GRBase: a new tool for data mining in the Gamma-Ray Burst archives

B. Gendre\textsuperscript{1,†}, A. Montieri\textsuperscript{1,2}, R. Primavera\textsuperscript{1,2}, G. Stratta\textsuperscript{1}, A. Antonelli\textsuperscript{1} and M. Capalbi\textsuperscript{1}
\textsuperscript{1}ASDC, via Galileo Galilei, 00044 frascati (RM), Italy
\textsuperscript{2}TELESPAZIO, Roma, Italy

Abstract. We present a tool dedicated to the data mining of gamma-ray burst afterglow data. This tool allows an easy browsing into the available data of GRB afterglows, and the extraction of scientific products related to these events (light curves and broad band spectra or spectral energy distributions). The tool is linked to the most up-to-date technology of the Virtual Observatory for data inclusion and output together with the standard FITS and ASCII formats. The tool is intended to be used with GRB afterglow data but can be easily extended to all kind of variable objects.

Keywords: Gamma-Ray Burst – data mining – product extraction

1. Introduction

Gamma-Ray bursts are unpredictable events occurring randomly in the sky (Vedrenne & Atteia 2009). Their exact nature is still elusive yet, but they are thought to be due to the accretion of a large amount of matter onto a newly born stellar mass black hole, formed either during the core collapse of a star, or the merging of two compact objects (Mészáros 2006). They consist of a main event, the burst itself, usually called the prompt phase, followed by an afterglow seen at all wavelengths for days to months after the start of the event.

The study of gamma-ray bursts is very complicated due to challenging observational constraints. These objects are fast transients. The prompt phase, which concentrate the larger part of the emitted energy, typically lasts a few tens of seconds
(Mészáros 2006). It is thus mandatory to observe them as soon as possible. These objects are also fast decaying, with a typical broken power law decay (with temporal decay index of the order of 1-2). Thus, instruments suited to observe them during the first minutes of the event (when in optical the event can reach a 8-10 magnitude) are not useful for later follow-up (when the event has reached a 25 magnitude). These two constraints explain why the gamma-ray burst observations are usually carried by automatic or robotic observatories. This however creates a difficulty for the gamma-ray burst scientists, which our tool aims to solve.

This paper is organized as follow. In Section 2 we present the limitations of the current observational strategy and our solution to solve it. In Section 3 we present our tool and its organization. In Section 4 we present the underlying database and the tools used to populate it. We finally conclude in Section 5.

2. Purpose of GRBase

The use of robotic or autonomous observatories is the rule in gamma-ray burst science. To date, a few hundreds of robotic observatories are active in optical and infrared bands (Vedrenne & Atteia 2009). These observatories are most often operated by a dedicated team Klotz et al. (such as the TAROT collaboration, 2009) and/or by a university for teaching purposes. As thus, the archive of these observations may be private. In fact, most of the information obtained by these teams are sent by the GCN network (Barthelmy 1998), once a first processing has been done. This leads to several problems:

- The observations are sometime published very late (current record holder is a publication of data about two years after the observations). This is a problem, as in some cases these observations could have ruled out some proposed explanation of the data.
- The observations are usually not completely advertised. In may be that only one data point is published, giving the impression that the burst is slowly decaying while in fact the data present flaring activities (see e.g. Gendre et al. 2011).
- The observations are not well cross-calibrated, or the user has to know all the technical setting of each instrument in order to correctly build a light curve from the data of several instruments. For instance, GROND and TAROT teams communicate measurements as magnitude without indicating that one is expressed in the AB system and the other in the Vega system.

Our tool has three aims, which will solve these points:

- Centralize the observation logs of all robotic instruments, of major observatories, in order to have the knowledge of all available observations for a particular object.
Figure 1. The interface of GRBase at the start of the tool. The control panel at left allows the user to choose the object they are interested in and to extract scientific products and/or information from the database. The visualization panel at right allows a concise view of available observations classified by time and frequency.

- List the available information for these observations, including when possible the photometric measurements, and the technical setting of the instruments.
- Ease the extraction of light curve and spectral energy distribution (or spectra) for the objects listed in the database.

3. Presentation of GRBase

The main interface of GRBase is shown in Fig. 1. It can be divided in two panels. The left one controls the tool, while the right one displays the information available.

3.1 Control panel

The control panel allows the user to interact with the tool and the underlying database. Indeed, while the tool is installed on the user’s computer, it is only a client to a central database that contains all the information available. The tool is object-oriented, i.e. the user’s first choice is to select an object (it is not possible to browse the sky for a position). From a practical point of view and/or to avoid repetitive entries of the same burst names, several lists of interesting bursts (that have a very peculiar property or an unusual high temporal sampling of data) will be managed by our team. Once the
object is chosen, the data related to it are downloaded from the database into the tool and presented in the visualization panel.

The control panel also allows the insertion of the date and/or band used for the extraction of light curves and spectral energy distributions. When extracting scientific products, the user is requested to choose between an exact extraction (based only on data points really present in the database) or an interpolation of data (in order to maximize the temporal or spectral coverage, at the expense of some model hypotheses).

The user can always insert her/his own data (which are never uploaded into the database) to increase the global coverage of the object of interest, and retrieve general information such as the position, distance, or duration of the object. All of these features are available either with the standard FITS/ASCII formats, but also in the new XML VOTable format used by the Virtual Observatory.

3.2 Visualization panel

The visualization panel presents in a synthetic view the available data ordered by time and frequency. Contrary to other tools, it does not directly present the measurements (i.e. light curves or spectra), but only the sampling factor of the data. This allows the user to make the best possible choice for the product extraction parameters when selecting the time and/or energy. The Fig. 2 presents a view of GRBase once an object has been chosen, with the Visualization panel activated.

The visualization panel also presents the data inserted by the user, with a lighter color to distinguish them with public data. The user always has the possibility of graphically discarding any data by simply clicking on it (the box will then disappear from the visualization panel) before any scientific product is extracted.

4. Database of GRBase

4.1 Organization

The main database of GRBase is physically located at the ASI Science Data Center, at Frascati (Italy). It consists of an archive of all publicly available data from GCNs and articles in a computer friendly format, inside a MySQL database. Assuming its object-orientation, the database is organized with several tables, the main one being the object table, which contains each object located in the database and the information needed to retrieve all measurements related to that specific object. It is foreseen to implement mirrors of this database in order to ease the use of the tool for distant countries.
Figure 2. The interface of GRBase once connected to the database, with one object inserted in the query field. The color coded rectangles allow an easy view of similar kind of data (X-ray, optical, infrared, sub-millimeter and radio). The tool also indicates those observations without measurement present in the database but known to the tool, as shaded boxes.

4.2 Database feeders

Initial tests of the tool and of the database were done manually. However, the very large amount of information to retrieve, process and to archive requires maximum robotization. We have developed for this purpose several small independent softwares, the database feeders that can populate the different tables of the database. They:

- search relevant publications,
- extract from a publication its NASA-ADS unique identifier,
- retrieve the publication in electronic form,
- extract the relevant information using some pre-processor like tools,
- and format and save this information into the database.

All of these feeders were first tested manually and quality controlled before being applied to the whole web. They all report any processing error to a human operator, who is in charge of processing them by hand when needed. Quality checks are performed on a random basis for improvement of the robots.
Figure 3. An example of an observation without any measurement present in the database. The tool indicates the team to be contacted for this information.

4.3 Interaction with GRBase

The GRBase tool can be used with no Internet connection to visualize only the user’s data, however for full use it has to be used on-line. The tool will make some queries to the database to obtain the list of objects, and the list of observations and measurements once the object has been chosen.

One important feature of GRBase is the ability illustrated in Fig. 3. If one observation does not contain a measurement, the tool will indicate the team that made the observation and who can provide the relevant information. This allows for all observations inserted in the database the highest possible impact and citation number.

5. Conclusions

We have presented the tool GRBase, aimed at easing the study of gamma-ray bursts. This tool is still in development phase, and any interested user can contact us for improvements and/or requests on the interface.

We request the interested robotic telescope operators to send us the log file of observations of their telescopes, so that we can insert them into the database (please contact us for format information).

We hope this tool will ease the use of GCN and article observational data for the study of those strange objects, the gamma-ray bursts.
We thank an anonymous referee for her/his help that improved the quality of this paper.

References

Mészáros, P., 2006, Reports on Progress in Physics, 69, 2259