EOSDIS

DADS

REQUIREMENTS

J. BERBERT
B. KOBLER

NSSDC Conference on Mass Storage Systems
NASA/Goddard
July 23-25, 1991

EOSDIS DADS REQUIREMENTS

by J. Berbert and B. Kobler

ABSTRACT

A brief summary of the EOSDIS Core System (ECS) DADS requirements is given, including the ECS relationship to EOSDIS Version-0, phased implementation of ECS, and data ingest, archive, and distribution daily data volumes anticipated at each of the 7 Distributed Active Archive Centers (DAACs).

EOS GOALS

The Earth Observing System Data Information System (EOSDIS) Data Archive and Distribution System (DADS) is part of the Earth Observing System (EOS) program. The EOS program goals are given in Fig. 1. In short the goals are to acquire, access, and analyze Earth Science data as NASA's contribution to the Global Change Research Program.

PHASED IMPLEMENTATION

The full capability of the EOSDIS DADS is built up in a series of steps, as indicated in Fig. 2. Version-0 (V0) implementation began in 1990 with an estimated data volume of 5 Terabytes (TB) and is expected to grow to 33 TB by 1994.

Version-1 (V1) and Version-2 (V2) are part of the separately funded EOSDIS Core System (ECS). A request for proposals (RFP) for the 10 year contract to build the ECS was released by the Government on July 1, 1991. During V1 implementation the ECS archive is expected to grow from 10 TB to 40 TB, and the number of active DADS is expected to grow from 1 to 7. the 3 DADS at GSFC, Langley, and Marshal are to be operational for the Tropical Rainfall Measurement Mission (TRMM), which is scheduled for launch in 1997. The V2 implementation of ECS is primarily to support EOS-A1 with its order of magnitude increase in data products the first year and a subsequent increase of about 330 TB, or one third Petabyte (PB), per year, thereafter.

EOSDIS V0

The contribution anticipated from V0 and the specific relationship of V0 to ECS are given in Fig. 3. It is anticipated that V0 will provide significant heritage to ECS through prototyping efforts and by working towards interoperability amongst existing data systems.

ECS SEGMENT AND DAACS

A logical system architecture for ECS is shown in Fig. 4 (taken from the ECS RFP Statement of Work (SOW)). The 3 ECS segments shown are the Flight Operations Segment (FOS), the Communications and System Management Segment (CSMS), and the Science Data Processing Segment (SDPS). The SDPS includes the Distributed Active Archive Centers (DAACs) and the Information Management System (IMS). A DAAC includes a Product Generation System (PGS) with a collocated DADS and a distributed part of the IMS.

DAAC LOCATIONS

The locations of the 7 DAACs are shown on the Fig. 5 map. They are at Goddard Space Flight Center (GSFC), Jet Propulsion Laboratory (JPL), EROS Data Center (EDC), Langley Research Center (LaRC), National Snow and Ice Data Center (NSIDC), Alaska SAR Facility (ASF), and Marshall Flight Center (MSFC).

DADS FUNCTIONS AND REQUIREMENTS

The 5 major DADS functions, namely Ingest, Archive, Process Orders, Manage System, and Distribution, are given in more detail in Fig. 6, along with some of the key performance requirements for ingest and distribution. A key performance requirement for ingest is to be capable of accepting Level-0 (L0) data from the Customer Data Operations System (CDOS) at a high data rate. Key performance requirements for Distribution are to provide data products ready for network distribution within an average of 5 minutes of receipt of product order, and ready for physical media distribution within 24 hours of receipt of product order. Also, the capabilities for both network and media daily distribution rate must be equivalent to daily ingest rate.

DADS INTERFACE

Fig. 7 is the Conceptual DADS Context Diagram taken form the RFP Requirements Specification. This illustrates the multitude of data exchange interfaces for DADS data ingest and distribution. Some entities on this diagram not previously identified are the Affiliated Data Centers (ADCs), Other Data Centers (ODCs), Earth Probe Data Systems (EPDSs), Science Computing Facilities (SCFs), and International Partners (IPs).

DATA VOLUMES PER LEVEL

In Fig. 8, the total DADS daily data volumes, for the platforms EOS-A1, TRMM, and SAR are given for the data processing levels, L0, L1A, L1B, L2, L3, and L4. SAR is the EOS Synthetic Aperture Radar (SAR) platform, which is a separately funded option on the ECS contract. These daily data volumes are taken from the ECS Requirements Specification, Appendix C. As can be seen from this figure, the amount of data to be ingested, archived, managed, and distributed expands significantly from the amount of L0 data received from CDOS. For this set of platforms, the expansion factor is 3.6.

The total daily data contribution from TRMM is 18 GB/day, or 6.6 TB/year, which is small compared to EOS-A1, but large enough to fill 6 StorageTek Near-Line Library Units (Silos) per year, each Silo containing 6000 3480 type cartridges. Moreover, the total daily data contribution form EOS-A1 is 895 GB/day, or about 50 times the TRMM contribution. SAR adds 591 GB/day, or about 33 times the TRMM contribution.

DATA VOLUME PER DAAC

In Fig. 9, the total DADS daily data volumes, for the same 3 platforms, are given for each of the 7 DAACs. The 5 DAACs at JPL, LaRC, NSIDC, ASF, and MSFC vary in size from 3.5 to 8.3 GB/day, which is equivalent to 1.3 to 3.0 TB/year, or 20 to 45 TB in the 15 year EOS data collection period. Thus, these 5 DAACs could be called Tera-DAACs.

The other 2 DAACs at GSFC and EDC are roughly 2 orders of magnitude larger, and with EOS-A alone, each grows to a size of 2 to 3 PB over the 15 year EOS-A lifetime, thereby qualifying as Peta-DAACs.

It should be noted that the data volumes given in Figs. 8 and 9 do not include additional data volume required due to backup of hard-to-replace data products and due to reprocessing of selected data sets. However, this is partially offset by the fact that CDOS provides the disaster backup for L0 data, so that it is necessary for the DAACs to archive L0 data for only a year.

MEDIA REQUIREMENTS AT GSFC FOR EOS-A1

In Fig. 10, the daily data volume of 489 GB/day at the GSFC DAAC for the EOS-A1 platform is converted into a daily media requirement for several types of physical

media, ignoring utilization efficiency. For 3480 type cartridges containing 200 MB of data, this translates into 2445 cartridges per day, enough to fill a 6000 cartridge StorageTek silo every 2.5 days. With the newer 3490 cartridges, having double the data density, it takes 5 days to fill the silo. The D1 and D2 tape technologies reduce the daily number of cartridges required by about 2 orders of magnitude. It is anticipated that technological progress toward higher density data recording will continue over the next 7 years prior to EOS-A1 launch, resulting in a physically smaller and more manageable archive at that time than would be possible with current technology.

TRANSFER RATES

A potential bottleneck for timely EOSDIS DADS operations, is in the available data transfer rates for read/write devices compatible with available storage media. Data transfer rates available for drives compatible with the types of media considered in Fig. 10 are given in Fig. 11. With a transfer rate of 3.0 MBps, as is available for 3490 type cartridges, a single image file of 327 MB requires 109 seconds to physical read, again ignoring efficiency factors. Technological progress toward faster data transfer rates may be achieved prior to EOS-A1, but progress in this area has not been as rapid as in the area of higher density data recording.

EOS (EARTH OBSERVING SYSTEM) GOALS ARE TO DEVELOP AND OPERATE:

a) An observing system to acquire essential, global Earth science data on a long-term, sustained basis and in a manner which maximizes the scientific utility of the data and simplifies data analysis.

b) A comprehensive data and information system to provide the Earth science research community with easy, affordable, and reliable access to the full suite of Earth science data from EOS and international partner observatories, NASA Earth Probes, and selected Earth science data from other sources.

c) As the cornerstone of the Mission to Planet Earth Global Change Research Program, an integrated scientific research program to investigate processes in the Earth System and improve predictive models.

i.e. TO ACQUIRE, ACCESS, ANALYZE EARTH SCIENCE DATA

Fig 1

PHASED IMPLEMENTATION

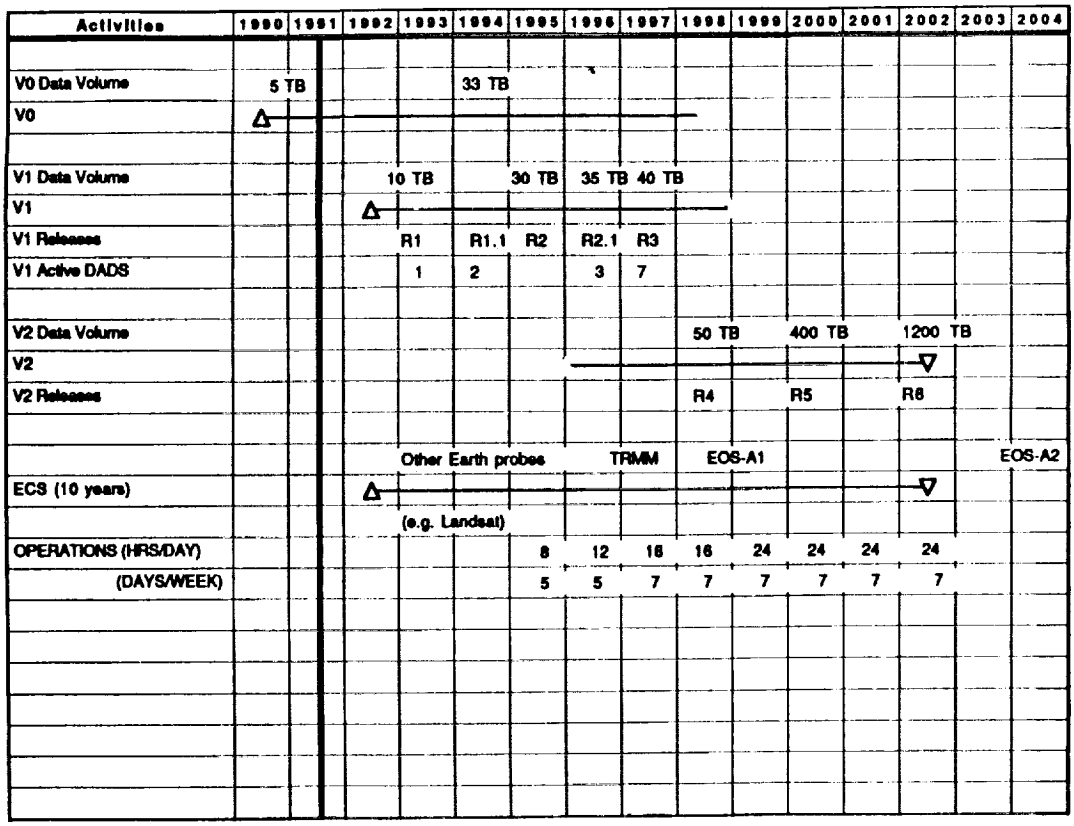

Fig 2

EOSDIS V0

V0 TO PROVIDE:

- **Interconnection of existing data systems at DAACs**
- **Prototyping of selected tasks in distributed IMS, networking, standards**
- **Some additional Earth Observation data sets to be added to the existing Data Systems under V0**

V0 RELATIONSHIP TO ECS

- **Provides early experience/information/results for potential inclusion in ECS design**
- **ECS contractor to connect ECS to V0 and provide a level of interoperability**
- **Selected data sets from V0 to be copied for inclusion into ECS**

Fig 3

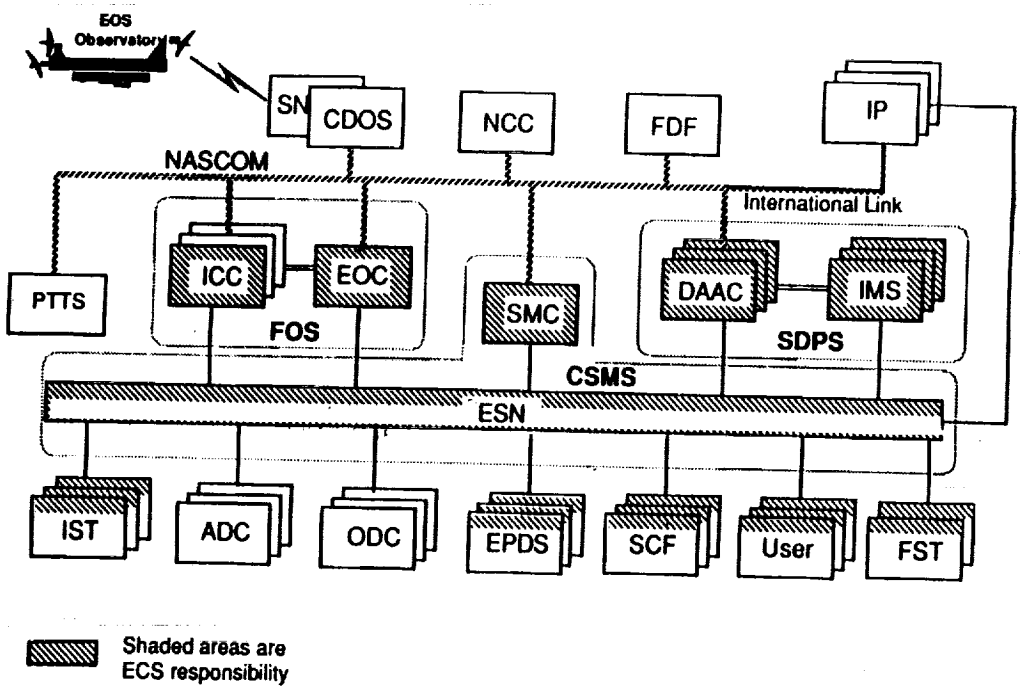

Figure 1.4.1 - 1. ECS Logical System Architecture

Fig 4

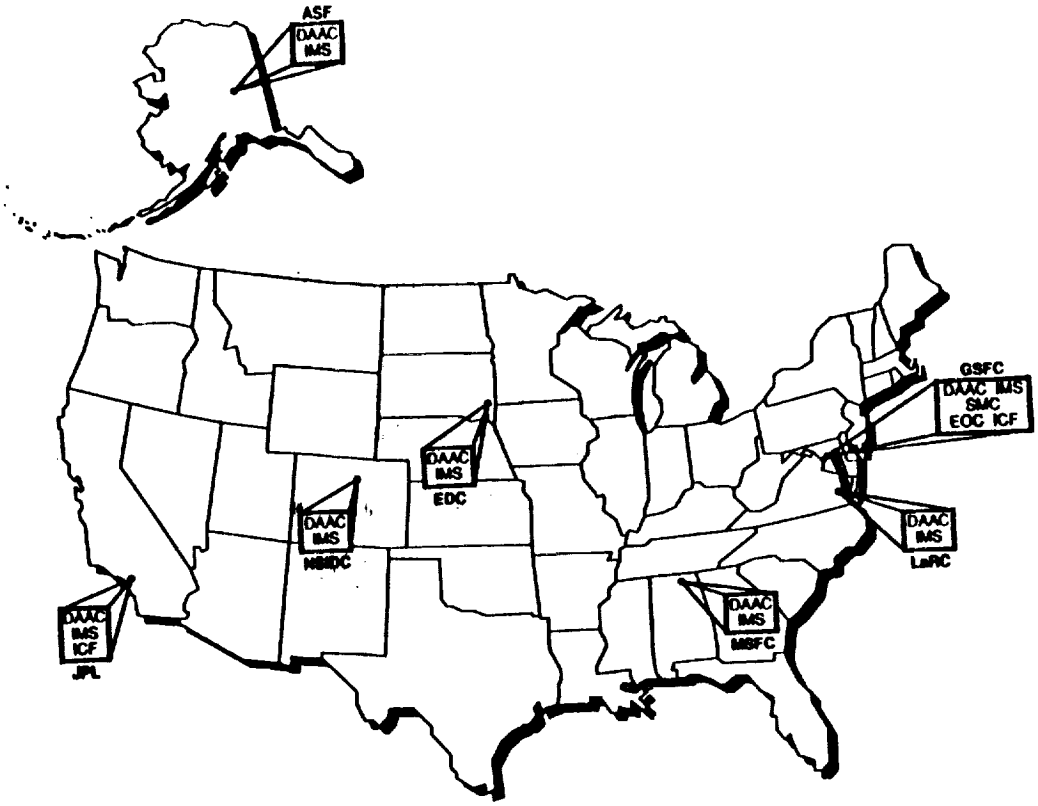

Figure 7.2.2-1 SDPS Physical Architecture

Fig 5

DADS FUNCTIONS

INGEST- Receive/Validate data products and data from CDOS , PGS, SCFs, other DAACs, ADCs, ODCs, EPDSs, IPs, Users, and others

ARCHIVE- Store data and data products on archive media

PROCESS ORDERS - Fulfill product orders provided by IMS, Retrieve data from archive, subset, reformat, stage for delivery. Support reprocessing

MANAGE SYSTEM - Monitor and report status and accounting information to SMC, operate File Storage Management System with hierarchical archive, schedule operations according to SMC directives, monitor media BER (10-12) and provide for data restoration/migration, backup selected data**

DISTRIBUTION - Distribute data and data products to PGS, SCFs, other DAACs, ADCs, ODCs, EPDSs, IPs, Users, and others via networks (5 minutes) and by Physical media (24 hours). Provide daily distribution rate capability equivalent to daily ingest rate

Fig 6

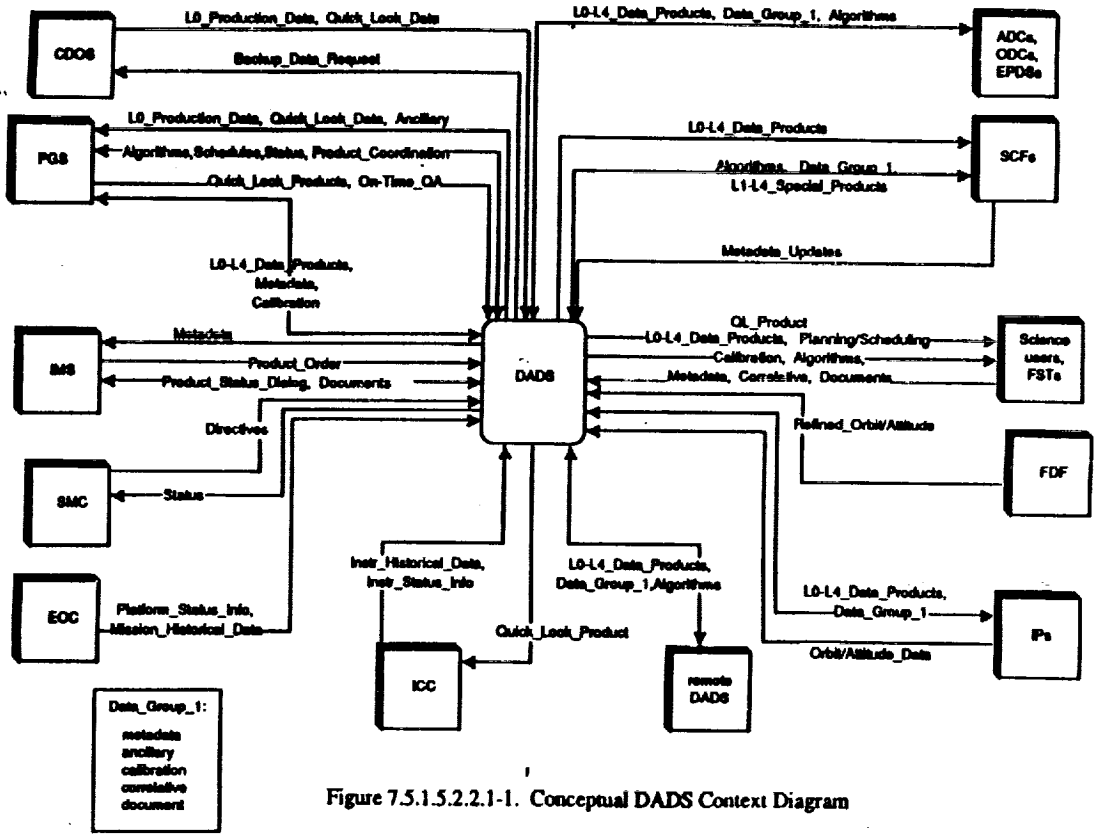

Figure 7.5.1.5.2.2.1-1. Conceptual DADS Context Diagram

Fig 7

DADS Daily Data Volume by Location (Log Scale)

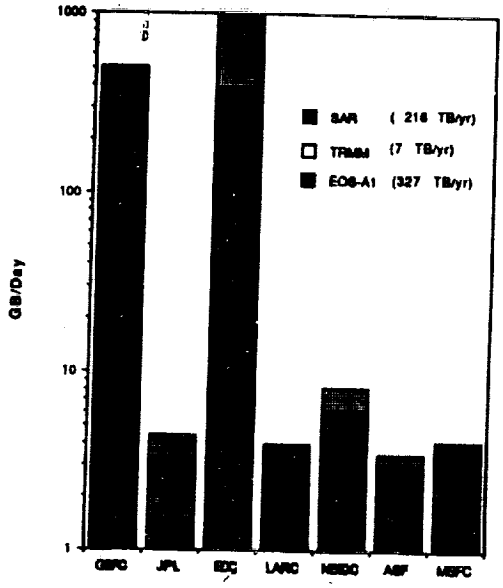

Fig 8

DADS Daily Data Volume by Location

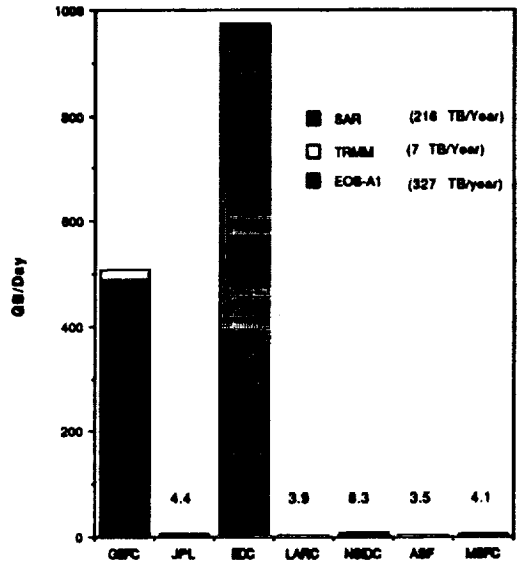

Fig 9

GSFC (EOS-A1) Daily Media Requirements

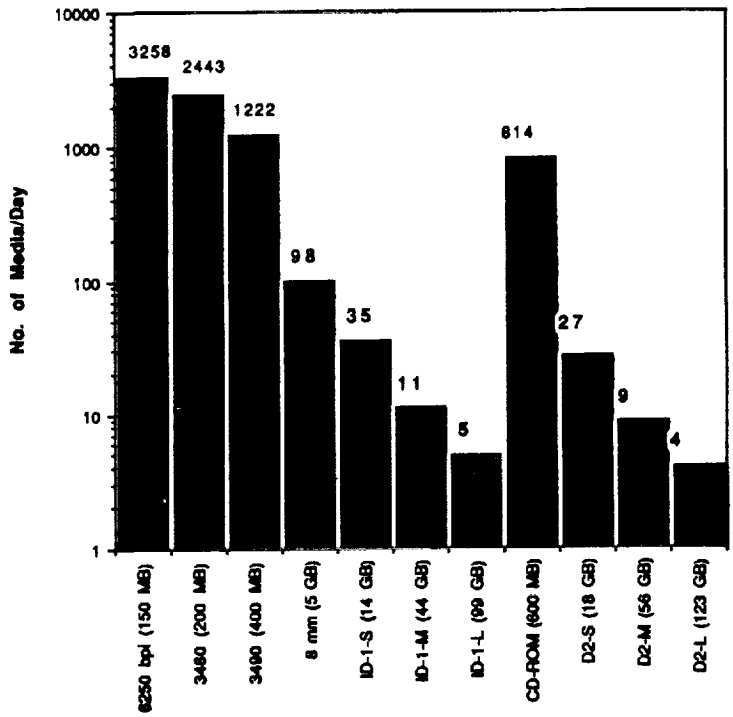

Fig 10

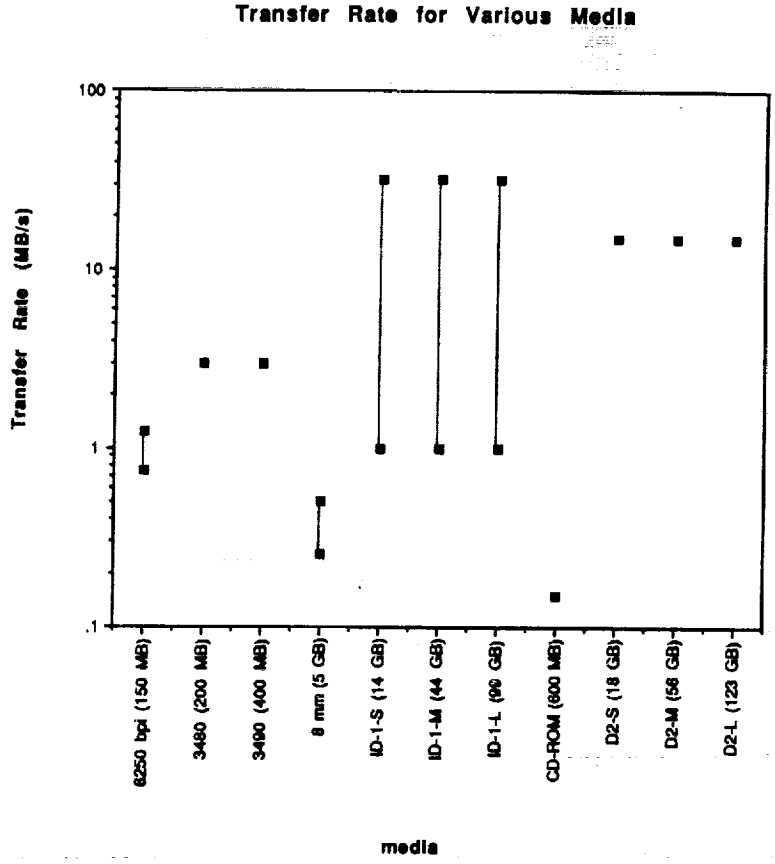
Transfer Rate for Various Media
Transfer Rate (MB/s)
100
10
1
.1
6250 bpi (150 MB)
3480 (200 MB)
3490 (400 MB)
8 mm (5 GB)
ID-1-S (14 GB)
ID-1-M (44 GB)
ID-1-L (96 GB)
CD-ROM (600 MB)
D2-S (18 GB)
D2-M (56 GB)
D2-L (123 GB)
media

Fig 11