

Title: Preservation of Digital Data from High Volume Data Repositories

Authors: Cynthia Hall Atkinson, Jet Propulsion Laboratory / California Institute of Technology and Christopher E. Isbell, United States Geological Survey

Abstract:

Preservation, archive, and distribution of data products are essential requirements within the functions of the Planetary Data System (PDS) as well as any other data archive organization. Traditionally within the PDS, data products are and have primarily been distributed via physical portable media such as Compact Disc (CD) or Digital Versatile Disc (DVD). However, due to increasing data volumes, data products are also now housed on local hard drive disks or disk farms and distributed through on-line Internet services. Developing and future technologies will impact current and future requirements for data distribution. Although on-line data distribution services are expected to be the primary method for product distribution to the science community, interim and long-term hard-media archives will be desired for most if not all distributed datasets.

In developing long-term hard-media archives for future Earth and Space Science missions, several issues must be addressed. Increased data returns result in very large datasets; for these data sets, current physical recording media are impractical, as storage requirements greatly exceed current recording capacities. Hardware, software, computer platform, media, and end-user considerations impact the design and format of data products intended for wide distribution; the technical and evolving nature of these and related factors require utilization of the latest and appropriate technologies in order to meet archive and distribution requirements in serving all data users. The ability to access data products across multiple computer platforms and varying operating systems is essential, true for both on-line and hard-media distribution.

Issues remain for current and near-term datasets intended for hard media archive. Continued evaluation of factors such as industry and file system formats, media quality and longevity, recording hardware and software, and media storage requirements is essential.

The PDS, in a collaborative effort with other government agencies and industry, has begun to research many of the evolving archive / distribution technologies and standards in an attempt to address many of the factors mentioned above and provide a viable solution to the development of a long-term hard-media archive for future Earth and Space Science missions.

Key areas:

1. Online Distribution of data

- provides seamless access to archive data from heterogeneous, geographically disperse data repositories and systems
- made available immediately upon delivery to the PDS
- eliminates immediate need of distribution of bulky CD/DVD packages
- BUT while online provides an efficient means of data distribution, it does not satisfy issues related to long-term preservation and catastrophic recovery

2. Preservation

- Increased Data returns - current media impractical
 - example of a mission with large data returns like Odyssey or MRO, X GBs of data returned...
 - current media capacity ranges from 4.3 GBs with DVD to 200 GBs with Hard drive
 - include chart with all media and capacity
 - current media types are
 - Magnetic tape and Hard drive
 - Environment - small fluctuations in temp and humidity affect

- Electromagnetic Fields - EMF sources, intensities and locations unpredictable
 - Tapes must be exercised several times before recording data and periodically during shelf life cycle
 - Hard drives must be prepared through low-level formatting
 - Tapes shelf-life is ~10-30 years
 - Hard drive shelf life is ~20-40 years
 - DVD
 - Variety of formats to include DVD-ROM, DVD-R, DVD+R, DVD-RAM, DVD-RW, DVD+RW
 - Rewritable options (RAM, -RW, and +RW) are not viable solutions for Utilize Universal Disc Format (UDF) file system standard
 - DVD-ROM has multiple layer capability but must be mastered at off-site vendor facility
 - DVD-R and DVD+R in widespread use due to low costs and in house writing capabilities; therefore provide best solutions (Note: PDS has not researched DVD+R option and as such will only discuss DVD-R)
 - * DVD-R consists of two formats, General and Authoring
 - * Authoring possibly phasing out
 - * General technologically supported by entire DVD industry
 - HD-DVD
 - Variety of contenders include Blu-Ray, Next Generation, HD-DVD, Blue-HD-DVD-1, Blue-HD-DVD-2, Red Laser HD-DVD-9 to name a few (maybe include table with varieties, attributes). Various factors will determine dominant format for eventual archive usage.
 - Media and Recorders only available in some markets so very limited use
- Hardware, Software, computer platform and media considerations (Note: Focus is on DVD-R as this is most widely used archive option for PDS at this time)
 - Research and evaluation of hardware and software is key for preparation and distribution of digital data as these factors impact the design and format, and are constantly evolving with advances in technology.
 - Understanding hardware and software configurations ensures the maximum (although not ultimate) compatibility across multiple computer platforms and varying operating systems
 - The PDS is confident properly formatted media will mount and be readable under most platforms and hardware configurations
 - Media readability can be affected by complex scenarios involving (but not limited to) hardware and/or operating system patches, version incompatibilities, or perhaps media parity errors
 - may not resolve all problems within a moving window "current requirements" time frame but can recognize the limitations
 - Industry and file system formats
 - The industry accepted standard for file systems on a DVD is UDF.
 - To ensure compatibility across numerous platforms it is recommended that the ISO 9660 file system also be included (as this has been the accepted standard for years (with CD production).
 - ISO Level 1 is used for "8.3" naming convention datasets

