Astronomical Data Analysis Software and Systems XVI **P3.09**
ASP Conference Series, Vol. XXX, 2006
R. Shaw, F. Hill and D. Bell, eds.

Chandra Data Archive Support for the Chandra Source Catalog

P. Zografou, P. Harbo, M. Tibbetts, D. Van Stone

Smithsonian Astrophysical Observatory, 60 Garden St., Cambridge MA 02138

Abstract. The Chandra Data Archive is being extended to include the Chandra Source Catalog consisting of a list of sources with their properties and their associated data products. The system supports periodic catalog releases as well as an evolving catalog with snapshots for any point in time. Catalog entries are associated with their corresponding observation level data which can be accessed through catalog searches. In this paper we present how the existing archive architecture accommodates the new requirements for catalog storage, searches, interface with the processing pipelines and access by external users.

1. Chandra Data Archive Overview

The Chandra Data Archive (CDA) includes a number of servers which manage the archive contents and respond to client requests (Figure 1). Servers of different types are dedicated to different tasks: archive servers manage data files including observing proposals, mission planning schedules and observation data products; database servers manage databases of proposals, Chandra users, telescope monitoring data, the observing catalog (OCAT), file metadata and archive operational data; Java/Web servers host the middle tier of Search and Retrieval (SR) applications; and FTP Servers provide alternative access to copies of the public data products.

User access to the archive is provided by client applications. Local client applications, including *arc4gl* for sending language commands to the archive servers, are used by the Chandra X-ray Center (CXC) operations to ingest and retrieve files. A Java client for SR, *ChaSeR*, is available for download and can be used to access the archive remotely. A number of web-based applications are also available for SR and CXC operational tasks.

The archive stores multiple versions of data files. Version dependencies among standard data products are stored in databases with file metadata and are used for the retrieval of consistent data distribution packages. All versions are available internally at the CXC. A single 'default' version is available to the public. Unlike data files, data in databases are not versioned and only their 'default' version is stored.

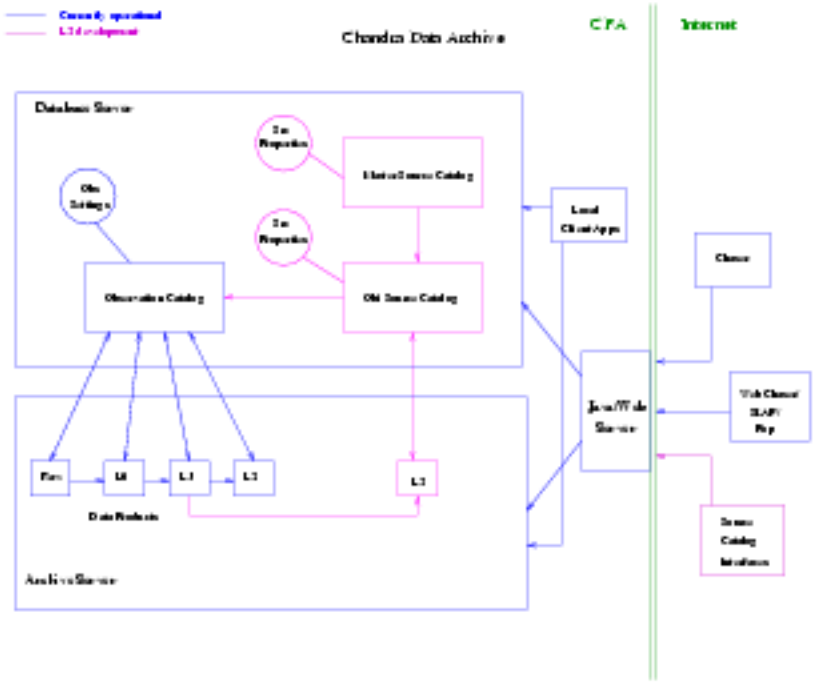

Figure 1. Archive Architecture.

2. Chandra Source Catalog Archive Requirements

The archive must store and make available multiple versions of the data that constitute the Chandra Source Catalog (CSC) including: observation data products of processing level 3 (L3) such as event list, image, exposure map; sources detected in individual observations (source-by-observation) including a source properties database and the L3 source data products such as image, spectrum, psf, light curve; and master sources merged from sources-by-observation and their properties database. Links between master source, source-by-observation and OCAT observations and links between L3 and lower level data products must also be maintained.

The archive must support: ingestion of new sources and updates of master sources from observations as they become public; catalog updates from reprocessing initiated at any level; periodic catalog releases; and catalog snapshots for any point in time.

The archive must provide adequate CSC physical storage for the duration of the Chandra mission with estimated CSC data volume of 0.6 TB per year of Chandra mission comprising: 1,000 observations with 18,000 observation level data files; 140,000 sources-by-observation with 5,500,000 source data files; and 140,000 master sources, initially same as sources-by-observation.

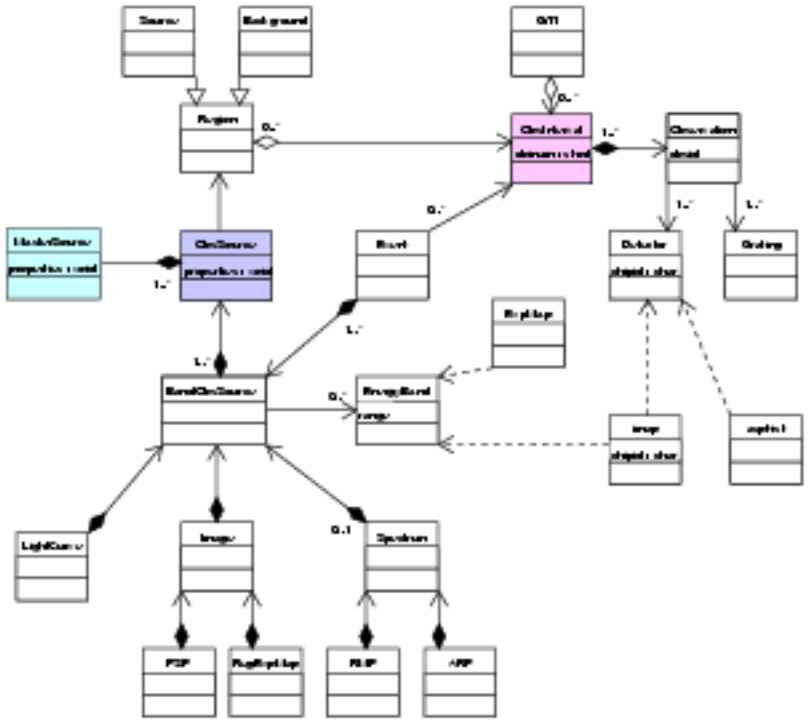

Figure 2. Source Catalog Entities.

3. Archive Architecture for the Source Catalog Support

The CSC is incorporated in the CDA with addition of databases and L3 products storage on dedicated CSC servers within the existing archive architecture. Lower level databases that are tightly coupled with the CSC are replicated to the CSC servers for easier access.

The *arc4gl* archive server language is extended with L3 options for access of the data products. Additional web-based user interfaces, using the existing Java/Web middle tier, are planned for browsing of the databases and data retrieval.

The CSC databases store all versions of sources and implement many-to-many relationships between master sources and sources-by-observation (Figure 2). Source-by-observation records are also linked to their observation records in OCAT. In order to meet the multi-version source requirement, the OCAT is enhanced to store multiple versions of data fields linked to the CSC. All versions of L3 data products are kept in the archive and have links to their associated source-by-observation versions.

In order to support periodic releases and snapshots, the CSC database records have *creation time stamp* and *catalog release label*. A user view of the catalog selects by label to see a release or by time for a snapshot. In both cases, the view selects qualified master sources, then follows the versioning links to sources-by-observation and associated data products.

Acknowledgments. This project is supported by the Chandra X-ray Center under NASA contract NAS8-03060.

References

Fabbiano, this volume, [O2.1]

Evans et al 2006, in ASP Conf. Ser., Vol. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco: ASP), 65

Zografou et al. 2006, in ASP Conf. Ser., Vol. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique (San Francisco: ASP), 711

Zografou et al. 2004, in ASP Conf. Ser., Vol. 314, ADASS XIII, ed. F. Ochsenbein, M. Allen, & D. Egret (San Francisco: ASP), 145