Chandra Data Archi e Support for the Chandra Source Catalog

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

Smithsonian Astrophysical Observatory

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 accomodates the new requirements for catalog storage, searches, interface with the processing pipelines and access by external users.

Chandra Data Archive Overview

- **Multiple Servers for different tasks** (Fig 1)

  1. Archive servers manage data files including proposals, mission planning schedules and data products.
  2. Database servers manage databases of proposals, users, telescope monitoring, the observing catalog (OCAT), file metadata and archive operational data.
  3. Java/Web servers host the middle tier for Search and Retrieval (SR).
  4. FTP Servers provide alternative access to copies of the public data products.

- **Client access**

  1. Local client applications are used by Chandra Xray Center (CXC) operations, including *arc4gl* to ingest and retrieve files from the archive servers.
  2. Remote Java client for SR (*ChaSeR*).
  3. Web based applications for SR and CXC operations tasks.

- **Data Versioning**

  - Multiple versions of all data files kept in the archive.
  - Version dependencies among products in the standard data distribution package for each observation are kept with file metadata.
  - Only the most recent versions are publicly available.
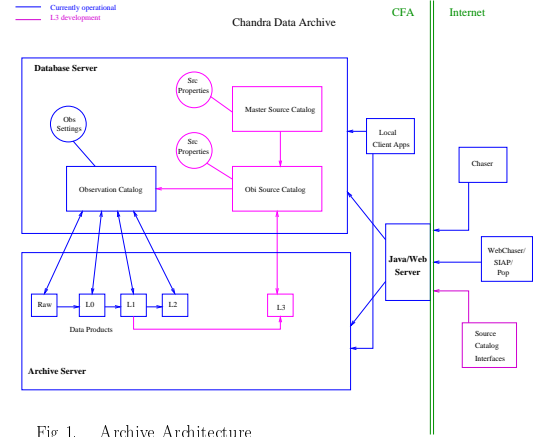 - No versioning of databases. Records are updated individually.

Fig 2. Source Catalog Entities

Fig 1. Archive Architecture

Chandra Source Catalog (CSC) Archive Requirements

- **Contents**

  Store and make available multiple versions of

  1. Observation L3 data products e.g. events, image, exposure map.
  2. Sources detected in individual observations (source-by-observation); properties database; L3 data products e.g. image, spectrum, psf, light curve.
  3. Master sources merged from sources-by-observation; properties database.
  4. Links between master source, source-by-observation, L3 and OCAT observations.

- **Operations**

  Provide support for

  1. Catalog updates from reprocessing.
  2. Master source updates from new public observations.
  3. Periodic catalog releases.
  4. Catalog snapshots for any point in time.

- **Data Sizing**

  Estimated L3 data volume for one year of Chandra mission

  - 1,000 observations, 18,000 observation data files.
  - 140,000 sources-by-observation, 5,500,000 source data files.
  - 140,000 master sources, initially same as sources-by-observation.
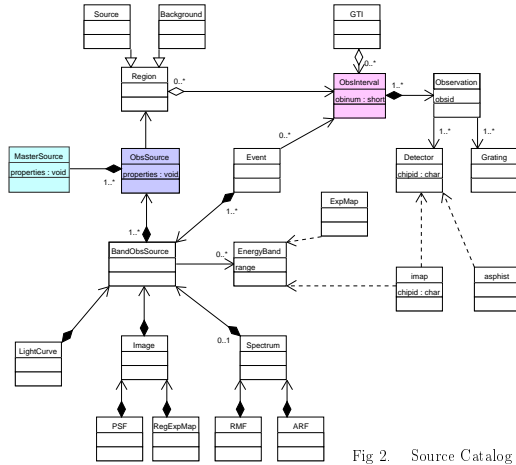 - Total data volume 0.6 TB

Archive Architecture for the Source Catalog Support

- **Data Management**

  - Dedicated archive and database servers for handling of large data volume w/o interfering with other CXC operations.
  - Replication of tightly coupled data e.g. OCAT on to the CSC servers.

- **Versioning and Data Links**

  - All versions of sources are kept in databases. Many-to-many relationship between source-by-observation and master source (Fig 2).
  - All versions of L3 observation records are kept in the database with links to observations in the OCAT. OCAT enhancements to store mutliple versions of observation attributes needed by the CSC.
  - All versions of L3 data products are kept in the archive with links to associated versions of lower level data products.
  - A matching set of L3 observation, source-by-observation and master source entity versions and their associated data products become available for retrieval simultaneously.

- **Archive Interfaces**

  - Extend the Archive server language, *arc4gl*, to ingest, browse and retrieve CSC data products.
  - Web based interface using middle tier Java/Web servers.

- **Releases and snapshots**

  - Database records have *creation time stamp* and *catalog release label*.
  - User view selects by label for a release or by time for a snapshot.
  - The view selects qualified master sources, then follows the versioning links to sources-by-observation and associated data products.

- **Development Status and Plans**

  - Internal archive release integrated with L3 pipelines including archive and database server components, data links and version dependencies, *arc4gl*.
  - Working on master source database and public interface requirements.
  - Product definitions are evolving, structures will need to adapt.