The XMM Science Archives

Jean Clavel
European Space Agency, Space Science Department, Astrophysics Division, VILSPA, Apartado 50727, 28080 - Madrid - Spain

ABSTRACT

An essential part of the Science Operation Center (SOC), the XMM science Archives will be the only repository of all XMM scientific data. All files relevant to a given observation, from the initial proposal up to the processed data and distribution logs are automatically stored, linked and catalogued such they can easily be retrieved. The archives also encapsulate the S/W and data-bases necessary to run XMM payload operations. The archives are the only interface between the SOC – located in VILSPA, near Madrid, Spain – the Mission Operation Center (MOC) in ESOC, Germany and the Survey Science Center (SSC) in Leicester, UK. Therefore they play a crucial role in the de-centralised XMM ground segment and ensure that all files, data-bases and software systems shared by the three centers are identical and under strict configuration control. All science data are stored in FITS format and accessible on line via a web based browser interface. The data-base management system uses ORACLE and support up to ten external users in parallel. The data remain private to the Principal Investigator (PI) for one year after they have been correctly processed and calibrated. After one year, they are released for access by the scientific community. PI's get automatically notified when their data have been processed and archived and they receive their data on CD-ROMS.

1. Introduction: a summary of the XMM ground segment data-flow

The XMM payload operations will be run by the Science Operations Center (SOC), located in the VILSPA ESA ground station, near Madrid Spain. Except for the pipeline data processing which is performed by the Survey Science Center (SSC) in Leicester, UK, the small SOC team is responsible for all the classical functions associated with an observatory type science mission: they issue the Announcement of Opportunity (AO) to the scientific community, receive, process and archive the proposals, provide technical assistance to the potential Guest Observers (GO), evaluate the proposal feasibility and provide technical guidance to the Observing Time Allocation Committee (OTAC), expand the proposals into a sequence of observations with all 3 instruments on-board XMM, schedule the observations, receive, re-format and archive the telemetry, monitor the execution of the observations in real-time and react to payload anomalies should they occur, dispatch the re-formatted telemetry (Observation Data Files, or ODF and Slew Data Files, SDF) to the SSC and receive back the processed data, archive and catalogued them, and finally distribute

the data to the GO on CD-ROM's. In addition, the SOC is responsible for the calibration of the instruments and, together with the SSC, for the maintenance and optimization of the Science Analysis System (SAS) which is used by the pipeline processing software in Leicester. The Mission Operation Center (MOC), located in ESOC, Germany has the overall responsibility for the XMM spacecraft (S/C) operations. For each 48 hours revolution, they provide the SOC with a skeleton schedule containing basic platform or ground activities (such as station hand-overs) necessary to support XMM operations. The SOC then fills-in the slots reserved for payload operations with a set of observations selected from the proposal data-base according to visibility constraints and scientific merits as defined by the OTAC. This so-called Preferred Observations Sequence (POS) is then exported back to the MOC who adds all necessary platform and ground activities (Enhanced POS, or EPOS) required to support the payload operations, such as slews or reaction wheels momentum dump and expands it onto a Central Command Schedule (CCS), i.e. a precise time-line of all the telecommands to be uplinked to the XMM S/C for that revolution. The CCS execution is under control of the MOC who as the overall responsibility for the safety of the XMM spacecraft, including its instruments. However, the SOC has the possibility to issue via the MOC payload telecommands in near real-time to e.g. optimize the on-going exposure, if necessary, or to react to payload anomalies when they occur. The raw telemetry is received and archived by the MOC and sent in real time to the SOC via a dedicated and redundant 128 kbps line. Similarly, there is a dedicated line between the SOC and SSC for the exchange of ODF and processed data and ancillary files (e.g. calibration).

2. Purpose and main functions of the XMM science Archives

In such a de-centralised ground segment, the XMM Science Archives play a critical role which is twofold:

- Of course, the main function of the Archives it to act as a repository of *all* XMM scientific observations. Therefore, all data resulting from or pertinent to a given observation get archived, catalogued and linked via pointers so as to be easily retrievable. In this role, the archives act as the main interface to the scientific community. Note that the XMM archives in the SOC are the *sole* repository of all XMM data.

- The SOC, MOC and SSC share a certain number of files, data-bases and software sub-systems which are essential to the successful execution of XMM operations. To ensure compatibility of the ground segment, it is therefore essential that any modification of a shared item at e.g. MOC be implemented simultaneously at the SOC. For this purpose, the Archive Management System (AMS) plays a central role in acting as the only interface between the SOC and the MOC on one side, and the SOC and the SSC on the other side.

2.1. Information content

A basic feature of the XMM archives is that essentially *all* information items relevant to a particular observation get automatically stored. These include:

1. The proposal, as submitted by the GO via the Remote Proposal Sub-system (RPS), including its associated observations and individual exposures. Note that each proposal contains the administrative information related to the GO, such as his mailing address which is required for CD-ROM mailing.
2. The expanded observations after ranking by OTAC and expansion by the SOC
3. The relevant scheduling files (e.g. POS) and central command schedule
4. The relevant telecommand history file listing all telecommands actually uplinked to the s/c to execute this observation and their execution outcome, successful or otherwise.
5. The Observation Log, and flags as generated during or immediately after the execution of an observation, indicating its success or otherwise and a first assessment of the data quality.
6. The Observation Data File (ODF) which contains the raw telemetry re-formatted as a FITS file, as well as the Slew Data File (SDF) containing instruments data acquired during slews from one astronomical target to the next.
7. Accurate attitude data providing the actual pointing coordinates of the XMM telescopes during the observation and the S/C slews, as reconstructed by the MOC
8. The Current Calibration File (CCF) used to calibrate and process the data; the calibration observations used to derive the CCF are implemented and executed by the SOC but stored in the archives in the same way as normal astronomical observations.
9. The pipeline data products as generated and validated by the SSC
10. Various flags indicating the status of a data product; for instance, once the data have been written onto a CD-ROM for shipment, a flag is set in the archives together with the distribution date.
11. Ancilliary files related to an observation, such as simulated data; note that several astronomical catalogs are also stored in the archives
12. In addition, all S/W sub-systems, data-bases, procedures and documents used in the ground segment at the time of the observation and their version number are automatically recorded and linked with the observation, for traceability purpose. In addition, the GO can request through the archives, the Interactive Analysis Software (IAS) sub-system to reduce his/her own data again in a more optimised fashion; the IAS shall contain all the functionalities of the SAS.

All the above items are linked via pointers so as to be easily retrievable. Most of items shall be of limited use to the average GO. However, the fact that they are easily accessible will be a great help to SOC staff, both during and after an observation. For instance, during an observation the SOC staff on duty as the possibility to recall the simulated data and compare them with the actual data from the on-going exposure. If required, he/she can also retrieve catalog data to assist in the on-line assessment of the observation. By storing all information in an easily accessible way, one also ensures that it is always possible to trace-back exactly what happened to a particular observation data-set, should that become necessary because of e.g. a ground system or on board anomaly (trouble-shooting).

2.2. The archives as the sole i/f between SOC, MOC, SSC and GO's

Distributed ground segments such as that of XMM, face the particular challenge that at any point in time, all parties must share the same information and the same tools. For instance, it will be necessary during the mission to update the CCD clock sequences of the instruments or the hot pixels table. Both items are part of the instrument on-board software which is maintained by the SOC. However, patches to the instrument software RAM are uplinked by the MOC. The SOC and the MOC also share the same payload telecommand and telemetry data-bases as well as the same software to process the telemetry and monitor the execution of XMM operations in real-time. It is therefore essential that the MOC and the SOC maintain the same version of the software and data-bases at any given time. For this purpose, the SOC and the MOC have to have a single interface, which is the Archives. In a similar way, the SOC and the SSC share the same set of Current Calibration Files (CCF) and SAS tasks used by the pipeline to process the observations and transform the telemetry into a scientifically meaningful set of data. The CCF are generated and maintained by calibration experts in the SOC but used by SSC to calibrate the data during pipeline execution. Again, the archives act as a single point of contact between SOC and SSC.

Within the SOC itself, it is also essential to maintain a strict configuration control of all sub-systems since these interface with each other. For instance, the Sequence Generator Sub-system (SGS) which schedules the observations takes its inputs from the proposal data-base (shared with the RPS and the Proposal Handling Sub-systems, PHS) and from the telecommands data-base (shared with the Payload Monitoring Subsystems, PMS, and the MOC). Similarly, once an observation has been successfully executed, a flag needs to be written into the proposal data-base to avoid that the observation gets re-scheduled. This is done automatically at the time the data get ingested into the archives. Since all the SOC sub-systems communicate with each other, it is essential that they get their input from one single place, namely the Archive Management Sub-system, AMS. Only in this way can one ensure the compatibility of software and data-bases at all times.

Broadly speaking, the AMS therefore acts as the "configuration controller" of all software items, data-bases and files, both within the SOC and, for shared items, within the ground segment

as a whole.

As explained previously, all the information pertinent to a guest observer (GO) – e.g. proposals, observations, administrative information – is also stored in the archives, and all data distributed to the GO originate from the archives. Hence, the AMS also acts as the sole operational interface with the scientific community.

Each "user" of the archives, be they within the SOC, in the MOC, in the SSC or in the community, has access privileges to the archives corresponding to his/her functionality. For instance, the SOC engineers have write access for what concerns the instrument on-board software, whereas the MOC can only read this item. Conversely, the MOC can write into the archives a new version of the telecommand data-base but the SOC can only read it. Similarly, a given GO can only read and retrieve his own observation, not that from others GO's, at least during the one year proprietary period. Access rights are under control from the SOC AMS manager.

3. Archives design features

The heart of the Archives – and in fact of the overall SOC system – is a solaris server which acts as the central repository of all data, the center of communication between all SOC internal and external sub-systems and also as the server from which the software required by the various SOC sub-systems (e.g. PHS, SGS, ODS, SAS, ...) is loaded. The Solaris server runs Solaris 2.6 and uses 2 Sun multi-processors, for redundancy. The data are stored using an Hierarchical Storage Mechanism (HSM) such that older or bulk data which do not require rapid access – e.g. data processed with a previous version of the pipeline – are stored on a slower medium than frequently used data for which the storage medium has not been decided yet but is likely to be CD-ROM juke-boxes.

The Archive Management System (AMS) browser server runs on two NT Pentium PC's. It uses an Windows NT 4.0 web interface and Oracle 7.x as data-base management System. The AMS can support up to 30 parallel "users", with twenty of them "internal". Internal users can be processes, e.g. the Proposal Handling System (PHS) running in the back-ground and reading proposals from the data-base, or human beings, i.e. SOC staff, interactively interrogating the archives. Ten slots are reserved for "external" users, i.e. scientist from the community browsing the Archives content. The system has been sized such that the maximum response time to a data-base query is 30 s for an internal user and 60 s for external users. The user's browser must be either Netscape Navigator (version 3 or higher) or Microsoft Internet Explorer (version 3.02 or higher). It must support Java Classes (1.0) and JavaScript and must have plug-ins such as FITS viewers that are required to view downloaded files.

There are 6 PC's and 6 CD writers for the actual writing of GO data on CD-ROMs. Dedicated printers are attached to the CD writer units which automatically generate labels with the GO address.

The different components of the archive communicate via a dedicated Local Area Network (LAN) protected from the outside world by a "firewall". Access for external users is provided via proxy accounts on a router located outside the firewall. Note that the MOC is considered as an internal user and has direct access to the LAN and archives via the dedicated MOC-SOC 128 kbps line.

4. The Archives as seen by the XMM scientific community

The archives are accessed via a link on the general XMM SOC web page, the URL of which is still undefined at the time of writing (November 1998). For the first visit, the user needs to register. For this purpose, he/she fills-in a form with his name, address, phone & fax number and e-mail address and submit it. Upon successful registration, he/she should receive a password or PIN number which will be required for all future accesses to the Archives. Using this password, the user may then logon to the archives and access the browser main page. The next logical step is to go to the filter page which allows to define selection criteria for searching the archives. Many selection criteria are supported, such as proposal number, observation number, data type, revolution number, start & end time of the observation ...In addition all standard FITS keywords are also supported. One can also defines ranges and concatenate criteria by logical operators. After pressing submit, the user is returned a list page with all the observations that fit the selection criteria. He/she can further narrow his selection by selecting manually only those observations from the list which he/she actually wants. Only one record per observation is shown at this stage. For each observation, the full information can be obtained by clicking on the corresponding record. The user also has the option to retrieve the Interactive Analysis Software (IAS) for further tailored processing of his/her data at his institute. Note that the calibration files applicable to the selected data-set are automatically appended to the request.

When satisfied with his selection, the user can then select the option "order on CD-ROM" and/or "download". The option for CD-ROM production and data downloading will only be active for those data-sets for which the user has access privilege. As mentioned previously, only the PI of a given proposal can access his/her own data during the first year proprietary period. He/she gets notified by e-mail as soon as his/her data are archived. In this case, a CD-ROM is automatically generated by the system without the need for the GO to submit a request. However, he/she may also want to immediately download the data via internet. Because of band-width limitations, it is unlikely that all data types will be authorized for downloading, at least initially. In practice, very large data-sets – such as ODF's (telemetry re-formatted as FITS files) – are probably too bulky for downloading. As the SOC gains experience with the Archives, it will fine-tune the system so as to allow as many products as practical to be transferable via the internet. When selecting the CD-ROM option, the user is presented a last time with his/her administrative data as he/she entered them at the registration stage. These should be checked carefully since the address which appears on the screen is the one which will be used to mail the CD-ROM. All de-archiving requests

are automatically acknowledged by an e-mail to the requester and the corresponding data flagged as “shipped” in the data-base.