Astronomical Data Analysis Software and Systems XV
ASP Conference Series, Vol. XXX, 2005
C. Gabriel, C. Arviset, D. Ponz and E. Solano, eds.

Building the Chandra Source Catalog

Ian Evans, Stephen Doe, Janet Evans, Giuseppina Fabbiano, Kenny Glotfelty, John Grier, Roger Hain, Margarita Karovska, Michael McCollough, David Plummer, Francis Primini, Arnold Rots, and Jonathan Slavin

Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, MA 02138, USA

Michael Wise

MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract. The *Chandra Source Catalog* will be the definitive catalog of all X-ray sources detected by the *Chandra X-ray Observatory*. For each X-ray source, the catalog will list the source position and a detailed set of source properties, including commonly used quantities such as source dimensions, multi-band fluxes, and hardness ratios. In addition to these traditional catalog elements, the catalog will include additional source data that can be manipulated interactively, including source images, event lists, light curves, and spectra from each observation in which a source is detected.

1. Introduction

Rapid improvements in network bandwidth and storage capacity over the last decade has resulted in an explosion in the availability of archival astronomical data from the world's space- and ground-based observatories. It is now generally recognized that archival data are an invaluable resource that can be used by the astronomical community for numerous scientific investigations beyond those originally planned when the data were acquired. At the *Chandra X-ray Center (CXC)*, we are developing the *Chandra Source Catalog (CSC)*, which is intended to provide simple access to *Chandra* data for individual sources (or sets of sources matching user-specified search criteria) to satisfy the needs of a broad-based group of scientists, including those who may be less familiar with astronomical data analysis in the X-ray regime.

2. Design Goals

The primary design goals for the *CSC*, based on analysis of numerous use-cases, are (1) allow simple and quick access to the best estimates of the X-ray source properties and *Chandra* data for individual sources with good scientific fidelity, and directly support medium sophistication scientific analysis on the individual source data; (2) facilitate easy searches and analysis of a wide range of statistical properties for classes of X-ray sources; (3) provide a user interface that supports *searching and manipulating the actual observational data* for each X-ray source *in addition* to the static properties that are recorded in the catalog; and (4) to the maximum extent possible, include all real X-ray sources detected in all publicly available *Chandra* datasets, while maintaining the number of spurious sources at an acceptable level.

3. Catalog Construction

Each pointed observation interval, or *ObI*, may include data for upwards of several hundred X-ray sources. Although it is natural to record the static catalog information about each source individually, the design goals effectively mandate that the *actual observational data for each source individually* be extracted from each *ObI* that includes that source, and be accessible through the catalog.

The bulk of the *CSC* construction is performed by a set of *CXC Data System (CXCDS) Standard Data Processing (SDP)* pipelines. Catalog construction is conceptually a two part process. First, the observational data for each *ObI* is processed to identify all of the X-ray sources and extract the per-source data and source properties. In the steady state case this happens when the dataset becomes non-proprietary. Second, each source detected in the first step must be merged into the current master catalog.

4. Per-ObI Pipeline Processing

Processing the data from a single *ObI* is performed in two steps. The first step handles the data for the entire *ObI*, detects sources and extracts data for each source for further processing. The second step processes the data for each detected source and extracts the source properties that will be merged into the master catalog.

The *Detect Sources Pipeline* comprises the following steps: (1) reprocess the *ObI* using the latest available calibrations and processing algorithms with conservative data cleaning criteria to minimize the false source rate; (2) generate full field exposure maps to correct the data for sensitivity variations across the field of view; (3) generate background maps in each energy band, and correct for instrumental effects that raise the background; (4) detect sources in broad, soft, medium, and hard energy bands using a wavelet source detection algorithm; and (5) identify source and background regions for each detected source for use in the source properties pipeline.

Similarly, the *Per-Source Pipeline* comprises the following steps, which are executed for each detected source and energy band: (1) extract the photon events in the rectangle bounding the background region for the source, and construct a "postage stamp" image, full resolution exposure map, fluxed image and flux-error image; (2) compute the point spread function (PSF) at the source off-axis position and apply the aspect blur; (3) perform a two dimensional spatial fit to further characterize the detected source; and (4) search for variability using the *Gregory-Loredo* test (Rots 2006) and generate a light curve. If the detector

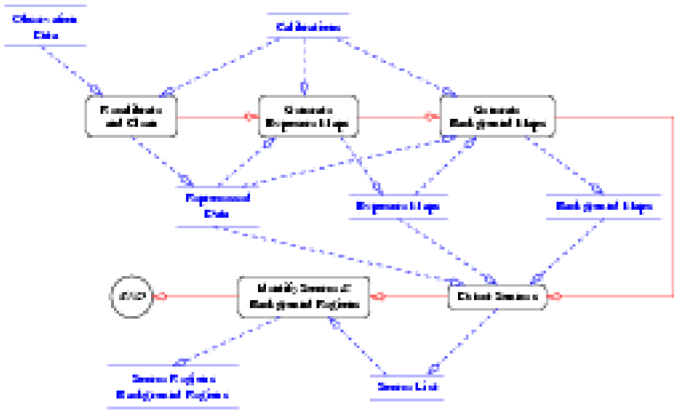

Figure 1. The *Detect Sources Pipeline* high-level flow.

is *ACIS*, we extract the source spectrum and attempt to fit the data with a power-law spectrum plus Galactic absorption.

5. Merge Pipeline Processing

The main functionality of the *Merge Pipeline* is to take the source properties extracted from an *ObI* for each source and merge them into the master catalog. The pipeline first performs a cross-match with the existing catalog, to determine if the source was detected in any previously processed *ObIs*. If no matching sources are found, then the new source properties are promoted to the master catalog. However, if one or more candidate matching sources are found, then the *Merge Pipeline* must identify which of those candidates match the current source. There is not a one-to-one relationship primarily because the *Chandra* PSF varies significantly across the field of view.

Once matching sources are identified, the master catalog source properties are updated by merging the source properties from the contributing *ObIs* based on a set of merging rules.

6. Pipeline Processing Infrastructure

For performance reasons much of the compute-intensive pipeline processing is performed on a multi-node *Beowulf* cluster running *Linux*. Certain tasks that rely on software that is not ported to *Linux* are restricted to executing on a *Solaris* machine. The pipelines are run under the auspices of the *CXCDS Automated Processing (AP)* infrastructure (Grier, Plummer, & Glotfelty 2006).

Pipeline processing is initiated for a pointing when the observation data become public. Merge pipeline processing is triggered for the set of sources identified by a per-pointing pipeline once the latter completes and the relevant data have been ingested into the archive.

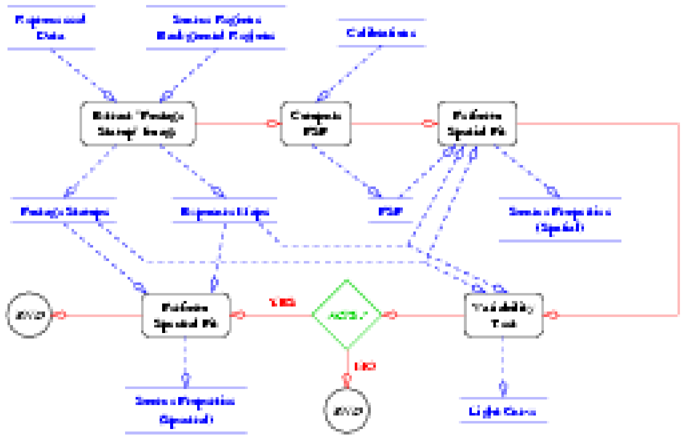

Figure 2. The *Per-Source Pipeline* high-level flow.

7. Catalog Storage and User Interfaces

Although the static source properties information that comprises the *CSC* will be stored in *SQL* databases, the per-*ObI* observational data for each source will be stored as *FITS* format files in the *Chandra Data Archive*. Using this approach instead of recording the data as database blobs leverages existing archive software and provides a file-based interface that is compatible with existing *Chandra* portable data analysis package *(CIAO)* tools, at the expense of more data storage overhead for the *FITS* file headers.

Updates to the master catalog will occur continually as new observations become public. To ensure traceability, a history of updates will be maintained so that the state of the catalog at any point in time is recoverable. *Snapshots* of the catalog will be obtained regularly and made available to the user community in addition to the live (continually updated) version.

Access to the catalog will be provided in two ways. First, a *graphical user interface (GUI)* will be provided that will allow users to peruse the catalog, perform queries on real or virtual columns, display results, and download data files. Second, a *scripting interface* will provide direct access to the catalog and data from users' data analysis scripts. Such scripts will be able to query the catalog and directly download data in file format or into memory, and then directly manipulate the data to perform further searches.

Acknowledgments. Support for development of the *CSC* is provided by the *National Aeronautics and Space Administration* through the *Chandra X-ray Center*, which is operated by the *Smithsonian Astrophysical Observatory* for and on behalf of the *National Aeronautics and Space Administration* under contract NAS 8-03060.

References

Grier, J., Plummer, D., & Glotfelty, K. 2006, this volume, [P.125]

Rots, A. 2006, this volume, [P.116]