ESA RSSD Strategy for Science Archives

Christophe ARVISET[1]

[1]European Space Agency, Research and Scientific Support Department

Science Operations and Data Systems Division

VILSPA, POBOX 50727, 28080 Madrid, Spain

Christophe.Arviset@esa.int

Abstract - In the past, traditional "per-project" approaches have led to each scientific mission developing its own data archive from scratch with specific architecture and user interface, without fully benefiting from experience and expertise built up in other projects. With the new transversal role of the Science Operations and Data Systems division and by using modern technology (Java, XML), efforts are now being made to re-use existing Science Archive projects to improve mission efficiency and cost. The example of the ISO Data Archive and the XMM-Newton Science Archive will be detailed, explaining the advantages of such an approach, which also leads to better prospects for long-term preservation of east access to astronomical data.

1. Old and New Development Approach

1.1 Developing Archives in the Past

In the past, science operations projects within ESA had a traditional structure where each team was working relatively independently from other projects. There was no major transfer of experience from one project to another, neither of technical expertise on existing systems nor of manpower resources. With this approach, each new project tended to "re-invent the wheel" as it could not really benefit from past experience. Furthermore, sometimes, identical requirements could lead to different implementations as continuity in core development team was lacking.

However, it must be admitted that perhaps the technology to support real knowledge transfer and re-use from old projects to new ones was not yet available. As a matter of fact, the web did not exist yet and systems were developed on proprietary hardware and operating systems. The lack of tools to develop cross-platform systems lead to that inefficient situation.

In this context, each project used to develop its own archive system on a closed and proprietary architecture, using specific media for data storage and data distribution putting heavy constraints on end users who wanted to get access to these data.

1.2 New Development Approach

A will to support the evolving needs of the community and to group together activities related to science operations for maximum effectiveness led to the creation of a "Science Operations and Data Systems" division within the "Research and Scientific Support Department" (RSSD). Key goals of this division are to improve the efficiency with which payload operations are prepared and executed and the resulting data captured, distributed and archived, and to provide continuity of expertise from concept to conclusion of a mission and also from mission to mission.

The Science Operations and Data Systems Division has the overall responsibility for scientific operations development and execution for astronomy missions within RSSD. In particular, it carries the role of developing science archives, providing easy access to mission data for the scientific community worldwide. It provides horizontal support within the department ensuring a more cost efficient approach through:

- Knowledge transfer between projects
- Manpower shared between projects
- Re-use of existing systems for new projects
- Easier and cheaper maintenance of existing systems

In the context of archives, these main goals can now be achieved, as the technology is now ready to support such an approach. All mission archives are now on-line on the World Wide Web where data distribution can be done via FTP in most of the cases. The web and Java allow deployment of system on a variety of platforms, independently of the platform the application has been developed on. Modular and flexible multi-tier and client –server architecture, together with the use of object oriented approach and XML allows configurability and re-usability of existing systems for other projects.

This new approach has been started successfully in the context of the ISO Data Archive and the XMM-Newton Science Archive as described in the following section.

2. One Example: IDA and XSA

2.1 ISO Data Archive (IDA)

The European Space Agency's (ESA) Infrared Space Observatory (ISO) is an astronomical satellite that was operational between November 1995 and May 1998. It operated at wavelengths from 2.5 to 240 microns, in the infrared range of the electromagnetic spectrum. Exciting new results on the cool universe came in practically all astronomical fields from comets to cosmology, with detecting water throughout the cosmos being one of the highlights. The satellite essentially consists of: a large liquid-helium cryostat; a telescope with a 60-cm diameter primary mirror; four scientific instruments and the service module.

The ISO Data Archive (IDA) has been developed at the ESA ISO Data Centre at Villafranca, Spain. It opened to the community in December 1998 and offers a state-of-the art archive facility to access ISO data products and auxiliary files. It is accessible at http://www.iso.vilspa.esa.es/ida/.
It contains all the ISO raw and fully processed, science and calibration data as well as all ancillary data (engineering, uplink and downlink data) for a total of about 400 GBytes stored on magnetic disks. Through a powerful and user friendly Java User Interface, over 1200 registered users have already downloaded the archive content more than five times since it was opened to the public. The IDA main characteristics can be summarized as:

- powerful and complex queries against the observations catalogue
- configurable results display, including product visualization tools
- customisable product retrieval via a shopping basket
- selection of product level for retrieval
- product retrieval via FTP
- inter-operability with other archives and applications

It has been designed in an open and flexible way, which allows configurability and re-usability for other archive projects.

2.2 XMM-NEWTON Science Archive (XSA)

XMM-Newton, ESA's X-ray Multi-Mirror satellite, is the most powerful X-ray telescope ever placed in orbit. It was launched by an Ariane 5 rocket from ESA's spaceport in Kourou, French Guiana, on 10 December 1999. With its unprecedented sensitivity it observes the X-ray sky, helping to solve many cosmic mysteries, ranging from extremely violent and exotic processes, such as enigmatic black holes, to the formation of galaxies. XMM-Newton also observes celestial objects within our Solar System, such as comets and planets.

The XMM-Newton Science Archive (XSA) has been developed at the ESA XMM-Newton Science Operations Centre at Villafranca, Spain. It has been available to the scientific community since mid April 2002 at http://xmm.vilspa.esa.es/xsa/.

It contains all the XMM-Newton raw and fully processed, science and calibration data as well as some ancillary data for an expected total of about 1 TBytes stored on magnetic disks. Based on the IDA architecture and code, it offers similar data query and retrieval (FTP and CDROM) facility through a user friendly User Interface. Proprietary data is accessible to the observation owner only while public data has unrestricted access.

2.3 An Open 3-tier Architecture

The IDA and the XSA were both built –by a common team– using the open 3-tier architecture described below. The main goal of this architecture is to separate the data from the presentation, which allow a more modular and flexible development.

As the data volume is not that big, data are saved on magnetic disks for fast access as a normal UNIX file system. From the data products, metadata is extracted and put in a Relational Data Base, SYBASE. Note that the data ingestion from the data producer and the metadata extraction are separate processes to allow new metadata data extraction when user requirements evolve.

The middle tier, also called the Business Logic, provide a transparent access to the data products and to the metadata. This key layer has been developed in Java and XML and resides on the archive server.

On the client side, several types of applications can be found. The standard IDA and XSA User Interface is a Java applet downloaded by the end user to access the archive content. Remote applications and other archives can also have access to the data and the metadata, bypassing the standard User Interface, by speaking to the Business Logic that will provide them with the required services via Java Server Pages. This architecture is especially powerful in the context of the Virtual Observatory initiatives worldwide where archives will all have to inter-operate together in a transparent manner for the end user.

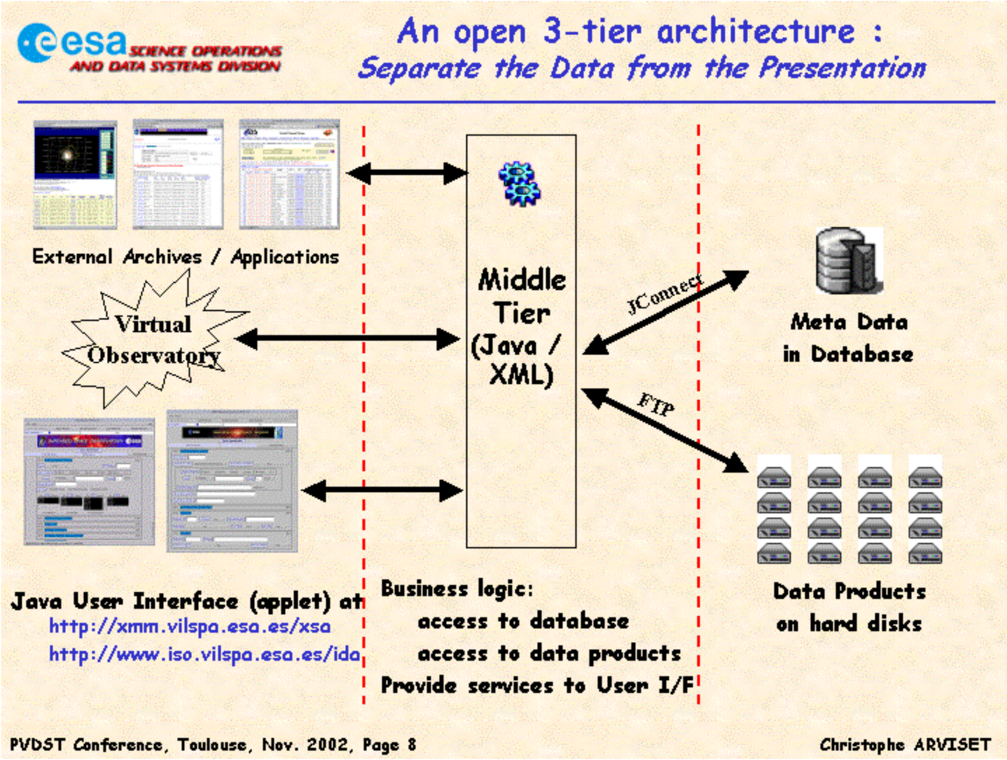


FIG. 1: Open 3-tier Architecture

This modular approach is extremely relevant for long-term maintenance or evolution of archives. As a matter of fact, it allows the complete substitution of a subsystem (like a port to some newer technology) without affecting the rest of the systems.

For example, if one decides to modify the storage media (from CDROM jukeboxes onto hard disks as it had become now affordable), this will not affect the database, nor the middle-tier or the client applications. Similarly, if one decides to move the database from e.g. SYBASE to ORACLE, the other sub systems will remain unchanged. The IDA and XSA standard User Interfaces are now Java based. In the eventuality that Java is going to disappear on the browser (most unlikely!), new user interface could be developed without modifying the back-end systems.

As shown, this flexible architecture will allow modular evolution of the archive if one of the technologies used by one subsystem becomes obsolete. This could be done without having to re-design and re-write the full archive.

2.4 Java and XML

In general, Java provided the multi-platform support, having the same code for all operating systems, realizing the dream of many developers: ***write once, run anywhere***. It provides easy deployment on user's browser while producing a clean and compact user interface that looks like an application, in contrary to the other web approach based on cgi-bin scripts.

Within the IDA and the XSA, Java is used for the middle-tier and the standard User Interface, which is distributed as an applet through a standard web browser. Java standalone application are also offered for people who would prefer this approach rather than the applet. Throughout the years, efforts have successfully been made to transfer code from the client to the business logic located on the server in order to improve performance of the overall application. Similarly this offers easier maintenance as modifications on the middle-tier does not require changes on the client side, which makes them transparent for the end user.

One can now go one step further by using XML to achieve multi-project support. In conjunction with the multi-platform capability of Java, the use of configurable XML files allows the same code to be used for various archives to reach ***write once, use anywhere*** capability.

Within the IDA and the XSA, XML is used in various ways. First, semi-dynamic XML configurable files are used to describe the appearance and associated action for each element (button, input field, etc..) of the user interface query panels. Similarly, other XML files defines the all graphic elements to be displayed as a result of a query. Second, the database is described with XML files which allows the Business Logic to make transparent access to the metadata, regardless of the RDBMS used (SYBASE, ORACLE, …). Third, data distribution request is made via a XML file, which is then mapped (through a DTD) to the data products files located on the data storage.

The use of both JAVA and XML have allowed to re-use the existing code of the IDA to develop the XSA with minimum software change, just modifying the XML configuration file. This resulted in having 95% of the Java code identical for the User Interface and Business Logic system between the IDA and XSA.

3. Advantages of this Approach

3.1 Data Storage and Metadata

IDA and XSA both use the well established FITS (Flexible Image Transport System) standard for the astronomical data products (http://fits.gsfc.nasa.gov/). The Planetary Mission Archive (in development, see below) will use the PDS (Planetary Data System) standard for planetary datasets.

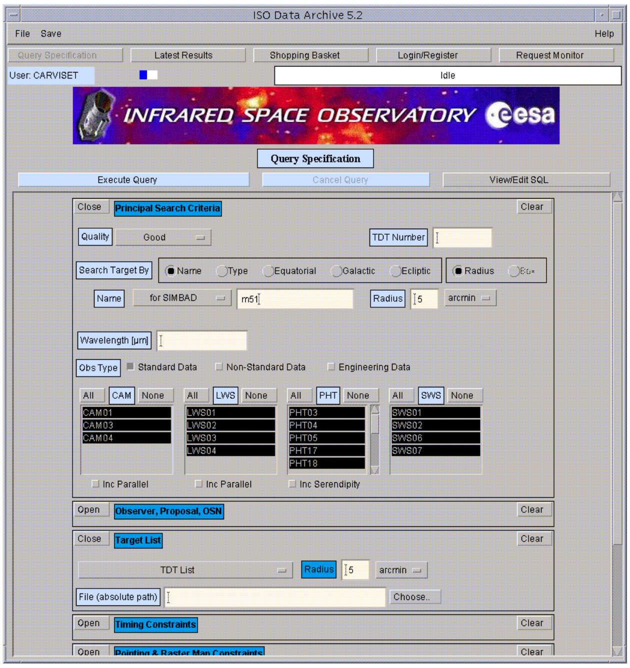
Using data product standards established by the scientific community ensures wide availability of tools for end user to handle and process the data. Furthermore, their wide acceptation and utilization makes sure that the archive data products will still be accessible and used in the long term.

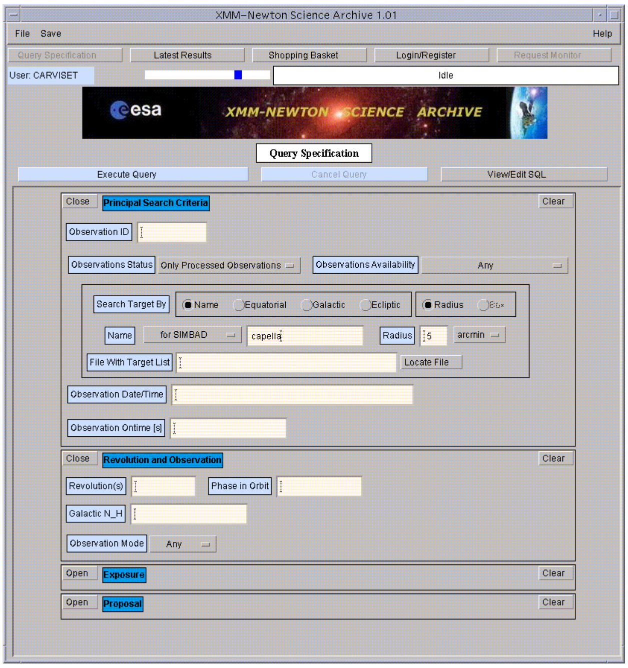
Extracting the metadata from the data products to ingest them into a database allows complex and fast queries on the archive content. In the long term, if metadata is well structured and well documented, porting the database to a new database is not a difficult task, in contrary to porting a software to an other programming language.

Using XML as way to describe the archive content (both for the metadata and the data product repository) makes easier its access from external applications and safer its long-term preservation.

3.2 Same Look & Feel for ESA Science Archives

Re-use of the same code for the IDA and the XSA~~,~~ user interfaces also offers similar Look & Feel. The user then immediately recognizes that he/she is accessing an ESA archive, leading to a clear ESA image within the community.





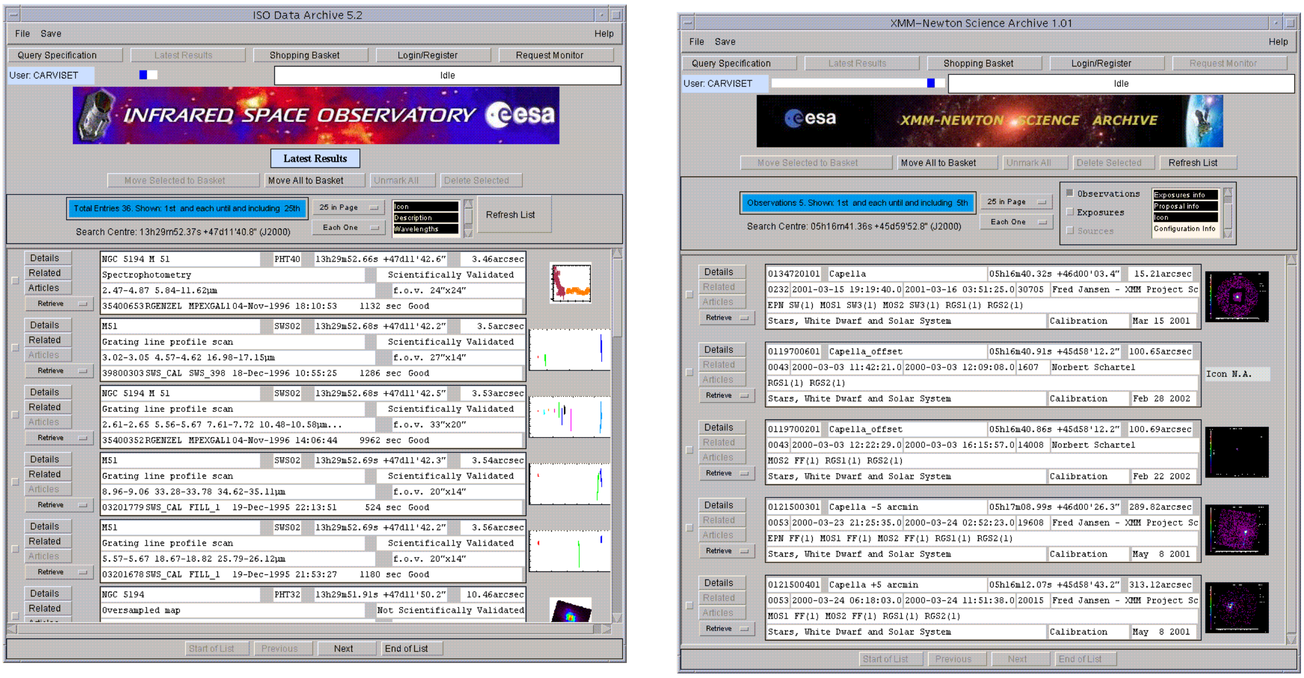


FIG. 2 : IDA and XSA query and results Displays

3.3 Share Knowledge and Manpower

The horizontal structure of the Science Operations and Data Systems Division has lead to the creation of an Archive Development Group in VILSPA, Spain. This group is providing support to various archive projects. Each member of the group is not dedicated to one archive in particular, but is working on a specific sub-system across all the archive projects. There are some Database engineers, some experts on User Interface and Middle-Tier and others on Data Storage, Data Distribution and Interoperability aspects.

This ensures permanent knowledge transfer from one project to another as people are working on several projects at the same time. Faster development is then made possible through this efficient manpower utilization. As the code is almost identical, problems fixed on one archive are fixed as well on the others when applicable which brings down maintenance costs.

This approach also permits a more flexible manpower allocation. Some “old” projects do not have enough money to maintain several software experts on distinct fields. But by sharing them with new projects, “old” project can afford to pay for fraction of expertise, while new projects benefit from the previous expertise and get development made faster and cheaper.

Furthermore, people remain strongly motivated by participating to new projects and discovering new technology, while at the same time still providing support part-time to some less challenging past projects.

3.4 Maintenance and Development Costs

Re-using same architecture from previous projects brings the advantage that less effort is required in designing new projects. Furthermore, one is ensured that the architecture is robust as it has been implemented and used successfully already.

Re-using same code obviously bring the overall development costs down in the short term. In the long term, it also ensure cheaper maintenance as fixes or improvement are applied to all archives at the same time. The variety of end users around the different scientific communities makes sure that problems are found (and later on fixed!) and makes the archive more robust.

Re-using the IDA experience allowed to build the XSA from scratch up to public release to the scientific community within a year. A user workshop established the User Requirements in March 2001 and XSA v1.0 was made public mid-April 2002. The XSA development costs represented only one third of the IDA to reach the first public version. Moreover, as IDA had already been used by thousands of users over several years, the first version of XSA inherited from the stability and reliability of the IDA. This approach should allow the much quoted paradigm: "**Better, Faster, Cheaper**" to be actually achieved in archive development and maintenance.

4. Future Plans

ESA RSSD strategy for developing Science Archive is to build an Archive Development group at VILSPA, Spain, within the Science Operations and Data Systems Division. This group will provide horizontal support to various missions archive projects. People will be work for several projects at the same time, ensuring knowledge transfer and re-use.

By using a flexible multi-tier architecture and modern technology (Java, XML), archive projects can re-use design and code from earlier projects, which brings down development and maintenance cost, while speeding up development timescales. The modular approach will allow evolution and port of subsystems without affecting the other systems, thus assisting in preserving efficient long-term access to the data from ESA's science missions.

This strategy has already put in place for the ISO Data Archive and the XMM-Newton Science Archive. It is being setup as well for the INTEGRAL Science Operations Centre Data Archive (for ESA internal use) and for the Planetary Mission Science Archive (International Halley Watch campaign, MARS EXPRESS, ROSETTA, SMART-1, …) and will be ready for future missions (HERSCHEL, PLANCK, …).

References

[1] C. Arviset, T. Prusti. *The ISO Data Archive*. ESA Bulletin, June 1999.

[2] C. Arviset, M. Guainazzi, J. Dowson, J. Hernandez, P. Osuna, A. Venet. *The XMM-Newton Science Archive*. New Visions of the X-ray Universe in the XMM-Newton and Chandra era Conference, November 2001.