

## Accessibility of Data Backup on CD-R after 8 to 11 years

P. M. A. van Ooijen, A. R. Viddeleer, F. Meijer, and M. Oudkerk

### INTRODUCTION

With the more widespread introduction of picture archiving and communication systems (PACS) about a decade ago, many institutions would perform long-term storage on archiving systems using tape which they believed to be a trusted and secure archiving method. However, this was disputed by others who set off to use more novel techniques such as the long-term storage on jukeboxes containing large amounts of compact disks, recordable (CD-Rs). For this storage, a major requirement was a long guaranteed lifetime of the CD-R media used. Different manufacturers claimed to meet these long-term storage requirements with their high-end CD-Rs, provided that optimal storage conditions are met. In this study, we wanted to evaluate whether or not CD-Rs could be a valid long-term storage medium for secure medical image data storage.

### MATERIALS AND METHODS

Although we already migrated to DVD storage years ago, an old CD-R archive with dates of burning the CD-R ranging from February 28, 1996 to November 11, 1999 was still available stored in boxes outside of their jewel cases. The CD-Rs were taken from the jukeboxes early 2000 when migration to DVD storage was completed and were never taken out of the boxes until this study commenced. The old archive consisted of 600 CD-Rs containing about 600 Mb of lossless compressed data per CD-R. All CD-Rs were, although not specifically medically certified, stated to be

dedicated for archival purposes by the vendor (InfoGuard™ protection CD-Recordable from Eastman Kodak Company, Rochester). The official position of Kodak on the lifetime is a guaranteed lifetime for InfoGuard protection system CDs of over 100 years. The infoGuard protection system consists of a very stable dye layer covered with a super-tough durability overcoat to help protect the surface of the media from scratches and fingerprints.

For testing the CD-Rs, a high-quality DVD writer (Plextor 716A, firmware revision 1.11, Plextor Ltd., Milpitas, CA, USA) was used.

Of the 600 CD-Rs, a random sample of 25 CD-Rs were taken to be diagnosed using a dedicated, freeware software tool (CD Tester, Profiler 3D, [www.profiler3d.de](http://www.profiler3d.de)) which tests the readability of the CD-R by opening all data and then checking all binary data, displaying any failures and, if possible, the cause of the failure.

Another random sample of 25 CD-Rs were physically analyzed by determining the low level C1 and C2 error count, as well as unrecoverable errors, using the Plextools Professional LE v3.13 analysis software with a constant read speed of eight times. C1 (in errors per second) indicates the

---

*From the Department of Radiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands.*

*Correspondence to: P. M. A. van Ooijen Department of Radiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ, Groningen, The Netherlands; fax: +31-50-3612286; e-mail: [p.m.a.van.ooijen@rad.umcg.nl](mailto:p.m.a.van.ooijen@rad.umcg.nl)*

*Copyright © 2008 by The Author(s). This article is published with open access at Springerlink.com*

*Online publication 4 November 2008*

*doi: 10.1007/s10278-008-9161-9*

block error rate, which consists of bit errors at the lowest level. C2 provides an indication of the use of error correction by the CD player, and occurrence of C2 errors usually indicates poor media quality or failure of a CD writer to produce a quality burn<sup>1</sup>. According to the help file enclosed with PlexTools, the existence of C1 and C2 errors on a CD-R is perfectly normal. When numbers are low, a good CD-R drive can correct for these errors. However, with higher C1 and C2 error count, the possibility of encountering errors that cannot be corrected increases.

To serve as a benchmark, other old CDs, without special protective coating, were deliberately scratched and exposed to sunlight in order to determine the correct performance of the CD tester software. Furthermore, the proper functioning of these tools was also evaluated by reading other old CD-Rs with known problems.

Finally, all 600 CD-Rs were scanned for DICOM files to retrieve selected data for a scientific study, and a selection of the data was uncompressed and transferred to a hard drive location using dedicated software tools.

## RESULTS

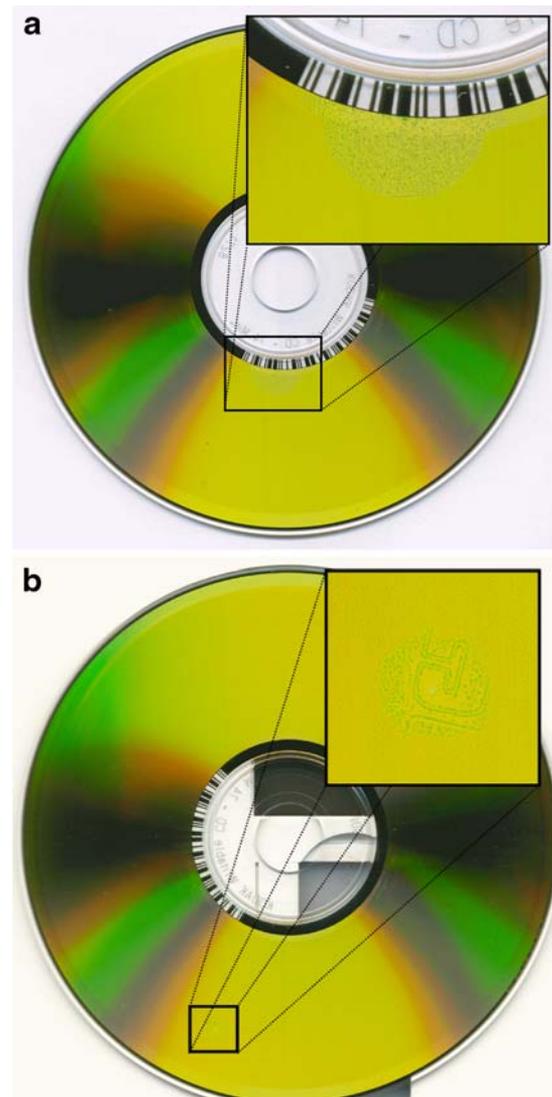
Using test CD-Rs, it was shown that minor scratching or damage on non-data parts did not cause any problems. With significant damage hampering the readability of the CD-R, either the CD-ROM drive failed to read the CD-R altogether or the CD tester software prompted with one or multiple errors on the CD. Furthermore, white label CD-Rs of about 5 years old already presented with reading errors when using CD Tester software.

A random sample of 25 actual archive CD-Rs were tested using CD Tester software with an average of 3,608 files per disk. Search speed was reported to be  $3,393 \text{ kb s}^{-1}$  on average. Mean amount of data stored on the CD-Rs was 564.17 Mb with a total of 89,997 DICOM files and 9.6 Gb. Of those 25 CD-Rs, none reported failures during the technical scanning.

The other random sample of 25 archive CD-Rs that were analyzed had an average age of 9.5 years. These CDs showed a remarkably low mean C1 error count ( $4,807 \pm 2,137$ ) and a mean C2 error count of 24.8. These more severe C2 errors were

present on six CD-Rs of the 25 CD-Rs, while the remaining CDs did not contain any C2 errors. No relation was found between CD age and the number of errors. All errors (C1 and C2) could be corrected by the DVD drive's error correction algorithms.

However, using the DICOM scanning tool on all 600 CD-Rs, nine CD-Rs were found containing



**Fig 1.** a Defects most likely caused by corrosion were found at the inner edge of the CD-R (*left*) in the shape of half a circle and in the middle of the CD-R (*right*) in the shape of a circle. b Defects most likely caused by corrosion were found at the inner edge of the CD-R (*left*) in the shape of half a circle and in the middle of the CD-R (*right*) in the shape of a circle.

severe read errors and could not be read (1.5%), while all DICOM data could be extracted from the remaining CD-Rs without any problems (98.5%).

Visual examination of the damaged CD-Rs revealed that the problems were caused by several problems. Corrosion of the reflective layer was visible in eight CD-Rs (Fig. 1), while only one CD-R was severely scratched at the opposite side, causing it to be unreadable. In four CD-Rs, the corrosion occurred at the exact same location at the inner ring of the CD-R and seemed to originate from the edge spreading out in a half circle, suggesting a defective sealing of the reflective layer (Fig. 1a). In the remaining four CD-Rs, the same corrosion spots were found, but in this case, in the middle of the CD-R and circular in shape with in the middle an apparent defect in the CD-R (Fig. 1b). In five of the eight CD-Rs, the defects appear close to, but not directly under, text written on the CD with a special CD marker.

The unique identification numbers of the different CD-Rs are given in Table 1. Apparently, information about this number is difficult to obtain from the manufacturer, but if we assume that the first two four-digit numbers determine a production batch, than five out of eight (63%) are produced in batch 8271 4051 and two out of eight in batch 8302 4201 (25%). The fact that all CD-Rs with defects other than severe scratching of which we know the burn date are burned during the period of February to July 1999 supports the suspicion that the failures were limited to certain batches of CD-Rs.

When scanning the nine CD-Rs with visible defects to determine the readability of the files on the CD-Rs, the overall percentage of readable files on the CD-Rs with problems is 62.0%. Percen-

tages per CD-R are given in Table 1. When assuming an average of 3,527 files per CD-R (based on the results in Table 1), in total, 12,071 out of the approximated 2,116,333 images (0.6%) of the total archive were lost due to the described problems.

## DISCUSSION

The main shortcoming of this study is the fact that the CD-Rs we examined were not stored under the optimal conditions as dictated by the vendor. However, we expect our procedure to be a very realistic way of storing old archive media in boxes on a shelf in the computer room in many hospitals<sup>2</sup>, and defects are reported to occur both in media still inside the jukebox and in media stored on the shelves<sup>2</sup>. Furthermore, scratching most probably caused by manual handling of the CD-R was the main reason for the failure in only one of the CD-Rs.

The defects in the eight CD-Rs could have been apparent already when they were first placed into the jukeboxes. However, no visual inspection of every CD-R was performed at that time. Perhaps, visual inspection of every CD-R or DVD-R added into a backup system could detect possible defective media in an early stage and prevent them from being used. However, without specific training of the staff responsible for inserting the CD-R or DVD-R into the system, it is likely that most defects will be missed during visual inspection. Furthermore, besides the difficulty of having people trained to perform the task, the visual inspection of hundreds of CD-Rs or DVD-Rs will be extremely time-consuming. Therefore, when

**Table 1. Properties of the Nine CD-Rs with Reading Problems**

CD no.	Defect location	Unique identification number	Readable files/total no. of files (%)	Date written
1	Edge	8271 4051 1994	0/3527 (0.0%)	Unknown
2	Edge	8271 4051 1978	0/3527 (0.0%)	Unknown
3	Inner Edge	8271 4051 1996	0/3527 (0.0%)	Unknown
4	Inner Edge	8302 4201 1188	3165/3340 (94.8%)	May 1999
5	Middle	8271 4051 1973	3427/3433 (99.8%)	February 1999
6	Middle	8271 4051 1898	2741/3025 (90.6%)	March 1999
7	Middle	9051 4021 1149	5179/5226 (99.1%)	July 1999
8	Middle	8302 4201 1167	2415/2874 (84.0%)	March 1999
9	Severe scratching	6083 1061 0347	2747/3266 files (84.1%)	January 1997

purchasing CD-Rs or DVD-Rs for backup reasons, the highest quality should be regarded as the main issue, not the lowest costs.

Although the defects often appear near the text written on the CD-R, the origin of the defects never appears to be close to the actual text, which makes it unlikely that the marker caused the defects that started the corrosion. Besides this, the defects are also close to the text and lines imprinted by the manufacturer on the front of the CD-R.

Furthermore, on their own website, the manufacturer states that they have never seen a disk fail because of the use of a felt-tip pen or marker, although they do advise to write in the clear area near the hub (center hole), away from any coated layers or recorded information when possible.

Simulated results of Stinson et al.<sup>3</sup> show that 95% of properly recorded discs stored at the recommended dark storage condition (25°C, 40% RH) will have a lifetime of greater than 217 years. Although only 1.5% of the disks in our study failed, this does not imply yet that the performance claimed by Stinson et al.<sup>3</sup> can be met in a real-life situation.

The findings also support the concept of migrating from old archive media to newer media when available and not keeping the old archive running next to the new one. Although the migration initially will have higher costs both because of the time investment and the higher storage demand on the new system, the availability of the image data on the long-term archive will be more reliable in the future. The old media could be stored on a shelf for emergency recovery when needed, but could also be discarded and destroyed. Reported experience from other institutions show that migration can be a very time-consuming and costly process<sup>2,4,5</sup>, and preferably, automated tools should be developed to perform the migration. However, although costs are high, the fact that reported data loss is high could justify these costs. Maass et al.<sup>2</sup> reported a data loss of 35% when transferring 16- to 13-year-old magneto-optical disk (MOD) data to a new digital linear tape (DLT) system. Jung et al.<sup>5</sup> reported a loss of 20% of only 3- to 4-year-old DLT image data from when migrating to a new system. No publications could be found that evaluated the performance of CD-R backup. However, our study seems to indicate that backup on CD-R is more secure than

backup on MOD or DLT, since we only found a loss of 0.6% of the data on CD-R.

Besides migration when replacing the storage media, it could also be advantageous to copy all data on CD or DVD every 5 years (as compared to every 2 years for tape archives) as proposed by Wirth et al.<sup>6</sup>. They state that CD and DVD have not proven their claim for long-term storage yet and that it is also required to check the complete archive every now and then to check for problems and to find and possibly correct defective media. This statement is supported by our findings.

## CONCLUSION

Claims of vendors of storage media about the lifetime are very high. However, one bad CD-R could already cause serious problems in medical care when the data contained on that CD-R are needed for patient follow-up. Our study shows that in the current PACS installations of 10–15 years old, problems with the CD-R archives can already occur even when all data are stored on guaranteed media.

Tests have shown that CD-Rs can be damaged already after 4–5 years, so storage on portable media such as CD-R is not as safe as suggested by the manufacturer. Even when using more expensive, guaranteed media, several discs became (partly) unreadable after 10 years. Using CD marker pens may speed up this process, but a direct relation between the writing location and the offset of the defects was not found.

In conclusion, archival on CD-R is a reliable method provided that migration to new media is performed regularly and strict storage conditions are met.

## OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## REFERENCES

1. Speer Bob, Explaining Common CD Errors. CD Mastering Services. <http://www.cdmasteringservices.com/digitalerrors.htm>. Accessed August 11, 2008.

2. Maass M, Kosonen M, Kormano M: Radiological image data migration: Practical experience and comparison of the costs of work. *Acta Radiol* 42(4):426–429, 2001
3. Stinson D, Ameli F, Zaino N. Lifetime of KODAK writable CT and photo CD media. Eastman Kodak Company 1995. <http://www.cd-info.com/CDIC/Technology/CD-R/Media/Kodak.html>. Accessed January 15, 2008.
4. Behlen FM, Sayre RE, Weldy JB, Michael JS: “Permanent” records: Experience with data migration in radiology information system and picture archiving and communication system replacement. *J Digit Imaging* 13(2 suppl.1):171–174, 2000
5. Jung H, Kim H-J, Kang W-S, Lee S, Kim S-R, Chang LJ, Kim J-H, Yoo SK, Kim K-H: Migration of medical image data archived using mini-PACS to full-PACS. *J Digit Imaging* 17(2):100–108, 2004
6. Wirth S, Treitl M, Villain S, Lucke A, Nissen-Meyer S, Mittermaier I, Pfeifer K-J, Reiser M: PACS: storage and retrieval of digital radiological image data. *Radiologe* 45:690–697, 2005 (German)