ESA RSSD Science Archives Inter-Operability and Virtual Observatory Services

Christophe Arviset, Isa Barbarisi, John Dowson, Nicolas Fajersztejn, Marco Freschi, Jose Hernández, Daniel Marin, Bosco Olabarri, Iñaki Ortiz, Pedro Osuna, Esther Parrilla, Jesus Salgado, Aurélien Stebe

ESA, Research and Scientific Support Department
European Space Astronomy Centre
PoBox 50727, 28080 Madrid, Spain

Christophe.Arviset@esa.int

Abstract. Within its overall responsibility for scientific operations development and execution for astronomy missions, the Science Operations and Data Systems Division of the European Space Agency's Research and Scientific Support Department (RSSD) carries the role of developing science archives, providing easy access to mission data for the scientific community worldwide. The Virtual Observatory (VO) is a world-wide initiative in astronomy which aim is to allow astronomers to perform new science by providing them with a "federation of astronomical archives and databases around the world, together with analysis tools and computational services, all linked into an integrated facility". Therefore, apart from the standard access to ESA archives via their powerful and user-friendly user interfaces, arose the need for direct machine-to-machine interoperability services, meeting the various protocols defined by the VO standards. ESA Scientific Archive open architecture has allowed easy integration of ESA Archives into the VO without requiring changes in the way data and metadata were stored for each project. By using translation layers, inter-operability services have been adapted to meet the VO standards. An example of a VO application in action will be given with VOSpec, a tool to handle VO-compatible spectra. VOSpec is making use of various VO standards and enables scientists to make new science, combining easily real observational data with theoretical data.

Science Archives at ESAC

Archives Projects at ESAC

ESAC, European Space Astronomy Centre (previously called VILSPA), located in Villafranca del Castillo, near Madrid, Spain is becoming the centre for all ESA Astronomy and planetary space missions. That will include the Science Operations Centre of such missions, as well as the scientific archives for data storage and data distribution to the Scientific Community.

The ISO Data Archive (IDA) has been opened to the community since December 1998 and offers a state-of-the art archive facility to access ISO data products and auxiliary files. It is accessible at http://iso.esac.esa.int/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 1400 registered users have already downloaded the archive content more than eight times since it was opened to the public

The XMM-Newton Science Archive (XSA) has been available to the scientific community since mid April 2002 at http://xmm.esac.esa.int/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 3 TBytes stored on magnetic disks. It also contains the XMM-Newton source catalogue, with powerful search criteria. 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. Furthermore, the XSA offers an on-the-fly reprocessing facility to allow the users the get the data calibrated with the latest version of the pipeline processing software.

ESA Integral Science Data Archive (ISDA) has been available for internal use since December 2003 and has been opened to the public in July 2005 at http://integral.esac.esa.int/isda/. It contains all the INTEGRAL raw and fully processed data for a current size of 4 TBytes stores on magnetic disks. Based on the XSA, it offers same type of functionalities, including various levels of searches as per INTEGRAL data organization (observations, observations data sets, science windows, …).

The Planetary Science Archive (PSA) is holding all ESA Planetary missions datasets. The first version was released in March 2004 at http://www.rssd.esa.int/PSA and it now contains historical GIOTTO data as well as more recent Mars Express, Rosetta and Smart-1 data. All future ESA planetary missions' data (Huygens, Venus Express, BepiColombo) will also be part of the PSA. Also based on the XSA architecture, the PSA is the central place for all ESA planetary missions and offers similar power search, view and retrieving capabilities. Furthermore, a PSA dataset Validation Tool has been developed to allow the Mission Instruments Teams to validate their Datasets against the PDS standards before they are ingested into the PSA.

One Team, Several Projects

Since 2002, a dedicated Science Archives Team has been created at ESAC. It provides support to various archive projects. Each member of the group is not dedicated to one project 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. 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.

Accessing the Data through User Interface

By default, users can access the archives through a user-friendly but powerful GUI at the URLs given above. Developed in Java and made configurable using XML, the Archives GUI main characteristics are the following:
- Powerful and complex queries against the missions catalogues (observations, or other more fine grained catalogues, like sources, instrument configuration, groups of data, etc…)
- Queries organized by panels, allowing hundreds of queriable and displayable parameters
- Public data accessible to everyone, proprietary data accessible to data owner only or privilege users
- Configurable results display with navigation buttons through the catalogue list
- Product visualization tools
- Quick data download via a single click or more complex and customizable data retrieval via a shopping basket
- Selection of data products level for retrieval (from raw to fully processed data)
- Product retrieval via FTP

An example of the XMM-Newton Science Archive User Query and Display panels as well as a data product is given in Figure 1.
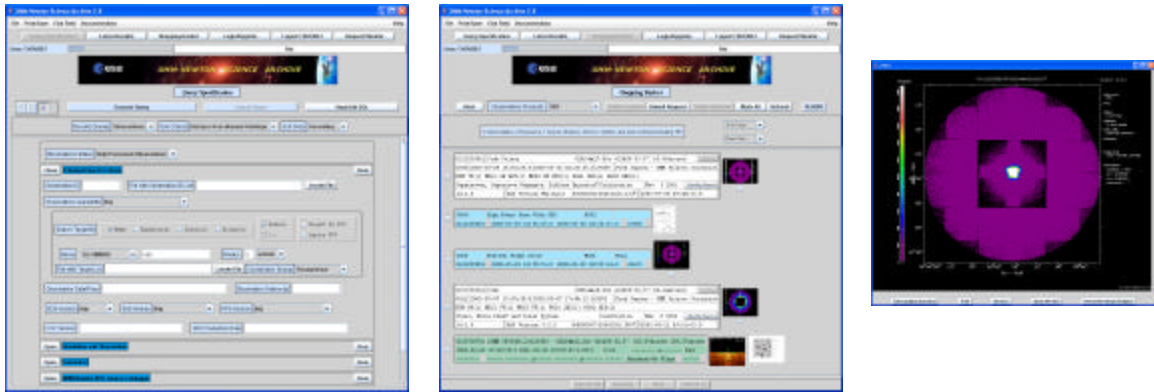
Fig. 1. XMM-Newton Science Archive User Interface

Accessing the Data through a scriptable interface

All ESAC archives have an open 3-tier architecture with a Project Business Logic providing transparent access to the Data products (usually saved on magnetic disks) and to the metadata (usually saved in a relational Database). The Archives GUI are making use of this Business Logic to perform queries and data retrievals to the archive.

Along the years, we have seen that although the existing archives GUI offer powerful queries functionalities, some expert users (eg Calibration Scientists) require a more direct access to the metadata and to the data products. To be able to perform some regular monitoring and calibration tasks in an automatic way, we created an Archive Inter-Operability (AIO) system which allows users to access data and metadata from a scriptable interface. The AIO system has been built as a complete new layer on top of the existing archive business logic (see Figure 2) to offer more services as required by the users:

- Access to archive metadata content via an XML scriptable interface
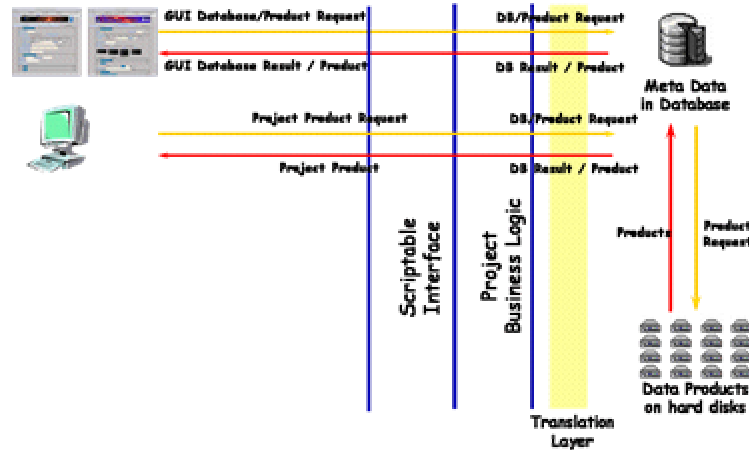- Access to data via HTML, XML scriptable interface or via a direct socket

Fig. 2. Scriptable Interface to the ESA Archives

This system allows the users to write scripts like :

```
foreach obs in (observation_list)
    AIOgetfile (obs, file_required)
    analyse_process file_required
    save results
    delete file_required
end
```

VO Compliant Archive Inter-Operability System

The Virtual Observatory is a world-wide initiative in astronomy which aims is to allow astronomers to perform new science by providing them with a "federation of astronomical archives and databases around the world, together with analysis tools and computational services, all linked into an integrated facility".

All astronomy VO projects around the world have formed the International Virtual Observatory Alliance (IVOA, http://www.ivoa.net/) to join their efforts to build an integrated facility. Various working groups have been setup and are defining inter-operability standards, such as :

- VOTable (XML based file format) for metadata and data exchange
- SIAP for Simple Image Access Protocol
- SSAP for Simple Spectra Access Protocol
- VOQL for VO Query Language
- Registry for storing VO resources
- VOSpace for virtual storage
- Etc….

ESA is participating actively in this VO initiative and has secured dedicated funds, aside the Archives funding but within the same team, to make sure ESA Archives are fully part of the VO. Hence, ESA-VO (http://esavo.esa.int/) project is part of the European VO project called EURO-VO (http://www.euro-vo.org/) and is as well part of the IVOA worldwide initiative.

These VO protocols are still under definition and evolution is to be expected. Therefore, there was a need to keep our archives inter-operability system flexible so we could adapt easily with existing and future VO standards.
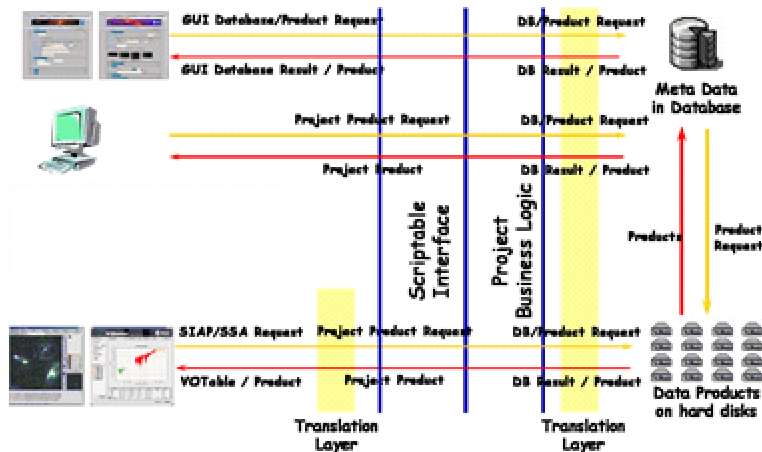

Fig. 3. ESAC Archives VO Compliant Architecture

By building "translation layers" using XML transformation rules (see Figure 3), we could easily adapt our existing AIO systems to these new VO standards. VO standards metadata query and data requests are

translated into our AIO metadata and data requests to be passed to our business logic. Similarly, metadata results and data products were translated from our system towards the VO standards.

This architecture presents the following advantages:
- Our existing archive back-end (data repository and metadata in the database) did not need to be modified to be published through the VO. These data are just transformed on-the-fly to meet the requirements of the VO protocols.
- When VO standards evolve, we will only have minor changes to perform to remain compliant with new standards versions, as we will just need to adapt our translation layers to these new standards.

VOSpec : A Tool to handle VO Compatible Spectra

Overview

Currently, some tools already exist in astronomy to display and analyze spectra coming from one mission. With the advent of the VO and the emerging interoperability standards, it has now become possible to develop a tool that is able to access spectra from various missions, display them together, allowing scientists to perform spectra analysis on a wider wavelength range.

VOSpec is already publicly available at http://esavo.esa.int/vospec/ and offers the following main functionalities:
- display and superimpose spectra from various VO resources around the world (see Figure 4)
- allow clever and powerful units conversion
- offer polynomial, black body and Gaussian line fitting
- display theoretical data together with real observational data
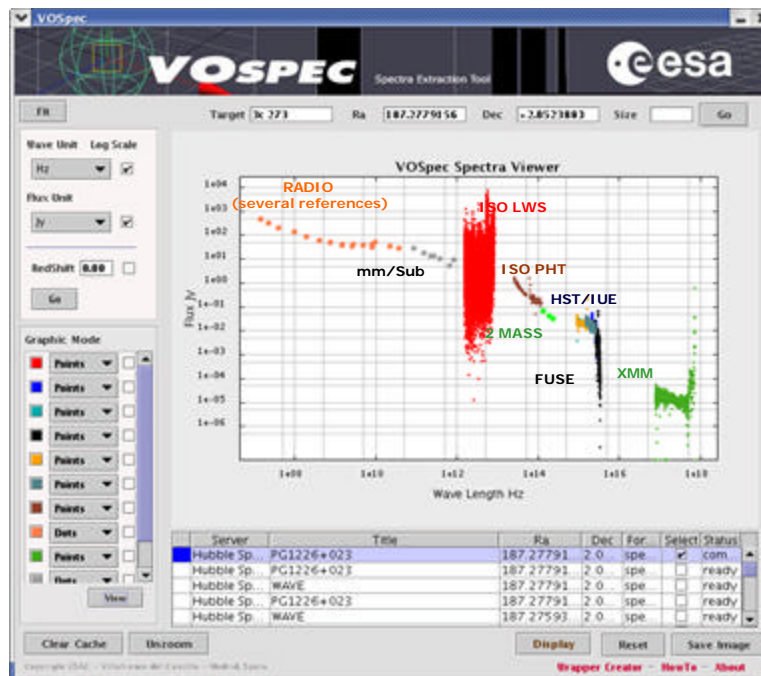- obtain spectra line data

Fig. 4. Analysis of Spectra Energy Distribution with VOSpec

From the beginning, VOSpec has been designed with VO in mind, making sure to build a tool that makes use of available VO interoperability standards and that is flexible enough so it could accommodate more functionalities when new standards are defined.

Finding VO resources and available spectra

When starting VOSpec, the user can define the target for which spectra have to be found. By clicking the GO button, VOSpec will transparently call the SIMBAD Web Services (located in the Centre des Données de Strasbourg in France) to resolve the target into coordinates which are usually used in astronomy for querying objects and sources.

Furthermore, VOSpec contacts with a VO Registry (e.g. the NVO registry in the USA or the ESA VO Registry in Spain) to ask for all the available SSA (Simple Spectra Access) Servers. These servers are VO servers which can serve spectra according to the SSA protocol. The user can select the servers he wants to query for that target and VOSpec will then connect with each of them to get the list of spectra available from them. Each server returns a VOTable which is an XML-based file format used in the VO world to exchange metadata and data. This VOTables contains the metadata information about these spectra, including a link (URL) to the real data. All these VOTables are compiled into a single one, displayed within VOSpec so the user can select the spectra he wants to see.

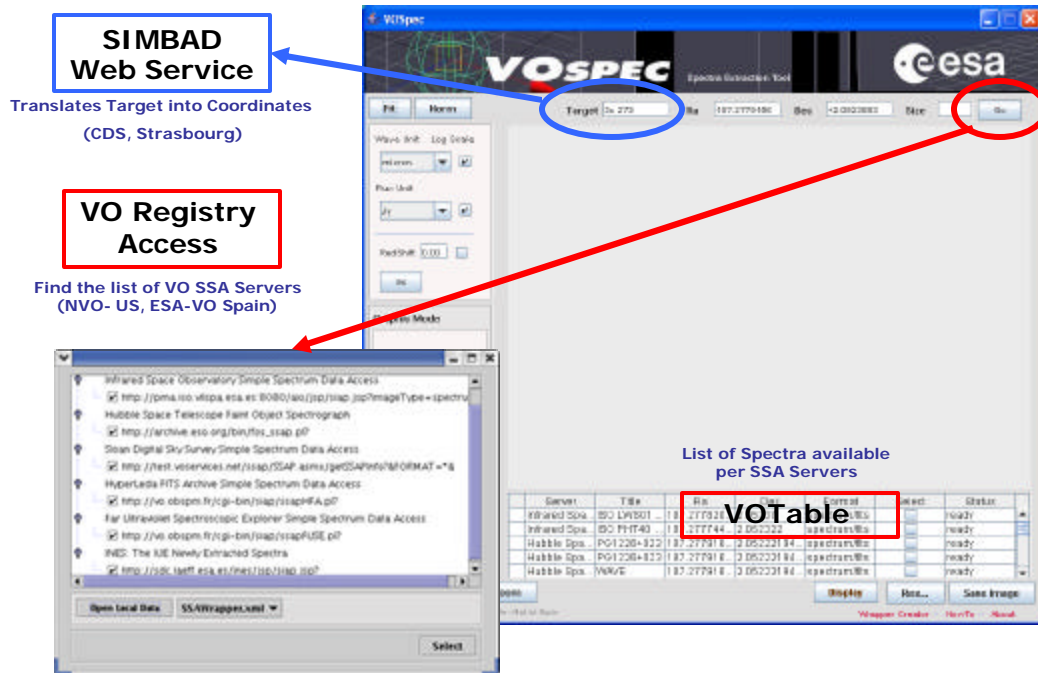The whole process described above can be see in Figure 5.

Fig. 5. Finding VO Spectra with VOSpec

Displaying VO spectra and additional services

Once the list of available spectra is displayed, the user can select the ones to be retrieved. For each of them, an other SSA call will be performed requesting the final data (spectra) from the corresponding remote SSA server, converting it on-the-fly to a common unit and then display all the spectra in the VOSpec main window as seen in Figure 4.

By September 2005, several data providers have deployed SSA servers for various missions:
- ISO from ESA, Spain
- IUE from LAEFF, Spain
- FUSE and HyperLedas from France
- HST from ST-ECF in Germany
- SDSS from JHU in the US
- Various other space missions from MAST in the US

As more data providers deploy SSA servers and register them in a VO Registry, they will automatically become available from VOSpec without any need to modify any software.

The tool also allows the users to make easy units conversion from one unit to another to allow the scientists to look at the data the way they like. Furthermore, there is also the possibility to fit the displayed line with well known lines in astronomy such as polynomial, black body or Gaussian lines.

VOSpec also offers the possibility to combine local user spectra with the ones obtained from the VO SSA server. To perform this, the local spectra must be in a SSA compliant format and VOSpec offers a tool to wrap local spectra into a SSA format so they can look as if they are coming from a VO SSA Server.

Additional services are also being defined, making use of other emerging VO protocols. In particular, collaboration with the Spanish Virtual Observatory project, effort is being put into making Theoretical Spectra available through SSA-like servers. Theoretical spectra are not normalized, but the tool offer the option to normalize these types of spectra to the real observational spectra so the scientist can check if the observed spectra corresponds to some existing theoretical spectra model.

Furthermore, work is being done in giving access to more detailed lines information by plugging the VO Simple Line Access Protocol (still under development) into the VOSpec.

Conclusions

ESA Science Archives are being developed by a single team located at the European Space Astronomy Centre, in Villafranca del Castillo, near Madrid, Spain. Through architecture and code re-use across various archive projects, development and maintenance of archives is more efficient and ensures long-term preservation of ESA science data holdings.

ESA Science Archives can be accessed through powerful and user friendly user interfaces, which is the main way for most of the world-wide scientists. Furthermore, scriptable interfaces have been developed for specific requirements of more expert users such as calibration scientists. An extra layer has been developed on top of these scriptable interfaces to adapt the ESA science archives to the Virtual Observatory standards in astronomy. By using clever translation layers, the archive back-end (data repository and databases) did not need to be modified to be accessed through the VO protocols.

The flexible and modular multi-tier architecture allows modification of one part of the system without affecting all the other parts. Additionally, its openness helps greatly in adapting it to new interfaces standards that can emerge, in particular the ones of the Virtual Observatory in Astronomy.

ESA Science Archives Team has also developed VO access to the ISO and XMM-Newton archives, as well as VO tools, like VOSpec, which enables the scientists to perform new type of science, thanks to the newly defined VO interoperability standards.

Acknowledgements