Chandra Data Archive Operations: Lessons Learned

Michael L. McCollough, Arnold H. Rots, and Sherry L. Winkelman

Smithsonian Astrophysical Observatory, 60 Garden Street, MS-67, Cambridge, MA 02138 U.S.A.

ABSTRACT

We present a discussion of the lessons learned from establishing and operating the Chandra Data Archive (CDA). We offer an overview of the archive, what preparations were done before launch, the transition to operations, actual operations, and some of the unexpected developments that had to be addressed in running the archive.

From this experience we highlight some of the important issues that need to be addressed in the creation and running of an archive for a major project. Among these are the importance of data format standards; the integration of the archive with the rest of the mission; requirements throughout all phases of the mission; operational requirements; what to expect at launch; the user interfaces; how to anticipate new tasks; and overall importance of team management and organization.

Keywords: Chandra, Data Archive, Space Based Observatory, Observatory Operations, Data Processing, Data Distribution, X-Ray

1. INTRODUCTION

The Chandra X-ray Observatory is one of the four NASA "Great Observatories". An important operational part of this mission was the establishment and operation of the Chandra Data Archive (CDA). Many of the experiences in developing and running the CDA may also be shared by other large missions in creating their archives. Presented in this article is a discussion of the lessons learned from CDA.

In this paper we review the development of the archive in the context of the Chandra mission. This includes an overview, preparations for launch, operations, and the unexpected developments. Discussions of the lessons learned in terms of data standards, archive integration, necessary requirements, operations, the realities of launch, user interfaces, new tasks, and the management of the archive are given. It is hoped this review will aid other missions in the establishment of their archives.

2. DEVELOPMENT OF THE CHANDRA DATA ARCHIVE

2.1. Chandra Mission

The Chandra X-ray Observatory (CXO) is an orbiting observatory, launched in July 1999, which is capable of performing high resolution spatial (sub-arcsecond) and spectral ($E/\Delta E \sim 1000$) observations at X-ray energies $0.2 - 10$ keV. The observatory has two main instruments, the Advanced CCD Imaging Spectrometer (ACIS) and the High Resolution Camera (HRC). These can be supplemented by two sets of transmission gratings used for high resolution X-ray spectroscopy.

The Chandra X-Ray Center (CXC) is responsible for the entire mission life cycle of Chandra observations: issuing annual "Calls for Proposals", receiving observing proposals, managing the peer review, Mission Planning (MP), flight operations, receipt and processing of telemetry, Verification and Validation (V&V), data archiving, and data distribution (to the guest observers and the public). The CXC is operated by the Smithsonian Astrophysical Observatory at the Harvard-Smithsonian Center for Astrophysics in Cambridge, MA, under contract with NASA.

Further author information: (Send correspondence to M.L.M); E-mail: mccolml@head.cfa.harvard.edu

This paper was published in *Observatory Operations: Strategies, Processes, and Systems*, edited by David R. Silva, Rodger E. Doxsey, Proc. of SPIE Vol. 6270, 62700O, (2006), and is made available as an electronic reprint with permission of SPIE.

2.2. Overview of the Archive

The CDA has a central position in the operation of the CXC. Its primary role is one of storage and distribution of all data products including those that users of the observatory need to perform their scientific studies using Chandra data. Equally important, it stores and maintains databases containing the data and metadata necessary for the various components of the CXC to perform their tasks relative to maintaining and operating the CXO.

In past missions it was not uncommon to separate the archive databases into two separate entities, one used to upload data to the spacecraft and another to receive the downloads from the spacecraft. This had the advantage of simple design, greater flexibility, and possibly a quicker and easier response to operational problems and issues. But this approach has the potential of problems with data integrity and discrepancies between the databases. For these reasons the CDA was implemented with its various databases linked.

2.2.1. Components of the Chandra Archive

The CDA can functionally be broken into seven components (see Rots et al. 2002 and Spitzbart et al. 2002 for a discussion of these components). These are represented by (see Figure 1):

- `Observation Catalog:` databases containing information related to observation configurations;
- `MTA Databases:` Monitoring and Trends Analysis (MTA) database ;
- `Processing Status Database:` database for tracking the processing history of data;
- `Data Products Databases:` databases which catalog the data warehouse;
- `Data Warehouse:` the physical disk storage area for the data;
- `Ingest Interface:` the method by which data products are ingested into the data storehouse;
- `Search and Retrieval (S&R) Interfaces:` provide access to the data.

2.2.2. Clients and Services

In creating and running an archive one needs to identify who the clients will be. It is important to realize that different groups may have very different needs and require very different responses from the archive operators. For CDA most users fall into one or more of the following groups:

- `Guest Observers:` They are the scientists that propose observations and use the data from the observatory to perform scientific analyses. They need to be able to access the data from a variety of platforms and access tools.
- `Astronomical Community (Archival Users):` Eventually all Chandra data become available to the public. The general astronomical community represents an important part of this public. Their experience may range from an expert user of X-ray data to being a relative novice. As such, the interface for these users may be somewhat different than for the typical Chandra observer.
- `Data Systems (DS):` The operations personnel in the DS branch are responsible for taking the data from raw telemetry, processing to final data products, and distributing the data (Evans, I. et al. 2006). There are also many intermediate products that the observer may never see, but that will have to be accessed by various members of DS. Among the tasks this group performs are: (1) data processing (Nichols et al. 2006); (2) V&V (Nichols et al. 2006); (3) operations software development and maintenance (Evans, J. et al. 2006, Plummer et al. 2006); (4) data calibration support (Graessle et al. 2006); (5) hardware support (Paton 2006); and (6) archive operations (Rots et al. 2002 and this paper).
- `Mission Planning:` This group is responsible for planning and scheduling the observations that will be executed on board the spacecraft. They need rapid access to the observation configuration and a current view of the status of observations (unobserved, scheduled, partially observed, observed).

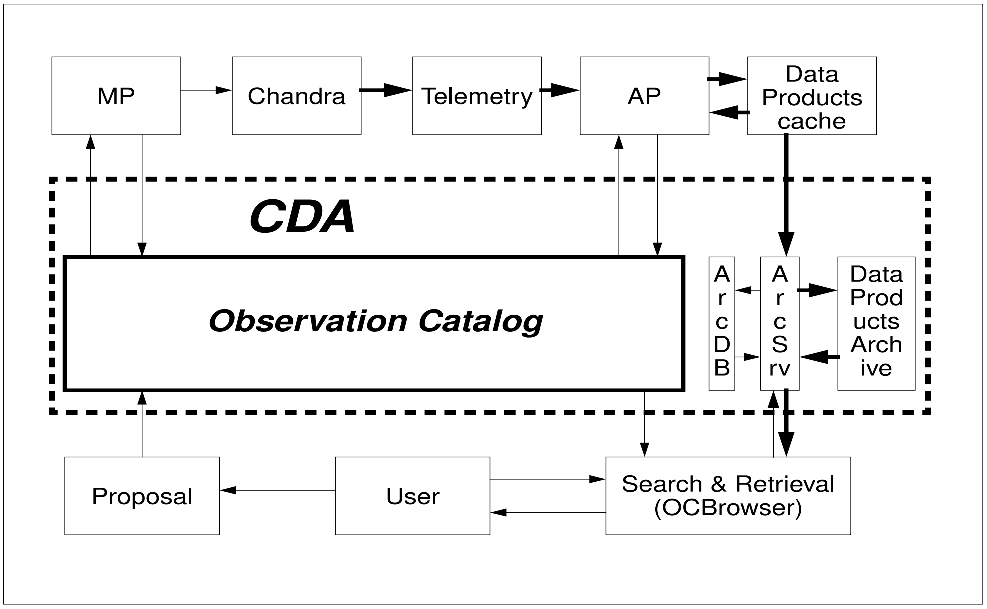

Figure 1. The figure shows schematically the CDA. The sizes of the boxes in the CDA say nothing about the data volume, but the thickness of the arrows indicates whether the data transfer volume is heavy or not. "Data Products" is the data warehouse, "ArcDB" is the data products databases, and "ArcSrv" is the interface server. The MTA databases are not shown.

- `User Support:` Members of this team act as the interface with the observer. They have the responsibility to make sure that the correct instrument parameters have been set for a given observation to help maximize the scientific return from the observation.
- `Instrument and Calibration Scientists:` This team has responsibility for the health and safety of the instruments and spacecraft. They very often need access to data not normally used by the general observer.
- `MTA:` The Monitoring and Trends Analysis (MTA) group is responsible for providing tools and information which allow other groups to make effective decisions on the production of quality science data from CXO.
- `Chandra Director's Office (CDO):` The Director's Office is responsible for the interaction of CXO with the entire astronomical community. Among their responsibilities are: soliciting new observing proposals, peer review, grants (distribution, tracking, and reporting), publication statistics, and reporting to the Chandra User's Committee and NASA management.
- `Project Management:` This includes CXC management and the NASA oversight of the CXO. They need to be continually updated on: observations performed, data volume, publications, and number of users of the data.

2.3. Preparations: Establishing the Archive before Launch

A limited version of the archive was implemented by the DS software group before the ground testing of the Chandra optics and focal plane instruments at the NASA Marshall Space Flight Center. The experience of the ground test provided a determination of the necessary data and metadata for the archive (Zografou et al.

1998). At this time it was realized that more careful thought needed to go into the design of the archive and its overall interaction with the various parts of the project. A full time scientist was needed to oversee the archive and provide contact with the other CXC players. Near the end of 1997 the position of archive scientist was established.

In 1998 the archive scientist worked with the various Chandra teams to establish the data format standards and determine the necessary metadata needed for the archive. A new version of the archive was established and testing was begun to determine the interfaces that would be necessary for the various users of the archive. At the end of this paper are the links to the CDA website, CDA data standards, FITS standards, and the recommended FITS standards for the high energy astrophysics community.

Roughly a year prior to launch the archive operations group (*arcops*) was established. The roles of this group consisted of archive scientist, archive system manager, database administrator, and data technicians.

2.4. Transition to Operations

In early 1999 simulations involving the archive were conducted as part of an end to end test of the entire system. It was at this time that the issue of the resource needs of the archive and the mode in which it would operate were finalized. Among the questions that were addressed:

- How much server capacity is needed? This was an estimation of what the load would be on the system both internally by CXC and from external users.
- How will the hardware be configured and where will it be located? This addressed two important concerns. (a) *Redundancy*: It had been decided earlier (Zografou et al. 1998) that there would be two servers. They would be located in different locations, reducing the likelihood of the loss of both servers. This would make recovery from a serious disruption of one of the two sites much easier. (b) *Load Sharing*: Having two servers also allowed the splitting of the load, with one server handling the requests from operations and the other server handling external users.
- What portion of the data should be kept online? Initial designs called for the majority of the data to be copied to DVD and only a small fraction of the available data kept online. But due to uncertainty in the DVD formats standards, a final decision was postponed.
- What should be the replenishment policy for data storage? It was decided to monitor the archive data volume and to have a policy which ensured that we would always have enough storage capacity for the next six months. With disk prices decreasing and capacity increasing it would be foolish to procure capacity far into the future. It was also decided to store the archive data on robust general purpose file servers. This hardware is more costly, but it more than offsets the personnel cost incurred with cheaper hardware solutions.
- What will the network be like? Given the different server locations, the large number of users, and the data volume, a careful evaluation of the needs was necessary. For a review of the network structure which was put in place see Paton (2006).
- What type of data security would be implemented? For a large data center data security is an important issue. For CXC there is layered network security with a firewall at the perimeter of the network (Paton 2006). In addition, an extra layer of security was implemented for the archive. The archive is isolated from the rest of the CXC system with special passwords, procedures, and privileges needed to access the archive data.
- What type of coverage and support would be necessary by the *arcops* team? This addressed issues such as emergencies, who would need to respond, and contact information of other members of the CXC. For the first month of operations it was decided to provide 24/7 support, to be reduced, or at least reconsidered after that period.

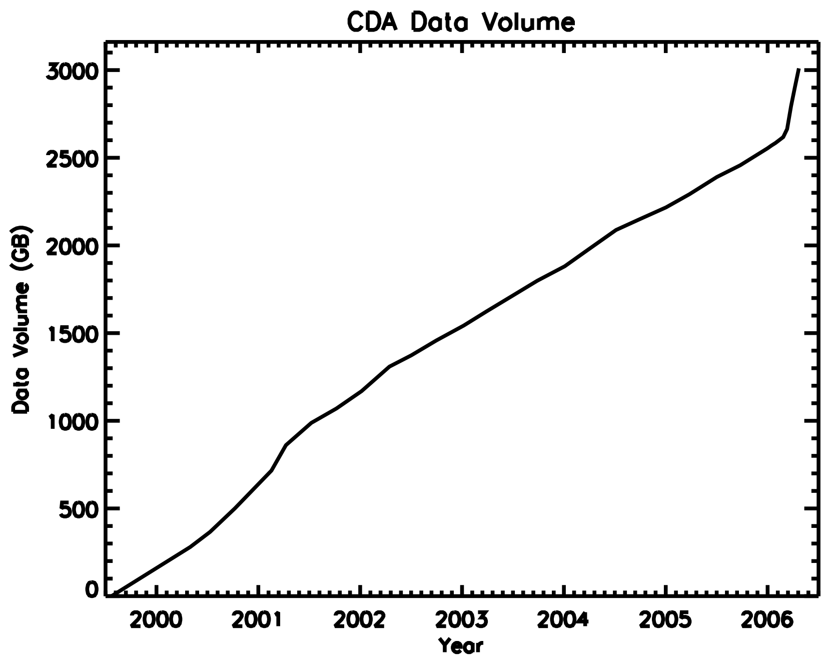

Figure 2. The figure shows the data growth of the CDA. It starts in July of 1999 and continues through late April of 2006. Note the increases in 2000/2001 and in 2006 that are the result of major reprocessings of all of the data accumulated in the Chandra archive.

2.5. Operations

Personnel were trained and prepared for the July 1999 launch of the spacecraft. With the beginning of operations started the real-life test. At this point feedback from operations became important to assess and identify which components were having problems. Estimates of the necessary software support were updated. The consequences of an ever-changing mission environment and corresponding changes in requirements began. As the mission progressed, the archive has steadily increased in size (see Figure 2).

2.5.1. Initial Issues

In the initial part of the mission, the *arcops* group dealt with the following issues:

- There were no user interfaces for search and retrieval of the data. This resulted in the creation of the *Provisional Retrieval Interface*.
- Data distribution was a very manual process. *Arcops* had to gather data products and V&V reports for the distribution of each observation.
- MP interfaces with the Observation Catalog had limitations and unexpected complications. Real-life operations forced a reconsideration of several of the assumptions made for the planning process. This often resulted in manual workarounds for various problems.
- It was also realized early on that the cost of disk storage was dropping with time. It became desirable and feasible to keep all of the data online rather than having to write it to DVD.

Over the next three years the manual components of these tasks became more automated.

2.5.2. Operational Environment

The operational environment has changed dramatically from the start of the mission to the present. The mission can be seen as a progression from an intense early phase of the mission to a more controlled later state. The early and current states can be summarized as follows:

- `Early Operations:`
  - *Intensive:* There was an immediate demand to address problems when they occurred. Failure to address problems could result in operations problems throughout the mission.
  - *Reactive:* Personnel had to respond rapidly to problems and software failures. In this early phase one had to make a constant effort to address problems to keep data flowing.
  - *Procedures Development:* This was the point at which procedures for accomplishing tasks were being defined and evolved to meet the real world environment of operations. This is where one discovered which things did or did not work. For things that did not work solutions had to be found.
- `Current Operations:`
  - *Branching Out:* In the initial part of the mission the effort was to make the archive work. As the mission matured there has been an effort to work on improving the services that are offered. This includes looking at, and taking on tasks, that both other CXC personnel and outside users find useful.
  - *Proactive:* The system has become well enough understood to be able anticipate problems and demands and try to address them before they have a major impact. This is a result of the knowledge gained in the running of the mission.
  - *Established Procedures:* With the knowledge learned, procedures have been defined and established. This helps both the current mission (making it easier for new personnel to be productive) and future missions.

2.5.3. Operational Software

With the start of operations various limitations of the software were found. When the software was used in the operational environment it was discovered that requirements were not complete or that software could not be reasonably used in a manner for which it was designed.

Also the response time from the software development teams is limited. They need to support several different teams. All the teams are competing for the time and programming resources that are available. For major problems there is a reasonably quick response. But there can be considerable delays for lower priority items.

For these reasons it is important that the archive operation team members have programming and scripting skills. In this way smaller tasks, which are important for daily operations, can be responded to rapidly. In addition, these programming skills are important to address the issues mentioned in the next section.

2.6. Unexpected Developments and Unfinished Business

Unanticipated tasks and new requirements have developed as the mission has progressed. Some tasks defined at the beginning of the mission changed substantially (in size and scope) from the original specifications or were not finished by the time of launch.

- `FITS Keyword Dictionary:` Well-defined data and metadata specifications are fundamental for the operation of a large observing facility. For the CXO data products, the FITS keywords have been gathered into a dictionary database which provides easy access to their use and meaning. The functionality, description of the database design, and the display tools for this dictionary is described in Winkelman et al. (2004).

- `Bibliographic Database:` Early in the mission, the CDA started development of a bibliography database, tracking publications in refereed journals and on-line conference proceedings that are based on Chandra observations, allowing users to link directly to articles in the ADS (Accomazzi et al. 2003) from the archive, and to link to the relevant data in the archive from the ADS entries (Rots et al. 2004). As a result a comprehensive database has been developed which is capable of tracking almost all mission-related publications and preserving a wealth of relevant information. This database allows cross-linking with the ADS, literature search from the archive, and the creation of metrics.

  The concept of the bibliographic database did not exist before launch. But it has become apparent that this is a service that the Chandra users have come to expect. The CDA was one of the first to recognize this, pioneer the field, and become the first major observatory with a complete database of this type.

- `Grants and Performance Information (Administrative):` In addition to the operations of the spacecraft and the distribution of data, it was realized that the archive could assist in handling many of the administrative details of the mission. This applies to determining when grants should be released to the guest investigators, when reports are due from the guest investigators, as well as providing information on the publications that are linked to a given grant, and overall publication statistics.

- `MTA Databases:` The MTA group does real-time monitoring, generation of alerts for mission critical components, long term trending of spacecraft systems, analysis of subsystems life expectancies, and anomaly resolution. At the beginning of the mission the archival needs for this group had not been fully established. A discussion of the MTA activities and their initial database can be found in Spitzbart et al. (2002).

- `Data Processing Status:` Due to the complicated processing history of CXO data, it became apparent that a database was needed to track this history on an observation by observation basis. Further complications were added when the archive also had to track different versions of the data for a given observation. The result was the *Processing Status Database* and the *Chandra Observations Processing Status Tool*. A description of this database and its associated tools can be found in Winkelman et al. (2002).

- `Ingest Monitoring:` The ingest monitoring database was designed to track the data checking operations that *arcops* performs for every file that is submitted for ingest to the archive. It tracks each file from submission of ingest by data operations to notification that the file can be deleted from the ingest cache area. Part of this tracking also includes verifying the checksum for all data files.

- `Download Monitoring:` In order to support regular operations, it was necessary to develop a database that records and monitors the user activities that affect the archive servers. This database provides information on the number of users that are browsing or retrieving data at any given time, the archive interfaces that they are using, and how much and what type of data are being downloaded. With this database archive statistics can be gathered and trends can be monitored. In deciding on the information collection strategies, a careful consideration of privacy issues was made. A description of this database and how it is used can be found in Blecksmith et al. (2003) and Winkelman et al. (2006).

- `Mirrors:` At the start of the mission it was decided in order to reduce problems with data volume and flow to Europe that a mirror site would be established at the LEDAS center in the U.K. This overall has worked well. But, this mirror being an exact copy of the archive, there are problems with keeping software licenses and the archive and its software current at LEDAS. Additionally there have been requests for similar data access from other sites around the world. To address this a system of *ftp mirrors* is being established at various locations around the world containing only the current version of the primary and secondary data products for public observations. The data in these sites are routinely updated by the archive software at CXC. The web based search and retrieval interface performs its search functions through access to the archive databases at the CXC. When the user requests a download, the server determines the nearest ftp site to the user and downloads the data from that location. An *ftp mirror* site has been established in India; sites in Japan, Italy, and a conversion of the U.K. site are being worked.

- `Chandra Source Catalog:` Another project where the archive plays a central role is the Chandra Source Catalog. The Chandra Source Catalog will be the definitive catalog of all X-ray sources detected by CXO

(Evans et al. 2006). For each X-ray source, the catalog will contain the source position and a set of source properties. In addition to the traditional catalog static database, the catalog will have a dynamic component (data objects) which will allow the user to manipulate interactively the source images, event lists, light curves, and spectra for each detected source. This must be made to work seamlessly with the current archive.

3. LESSONS LEARNED

The creation and running of the CDA has been a challenging and rewarding endeavor. From our experience there are several important points we would like to note and expound upon in the expectation that they will be helpful to others.

3.1. Data Format Standards

It is crucial that data format standards (though not necessarily the formats themselves) be defined in the very early stages of the project. What has tended to happen is that these standards get defined when the project has started writing the software that will produce data products for users, i.e., at best two years before launch. At that point, instrument teams have been testing their hardware using a variety of convenient formats. This has several negative impacts:

- The software used for producing user data products is not properly tested, and certainly not by people who understand the instruments.
- The software that has been tested and used by the instrument teams and is known to work, is unusable for this purpose since it depends on non-standard data formats.
- Instrument team personnel cannot provide much help in solving problems in the "official" software since they are unfamiliar, both with that software and with the data formats it uses; cross-checking against instrument team software is impossible for the same reason.

All this can be avoided if the data format standards are defined very early on, if the instrument teams commit to using them for their pre-flight (or pre-deployment) software, and if formats get defined according to these standards as the instrument teams need them.

As part of our format standards we required all FITS files to be checksummed. This feature has enabled us to verify the integrity of our products through a simple procedure.

3.2. Integration of the Archive

Another early decision concerns the exact scope of the archive and its databases: which project's components are part of it, or controlled by it, and which are not. Specifically, this requires (at a rather high level) setting the granularity of the activities that are controlled by the archive's databases and how much detail the archive is aware of. More about this in the next section.

3.3. Requirements

Development of the archive is, of course, driven by requirements and these requirements need to be formulated explicitly. The approach to requirements should take the following points into consideration:

- The requirements on requirements need to be formulated early, because they are needed before the design can start.
- These need to be obtained from all players involved. Requirements need to be complete before the design process starts, particularly because changes in database design are more costly than changes in general software design.

- However, one needs to be aware that requirements can (and will) change during implementation and this should be taken into account in the design. It means that there needs to be close contact between the archive designers/implementors and their clients.

- Even so, things are certain to change upon deployment and in the operational phase. Real-life operations are always different from what we think they will be. The archive needs to be responsive to changing requirements and the design needs to support it.

- As noted in the previous section, the granularity of the archive components needs to be carefully considered. At the level of the project's players this means that one needs to carefully weigh which of the players' activities are controlled through the archive and which are performed by the players outside the archive's domain. If there are archive connections (e.g., a player's "private" activity that requires information from the archive), the interfaces need to be defined specifically. An example in our case concerned the mission planning software: it seemed a sensible idea to lock proposed observations that a mission planner was considering; in practice, it did not quite work out that way since several planners were always working simultaneously on different weeks and they got in each other's way by locking observations that might end up in one week or another.

3.4. Operations

When preparing for live operations the following is a list of issues that should be addressed before going live:

- `Staffing:` Clear work efforts should be estimated and appropriate staff levels sought. It can be expected early in the mission that extra time and effort will be needed. But there should be a transition to normal operations. In staffing it is important to realize that the workload of the archive will not decrease with time and that the archive will outlive the rest of the mission.

- `Training:` The personnel should be well trained in the routine tasks that need to be performed. In real operations new important tasks will rapidly present themselves. Hence the better the group is able to deal with known tasks the more able it is to address new problems and tasks. It is also desirable that there be sufficient cross-training to prevent single point failures.

- `Documentation:` This is the time when written procedures and operations manuals should be developed. It is desirable that some of these documents be a "living" document (using technology like a *Wiki*) so it can be relatively rapidly updated to adjust the procedures to new requirements and a changing work environment.

- `Hardware Evaluation:` There should be a complete evaluation of the hardware needs and status before operations. How much hard disk space do we have (and need), what server capacity do we need, etc. Identifying potential hardware bottlenecks and addressing them well before going live will prevent problems.

- `Software Testing:` Run as many test cases as possible before going live. Setup and run test servers and databases. These should be used constantly throughout the mission to test new versions of the archive software as the system and mission evolves with time. This is the point where standard testing procedures and plans are developed. The system will go live at a certain time no matter how much testing has been done (ready or not here it comes).

- `Shadow System Testing:` Some software builds will be complex enough and be very difficult to recover from if mistakes are discovered after installation. For these software builds it is desirable to test the software through a shadow archive and shadow processing system. The shadow system starts with a snapshot of the operational archive and a processing system with the new software. Data is then processed through both the normal processing system and the shadow system. A comparison is then made between the results of current processing system with the new system and discrepancies are noted. Similarly, the production and shadow archives are checked for differences.

- `Risk Assessment:` A risk assessment should be done for the archive. What will happen if there is a power failure? Can the expected data flow always be handled? If certain archive software fails are there work arounds that will allow the necessary tasks for spacecraft operations to continue? These and other problems should evaluated as launch approaches.
- `Emergency Plans and Response:` There should be in place plans and responses to issues found in the risk assessment. Address such issues as absences, on-call schedules, and contact information.

3.5. Managing Proprietary Data

Most of the Chandra science data is subject to a proprietary period of one year, though some observations have shorter proprietary periods and others have none. This requires solutions in three areas:

Authentication and authorization: We instituted three classes of archive accounts, each with simple password authentication:

- Privileged mission personnel accounts that provide access to all data and to certain databases; permission roles are granted on an individual basis.
- Principal Investigator (PI) accounts that authorize access to all data that are associated with that PI's proposals.
- Proposal accounts that authorize access to all data associated with a single proposal; this allows a PI to share data access with collaborators on a selective basis.

Protecting the proprietary nature of data that are to be downloaded: We rejected the solution to encrypt the downloads since it requires users to install extra software. Instead, we opted for placing download data in an anonymous ftp staging area behind a main directory that is not world-readable, in a layer of directories with random names: the URL that is provided to the user effectively acts as a password.

Public release of data upon expiration of the proprietary period: One should be aware that it is essential to be able to adjust and override public release dates. And when it happens (note: not *if*, but *when*) that data are accidentally released, one needs a recovery procedure: withdraw the data from public view, and identify and contact users who obtained them during the open window; note that this puts a requirement on download logging capabilities.

3.6. Launch: Handling Things that Did Not Get Done

Rest assured, when deployment or launch comes about, there will be quite a few components that are not yet ready. In our case it was, among others, a user data retrieval interface, a data distribution system, and the V&V system. The archive operations team will need to fill that void and it had better be prepared for it. This involves taking a careful inventory of missing pieces shortly before deployment, a weighing of alternatives, and a balanced allocation of resources. Above all, it means making sure that the team has the right skills set to match the tasks (as well as matching the solutions to the available skills).

Early in the mission there may be times that heroic efforts may be needed in order to make the system work and avoid potential disaster. But as the mission progresses this should become much less. If this is not the case a careful evaluation of your tasks and staffing should be made.

3.7. User Interface

The simple truth is that the interface used by outside archive users will be ever-evolving. The technology evolves and the users' expectations change (become more demanding) over time: an interface implemented seven years ago would not be acceptable today. What this means is that one should not expect the archive's search and retrieval interface ever to be finished and one should budget for this.

3.8. New Tasks

In a similar vein, developments in the user community will present new tasks and responsibilities for the archive that were not yet on the horizon when the original requirements were drawn up. In our case, the bibliographic database is an example: ten years ago, people might have dreamt of it, but did not expect it; today, it is an obvious requirement. The Virtual Observatory is another example. Again, one needs to budget future resources for this.

3.9. Management and Team Organization

Any operational environment that interfaces directly with users and the general public needs to be flexible, agile, and responsive on short notice. Quick, temporary fixes are called for and not only requirements, but also priorities, shift continuously. To be able to respond to this environment requires a rather tight interaction between the operations side and the development side of the archive. Optimal communication is also crucial in order to reconcile permanent development solutions with temporary quick fixes. We strongly recommend that these aspects be considered in future missions. One option would be to bring both teams under a single archive management, but such a move would require careful management of interfaces with other software development teams.

4. CONCLUSIONS

After over six years of operations we have gained considerable experience in the construction and operating of an archive for a major NASA project. It has been taken from concept to a fully functional archive serving international clientele. We have presented a range of lessons we have learned in the process of creating and running the the CDA. Among those which we feel are most important are:

- `Data Format Standards:` Data formats must be established as early as possible. This will drive the interconnectivity of the mission and how well things work together.
- `Requirements:` The requirements are needed as early as possible. But flexibility needs to be build in. As the mission evolves they need to change to meet the demand.
- `Integration:` The archive must be fully integrated into the mission. An early decision needs to be made of what things are included in the archive and what things are not. Major changes in this balance will be difficult if not impossible to realize later on.
- `Staffing and Skills:` One must have the correct staffing along with the appropriate skills. This not only includes database and science skills but programming skills as well.
- `New Interfaces and Tasks:` The user interfaces will be constantly changing, driven by changing needs and by technology. New tasks will constantly present themselves and there will always be a need to be ready for them.
- `Management:` It is important to have good management of the team. It is also important to recognize who should be part of the team and what are their responsibilities.

We hope that this analysis will be helpful to future missions and benefit the development of their archive operations.

ACKNOWLEDGMENTS

Support for development of the Chandra Data Archive 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 NAS8-03060.

5. REFERENCES

1. A. Accomazzi, G. Eichhorn, C. Grant, M. Kurtz, and S. Murray, S., " ADS Web Services for the Discovery and Linking of Bibliographic Records", *Astronomical Data Analysis Software and Systems XII*, H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds., ASP Conf. Ser. **295**, pp 309-312, 2003.
2. E. Blecksmith, S. Paltani, A. Rots, and S. Winkelman, "Chandra Data Archive Download and Usage Database", *Astronomical Data Analysis Software and Systems XII*, H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds., ASP Conf. Ser. **295**, pp 283-286, 2003.
3. I. Evans, S. Doe, J. Evans, G. Fabbiano, K. Glotfelty, J. Grier, R. Hain, M. Karovska, M. McCollough, D. Plummer, F. Primini, A. Rots, and J. Slavin, "Building the Chandra Data Source Catalog", *Astronomical Data Analysis Software and Systems XV*, C. Gabriel, C. Arviset, D. Ponz and E. Solano, eds., ASP Conf. Ser. **351**, (in press), 2006.
4. I. Evans, M. Cresitello-Dittmar, S. Doe, J. Evans, G. Fabbiano, G. Germain, K. Glotfelty, D. Plummer, and P. Zografou, "The Chandra X-Ray Observatory Data Processing System", (these proceedings), 2006.
5. J. Evans, M. Cresitello-Dittmar, S. Doe, I. Evans, G. Fabbiano, G. Germain, K. Glotfelty, D. Hall, D. Plummer, and P. Zografou, "The Chandra X-Ray Center Data System: Supporting the Mission of the Chandra X-Ray Observatory", (these proceedings), 2006.
6. D. Graessle, I. Evans, K. Glotfelty, X. He, J. Evans, A. Rots, G. Fabbiano, and R. Brissenden, "The Chandra X-Ray Observatory Calibration Database (CalDB): Building Planning and Improving", (these proceedings), 2006.
7. J. Nichols, C. Anderson, P. Mendygral, and D. Morgan, "Chandra Data Processing: Lessons Learned and Challenges Met", (these proceedings), 2006.
8. L. Paton, "Chandra Hardware and Systems: Keeping Things Running", (these proceedings), 2006.
9. D. Plummer, J. Grier, and S. Masters, "The Chandra Automated Processing System: Challenges, Design Enhancements, and Lessons Learned", (these proceedings), 2006.
10. A. Rots, S. Winkelman, S. Paltani, and E. DeLuca, "Chandra Data Archive Operations", *Observatory Operations to Optimize Scientific Return III*, P. Quinn, ed., Proc. SPIE **4844**, pp. 174-181, 2002.
11. A. Rots, S. Winkelman, S. Blecksmith, and J. Bright, "The Chandra Bibliography Database", *Astronomical Data Analysis Software and Systems XIII*, F. Ochsenbein, M. Allen, and D. Egret, eds., ASP Conf. Ser. **314**, pp 605-608, 2004.
12. B. Spitzbart, S. Wolk, and T. Isobe, "Chandra Monitoring, Trending and Response", *Observatory Operations to Optimize Scientific Return III*, P. Quinn, ed., Proc. SPIE **4844**, pp. 476-484, 2002.
13. S. Winkelman, A. Rots, S. Paltani, and D. Hall, "Tracking the Processing Status of Chandra Observations", *Observatory Operations to Optimize Scientific Return III*, P. Quinn, ed., Proc. SPIE **4844**, pp. 485-492, 2002.
14. S. Winkelman, A. Rots, K. Michaud, and S. Blecksmith, "Chandra FITS Dictionary", *Astronomical Data Analysis Software and Systems XIII*, F. Ochsenbein, M. Allen, and D. Egret, eds., ASP Conf. Ser. **314**, pp 133-136, 2004.
15. S. Winkelman, A. Rots, A. Duffy, S. Blecksmith, and D. Jerius, "Where Do the Data Go? An Analysis of Chandra Data Dissemination", *Astronomical Data Analysis Software and Systems XV*, C. Gabriel, C. Arviset, D. Ponz and E. Solano, eds., ASP Conf. Ser. **351**, (in press), 2006.
16. P. Zografou, S. Chary, K. DuPrie, A. Estes, P. Harbo, and K. Pak, "The ASC Data Archive for the AXAF Ground Calibration", *Astronomical Data Analysis Software and Systems VII*, R. Albrecht, R. N. Hook, and H. A. Bushouse, eds., ASP Conf. Ser. **145**, pp 391-394, 1998.

6. LINKS

– http://cxc.harvard.edu/cda/
– http://cxc.harvard.edu/cda/doc.html
– http://fits.gsfc.nasa.gov/
– http://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html
– http://adswww.harvard.edu/index.html