THE US ROSAT PUBLIC DATA ARCHIVE

M. F. Corcoran
(NASA/GSFC-LHEA & USRA)

last update: 27-September-2001

Contents

Abstract

This article summarizes how *ROSAT* observations were proposed, observed and archived. This article is useful to help *ROSAT* archive users understand how data entered the US archive, how datasets were named, and how to find and access archived *ROSAT* data.

For help please e-mail rosathelp@olegacy.gsfc.nasa.gov or request@olegacy.gsfc.nasa.gov

1 Introduction

By international agreement, the *ROSAT* Public Data Archive (USRPDA) officially opened on November 1, 1992 with the release of the first set of processed data. The following article describes how to identify, request and retrieve data from the US archive. This article assumes the reader to be familiar with the form and content of *ROSAT* data sets for the PSPC and the HRI. Users unfamiliar with *ROSAT* data products should consult the *ROSAT* User Handbook, available online at http://heasarc.gsfc.nasa.gov/docs/rosat/ruh/handbook/rosathandbook.html

The *ROSAT* Public Data Archive is located at the Goddard Space Flight Center in the US, at the Max-Planck-Institut für extraterrestrische Physik (MPE) in Germany and in the UK at Leicester University. In the US, storage of and access to the data was originally provided by the National Space Science Data Center (NSSDC) Data Archive and Distribution Service (NDADS). On 18 Apr 1997, a memo of understanding between the HEASARC and the NSSDC was signed in which prime responsibility for serving *ROSAT* data to the public was moved from the NSSDC to the HEASARC, with the NSSDC having ultimate responsibility for the deep *ROSAT* data archive. *ROSAT* data are now kept on line as part of the HEASARC archive. The US archive and the German and UK archives each contain complete sets of all public observations.

The *ROSAT* Public Data Archive contains processed photon events lists, images and housekeeping and calibration information. Originally, none of the Data Archives contained source lists, derived spectra or lightcurves produced by the standard processing system (SASS). These science products where considered rather unreliable for general use without prior screening by the *ROSAT* project. Screening of these results was completed in August 2001 and these result products have now been released to the *ROSAT* archives at MPE, in the US, and in the UK. By mutual agreement between the USRSDC and MPE, data were archived in the format in which it was received by the principal investigator (PI). This meant that observations made by US observers and processed by the US *ROSAT* Science Data Center (USRSDC) werw released to the archive by the USRSDC, while MPE was responsible for releasing all data processed at MPE.

2 The *ROSAT* Observing Cycle

This section discusses the *ROSAT* observing cycle, from target proposal, to observation completion, to release of the data to the archive.

2.1 Approved Observing Programs

The *ROSAT* timeline of observations was based on a competitive selection of targets proposed for by members of the astronomical communities in the US and in Europe. The review cycle usually ran for 1 year, so that accepted targets would be scheduled for observation the year after the the proposal review was held. There were 8 announcement of opportunities (AOs) for *ROSAT* observing proposals. The ninth observing opportunity was announced (on 1998 May 13) but unfortunately *ROSAT*'s operational life ended (on 12 Feb. 1999, 09:18:52 UT) before proposals were submitted. Proposals could and most often did contain more than one target. Each proposed target was assigned a priority by the proposal peer review committee. The following target categories, or priorities, were used:

- Priority A targets: these targets represented the best use of *ROSAT* and offered the highest science return
- Priority B targets: these targets represented excellent use of *ROSAT* and offered high science return

- Priority C targets: these targets represented potentially good use of *ROSAT* but were not as compelling as the priority A and B targets. These targets would be used to fill gaps in the *ROSAT* observing timeline.
- Rejected targets: In the opinion of the review committee these targets were not ranked as highly as the priority A, B and C targets, and would not be observed in the current timeline.

2.2 Observation Scheduling

Targets granted an "A" or "B" priority were guaranteed to be included in the *ROSAT* observing timeline until they received 70% of the observing time requested by the proposal or until the mission ended. Priority "C" targets were not guaranteed to be placed in the *ROSAT* observing timeline at all; however "C" targets might be place in the timeline if doing so would improve the observing efficiency of *ROSAT*. Priority "C" targets would not receive more than 70% of the requested observing time. However, proposers could (and often did) repropose observation of priority "C" (and/or rejected targets) in subsequent observing cycles.

ROSAT observations were scheduled in the timeline based on target positions. Targets within the same *ROSAT* field of view which were proposed for in different observing programs would be scheduled as individual pointings. Each pointing was assigned a 6-digit *ROSAT* OBSERVATION REQUEST number, usually called the **ROR number**. The first digit of the ROR designated the type of target being observed: 0 for mispointed observations, 1 for calibration, 2 for stars, 3 for cataclysmic variables, 4 for X-ray binaries, 5 for supernova remnants, 6 for galaxies, 7 for active galaxies, 8 for clusters of galaxies, and 9 for solar system objects, surveys and miscellaneous objects.

2.3 Data Processing and Distribution

For each target (marked by an individual ROR number) the principal investigator would receive a package (either a tape or, later, CDROM) containing the *ROSAT* X-ray data for that particular pointing at that particular target. A *ROSAT* dataset, as distributed to the principal investigator would consist of a set of files which contained all data processed by SASS for that particular pointing. A given *ROSAT* dataset would be designated by an individual sequence number composed of the ROR number and the "add-on" suffix, and each distributed *ROSAT* dataset would contain:

- Data Files: these files contain X-ray photon lists, images, and spacecraft housekeeping, status, and calibration information; and
- Results Files: these files contained lists of X-ray sources derived from the list of accepted X-ray photons, X-ray background maps, X-ray source spectra (for bright sources) and/or X-ray source lightcurves.

At the start of the *ROSAT* Mission the US and German data centers processed and distributed *ROSAT* data in distinct data formats. This caused difficulties for users who wanted to analyze together *ROSAT* data obtained by US and German principal investigators. Starting in 1994 April, processing centers in the U.S. and Germany began distributing data in a common format, the so-called "Rationalized Data File" (RDF) format. Currently all data in the final version of the *ROSAT* archive is in the RDF format.

Data were distributed to principal investigators soon after they were processed. This means that a number of targets had multiple release dates; for example, the initial pointing of a target might be observed, processed and distributed months before the first "add-on" pointing. This originally caused some confusion for archive users, since it was not clear on which date the completed data were distributed.

Starting in Fall 1993, all data sets previously processed by the USRSDC and MPE were reprocessed by a new version of the processing software (SASS REV1) designed to correct some bugs in the previous (REV0) version of SASS. REV0 Data sets were reprocessed by SASS REV1 and re-distributed to the original principal investigator who retained sole data rights to the reprocessed data for a period of 6 months. The 6 month proprietary period only applied to REPROCESSED data sets (i.e. those data which were previously processed under SASS REV0 and now reprocessed with SASS REV1); new data sets processed only by SASS REV1 still retained the usual 1-year proprietary period. Some of these REV0 data were already released to the public archives. Though the community had no immediate access to the REV1 versions of these data, the SASS REV0 versions were still available for access through the public archive. At the end of the 6 month proprietary period, the REV1 versions were released to the public archive, replacing the REV0 version.

At the start of REV1 processing data were still being distributed in the non-RDF formats. However with the start of SASS/REV2 processing, data were finally distributed in the RDF format by both the GSFC and MPE data centers. US, UK and German data in the RDF format now constitute the official *ROSAT* public archive. Archive users in general are not able to readily access the REV0 data sets if they have been superceded by an RDF version, though in some compelling special cases users can have access to the REV0 data.

There have been rare cases when corrupted data have been sent to a PI. In these cases, data have been reprocessed and redistributed to the PI. It was the policy of the *ROSAT* Public Data Archive that the corrupt data did not enter the archive. In these cases the official public release date for the data was 54 weeks after the good data were distributed.

3 Release of Completed *ROSAT* data to the US *ROSAT* Archive

In most cases, completed observations were released to the public data archive 54 weeks after the data were shipped to the principal investigator. (There were some cases in which a principal investigator might waive the proprietary period, if, for example, a particular observation consumed a large amount of observing time and if the observation might be of special interest to the general community of observers.) An observation was considered completed if

1. the amount of accepted time (as determined by the Standard Analysis System Software) for a given ROR of a priority A or B target was at least 70% of the time alloted to the target by the peer review; or

2. the observation clearly would not receive any additional amount of observing time in the future (for example, due to the demise of instrumentation).

Observations which are completed under the second criterion include priority C targets (those targets which were accepted by the peer review committee on a "best-effort" basis) which received less than 70% of the allotted observing time but which would not be guaranteed to be scheduled for additional observation, or priority A, B or C targets which would not receive further observing time due to the demise of the instrument(s) or satellite.

In general observation of an individual target (or in other words a single ROR number) would be broken up into individual pointings. This naturally happened because of earth occultation of the source (since *ROSAT* was in a low earth orbit), but also because most sources had a finite observing window. If a given target observation could not be completed within one observing window, it would be scheduled for observation in a following window. In the early phases of the mission, individual pointings from different observing windows would be combined into one complete observation package. It was realized that doing this merging actually degraded the spatial resolution in the combined dataset, due to problems solving aspect in different observing windows, using different sets of guide stars. Eventually individual pointings belonging to a single ROR but obtained in different observing windows were no longer combined. These subsets of a given pointing were termed the "add-on" pointings. The amount of time a target (or more precisely, a single ROR) received is then the sum of the amount of accepted time (determined by the standard processing system SASS) derived for all the individual pointings (the initial pointings plus any "add-on" pointings).

The US processing center (housed in building 28 at GSFC) was responsible for the release to the public data archive of all datasets belonging to principal investigators associated with the US *ROSAT* program. The German *ROSAT* data processing center at MPE was responsible for the release of all datasets belonging to principal investigators associated with the German/UK *ROSAT* programs. The USRSDC and the *ROSAT* data center at MPE had sole responsibility for determining completeness of an observation for US and European observations, respectively, and final responsibility to determine whether an observation would be released to the archive. Nearly all completed observations were released to the archive. Released observations even included those for which there were no accepted times according to SASS processing (hence no good science photons) if there were rejected photons. However datasets in which there were neither good science photons nor rejected photons produced as a result of standard processing were not released to the archive.

4 Retrieving Data from the *ROSAT* Data Archive

4.1 Organization of the US *ROSAT* Archive

The *ROSAT* archive is stored on a magnetic disk farm administered by the HEASARC. The archive is a standard unix file system. PSPC data can be found at

`ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data`

while HRI sequences are located in

`ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data`

ROSAT data are stored within directories named by sequence type as follows:

Sequence Type	Description
000000	Mispoints
100000	Calibration observations
200000	Normal stars
300000	Cataclysmic Variables
400000	X-ray Binaries
500000	Supernova Remnants
600000	Galaxies
700000	Active Galaxies
800000	Clusters of Galaxies
900000	Solar Systems, Surveys and Miscellaneous

Within the sequence-type subdirectories, each ROSAT data sequence is stored in a subdirectory named by the sequence id. Each sequence subdirectory contains a number of FITS files in RDF format containing the public *ROSAT* data for that sequence.

4.2 Retrieving data via anonymous FTP

For each sequence type listed above a table of contents file in HTML format is stored within the `processed_data` directory, which may be used to identify sequences of interest. The table of contents file has the form `sequence_type.html`, for example

```
ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data/800000.html
```

To aid the user these table of contents files are also labelled with a descriptive name, for example

```
ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data/clusters_of_galaxies.html
```

ASCII lists of all public archived data are available via ftp from

```
ftp://legacy.gsfc.nasa.gov/rosat/doc/archive/lists/rosarchive_final.pos.gz
```

for the list of archived sequences sorted by declination and

```
ftp://legacy.gsfc.nasa.gov/rosat/doc/archive/lists/rosarchive_final.seq.gz
```

for the list of archived sequences sorted by sequence number

Once a sequence of interest is identified the sequence data directory can directly be accessed by ftp, and data files of interest can be transferred to the user's computer.

For example,users who may be interested in *ROSAT* PSPC observations of normal stars can access a list of such observations at `ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data/normal_stars.html`

The first entry in this file is a pointing at V410 Tau with a sequence identification number of RP200001A00. Data for this sequence is archived at

`ftp://legacy.gsfc.nasa.gov/rosat/data/pspc/processed_data/200000/rp200001a00/`

FTP from the unix command line can also be used to download data from the *ROSAT* archive. Using command line ftp, users can tar an entire data directories for download. An example, in which the entire directory associated with the *ROSAT* observation 800646a01, is shown below:

```
your_machine% ftp legacy.gsfc.nasa.gov
Connected to legacy.
Name (legacy.gsfc.nasa.gov:einstein): anonymous
331 Guest login ok, send your complete e-mail address as password.
Password:
ftp> cd rosat/data/pspc/processed_data/800000
250 CWD command successful.
ftp> dir
200 PORT command successful.
ftp> bin
200 Type set to I.
ftp> get rp800646a01.tar
200 PORT command successful.
150 Opening BINARY mode data connection for /bin/gnutar.
```

Specifying a directory with a "`.tar`" extension, the entire directory will be `tar'ed` and transferred to the user's local disk.

4.3 Using BROWSE to Access the ROSAT Archive

BROWSE is a web-based interface to catalogs and data archives supported by the HEASARC. BROWSE can be used to search the *ROSAT* catalogued datasets by position, source name, or table parameter (like observation start date, end date, exposure time, etc.) and access and download data directly from the archive. The BROWSE interface can be accessed at

`http://heasarc.gsfc.nasa.gov/db-perl/W3Browse/w3browse.pl`

The main ROSAT observation catalogs are the *ROSAT* Master Observation catalog (ROSMASTER) and the *ROSAT* Archival Data catalog (ROSPUBLIC). There are also a nubmer of *ROSAT* source catalogs available for searching with BROWSE. The "basic" BROWSE search facility can be used to search for *ROSAT* data for a given object or celestial coordinate by inputting the name or coordinate of the object and clicking the *ROSAT* checkbox under "X-ray Missions". Advanced searches can allow searching the catalogue of archived observations by position, source name and/or parameter.

After the search is done, the results can be displayed in an HTML table. The results can be sorted by any displayed parameter. Data products can be displayed as clickable links in the web browser or directly downloaded as a tar file.

For detailed information on using BROWSE, see that BROWSE help page at

`http://heasarc.gsfc.nasa.gov/W3Browse/w3browse-help.html`

5 For more information

Please contact


```
Dr. Michael Corcoran
Code 668
Goddard Space Flight Center
Greenbelt      MD 20771
(301) 286-5576
corcoran@barnegat.gsfc.nasa.gov
```