



IRAM

Revision: 2
2003-11-13

Contact Author
P. Dumontroty

Institut de RadioAstronomie Millimétrique

Owner P.Dumontroty, F.Morel, A.Perrigouard,

Keywords:

Approved by:

A.Perrigouard

Date:

Dec 2003

Signature:

Change Record

REVISION	DATE	AUTHOR	SECTION/PAGE AFFECTED	REMARKS
1	2003-11-24	P. Dumontroty		
2	2004-04-22	P. Dumontroty	Remove DAT references	

Content

1 INTRODUCTION 4

2 ARCHIVING MEDIA..... 4

3 ARCHIVING FLOW 4

Plateau de Bure Archiving Process

This document explains, how we archive the data, which are generated on the Plateau de Bure when a project has been observed.

1 Introduction

When astronomical source is observed, raw data are collected using the CLIC software and they are put in a file corresponding to the project and the date of the observation. All the raw data are stored in such files to be used later for data calibration, reduction and finally to create maps and scientific results. Those data are very important because they are the source of all scientific results and are for this reason archived on several different kind of media.

2 Archiving media

To ensure maximum data security, preventing accidental destruction due to media longevity, fragility, human error or fire, the data are archived physically on 3 different places. Before Apr 21,2004 a copy is also made on DAT actually stored on the PdB.

- ❑ The first set of data is stored on DVD-Rs, located at IRAM Grenoble.
- ❑ The second one on a fileserv (datasrv2) located on the Plateau de Bure.
- ❑ And finally the third set is stored on an identical fileserv (datasrv1) located at IRAM Grenoble

DVD-R media are used for their longevity (about 20 to 100 years) and because it is easy to recover data on that kind of media from any PCs having a DVD reader. For accessing the raw data, the astronomer never uses DVD-R since they can access the data directly through the fileserv using the *getproj* command.

The fileserv ensure quick and easy access to the raw data and they avoid media manipulation, which is one of the most important causes of data destruction. The fact that two identical servers are located in two different places guaranties data availability and security against fire and hardware problem.

3 Archiving flow

All the process starts on **bure1**: the real-time HP-UX workstation. An astronomical source is observed using the CLIC software and the data are stored from time to time and scan by scan into an IPB file with a name corresponding to the concatenation of a project name and the date of the day (i.e. 22-oct-2003-mb09.ipb). The observational parameters are also stored at the same time in two files of the same name with the extension **.rdi** and **obs**. Those files (data and observational parameters) are located in the directory /data/CURRENT of bure1. At the same time, a procedure CLIC @COPY copies each new scan to bure2 into the same directory name: /copy/CURRENT. When an observation is closed, those 3 files are moved respectively on **bure1**, from /data/CURRENT to /copy/COPIED and on **bure2**, from /copy/CURRENT to /copy/COPY.

All the other operations are done using the *savedata* program that is normally run each day by the operator on **bure2**. *Savedata* really does the data archiving on the four different places mentioned above. It chronologically does the following steps:

1. Gets IPB data of the specified day (yesterday by default) from the directory /copy/COPY and, at the same time, moves and copy them respectively into the directory /copy/DVD and /PdBDATA/ARCHIVE/CURRENT (which is physically located on the file server datasrv2).

When Data are moved from */copy/COPY* to */copy/DVD*, the name of each file is translated to conform ISO standard syntax. The name convention we used is: **AMjjproj.ext** where

- A is the year: A <= A <= Z <-> 1990 <= A <= 2015
- M is the month: 1 <= M <= C (Hexadecimal value)
- jj is the day: decimal value
- proj is the project name: 4 alphanumeric or numeric characters max
- ext is an extension: 3 characters max to specify IPB, RED or OBS file content.

At that point, raw data are stored simultaneously on bure1 (the real-time HP-UX workstation), **bure2** (in the */copy/DVD* directory) and the **datasrv2** file server (in the */PdBDATA/ARCHIVE/CURRENTDVD* directory).

2. If */copy/DVD* is full (4.5 Go reached) the data are copied to DVD-R (using the cdrecord-ProDVD tools) and finally are moved to the */copy/ARCHIVED* directory.

On the **datasrv2** file server, the files are moved from the */PdBDATA/ARCHIVE/CURRENTDVD* directory to their final directories. Two main principles are used:

- Keep the DVD-R structure (organization) to recover easily data from DVD-R if needed
- Sort data by project, which is easier to recover data by astronomers.

The first one is assumed by creating a directory structure that correspond, to each CDs/DVDs archive and where each file is a link pointing to the physical file located in its respective project directory. The */PdBDATA/ARCHIVE/* directory contains **PDBnnnnn** directories corresponding to each CD/DVD label and */PdBDATA/PROJECT* directory contains **YNNN** project directories. When moving files, two kinds of file are considered according to their extension name:

- RAW data <file>.IPB, <file>.RDI, <file>.OBS
- Other files

The files, which are not RAW data of type (ipb, obs or rdi), are kept on: */PdBDATA/ARCHIVE/CURRENTDVD*.

The RAW data files are analyzed to extract from their name, a project or a subproject. If it is not possible, the files are put on */PdB/PROJECT/MISC/<filedir>* directory, where *<filedir>* is the file name without the first 4 characters (used to encode a date) and characters following a dot which are considered as extension file.

If the files are really RAW data file, they are put on */PdBDATA/PROJECT/<PPPP>* and a link is created on */PdBDATA/ARCHIVE/CURRENTDVD* to keep the DVD structure.

Each time a DVD recording process is done, the */PdBDATA/ARCHIVE/CURRENTDVD* is rename to the new DVD label: */PdBDATA/ARCHIVE/PDBnnnnn*.

The file convention naming is: **DDDDPPPP.ext** where **DDDD** is an encoded date and **ext** is the file type (IPB, RDI, OBS). **PPPP** is normally the project or subproject name.

The standard project/subproject convention naming (**PPPP**) is either:

project: **Y0nn** where **Y** is a Year encode digit (A->1990), 0=zero, **nn** an hexadecimal uppercase number.

subproject: **Yann** where **Y** and **nn** are like above and **a** an upper case alphabetical letter starting with A and indicating a subproject number.

All other syntaxes are considered obsolete and have for consequences to redirect the files in the MISC directory and subdirectory: **PPPP**. An exception with the project name like: **Y-nn**, **Y--n**, **Ya-n**, **Yann** and like **HOLO**, **BASE**, **FLUX**, **BAFL**, **SOUR** and **VLBI** which are considered valid.

Operations are logged in the log file */PdBDATA/LOG/YYYY* where **YYYY** is the current year.

Now, raw data are stored simultaneously on DVD-R, **bure2** (in the directory */copy/ARCHIVED*) and the **datasrv2** file server (in the */PdBDATA/ARCHIVE/CURRENT* directory).

Some log files are also created to follow the process and to recover later the data.

<i>/copy/COPY/savedata.log</i>	All operations done by savedata.
<i>/copy/COPY/savedata.last</i>	Last DVD-R.
<i>/copy/COPY/<volname>.LIST</i>	All files saved on DVD-R.

- Each night (starting at 8:00) the data are automatically transferred from the Plateau de Bure file server **datasrv2** to the IRAM Grenoble file server **datasrv1**, using rsync, which ensures the complete duplication and mirroring of the two file servers. To avoid transfer bottleneck during the day, it is important to start regularly (each day) the command *savedata*.

At this point, the data are available and accessible, by project, to the astronomers at Grenoble Astronomers, using the command *getproj* on any RedHat PCs.

The following is the flow card of raw data migration.

