To build release 2.0 of the Chandra Source Catalog (CSC 2.0), we require scientific software tools and processing pipelines to evaluate and analyze the data. Additionally, software and hardware infrastructure is needed to coordinate and distribute pipeline execution, manage data in, and handle data for Quality Assurance (QA) intervention. We also provide data product staging for archive ingestion.

Release 2.0 utilizes a database driven system for operations and production. Included are five instances of the Automatic Processing (AP) system (Source Detection, Master Match, Source Properties, Limiting Sensitivity and Convex Hulls). This infrastructure: executes the data analysis tools that are organized into pipelines that process the data in order to determine the source positions and properties that populate the catalog, produces the tabulated properties for the sources along with corresponding file-based data products (in FITS format) which are stored in the archive and accessible to the user, and allows metadata to be used to manage processing. This database: defines the "CSC Elements" (stacks and ensembles) for the five CSC Modules for processing, records the status of processing of the CSC Elements in each CSC Module, regulates the Catalog Automatic Processing (Catalog AP) of the CSC Elements between and within the CSC Modules, and The HPCC supports the data intensive pipeline processing.

The original instance of the Automatic Processing system, called "Standard AP," is used to process observation data from the Chandra Observatory. Standard AP: determines the order of pipeline processing, prepares the pipeline inputs, executes the pipelines, handles the data product outputs, handles the retrieval and ingest of data products to the data archive, and "Catalog AP" is an extension of Standard AP. In addition to basic pipeline functions, for CSC 2.0 processing, Catalog AP: distributes data and pipeline processing on an HPCC, provides Quality Assurance (QA) intervention and associated re-processing prior to archive ingestion. There is an instance of Catalog AP for each of the CSC Modules.

A "CSC Stage" is a CSC Module processing component that kicks off one or more pipelines (e.g., Source Properties Master). The "Process Manager" distributes the set of pipelines for the CSC stage over the HPCC (e.g., one pipeline per source). The HPCC executes these pipelines in parallel.

This work has been supported by NASA under contract NAS 8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center.

Abstract

To build release 2.0 of the Chandra Source Catalog (CSC 2.0), we require scientific software tools and processing pipelines to evaluate and analyze the data. Additionally, software and hardware infrastructure is needed to coordinate and distribute pipeline execution, manage data in, and handle data for Quality Assurance (QA) intervention. We also provide data product staging for archive ingestion.