Full length article

GOLIA: An INTEGRAL archive at INAF–IASF Milano

A. Paizis a,∗, S. Mereghetti a, D. Götz b, M. Fiorini a, M. Gaber c, R. Regni Ponzeveroni a, L. Sidoli a, S. Vercellone d

a INAF–IASF, Sezione di Milano, Via Bassini 15, 20133 Milano, Italy
b AIM/Service d’Astrophysique- CEA Saclay DSM/Irfu, Gif sur Yvette, France
c Former ISDC, Chemin d’Ecogia 16, 1290 Versoix, Switzerland
d INAF–IASF, Sezione di Palermo, Via Ugo La Malfa 153, 90146 Palermo, Italy

Article history:
Received 27 November 2012
Accepted 9 February 2013

Keywords:
Astronomical databases: miscellaneous
Gamma rays: general
Methods: data analysis
Surveys
X-rays: general

Article Info

A B S T R A C T

We present the archive of the INTEGRAL data developed and maintained at INAF–IASF Milano. The archive comprises all the public data currently available (revolutions 0026–1079, i.e., December 2002–August 2011). INTEGRAL data are downloaded from the ISDC Data Centre for Astrophysics, Geneva, on a regular basis as they become public and a customized analysis using the OSA 9.0 software package is routinely performed on the IBIS/ISGRI data. The scientific products include individual pointing images and the associated detected source lists in the 17–30, 30–50, 17–50 and 50–100 keV energy bands, as well as light-curves binned over 100 s in the 17–30 keV band for sources of interest. Dedicated scripts to handle such vast datasets and results have been developed. We make the analysis tools to build such an archive publicly available. The whole database (raw data and products) enables an easy access to the hard X-ray long-term behaviour of a large sample of sources.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The INTErnational Gamma-Ray Astrophysics Laboratory, INTEGRAL (Winkler et al., 2003), is a medium sized ESA mission successfully launched on October 17th, 2002. Its payload consists of two main gamma-ray instruments, the spectrometer SPI (Vedrenne et al., 2003) and the imager IBIS (Ubertini et al., 2003), covering the 15 keV–10 MeV band. IBIS is a high angular resolution gamma-ray imager optimized for accurate point source imaging and for the continuum and broad line spectroscopy. It consists of two layers, the lower energy one (IBIS/ISGRI, 15 keV–1 MeV, key parameters in Table 1, Lebrun et al., 2003) and the higher energy one (IBIS/PIcSIT, 0.175–10 MeV, Labanti et al., 2003). Co-aligned with SPI and IBIS are two X-ray monitors JEM-X (4–35 keV, Lund et al., 2003) and an optical monitor OMC (500–600 nm, Mas-Hesse et al., 2003). INTEGRAL nicely links the energy band of pointed soft X-ray missions (such as Chandra, XMM-Newton) to high energy gamma ray space missions such as AGILE, Fermi and ground based TeV observatories.

Operating since 2002 and thanks to the synergy of its complementary instruments, INTEGRAL provides an impressive quantity of high quality data for the study of key science areas, such as nucleosynthesis, diagnostics of supernovae and supernovae remnants, compact sources (black holes, neutron stars, white dwarfs), electron–positron annihilation emission, super-massive black holes, the cosmic X-ray background, gamma ray bursts, etc (see Winkler et al., 2011, and references therein for a recent review). At the time of writing a total of about 8.5 years of data have become public and are available to the scientific community for exploitation.

INTEGRAL instruments have very wide fields of view. Even considering the fully coded field of view, where the signal to noise ratio is maximal, that is smaller than the total telescope field of view, we obtain fields of about 9°×9° for IBIS, 16° (corner to corner) for SPI, 5° (Ø) for JEM-X and 5° × 5° for OMC. Hence a complete study of a given source (or class of sources) requires huge datasets to be downloaded each time.

In order to increment and ease our exploitation of INTEGRAL data, we undertook the task of preparing and maintaining an INTEGRAL archive, GOLIA (Giant On-Line INTEGRAL Archive). GOLIA’s aims are twofold: provide a local database of the available public data to facilitate personalized analysis of already downloaded data, and offer easy-to-browse IBIS/ISGRI data products for a quick and efficient view of the hard X-ray sky. All of which is locally and interactively available at INAF–IASF Milano. The scripts we have used to build GOLIA are publicly available, as described in Section 4.1.

2. A walk through GOLIA

The basic structure of GOLIA is shown in Fig. 1 and is described below.

**Fig. 1.** Basic structure of GOLIA. The black solid arrows show the steps made by the archive **Owner** (indicated as ‘O’), while those shown with dashed red arrows are performed by the archive **User** (indicated as ‘U’). The Download, Process and Post-process steps are shown. All the scripts of these three phases are publicly available (see Section 4.1). The tools launch\_IMA, launch\_LCR, ima\_pick, readarcres and readnew have been developed specifically for our archive. See Section 2.

**Table 1**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>IBIS/ISGRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy range</td>
<td>15 keV–1 MeV</td>
</tr>
<tr>
<td>Detectors</td>
<td>16384 CdTe</td>
</tr>
<tr>
<td>Spectral resolution (FWHM)</td>
<td>8 keV @ 100 keV</td>
</tr>
<tr>
<td>Field of view (fully coded)</td>
<td>9° × 9°</td>
</tr>
<tr>
<td>Angular resolution (FWHM)</td>
<td>12''</td>
</tr>
<tr>
<td>Source location (radius)</td>
<td>1'' for S/N = 30</td>
</tr>
<tr>
<td>Continuum sensit. @100 keV</td>
<td>2.9 × 10⁻⁶ ph cm⁻² s⁻¹ (3σ detection in 10⁵ s.ΔE = E/2)</td>
</tr>
<tr>
<td>Absolute timing accuracy (3σ)</td>
<td>61 μs</td>
</tr>
</tbody>
</table>

**2.1. Building the archive (archive Owner)**

A detailed description of the INTEGRAL Off-line Scientific Analysis (OSA) software package, used to analyze INTEGRAL data, is beyond the scope of the paper and we refer to the Documentation web page¹ of the ISDC Data Centre for Astrophysics (former INTEGRAL Science Data Centre, Courvoisier et al., 2003). We recall here one key point: INTEGRAL data are organized into the so-called “Science Windows” (ScWs, that for the purposes of this work can be considered as corresponding to a pointing at a fixed sky position, ∼2 ks). During the scientific analysis, all the ScWs belonging to the same observation are grouped together to form an “Observation Group”. The OSA package performs tasks in a modular way, i.e. it first applies photon energy correction (COR level) to all the ScWs belonging to the Observation Group, then it performs the second step, the Good Time Interval computation (GTI) for all the ScWs, and so on. When the analysis of a single observation is to be done (e.g. a few days/weeks of data) this approach is efficient since it simplifies tasks such as mosaicking or light-curve collection. But since a global analysis of all the public data means handling several tens of thousands of pointings (currently 83733), we chose a single ScW based approach: each Observation Group consists of a single ScW and the analysis loops over all the available ScWs. Basically, the analysis is run at the level of the smallest available “data entity” (the ScW), with no a-priori grouping chosen by the archive **Owner**. A-posteriori grouping of ScWs, for mosaicking and light-curve collection, will then be possible as chosen by each archive **User**. Furthermore, if the analysis of one single ScW happens to fail (for any given reason, missing data, no GTI available, etc.), the analysis will continue and be completed for the remaining ScWs, independently. The implementation of such a script can be done in many ways, in our case Shell scripts with a simple “for each” loop have been chosen.

Our IBIS/ISGRI processing and post-processing steps (all performed by the archive **Owner**) are described below. All the scripts described in this section are publicly available, see Section 4.1.

1. **Download**: all the INTEGRAL public data are downloaded via the ISDC dedicated web page².

2. **Process**: the standard OSA 9 software package³ is run on all public data of the IBIS/ISGRI instrument, looping over ScWs via the Shell scripts launch\_IMA.sh and launch\_LCR.sh. We decided to run the image analysis (IMA) on the whole archive first, to have a quantitative view of the results, and then run the light-curve extraction (LCR) part on it. The scripts can however be merged to perform IMA and LCR for each ScW. The following steps are performed:

2-1 **Image reconstruction (IMA)**: for each pointing a standard analysis is performed in four energy bands: 17–30, 30–50, 17–50 and 50–100 keV. The results of this analysis are the starting point for pointing-based, ∼2 ks, light-curves (for all sources of the input catalogue) and maps/mosaics for a deeper analysis.

2-2 **Catalogue creation**: for each ScW, a catalogue is created merging the bright sources discovered from the previous IMA

---


³ The latest version of the software package can be found at [http://www.isdc.unige.ch/integral/analysis#Software](http://www.isdc.unige.ch/integral/analysis#Software).
step (with detection significance higher than 3) and sources of interest, regardless their detection level, be they new or already known. This allows to detect short bursts from interesting sources that would be otherwise missed in a single pointing (~2 ks). Successful tests have been made on known GRBs.

2. Light-curve extraction (LCR): 17–30 keV standard 100 s bin light-curve extraction is performed for all the sources of each ScW catalogue, i.e. thanks to step (2-2), also including sources of interest not necessarily detected in the ~2 ks image (step 2-1).

2-4 Clean and close down: relevant files are zipped and intermediate levels cancelled.

3. Post-process: for each pointing, the IMA analysis step produces images in several bands and lists (FITS files) with position, flux and significance of the sources. Since map and source list files are derived for each ScW, all the information about a given source is spread over many ScWs. The OSA tool src_collec permits to collect IMA source results into a single table. However, it enables to retrieve results for one source at a time only and the results are saved in a FITS table. We have developed a PERL script (ima_pick.pl) that extracts in one go single ASCII files for all the sources present in the archive, one file per source. The tool can be run in a cumulative way (the ASCII files can be updated with the new incoming results). Each ASCII file contains the results of the imaging step and is the starting point for quick visualization by the Users.

We note that the ima_pick script collects results based on the source name, creating a SourceName.dat file. Every time the analysis software (OSA) detects a new source, the name NEW_1 is given (with incremental number for more new sources), hence in the whole catalogue we obtain several NEW_1, NEW_2, etc., sources that are not the same, but have different positions. All these sources will be collected in the same ASCII file by ima_pick and need to be disentangled based on their coordinates.

The introduction of the Post-process step, via the ima_pick tool (not part of the OSA package), leads to a huge simplification in the visualization of the image products. The produced ASCII files are a very easy-to-handle starting point and indeed whole-archive (8.5 year) light-curves can be extracted in matter of seconds for whichever source originally in the catalogue.

2.2. Using the archive (archive User)

Given the main steps performed by the archive Owner (Download, Process, Post-process), each User can access GOLIA in many ways. The mere Download, done by the Owner, allows the User to access the entire INTEGRAL public data for a personalized analysis, avoiding multiple time and space consuming downloads.

Once the Owner has completed the Process step, the User can already attain analyzed ScW images to build customized mosaics and source searches via a deeper analysis e.g. aimed to dim population studies. For brighter sources (or sources of interest) overall light-curves binned at 100 s can be collected with the ISDC tool ic_pick (see Fig. 5 and Section 3.3). The stacking of the Owner ScW-based products of the “Process” step (be they mosaics or 100 s bin light-curve collection) can be time consuming for the User if a very large number of ScWs is involved, as in standard OSA usage.

Finally, once all the key IMA parameters/results are saved in the ASCII file via the Post-process step, it is immediate to plot them in different ways to have a quantitative view of the database. As an example, in Fig. 2 we show the detection significance (DETSIG) of the Crab in the whole archive in two energy bands, as a function of the off-axis angle, and the evolution of the latter with time, expressed in INTEGRAL Julian Date, JJD.\(^4\)

---

4 JJD = MJD – 51544.
Fig. 3. Whole-archive ∼2 ks light-curve obtained for IGR J11215-5952. The plot shows all the INTEGRAL/IBIS public available data: 2442 pointings, 4.8 Ms, from revolution 0030 (Jan 2003, IJD ∼1107) to 1060 (June 2011, IJD ∼4190). Sigma ≥ 5 (horizontal line) is a detection.

periodicity is the orbital period of the system. At that time we could not exclude that the true periodicity is half of this period (as indeed was later discovered, see Sidoli et al., 2007), because of the lack of INTEGRAL coverage of the source field of the first SFXT showing periodic outbursts.

This discovery triggered by our archive has led to further investigations within our team, including the development of a stellar wind model for OB supergiants to explain the transient and variable nature of SFXTs, and its application to all public INTEGRAL observations from 2003 to 2009 for a sample of 14 SFXTs (Ducci et al., 2009, 2010).

3.2. Low mass X-ray binaries (LMXB)

Visual inspection of light-curves such as Fig. 3 for persistently bright neutron star (NS) LMXBs (e.g. the Z source GX 5–1) has been the starting point for the systematic study of the evolution of the transient hard-tails in the spectra of these sources. Indeed hard tails dominate the X-ray emission above 30 keV, in an otherwise soft spectrum, and they result in a source detection in the 30–50 keV panel of Fig. 3 (e.g. Paizis et al., 2006a; Farinelli et al., 2008; Mainardi et al., 2010). An example for the transient tail can be seen in Fig. 4 for GX 5–1 (JEM-X and IBIS/ISGRI data). The origin of such tails dominating the spectra above ∼30 keV is still debated. Paizis et al. (2006a) proposed for the first time a qualitative unified physical scenario to explain the spectral evolution of NS LMXBs, including the peculiar transient hard tail. As a following step, in the quest to study in a quantitative way the evolution of the parameters describing the innermost physical conditions of NS LMXBs, a new Comptonization model was developed (comptB, Farinelli et al., 2008), and successfully applied to a sample of bright NS LMXBs studied with INTEGRAL (Mainardi et al., 2010).

3.3. Rapid (∼100 s) variability

In the archive, 100 s bin light-curves in the 17–30 keV band are available for all the sources detected in the IMA part (2 ks pointing basis) as well for sources of interest. Among others, we have extracted a whole-archive light-curve for the high mass X-ray binary 4U 1700–377, for a total of 18 Ms, sampled on a 100 s basis. Two different time zooms of the obtained light-curve (2 months, 0.8 days) are shown in Fig. 5. This complete light-curve is the starting point for a detailed study of the fast variability over 8.5 years of INTEGRAL data in this peculiar source (Sidoli and Paizis, in preparation).

3.4. New sources

As discussed in Section 2.2, we have developed an interactive tool that reads the NEW source ASCII files created by ima_pick, grouping them by coordinates (readnew).

Fig. 6, lower panel, shows multiple detections (from different bands and pointings) of a new source, the recently discovered transient MAXI J1659–152 (a Black Hole candidate, Negoro et al., 2010, 2012), not yet included in the INTEGRAL General Reference Catalogue, hence labelled as new, NEW_1, by the software. Fig. 6, upper panel (text), shows some stages of the interactive process, with the description of the different symbols and region of the sky selected. A zoomed-in IBIS/ISGRI image of one of the multiple pointings of Fig. 6 can be seen in Fig. 7 where MAXI J1659–152 is labelled as NEW_1 by the software. The well known NS LMXB Sco X-1, ∼9° apart, is also visible and appears much softer than MAXI J1659–152: indeed, while in the softer band Sco X-1 is the brightest source of the field (left panel, 17–30 keV), in the harder band Sco X-1 is not even detected (detection significance about 2), while MAXI J1659–152 is still clearly there (detection significance about 10, 50–100 keV, right panel).

We are able to trace back in the INTEGRAL/IBIS archive sources that were discovered recently (i.e. later than the creation of the
Fig. 4. Unabsorbed $E(F)$ spectra with superposed best-fit $w_{\text{abs}} \cdot \text{compTB}[1] + \text{compTB}[2]$ model (upper panels) and residuals in units of $\sigma$ between data and model (lower panels) for GX 5–1. Different line styles represent single components of the best-fit model. Dashed-line: pure thermal Comptonization for $\text{compTB}[1]$. Dashed-dot line: bulk + thermal Comptonization (upper spectrum) or simple BB (lower one) for $\text{compTB}[2]$. Solid line: total spectrum (from Mainardi et al., 2010).

$\textsc{INTEGRAL}$ General Reference Catalogue used, 2010 April 8th\(^6\), as well as to discover previously unnoticed new $\textsc{INTEGRAL}$ sources, such as IGR J15283–4443, discovered by our systematic analysis of all $\textsc{INTEGRAL}$/IBIS public data (Paizis et al., 2006b) and later on confirmed by Swift follow-up observations (Rodriguez et al., 2010).

4. Build your own GOLIA

We report the main characteristics of GOLIA (disk space, costs, time employed, etc.), in order to give a quantitative idea of the effort involved in such a work. The whole analysis has been performed on a single server HP ML330G6, openSUSE 11.4, two processors Xenon Quad-Core E5506, 2.13 GHz and RAM 8 GB (June 2011 cost $\sim$2300 Euro). A total of 12 external disks of 1 Tb each have been purchased to store the data and the results (213 Euro each Tb). Currently a total of 83733 pointings have been analyzed, i.e. starting revolution 0026 (December 2002) to 1079 (August 2011), for a total of 8.5 years of data. Disk space is needed for the raw data while the results themselves occupy a smaller fraction ($\sim$1.3 Tb for 8.5 years of results versus about 4.8 Tb of 8.5 years of data—excluding SPI raw data).

\(^6\) We note that on 2012 November 15th, a new catalogue has been delivered by the ISDC, including new sources reported in the literature: http://www.isdc.unige.ch/inTEGRAL/science/catalogue.
structure allows it) will result in a much lower processing time. The addition of further data products (more energy or time bins, spectral extraction or inclusion of other instruments) will of course increase the CPU time as well as disk space needed.

We have decided to begin extracting the results for IBIS/ISGRI because we believe that it is a very good starting point to investigate the less explored hard X-ray sky (with respect to the softer JEM-X range), with a good (12′ FWHM) angular resolution to disentangle sources in the Galactic plane and bulge (with respect to SPI, 2.5′ FWHM). Furthermore, we have also been triggered by our experience in the IBIS/ISGRI data analysis that allowed us to handle large datasets while choosing few energy and time bins, still capable to unveil interesting source behaviour (softening versus hardening, unexpected variabilities, transient behaviour, etc.) but yet within acceptable running time.

Automatic spectral extraction, though applicable with our scripts, would have been extremely time consuming given our single-server setup. Furthermore, we decided not to perform it since the feedback with our potential users, and our previous experience on whole-archive spectra, suggested that source spectra should be re-extracted with optimized energy binning to be meaningful for a detailed analysis.

Once the setup is up and running, the maintenance of the archive is not an issue and neither is its regular update. Maps, light-curves, raw data, fits and ASCII files can all be accessed by the users, enabling a personalized usage of the archive and triggering further detailed investigations such as the ones previously described.

4.1. Our scripts

All the INTEGRAL public data can be freely downloaded via the ISDC dedicated web page (see footnote number 2). A tar file with the scripts we used to build GOLIA (Process and Post-process steps, i.e. the ones performed by the archive Owner, see Section 2.1 and Fig. 1) and a short description (README) can be downloaded from our GOLIA web page.7 These scripts are the Shell scripts launch_IMA.sh (and its sub-routine analysis_IMA.csh), launch_LCR.sh (and its sub-routine analysis_LCR.csh) used to launch OSA looping over ScWs, and the PERL script ima_pick.pl, used to “re-format” the IMA results. Together with OSA, these are the tools we used to build GOLIA, they use freely available interpreters and need no compilation.

Once the archive is available, together with all the ASCII files created by ima_pick, browsing and visualization of the IMA results can be easily performed with any software that reads ASCII files. Mosaics of the images or collection of the 100s light-curves can be done as described in the ISDC documentation.

5. Conclusions

There are several on-going international collaborations and efforts to exploit and share INTEGRAL data and results with the scientific community (to some of which we have participated as well). We recall here the Galactic bulge monitoring programme (Kuulkers et al., 2007),8 the INTEGRAL Spiral Arms collaboration,9 and the INTEGRAL Galactic plane scan,10 all of which deal with the analysis of IBIS/ISGRI and JEM-X data of specific sets of data (i.e. portions of the sky). Among the all sky surveys we recall the IBIS/ISGRI catalogues by Krivonos et al. (2007)11 and Bird et al. (2010). An up to date collection of all the sources discovered by INTEGRAL can be found at the INTEGRAL IGR Source page.12 Last but not least, the ISDC has developed and is maintaining an all INTEGRAL instrument archive, known as HEAVENS13 that collects light-curves, images and spectra for a given source or position in the sky, from the public data.

Notwithstanding all the above, we believe that having a complementary local archive of INTEGRAL data and products that is easy to access and to handle (be it for only one instrument) is an important way to boost scientific knowledge. It is open to suggestions and ideas of the local users in a close and frequent feedback process that maximizes the scientific outcome. Developing and maintaining a database enables complete control over what is being done and how; raw data and products are accessible at all levels (i.e. analysis steps), entire sky maps are available (not just a catalogue or a zoom on the target, but complete workable FITS files) and the results of all the detected sources can be retrieved very quickly, regardless the time interval; e.g. long term light-curves can be easily extracted also by non-INTEGRAL experts (a matter of a few seconds for the 2 ks binning).

A systematic overview of different classes of sources can be performed, stimulating new investigations and collaborations, as shown in Section 3, and enabling interesting discoveries, determination of trends, as well as providing a complementary picture to other on-going missions.

8 http://integral.esac.esa.int/BULGE/.
9 http://spgs.ssl.berkeley.edu/~bodaghee/isa/.
10 http://gpsiasf.iiasf-roma.inaf.it/.
12 http://irfu.cea.fr/Sap/IGR-Sources/.
13 http://www.isdc.unige.ch/heavens/.
Acknowledgements

This work is based on observations with INTEGRAL, an ESA project with instruments and science data centre funded by ESA member states (especially the PI countries: Denmark, France, Germany, Italy, Spain, and Switzerland), Czech Republic and Poland, and with the participation of Russia and the USA. AP thanks S.E. Shaw, B.E. O’Neill and the ISDC staff since most of the know-how needed to set up such an archive was gained during her years at the ISDC (2000–2005). We acknowledge the Italian Space Agency financial support via contract INTEGRAL ASI-INAF I/033/10/0. We thank the anonymous Referee for his/her comments and suggestions, which improved the quality of the paper.

References


