

Status of the Chandra Source Catalog Project

The Chandra Source Catalog (CSC) project recently reached a significant milestone, completing a science review conducted by an international committee of experts in X-ray astronomy and astronomical catalog construction. The committee described the development of the Chandra Source Catalog as “an important, useful, and exciting project” that is in addition “blazing a path for other facilities.”

The CSC will be the definitive catalog of all X-ray sources detected by the Chandra X-ray Observatory, and 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. 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 elements, the catalog will include additional source data that can be manipulated interactively, by the user, including images, event lists, light curves, and spectra for each source individually from each observation in which a source is detected.

Design Goals

The primary design goals for the CSC are to (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 tabular properties that are recorded in the catalog; and (4) include all real X-ray sources detected down to a predefined threshold level in all of the public Chandra datasets used to populate the catalog, while maintaining the number of spurious sources at an acceptable level.

Catalog Releases, Characterization, and User Interfaces

The CSC will be released to the user community in a series of increments with increasing capability. The first release of the catalog is expected to include a subset of the elements described here. Each release of the catalog, including the first, will be accompanied by a detailed characterization of the statistical properties of the catalog to a well defined, high level of reliability. Key properties that will be characterized include limiting sensitivity, completeness, false source rates, astrometric and photometric accuracy, and variability information.

Both the tabulated source properties and the individual pointed observation source data (source images, event lists, etc.) that comprise the CSC will be stored in the Chandra Data Archive. The former will be recorded in SQL databases, and the latter will be stored as FITS files. Using this approach leverages existing archive software and provides a file-based interface that is compatible with existing CIAO tools. To ensure traceability, a history of updates will be maintained so that the state of the database at any point in time is recoverable.

The contents of each catalog release will be carefully controlled. They will include the subset of sources extracted from the database that pass quality assurance checks and conform to the characterization requirements established for the release.

Access to the catalog will be provided in two ways. First, a graphical user interface will be provided that will allow users to peruse the catalog, perform queries on real or virtual columns, display results, and

download data files. The user interface will comply with published Virtual Observatory standards such as SIAP, VOTable, and ADQL. Second, an application programming interface will provide direct access to the catalog and data from users' data analysis scripts. Such scripts will be able to query the catalog, directly download data, and manipulate the data to perform further searches. The first release of the catalog is expected to include a simplified user interface that does not include all of the capabilities described here.

Catalog Construction

Each pointed observation may include data for tens to hundreds of X-ray sources within the field of view. In addition to recording tabular catalog information about each source individually, the design goals mandate that the actual observational data for each source individually be extracted from each pointed observation that includes that source, and be accessible through the catalog.

The bulk of the CSC construction will be performed by a set of CXC Data System processing pipelines that run using the same automated processing (AP) infrastructure that is used to run standard data processing. For performance reasons, the compute-intensive pipeline processing will be done on a multi-node Beowulf cluster running Linux, and the AP infrastructure has recently been modified to support this capability. Most of the catalog processing steps are performed using existing CIAO data analysis software, although several new tools have been or are being developed for certain tasks that cannot be done using the current CIAO tool set. These new tools will be included in future CIAO releases.

Catalog construction is conceptually a two part process. First, the observational data for each non-proprietary pointed observation is processed to identify the X-ray sources and extract the per-source data and source properties. Second, the source properties for each source that is detected in more than one pointed observation must be reconciled.

Pointed Observation Pipeline Processing

Processing the data from each pointed observation is performed in two steps. The first step handles the data for the entire field of view, 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 ultimately be merged into the master catalog.

The Detect Sources Pipeline reprocesses each pointed observation using a defined set of calibrations and processing algorithms. Full field exposure maps and background maps are generated, and sources are detected in several energy bands. Source and background regions for each detected source are determined for use in the subsequent Per-Source Pipeline.

The Per-Source Pipeline executes for each detected source. The pipeline extracts the photon events in the rectangle bounding the background region for the source, and construct a “postage stamp” image of the region together with a full resolution exposure map and an image of the point spread function (PSF) at the source position. Various spatial, spectral, and temporal source properties are extracted at this point, both directly from the data, and by performing model fitting (for example, fitting the PSF to the postage stamp image) where appropriate.

Merge Pipeline Processing

The main functionality of the Merge Pipeline is to take the source properties extracted from each pointed observation in which the source is detected and merge them together for inclusion in the catalog. The pipeline first performs a cross-match to determine if the source was detected in any other pointed

observations. If no matches are found, then the new source properties are promoted. 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 source information is updated by merging the source properties from the contributing pointed observations based on a set of merging rules.

-- Ian Evans for the Chandra Source Catalog project team