

## The *INTEGRAL* Archive System

M. T. Meharga, P. Binko<sup>1</sup>, K. Pottschmidt<sup>2</sup>, M. Beck, R. Walter  
INTEGRAL Science Data Centre (ISDC)<sup>3</sup>, Chemin d'Écogia 16,  
CH-1290 Versoix, Switzerland

T. McGlynn

LHEA, Code 660, NASA's GSFC, USA

**Abstract.** We present the archive system developed for the long term storage and distribution of the data provided by ESA's International Gamma-Ray Astrophysics Laboratory (*INTEGRAL*). The unique properties of *INTEGRAL*'s data required the development of some new features (compared to standard archive components). We give a short overview of the main components of the system - namely the data ingestion software, the data organization concept, the archive database, the data distribution pipeline, and the modified Browse data access interface.

### 1. Introduction

In this short paper, we describe briefly the archive and distribution system of the data provided by ESA's International Gamma-Ray Astrophysics Laboratory (*INTEGRAL*, launched October 2002). As shown in Figure 1, the entry point of any data in the archive is the Ingest application - (1) in Figure 1 - (section 3.) and the Ingest data source is the ISDC system pipeline, see Beck, M. et al. 2004 for more details about the latter. During data ingestion, Ingest generates various metadata which are stored in a relational database, the Archive Database or ADB - (2) in Figure 1 - (section 4.). The standard way for the astronomical community to access the ADB is via the Browse facility - (4) in Figure 1 - at the ISDC, a modified HEASARC<sup>4</sup> Browse (section 5.2.). Different methods for triggering a distribution as well as for transferring the requested data are available. One of them is from Browse. Once the user has selected the needed data products in Browse, the data distribution pipeline (section 5.3.) is triggered. The distribution pipeline builds and provides the complex dataset of all related science and auxiliary files necessary for further analysis of observations.

---

<sup>1</sup>SYNSPACE AG, Rue de Lyon 114, CH-1203 Geneva, Switzerland

<sup>2</sup>Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, 85748 Garching, Germany

<sup>3</sup><http://isdc.unige.ch/>

<sup>4</sup>HEASARC, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

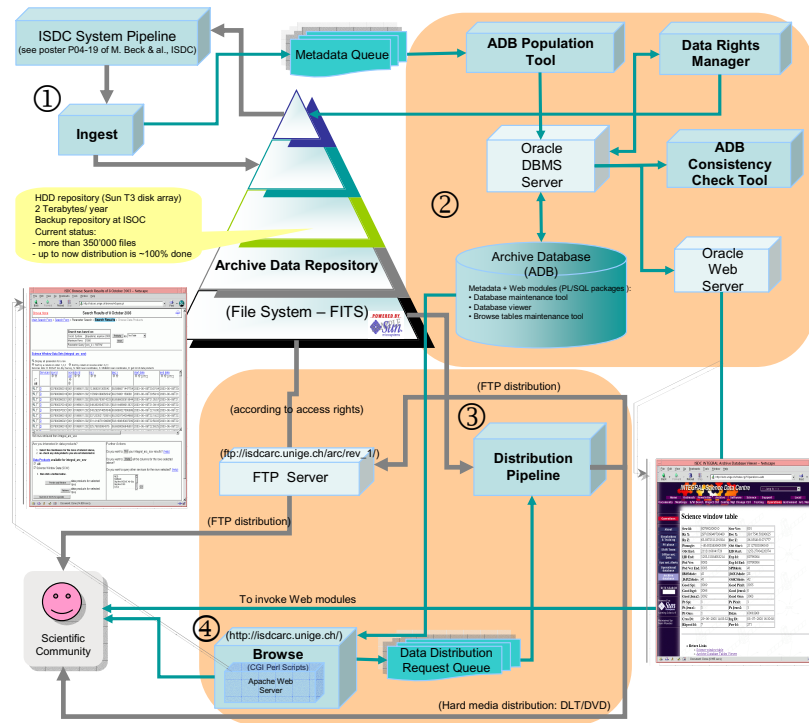


Figure 1. A synoptic diagram of the INTEGRAL archive system

## 2. The Archive Data Repository

Because of the *pointing-slew-pointing* dithering-nature of INTEGRAL operations, each observation of a celestial target is actually comprised of numerous individual S/C pointings and slews (S/C maneuvers to the next pointing). In addition, there are engineering windows (no scheduled observation periods), yet the instruments still acquire data. The ISDC generalizes all of these data acquisition periods into **Science Windows** (ScWs). An **Observation Group** (OG) is defined as any group of Science Windows used in the data analysis. The observations scheduled in the INTEGRAL observing program will be used to define observation groups (Standard OGs). The archive data repository structure has a high level directory structure as follows:

- scw/ contains the results of processing on a per ScW basis
- obs/ contains the results of processing on a per OG basis
- aux/ contains auxiliary data products
- cat/ contains observational catalogs necessary for data analysis
- idx/ contains ISDC index files used for fast searching of data
- ic/ contains data concerning the instrumental calibrations and operational characteristics

See Pottschmidt, K., et al. 2002, for more details about the archive data repository structure.

### 3. Archive Ingest

The Ingest component stores the data in the archive data repository fulfilling four main functions:

- copying of the data sets to the correct place, the data must belong to one of the classes mentioned above (with the exception of the index groups)
- adding the version number to the file names of the files in the scw/, obs/, and aux/ (partially) directories, no files may be deleted
- validation, i.e., the extraction of metadata for the ADB
- indexing of the data sets (idx/ directory)

In order to accept and process ingest requests continuously, there are two daemons available: the passive ingest daemon, which looks for trigger files placed into a predefined directory by an external process and creates the according entry in the ingest request queue file, and the ingest request queue daemon, which looks for new entries in the ingest request queue file. If the latter finds such entries, the ingest tool for the requested data class will be fully executed, including all points listed above.

### 4. Archive Database

The archive database as a part of the system archive and distribution system is an intermediate agent between the users (via SQL interface, Oracle Web forms, or Browse) and the archive repository which insures an efficient and fast access to the data. The archive database stores two types of data (generated by Ingest):

1. administrative data related to the observations, i.e. proposal data, observation properties, and
2. metadata on the archive repository content i.e. descriptions on data files and their locations, used in particular by Browse.

The two kinds of data are inserted in the ADB by the ADB population tool (Browse tables are updated consequently by database triggers). The content of the database can be viewed or updated through the web (Oracle Web Server) by the three applications (resp.): the ADB tables viewer, the ADB tables maintenance tool and the Browse tables maintenance tool (Oracle PL/SQL packages). The status of the data access rights is maintained in the ADB. The data rights manager allows the maintenance of the data access rights (file systems permissions) according to a data rights policy defined by ISDC. The consistency check tool performs consistency check between data files of the archive repository and their metadata stored in the database. Actually, the consistency check is performed between the database and the metadata queue after its regeneration by Ingest.

### 5. Accessing *INTEGRAL* Archive

#### 5.1. Direct Access

The data in the archive repository can of course be accessed directly as far as allowed by the data access rights. In practice, this is especially of interest for

the different projects organized in the context of the guaranteed time program: access to those private survey data of and for the ISWT is organized via UNIX group access permissions.

## 5.2. Browse

Browse is a Web application developed by HEASARC. It provides access to the catalogs and astronomical archives of HEASARC. Browse is adopted for *INTEGRAL* archive distribution through the Web. The unique properties of *INTEGRAL*'s data (large field of view, coded mask imaging technique, complex auxiliary information, multi-version data) triggered the development of some additional features for Browse:

1. support of multiple coordinates for the same observation group and
2. support of multiple repositories for the same mission.

In addition to the option for external users of triggering the data distribution pipeline via the modified Browse facility (available soon).

## 5.3. Data Distribution Pipeline

The data distribution pipeline distributes proprietary and public data products to several classes of astronomical community users (PI guest observers, ISWT members, general public), (800 Mb/ day). Management and control is provided by an OPUS environment which allows:

1. processing of up to seven distribution requests simultaneously,
2. handling of FTP, DVD and DLT requests in an independent way,
3. handling of simultaneous copy and verification processes.

The ISDC routinely triggers the distribution for PI guest observers as soon as their observations are completely processed. The distribution pipeline creates compressed data files (tarred and gzipped) containing the related repository subsets. Depending on the method specified in the distribution request (proposal administrative data), these files can then either be transferred via FTP or on hard media.

## 6. Conclusion

The *INTEGRAL* archive system is in production at ISDC (as well as at ISOC) since November 2002 without major problems. About 6 GB are archived for every ~3day revolution of *INTEGRAL*. Up to now the distribution of PI guest observers data is almost 100% completed. The future challenges include the performance improvement of each of its components, in particular Browse.

## References

- Beck M. et al. 2004, this volume, 436
- Pottschmidt, K., Binko, P., Meharga, M. T., Ouared, R., Walter, R., Courvoisier, T., 2002, Symposium “*Ensuring Long-Term Preservation and Adding Value to Scientific and Technical data*”, Toulouse, France