Evaluation of DVD-R for Archival Applications

Michael D. Martin
Jet Propulsion Laboratory
Mail Stop 171-264
4800 Oak Grove Drive
Pasadena, Ca. 91109
818-354-9342
Fax: 1-818-393-3405
E-Mail: mmartin@jpl.nasa.gov

Jason J. Hyon
Jet Propulsion Laboratory
Mail Stop 171-264
4800 Oak Grove Drive
Pasadena, Ca. 91109
818-354-0730
Fax: 1-818-393-3405
E-Mail:

This paper presents the results of DVD-R hardware, software and media evaluations and describes the production of a sample archive collection using DVD-R technology.

Evaluation of DVD-R for Archival Applications

Michael D. Martin
Jet Propulsion Laboratory
Mail Stop 171-264
4800 Oak Grove Drive
Pasadena, Ca. 91109
818-354-9342
Fax: 1-818-393-3405
E-Mail: mmartin@jpl.nasa.gov

Jason J. Hyon
Jet Propulsion Laboratory
Mail Stop 171-264
4800 Oak Grove Drive
Pasadena, Ca. 91109
818-354-0730
Fax: 1-818-393-3405
E-Mail: jhyon@pop.jpl.nasa.gov

For more than a decade, CD-ROM and CD-R have provided an unprecedented level of reliability, low cost and cross-platform compatibility to support federal data archiving and distribution efforts. However, it should be remembered that years of effort were required to achieve the standardization that has supported the growth of the CD industry. Incompatibilities in the interpretation of the ISO-9660 standard on different operating systems had to be dealt with, and the imprecise specifications in the Orange Book Part II and Part III led to incompatibilities between CD-R media and CD-R recorders. Some of these issues were presented by the authors at Optical Data Storage '95. The major current problem with the use of CD technology is the growing volume of digital data that needs to be stored. CD-ROM collections of hundreds of volumes and CD-R collections of several thousand volumes are becoming almost too cumbersome to be useful.

The emergence of DVD-R technology promises to reduce the number of discs required for archive applications by a factor of seven while providing improved reliability. It is important to identify problem areas for DVD-R media and provide guidelines to manufacturers, file system developers and users in order to provide reliable data storage and interchange. The Data Distribution Laboratory (DDL) at NASA's Jet Propulsion Laboratory began its evaluation of DVD-R technology in early 1998. The initial plan was to obtain a DVD-Recorder for preliminary testing, deploy reader hardware to user sites for compatibility testing, evaluate the quality and longevity of DVD-R media and develop proof-of-concept archive collections to test the reliability and usability of DVD-R media and jukebox hardware.

DVD-Recorder Evaluation.

The DDL purchased the Pioneer DVD-S101 in March 1998. The unit came with the Prassi Software, Inc. "DVD Rep™" pre-mastering and recording software for the Windows 95 and Windows NT. This software produces UDF Bridge format discs (UDF and ISO-9660 compatible). During the evaluation

phase the software was running on a Pentium 90 class machine with 32 megabytes of RAM. Pre-mastering a single volume could take as much as 12 hours under this configuration, largely due to the algorithm used for doing ISO-9660 8.3 file name generation for volumes with a large number of files. In January 1999 the pre-mastering hardware was upgraded to a 450 megahertz Pentium with 256 megabytes of RAM and the pre-mastering time dropped to less than 30 minutes per volume.

Recording operations were unreliable under early versions of the Pioneer firmware. Disc recording errors were encountered frequently at the beginning of the recording operation, rendering the DVD-R media unrecordable. The recorder was returned to Pioneer for "calibration", but that procedure did not eliminate recording errors. Only with the acquisition of Pioneer firmware version 1.07 did recording operations stabilize.

DVD Reader Deployment.

At the time we began testing the DVD-Recorder the only DVD-ROM readers available were IDE interface readers for PC machines. Unfortunately, only about one-third of our user community relies on PCs, the rest using a roughly equal mix of Sun workstations and Macintosh hardware. In the fall of 1998 the first SCSI interface DVD-ROM reader became available, the Pioneer DVD-302. About a dozen readers were purchased and deployed to six geographically distributed sites that make up NASA's Planetary Data System. In order to use the readers on these systems the operating systems of the host computers had to be upgraded to Solaris 2.6, Macintosh System 8.1 and Windows 95, NT 4.0 or 98. The installations were accomplished smoothly and each site was provided with customized DVD-R discs to allow them to test the hardware with their own data sets.

A Pioneer DRM-1004V40 DVD jukebox and iXOS Software AG "JUKEMAN" software was procured for the proof-of-concept archiving applications. This combination of hardware and software was installed on a Sun workstation and has proven to be reliable and easy to use.

DVD-R Media Evaluations

We have received and recorded media samples from several vendors. These samples were tested by Mark Worthington of CD Associates, Inc using the DVD100MG tester. According to Mark's evaluation, DVD-R media measurements do not differ significantly from DVD-ROM measurements and DVD-R media is expected to perform equally well in DVD-ROM drives. The best indicator of problems is a high parity inner block error correction rate. The first samples that were tested provided excellent results and convinced us to go ahead with an archive prototype.

More recently, eleven discs have been tested. Only three of these discs pass on all measurements. Seven discs exhibit block errors above the maximum allowed as well as other error types. Three discs fail on minimum asymmetry and three more have multiple failures, probably caused by manufacturing defects on the media. While none of these defects causes read-errors, we need to work with the media vendors to understand what is causing this high incidence of problems. It is clear that much more media testing needs to be done before we can feel confidant about the quality of the DVD-R media currently on the market. We also plan to provide these media samples to the National Media Lab for longevity testing.

DVD-R Archive Production

In order to evaluate DVD-R media for archival applications it was decided to develop two proof-of-concept test cases. The 400-gigabyte SkyMorph database has been generated by the Near Earth Asteroid Tracking (NEAT) program. This program has collected tens of thousands of images in visible wavelength with substantial sky coverage and temporal sampling of 15 minutes, days, months and years in a search for near Earth objects. The second test case is the production of an on-line copy of 850 CD-ROM discs generated by the Planetary Data system over the past decade. This will provide on-line access to nearly a complete collection of digital data from NASA's Planetary Exploration Program and allow experimental data mining techniques to be applied. These two test cases will allow us to evaluate a large number of discs in archival applications which will also benefit the NASA science community.

The SkyMorph collection is stored on two dozen 36 gigabyte hard disk drives. Each hard disk drive contains ten to fifty days worth of observations. The amount of data taken per day varies greatly, from several megabytes to six gigabytes. In order to fill each DVD-R, observing days are split over DVD-R volumes. The contents of a hard disk are downloaded to the pre-mastering system using FTP. The 35 gigabytes transfer takes about one minute to initiate and requires about 3.5 hours to complete. Partitioning the data into DVD-R sized partitions takes about 30 minutes for 9 DVD-R volumes. Pre-mastering each volume takes about one minute to set up and about 30 minutes to complete. Recording takes one minute to set up and about 60 minutes to complete. After recording, each disc is labeled with the volume id and tested quickly in a DVD-ROM reader on the pre-mastering workstation to assure that the contents match the source (requiring about one minute). Thus the total labor required to produce 9 discs is about one hour and the elapsed time about 17 hours.

After nine volumes are recorded, the set of volumes are inserted in the DVD jukebox and mounted, using a web-based interface to the jukebox software. Finally, a perl script is run which creates virtual directories on the server hard disc and then creates symbolic links for every file on the DVD-R discs. This allows users to access the logical volumes independent of what physical volumes they are stored on. During the first six weeks of production, about 60 discs have been recorded with no failures due to hardware or software. It is estimated that the storage media cost for the 400 gigabyte archive will be about $4,300.

Conclusions and Future Plans

Our experiences with DVD-R hardware and media have been both frustrating and rewarding. While this media has incredible potential for archival storage applications, the hardware and media errors we have experienced in our evaluation must be eliminated. We plan to continue recorder, media and reader evaluations as well as cross-platform testing of the Universal Data Format (UDF). We eagerly anticipate the release of new 4.7 gigabyte DVD-Recorders, both to support archiving applications and to produce check discs of DVD-ROM titles under development.

Acknowledgments