF generation - Implications on data systems

10:00   BREAK

Interoperability user requirements - near term
(continued)

12:00   LUNCH

13:00   Panel discussion of DIF creation   James/Smith/Gaffney
          experiences, suggestions, and comments

13:45   Master Directory Build 0 Demonstration   Johnson/Stocker

14:15   DIF issues
          (continued from Monday after overnight
          reflection)

15:00   BREAK

15:15   Master Directory Plans
               Directory searching procedures
               Keyword searching
               Other searching
               Information display
               Database expansion plans
               Plans/needs for Build 1

16:00   Advisory Group private session   Building 301 Room 427
        Working Group session   Building 198 Room 102

Thursday, Jan. 14th, Building 264 Room 739

08:00   COFFEE

08:30   Advisory/Working Group Summation/Recommendations

09:00   Master Directory Plans (conclusion)

09:30   DIF as a standard
        Current status and plans for revision
        Interagency interest
        Possible standardization routes

09:45   Proposal for Interoperability AGU Session   Abreu

10:00   Lexicon Working Group progress report   King

10:15   BREAK

10:30   Advisory Group private session   Building 301 Room 427
        Working Group session   Building 264 Room 739

11:30   Joint wrap-up session   Building 264 Room 739

12:00   ADJOURN

APPENDIX B - WORKSHOP ATTENDEES

Vincent Abreu
Space Physics Research Laboratory
University of Michigan
2455 Hayward St.
Ann Arbor, MI 48103   SPRLC::ABREU

James W. Brown
Mail Stop 301-433
Jet Propulsion Laboratory
4800 Oak Grove Drive   [JWBrown/NASA]NASAMAIL
Pasadena, CA 91109   TONYS::JWB

Tom Ciciarelli
804 National Center
USGS
Reston, VA 22092   NSSDCA::CICIARELLI

Peter Cornillon
University of Rhode Island
Graduate School of Oceanography
South Ferry Road   [P.Cornillon/OMNET]MAIL
Narragansett, RI 02882   URI::PETE

Elaine Dobinson
Mail Stop 301-320
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   JPLPDS::EDOBINSON

Sandy Dueck
Mail Stop 230-201DN
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   [SDueck/NASA]NASAMAIL

Mark Friedl
Dept. of Geography UCSB
Santa Barbara, CA 93106

Brad Gaffney
Mail Stop T1206-D
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   STANS::BCG

Pat Hogan
Mail Stop 506-232
Jet Propulsion Laboratory
4800 Oak Grove Drive

Pasadena, CA 91109   TONYS::PDH

Carroll Hood
SAIC
400 Virginia Avenue, S.W. Suite 810   [CHood/NASA]NASAMAIL
Washington, DC 20024   NSSDC::HOOD

Frank Islam
STX
4400 Forbes Blvd.
Lanham, MD 20706   [RShapiro/GSFCMAIL]GSFC

Mary James
Science Applications Research
4400 Forbes Blvd.   MJames/GSFCMAIL]GSFC
Lanham, MD 20706   NSSDCA::MJAMES

Helmut Jenkner
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218   SCIVAX::JENKNER

Ted Johnson
Science Applications Research
4400 Forbes Blvd.
Lanham, MD 20706   NSSDC::JOHNSON

Dennis Joseph
National Center for Atmospheric Research
P.O. Box 3000 Boulder, CO 80307

Tim Kaufman
Mail Stop 301-440
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   TONYS::TMK

Joe King
Code 633
NASA/GSFC
Greenbelt, MD 20771   NSSDCA::KING

Chuck Klose
Mail Stop 169-236
Jet Propulsion Laboratory
4800 Oak Grove Drive   [CKlose/NASA]NASAMAIL
Pasadena, CA 91109   STANS::JCK

Mike Martin
Mail Stop 233-208
Jet Propulsion Laboratory
4800 Oak Grove Drive   [MikeMartin/NASA]NASAMAIL
Pasadena, CA 91109   JPLPDS::MMARTIN

Richard Masline
Mail Stop 301-490
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   TONYS::RCM

Blanche Meeson
Code 634
NASA/GSFC   [BMeeson/GSFCMAIL]GSFC
Greenbelt, MD 20771   NSSDCA::MEESON

Carol Miller
Mail Stop 301-440
Jet Propulsion Laboratory
4800 Oak Grove Drive   [CMiller/J.P.L.]TELEMAIL
Pasadena, CA 91109   TONYS::CLM

Eni Njoku
Code EE
NASA Headquarters   [ENjoku/NASA]NASAMAIL
Washington, DC 20546   STANS::EGN

Jim Pfaendtner
Code 611
NASA/GSFC
Greenbelt, MD 20771   [CPeslen/GSFCMAIL]GSFC

Carolyn Robinson
Science Applications Research, Inc.
4400 Forbes Boulevard   [Ombudsman/GSFCMAIL]GSFC
Lanham, MD 20706   NSSDC::OMBUDSMAN

George Saxton
NOAA/NESDIS/DAPO FB4
Suitland, MD 20236   NSSDCA::HUNOLT

Ralph Shapiro
STX
4400 Forbes Blvd.
Lanham, MD 20706   [RShapiro/GSFCMAIL]GSFC

Sheldon Shen
Mail Stop 301-440
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109   TONYS::SHS

Elizabeth Smith
Mail Stop 169-236
Jet Propulsion Laboratory
4800 Oak Grove Drive   [EASmith/NASA]NASAMAIL
Pasadena, CA 91109   STANS::EAS

Jim Thieman

Code 633
NASA/GSFC
Greenbelt, MD 20771

[JThieman/GSFCMAIL]GSFC
NSSDC::THIEMAN

Stephen Ungar
Code 623
Goddard Institute of Space Studies
2880 Broadway
New York, NY 10025

Bitnet: CCSGU@NASAGISS
[SUNGAR/GSFCMAIL]GSFC

Nagi Wakim
Science Applications Research
Code 634   NASA/GSFC
Greenbelt, MD 20771

NSSDCA::NWAKIM

Ray Walker
Institute of Geophysics and Planetary Physics
University of California, Los Angeles
Los Angeles, CA 90024

[RayWalker/NASA]NASAMAIL
UCLASP::RWALKER

Phil Zion
Code EC
NASA Headquarters
Washington, DC 20546

[PZion/NASA]NASAMAIL
DAFFY::PMZ

Victor Zlotnicki
Mail Stop 169-236
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

[V.Zlotnicki/OMNET]MAIL
TONYS::VZ

APPENDIX C - ADVISORY GROUP MEMBERSHIP

NAME	AFFIL.	DISC./AGENCY
Raymond Walker, Chair	UCLA	Planetary
Vincent Abreu	U. of Mich.	Upper Atmosphere
Thomas Ciciarelli	USGS	Federal Agency
Peter Cornillon	URI	Oceans
Robert Gold	JHU/APL	Solar-Terrestrial
Helmut Jenkner	STScI	Astronomy
Dennis Joseph	NCAR	Federal Agency
James Pfaendtner	GSFC	Lower Atmosphere
Stephen Ungar	GISS	Land Science

APPENDIX D - WORKING GROUP RECOMMENDATIONS

1. The CI group can generate a guidelines document for interoperable directories, catalogs, and inventories giving the concepts or capabilities which should be common. The method of implementation is left to the developer. It should be understood that compliance with these guidelines will not yield a 'virtual data system'.
2. Use of common software for master directories in the government agencies may be feasible and worth pursuing.
3. A timeline showing planned development for each of the data systems would be useful in working toward coordinated CI development.
4. Catalog interoperability should not include all that is necessary for data system interoperability.

1. Responsibility for correcting or updating DIF files lies with the originating data center unless agreement is made to pass the responsibility to another data center or NSSDC. SAIC may be the "originating data center' for many of the DIF files resulting from their earth science data set identification.
2. The next version of the DIF manual should occur after the Build 0 evaluation and testing period. Expansion of the database should proceed slowly during this period and emphasize DIF files useful for testing. Presently existing DIF files and new ones created before the next version are likely to require many changes.
3. As a part of the evaluation and testing we would ask for a written critique from the advisors indicating consensus agreements on the needed changes to the directory and the DIF.
4. We should agree on a time when a 'final' version of the DIF structure is to be specified. All changes to the DIF after that time must be 'upwardly compatible' and not require changes to existing DIF files. Database expansion can proceed more rapidly at that point.

APPENDIX E - ADVISORY GROUP REPORT

Report of the NASA Catalog Interoperability Advisory Group

The NASA Catalog Interoperability Advisory (CIA) group met at the Jet Propulsion Laboratory (JPL) in Pasadena, California on January 12 to 14, 1988. This was the second meeting of the CIA. The purpose of the meeting was to review the progress since the first CIA meeting in the design and development of the Master Directory (MD) and to participate in the planning of future work on the system.

At its first meeting the advisory group developed several recommendations concerning the future directions in the cataloging and directory design efforts. The advisory group recommended:
1. The Master Directory should be separated from the original Central Online Data Directory (CODD). It was further noted that the CODD was a good start toward a solar-terrestrial data catalog and should be developed into one.
2. The Master Directory should be expanded to include not only the data in NASA archives other than NSSDC but should include information about the holdings of other agencies.
3. The NSSDC should prepare a simple document that defines the system requirements for the Master Directory.
4. The NSSDC should immediately undertake a study to determine the ways to provide access to data from other agencies.
5. The NSSDC should begin studies to plan for future data systems with increased interoperability.

During the approximately 6 months between the two CIA meetings, the design and development team for the MD made considerable progress in carrying out the first four recommendations. The MD development effort has been separated from the CODD and in a separate action plans have begun for a Solar-Terrestrial Data System. An important part of the second meeting was a series of presentations by representatives of NOAA, USGS and NCAR at which the involvement of these agencies was discussed. The development team has begun writing a Functional Requirements Document describing in detail the requirements for the first version of the MD. A key element in providing the MD with the required breadth is the concept of a Directory Interchange Format (DIF). The DIF is the mechanism through which data will be communicated between the data systems and the MD. During the period between the two meetings the definition and contents of the DIF matured and the first DIF's were used in the prototype catalog.

During the three day meeting the advisory group met three times in executive session and developed a set of principles and recommendations for the development and implementation of the MD in the immediate future. The advisory group made six major recommendations.
1. The data bases included in the MD must be accessible to the scientific community.
2. Users of the combined data systems starting with the MD and extending to the discipline data systems must be allowed to locate those data bases with ease.
3. Within NASA the disciplines should try to unify their approaches to data base cataloging.
4. Build one of the MD should be implemented and fully populated as soon as possible.
5. Wide science user participation should be encouraged in build one and the developers should use the feedback from the community to improve the system.
6. The build one system should have a detailed evaluation by the CIA after 6 months of use.
In the following section we have discussed each of these recommendations in detail. Finally in the last section we present several very detailed recommendations for studies to be carried out as a part of build one and outline some concerns about the implementation of the MD.

Detailed Recommendations

1. The data bases included in the MD must be accessible to the scientific community.

   The advisory group was concerned that data listed in the Master Directory might not actually be available to scientific users. No data bases should be included in the MD which are not readily available for users through the path indicated in the DIF. This concern arose because to a large extent many of the NASA data systems are still in the demonstration stage. Either they are still in development and have only preliminary versions online or they have not been expanded since they stopped being "pilots" and became official data systems. The committee recommends that the NASA data systems be populated as fully as possible and that the MD not accept any DIF until the corresponding data can be obtained through the data system.

2. Users of the combined data systems starting with the MD and extending to the discipline data systems must be allowed to locate those data bases with ease.

   The MD and the data systems under it must provide a clear path to the data in the catalogs. A user should be able to start at the MD and to end up by ordering the data required for a given study. Care should be taken at all levels of the system to make this as easy as possible. This is important because at least initially the MD will be connecting dissimilar data systems.

3. Increase interaction between disciplines by unifying the discipline oriented data base catalogs.

   This recommendation is an extension of the previous one. Initially the catalogs will be dissimilar simply because they were created at different times and represent different disciplines. The existing NASA Oceans, Climate and Planetary Data Systems have different cataloging schemes and user interfaces. As we strive for interoperability efforts should be made to evolve toward a more common system at least at the catalog level.

4. Build one of the MD should be implemented and fully populated as soon as possible.

   We believe that it is very important to deliver a working version of the MD as soon as possible. The development of a scientific data system is an iterative process. Developers design and build working prototypes of the system and then let users comment on it, criticise it and even break it. The development team then builds an improved version which meets the users objections. Only this way can a system be built with which the user community is comfortable. Only this way can the system adjust to the changing needs of its users. Once a prototype is working, it should be made accessible to as wide a population of scientists as possible. For this iterative process to work the system must be minimally useful from the beginning. This means that it must be populated with a sufficiently large data base to be useable. The advisory group feels that too often effort that should be devoted to making the current system work is actually devoted to designing the next generation. The end result is that litltle is leaned from the prototype and many of the same mistakes propogate to the "improved" system. For this reason, the advisory group suggests that no effort be devoted to the next generation until this generation, Build one, has been developed and used, not simply tested.

5. Encourage wide science user participation in build one and use the feedback from the community to improve the system.

   This recommendation is an expansion of the previous one. To adequately test the system we need as large a user base as is practicable. A small committee like the CIA cannot hope to provide the range of user experiences needed to fine tune the system. This requires many users with many problems.

30

6. The build one system should have a detailed evaluation by the CIA after 6 months of use.

The advisory committee will monitor the experience of the MD users during the first 6 months of use and will prepare a written evaluation of the system and suggest future developments.

Detailed Considerations

In addition to the broad recommendations discussed above the advisory group listed several rather detailed concerns. These are primarily concerned with understanding the scope of the MD effort. In particular the committee would like the developers to undertake a study of the scope of the MD effort and its relationship to the other data system activities.

1. We would like the development team to conduct an assessment of the possible contributions to the MD. This assessment should include determining the possible sources of the DIF's and the number of DIF's to be expected from each source.

2. The development team should also try to estimate the number of accesses of the system. They should try to understand who will use the system, and what resources will be required. The report on the study should clearly indicate how the estimates were made.

3. As part of the study the development team should review the requirements on all data systems involved in the MD effort and should determine their functionality. This includes the NASA data centers, the USGS, NEDRIS and NOAA.

The advisory group also had some concerns about the details of the implementation of the MD. These are listed below.

1. When including metadata in the MD careful consideration should be given to the aggregation of data. Separate entries should really reflect different data.

2. Members of the scientific community should be involved in the writing of the DIF's. This will help assure that the information is correct. Either the Principal Investigator or other knowledgeable person should be involved in writing or checking the DIF.

3. The development team should develop a plan for the maintenance of the DIF's. This plan should include a mechanism for updating DIF's that are no longer current and for removing DIF's for which data is no longer available.

4. In the prototype of the MD the user interface was developed by using the Transportable Applications Executive (TAE). Several of the committee members with experience with TAE seriously questioned the whether TAE is the appropriate system to use in this application.

NASA National Aeronautics and Space Administration	**Report Documentation Page**	

1. Report No. NASA TM-100701	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Report on the Second Catalog Interoperability Workshop January 12-14, 1988	5. Report Date April 1988
	6. Performing Organization Code 633.0

7. Author(s) James R. Thieman and Mary E. James	8. Performing Organization Report No. 88B0182
	10. Work Unit No.

9. Performing Organization Name and Address Goddard Space Flight Center Greenbelt, Maryland 20771	11. Contract or Grant No.
	13. Type of Report and Period Covered Technical Memorandum

12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546-0001	
	14. Sponsoring Agency Code

15. Supplementary Notes

Mary E. James is employed by Science Applications Research Corp.; Lanham, Maryland.

16. Abstract

This paper details the events, resolutions, and recommendations of the Second Catalog Interoperability Workshop which was held at the Jet Propulsion Laboratory in January, 1988. This workshop dealt with the issues of standardization and communication among directories, catalogs, and inventories in the earth and space science data management environment. The Directory Interchange Format, which is being constructed as a standard for the exchange of directory information among participating data systems was discussed. Involvement in the Interoperability effort by NASA, NOAA, USGS, and NSF was described, and plans for future interoperability were considered. The NASA Master Directory prototype was presented and critiqued and options for additional capabilities were debated.

17. Key Words (Suggested by Author(s)) catalog, directory, inventory, interoperability, standards interchange, master directory	18. Distribution Statement Unclassified – Unlimited Subject Category 62

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages	22. Price

National Aeronautics and
Space Administration

Washington, D.C.
20546

Official Business
Penalty for Private Use, $300

NASA