

BiSON

Birmingham
Solar-Oscillations
Network

TECHNICAL REPORT NO. 231

The BiSON CD-ROM Data Store

John Allison, William J. Chaplin, Yvonne Elsworth, Steven Hale, and
Brek A. Miller

The University of Birmingham, Edgbaston, Birmingham B15 2TT

2004 May 11

This technical report series is published by:



**THE UNIVERSITY
OF BIRMINGHAM**

High-Resolution Optical-Spectroscopy Group

School of Physics and Astronomy
The University of Birmingham
Edgbaston, Birmingham B15 2TT, United Kingdom
Telephone: +44-121-414-4551 FAX: +44-121-414-1438

The BiSON CD-ROM Data Store

John Allison, William J. Chaplin, Yvonne Elsworth, Steven Hale, and
Brek A. Miller

The University of Birmingham, Edgbaston, Birmingham B15 2TT

2004 May 11

Abstract

The organization of the BiSON CD-ROM Data Store is described. The procedures for adding data to and maintaining the Complete DAT Archive are prescribed.

1 Introduction

For many years the High-Resolution, Optical-Spectroscopy Group (HiROS) relied on Panasonic optical disk drives and optical disk cartridges for its mass storage requirements. Many other groups that require space to store large amounts of data choose Exabyte or other tape devices. In our case, the random-access capabilities of the optical-disk technology were what motivated us in our choice. The disadvantages associated with this technology are smaller storage capabilities and extremely high cost.

Long ago we found that it was becoming increasingly difficult to locate the cartridges required by the older drives. HiROS was forced to make a decision to move away from the Panasonic optical disk drives to more modern technology. The next choice was Panasonic PD disks. It was not long before these too were consigned to history.

The old Panasonic optical disks are now kept [1] in storage.

Presently, we have turned to CD-Rs for our data storage requirements. These can be read using ordinary CD-ROM drives found in nearly every PC manufactured today. We expect that CD-ROM reading ability will be present in new computers for some time to come.

Table 1: Five Archives

	<i>Complete DAT</i>	<i>Useful DAT</i>	<i><2003 RES</i>	<i>New RES</i>	<i>Useful RES</i>
<i>Collation</i>	station year	year	station year	station year	year
<i>Contents</i>	DAT RAW LOG CMP CTL CON Inf DATA /doc /bin <i>ReadMe.Txt</i> <i>cdrom.md5</i>	CMP CTL	RES <2003 RESID	RES	new-name RES
<i>Properties</i>	verifiable		sealed		

2 Five Archives

The CD-ROM store is divided into five archives as shown in Table 1.

The collation type for each archive is shown in the second row of the table above. “station/year” means that the data are first collated by station, then within each station the data are further divided up by year. A year of data from one station will not be split up between discs. For example, the first disc from Narrabri will contain Narrabri data beginning with 1991. It will also contain as many full years of data from the following years as possible. But it will not contain half a year of data. If the first disc holds Narrabri 1991, '92, '93, and '94, then the second disc will contain Narrabri '95, '96, '97,...

The collation type “year” means that whole years from all stations are grouped together and must be placed on the same disc. More than one year may go on the same disc. For example, the first disc may contain 1976 from all stations, 1977 from all stations, up to 1993 from all station. The next disc will start with 1994 from all stations and continue from there. For some of the archives, we hope that all stations and all years will fit on one disc.

3 The Complete DAT Archive

The *Complete DAT* archive consists of four CD cases containing CD-ROMs. Each case contains six “active” CD-ROMs and zero or more “sealed” CD-ROMs. There is one active CD-ROM for each station.

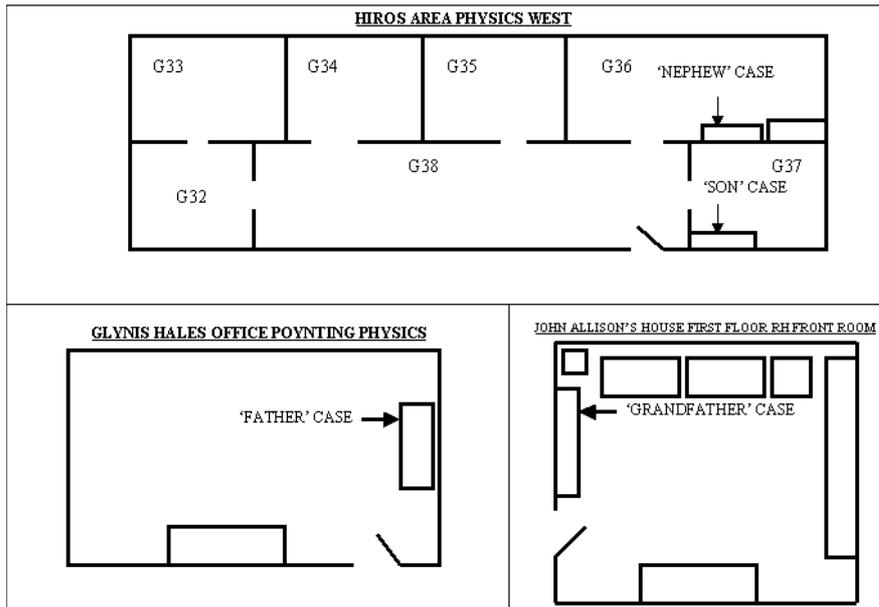


Figure 1: A map showing where to find the CD cases.

Each active CD-ROM contains zero or more complete old years of data. It will also contain part of the current year of data. Each sealed CD-ROM will contain one or more complete old years of data.

For each station, we add data to the appropriate active disc, one month at a time, until it fills up. At which point we seal the disc and start a new active disc. We arrange this so that years are not split between discs.

Data are added to the active CD-ROMs once every month using the son/father/grandfather method as follows. Assume that we are presently adding data for month M to the archive. This is probably because we are right now in month $M + 1$. At the beginning of the operation, the son discs contain data up to month $M - 1$.

All data from the son disc are copied to a work area on the hard disk and are checked against the checksum file *cdrom.md5*. The new month M of data is added to the work area and a new *cdrom.md5* file is created. The work area is written to a new CD-ROM; for the moment, we call this a nephew disc and place it in a “waiting case” called the nephew case.

After we do all six stations this way, we have a full nephew case. At this point, we rotate all of the cases. The nephew case becomes the new son case. The old son case becomes the new father case. The old father case becomes the new grandfather case. Now, the son, father, and grandfather cases hold data up to M , $M - 1$, and $M - 2$ respectively.

The active discs from the old grandfather case become great-grandfather discs and are removed from the case and buried. This now-empty (except for sealed discs) case will be used as the nephew case next month.

To summarise then: if the last month for which a complete set of data exists is M , then the son case contains all previously created sealed discs plus the active discs for all months up to

and including M . Similarly, the father and grandfather cases contain all previously sealed discs, plus the active discs for $M - 1$ and $M - 2$ respectively. If $M - 1$ is December, then a new sealed disc may be added to the then nephew case, if the January data causes the active disc to fill up. At such times, the new sealed disc should also be added to the grandfather case, so that on rotation the sealed discs in the now son and nephew cases are *always* the same.

Figure 1 shows where to find the CD cases.

The nephew and son discs are kept in the HiROS group office suite; the son discs are in G37 (ex darkroom), and the nephew discs are in G36 John’s office. The father discs are the first backup set. In principle these should be kept in the Poynting building, and are currently located in Glynis Hale’s office room S17. The grandfather discs are the second backup set and are kept in John’s house. The four CD cases are shown schematically in Figure 2.

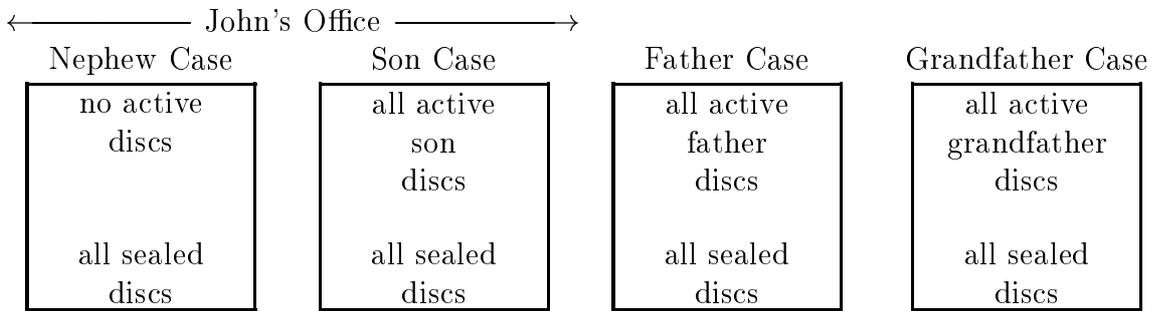


Figure 2: The four CD cases.

To continue the family analogy, there is a slight possibility of infanticide. In the event that data at any station is not received for more than three months owing to a fault, or a communication problem, the active son disc for that station would migrate to the grandfather case, be buried and cease to exist in any generation. In such an event the disc should be retained and treated as a sealed disc, until data is received again. A separate copy of this “quasi-sealed” disc should be placed in each CD case. (A clarification of this procedure is given in the Appendix.)

Once every year, one new copy of each sealed disc is made and the oldest copy is buried. All discs are checked against the checksum file at this time.

References

- [1] KONSTANTIN I. MARCHENKOV AND BREK A. MILLER. Panasonic optical disk cartridges and the IBM-PC. *BISON Technical Report Series*, Number 185, High-Resolution Optical-Spectroscopy Group, Birmingham, United Kingdom, July 2002.

A Appendix

Meanwhile in another space-time continuum...

It is April. In spite of the beginning of spring, the weather is bleak. In Carnarvon there has been no data for a month. Both Les and Brek are despondent. The state of the active CDs in cases is as follows (Carnarvon only):

Nephew case: No active discs
Son Case: January and February
Father Case: January
Grandfather Case: No active discs (but sealed discs up to December)

John brings back the Grandfather case and buries all the active disks for the other stations (with a heavy heart). He adds the data for March to the work area on the hard disk and creates the checksum file.

“Ah!” he ejaculates, inadvertently waking Bernie, Graham, and Rosy, (they remonstrate) “there is no data for March for Carnarvon. OK, I’m going to create the new son disks for March in the other stations, and create a new son disk for Carnarvon which still only goes up to February.”

[“Be quiet! / Silencio!” say his room-mates.]

So now we have, after John has moved the cases around:

Nephew case: No active discs
Son Case: January and February
Father Case: January and February
Grandfather Case: January

“If there’s still no data at the end of April,” whispers John, “I’ll end up with 3 identical active discs, which will confer on them the quasi-sealed status until data starts coming in again.”

Bernard looks up. “John,” he says, “shouldn’t you get out more?”

“My work here is done,” says John, “I must return to my own time.” He jumps into a convenient wormhole and comes back to that happy land where Carnarvon has data for March (thanks to George).

THE END