

## **The XMM-Newton SSC Database: Taking Advantage of a Full Object Data Model**

L. Michel, C. Motch

*CNRS, UMR 7550, Observatoire Astronomique de Strasbourg,  
Strasbourg, France*

C. G. Page, M. G. Watson

*X-ray Astronomy Group, Department of Physics and Astronomy,  
University of Leicester, Leicester LE1 7RH, UK*

**Abstract.** One of the main responsibilities of the Science Survey Centre (SSC) of the XMM-Newton satellite, an X-ray observatory launched by the European Space Agency in 1999, is to carry out a systematic analysis of the entire scientific data stream. Products resulting from the pipeline processing are shipped to the guest observer and eventually enter the XMM-Newton archive. In addition, the SSC compiles a catalogue of X-Ray sources and provides an identification for  $\sim 50,000$  new sources detected each year. In order to check product quality and to support the catalogue and source identification programmes, all SSC-generated products are stored in a database developed for that purpose. Because of the large number of transversal links, our data model was difficult to map into relational tables. It has therefore been designed with object oriented technology for both user interface and data repository, and based on an object-oriented DBMS called O2. The database is a powerful tool to browse and evaluate XMM-Newton data and to perform various kinds of scientific analysis. It provides on-line data views including relevant links between products and correlated entries extracted from many archival catalogues and also links to external databases. Besides browsing, the web-based user interface provides facilities to select data collections with any constraints on any keywords but also with constraints on correlated data patterns.

### **1. Database Overview**

The SSC database contains all data products resulting from the pipeline processing of the photon-event lists and other raw data from the XMM-Newton spacecraft. The products from a typical observation include  $\sim 100$  FITS files and  $\sim 400$  other files (HTML, PDF, etc.), and occupy  $\sim 400$  MB. These data files are grouped by observations and contain both observational data (graphical products, tables, spectra, images and event lists) as well as extractions from astronomical archival catalogues generated by the cross-correlation (ACDS) with the archives at NED and at CDS in Strasbourg. They also include the catalogue

of X-Ray sources compiled by the SSC. An overview of the pipeline structure can be found in Fyfe et al. (2001) and detailed product descriptions in Osborne (2000).

## 2. The Common Data Model (CDM)

All data products and data containers (e.g., instrument exposures) are modeled with a hierarchy of classes, the Common Data Model (CDM). Classes contain atomic attributes (position, flux, ...) and references to related objects such as correlated sources. Class methods are in charge of both content update and content representation (see Figure 1).

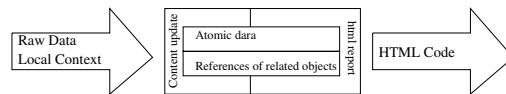


Figure 1. CDM class.

Consistency between data and GUI is easily maintained since instances manage their own contents. Persistence is managed by O2C, the 4G language provided with O2. All DBMS features (transactions, caching, indexing, ...) rely on the O2 engine in a transparent way for the developer. Any transient object is automatically made persistent whenever it is referenced by another persistent object. The same code may work with either persistent or transient objects.

The Common Data Model uses the inheritance mechanism widely. Objects of different classes can be seen as instances of one super-class (e.g., sky object) when they are handled in collections (such as for queries). Thanks to the late binding mechanism, they are however considered as instances of their real class when they are accessed individually (e.g., to read their content).

All FITS headers are instantiated into the database (one attribute per keyword). Only FITS table extensions are exploded. Each table row is represented by one instance which belongs to a collection modeling that particular table. FITS file blobs as well as graphical and HTML products are stored on the external file system and referenced into the database using URLs.

Pipeline products include lists of entries extracted from various astronomical catalogues (part of cross-correlation products) and having positions matching those of EPIC X-ray sources, or for a subset of catalogues, being located in the observation's field of view. A single archival source may in some cases be correlated with several X-ray sources detected in one or more observations. This information is very relevant for astronomers and must enter the database. Archival sources are stored in specific collections where their uniqueness is ensured. Objects representing archival sources will own as many references to X-ray instances as they have correlated X-ray sources. This is a good example of an N to M relationship which is much simpler to manage with an OO database than with a relational system. There is no table join to deal with, just vectors for object references. Furthermore queries can easily include constraints on vectors patterns.

Ingested products are modeled by more than 300 specific classes. Managing the code of so many classes by hand is not realistic especially since the data

formats may evolve during the mission. The data-loader deals with this task. Product compliance with data schema is checked at ingestion time and classes are updated or even created following predefined templates when required.

### 3. Graphical User Interface

Each HTML page is actually the HTML representation of one instance (see Figure 2). Atomic values are shown as such whereas related objects are represented by self-generated anchors. The database can be totally browsed by just following these links. All FITS keywords are listed and image and table FITS extensions can be displayed by the FITS previewer FIBRE (a Fortran90 CGI script which uses FITSIO and PGPLOT libraries).

Source 1350-03433180

Archival Catalogue: USNO 11252A (11529\_ent) [dss2]

String Keywords  
 USNO\_A2\_0: 1350-03433180  
 ACTTag:  
 MTag:

Numeric Keywords

SRC_NUM:	24	CAT_RADEC_ERR (arcsec)	1.000	RA_ICRS_ (deg)	03:26:05.99	Rmag (mag)	18.1
CAT_RA (deg)	03:26:05.99	D_EPIC_CAT (arcsec)	1.791	DE_ICRS_ (deg)	+49:00:06.28	Epoch (yr)	1954.768
CAT_DEC (deg)	+49:00:06.27	CAT_NUM:	210	Bmag (mag)	19.1		

Correlated EPIC Sources

Epic Summary Source	XMM Catalogue Entry	Summary Source List	Illi bill	Distance	Source Lists	Obs having the source in their FoV
[RA 03:26:06.22 DEC +49:00:05.94]	[RA 03:26:05.22 DEC +49:00:05.94]	0101440101_V01_OBSMLJ	147.2917 -6.4154	1.8"	[0101440101_V01_OBSMLJ] (173_ent)	[0101440101_V01]

Logout | New Password | Session Report

Mon Nov 4 10:28:11 2002 Laurent MICHEL logged on

Figure 2. Detail of an archival source.

Users may set-up queries on the properties of X-ray sources, archival sources, observations and exposures of all instruments. Constraints may be put on any atomic attribute and on some related objects. In addition, one may also apply constraints on the correlation patterns between X-ray and archival entries. The number of constraints contained in queries is not limited. An example, in pseudo-code, of the type of query which can be handled is given below:

```
select X-ray sources
  having a hardness ratio 2 in the range of 0.5 to 1.0
  and detected in observations having a duration > 10000 sec
  done by "I. Newton" or by "G. Galileo"
  and correlated with USNO entries at less than 3".
  but not correlated with any SIMBAD or NED entry.
```

Query results are stored into the database and can be displayed (see Figure 3) again at any time. This feature makes the possibly long response time for very complex queries more acceptable. User selections are kept from one session to the next.

### 4. Interoperability

The system allows the field of view of the EPIC instruments and source positions to be overlaid on any external image using Aladin facilities (Bonnarel et al.

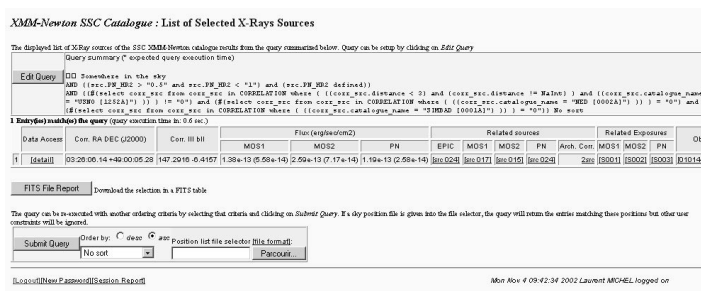


Figure 3. Selection of X-Ray sources.

2001). VizieR may also be queried for any X-ray or archival source. In addition, HTML pages stored in the database include links forwarding real time queries to SIMBAD and NED. Finally, users can download their selections as FITS tables allowing further local processing.

## 5. Prospects: Future of OODBMS

The system currently installed at Leicester manages  $\sim 200,000$  X-ray sources coming from over 2,000 observations and correlated with  $\sim 400,000$  archival sources. The database volume is about 30 GB with 400 GB of external files. Performance is good especially for complex queries. Only SSC members can currently use this database, but a version dedicated to the SSC XMM-Newton catalogue will soon be opened to the community. Although O2 matches our needs well, after a series of company take-overs, the product was withdrawn from the market in 2000 and is no longer under development. Software support continues from Oxymel, a company created by former O2 developers.

Since O2 has no future, we are actively seeking an alternative DBMS. Unfortunately object-oriented DBMS have failed to gain market share and we may have to move to a relational system. We have derived considerable benefit from transparent persistence, inheritance, abstract types and N-M relationships. These features can be implemented in relational systems with an object mapping layer provided we accept reduced flexibility and certainly more complex set-up.

## References

- Bonnarel, F., Fernique, P., Genova, F., Bienaymé, O., & Egret, D. 2001, in ASP Conf. Ser., Vol. 238, Astronomical Data Analysis Software and Systems X, ed. F. R. Harnden, Jr., F. A. Primini, & H. E. Payne (San Francisco: ASP), 74
- Fyfe, D.J. et al. 2001, to appear in the Proceedings of the Symposium on 'New Visions of the X-ray Universe in the XMM-Newton and Chandra Era', 26-30 November 2001, ESTEC, The Netherlands
- Osborne, J. 2000, SSC-LUX-SP-004, available at <http://xmm.vilspa.esa.es/>