# ESA SKY: A TOOL TO DISCOVER ASTRONOMICAL DATA FROM ESA MISSIONS

*Jesús Salgado* , Bruno Merín* , Fabrizio Giordano* , Deborah Baines* , Elena Racero* , Belén López Martí* , María H. Sarmiento* , Raúl Gutiérrez*, Pilar de Teodoro*, Christophe Arviset**

*ESAC Science Data Centre, ESAC, Madrid

## ABSTRACT

In the last years, the amount of data and resources of astrophysical data available for the scientific community at the different project science archives at ESAC is growing quite fast. The project archives offer a quite detailed and easy access to all mission data. However, the multi-wavelength discovery of data could be cumbersome and it usually requires specific knowledge of mission dependent language.

The ESAC Science Data Centre (ESDC) is working on a science-driven discovery portal, called ESA Sky, that allows the exploration of the astronomical resources (almost all the images and catalogues from ESA missions at the present stage) using a simple, intuitive and project agnostic portal.

Using techniques like visualization of multi-order all-sky mosaics based on HEALPix (HiPS), missions coverage (MOC), observational footprints, TAP services on common data models for fast and performant searches, DB geometrical indexes, internal connections between databases and wrappers around the project archives to download the final science ready data allows the handling of big amounts of data in a simplified way.

We will present the recently released first version of this tool, technologies used and future plans.

***Index Terms*****—**big data exploitation, heterogeneous data sources, interoperability and standards, linked data and semantics, space science, visualization and visual analytics

## 1. INTRODUCTION

ESA Astronomical, Planetary and Heliophysics missions have their data archives at ESAC, Madrid. The different archives cover huge wavelength ranges and are an important resource of knowledge for the science community. In order to make this data available, the ESAC Science Data Centre (ESDC) [1] provides a set of project archives where the experts and scientists linked to a certain mission can access the data in an easy and powerful way. The different archives cover web based user interfaces to RESTful [2] interfaces, to allow the integration of the data access from different missions from the command line.

However, the different archives are focused on the discovery of data for particular missions, by using advanced query parameters based on the instruments and using project dependent jargon.

In 2014 a set of scientists and engineers at ESAC embarked on a task to design and implement a general multi-mission interface prototype that could fulfill the following basic goals:

- Allow a discovery interface, mainly based on the access to multi-wavelength data from different missions in a transparent way
- Implement an accurate discovery of observations using geometrical queries
- Use of project agnostic language

Using this paradigm, an application able to interconnect all the different archives using a discovery interface was developed as a prototype and presented internally to ESA members and projects.

There was the general agreement to convert this prototype into an operational application and, in Autumn 2015, the first beta version of the ESA Sky interface was presented to the general scientific community during the ADASS XXV conference in Sydney.

## 2. ARCHITECTURE

ESA Sky reuses modules and protocols defined by the IVOA (International Virtual Observatory Alliance) as a probe that the libraries and standards provided by this organization help on the design and implementation of archives modules. The application is composed of a client that makes use of the CDS Aladin Lite [3] module to visualize All-Sky maps.

The All-Sky HiPS [4] maps have been generated for all ESA missions by collaboration with the different project experts and selecting the best possible input data. HiPS are conceived as multilevel all sky maps where the data that travels from the server to the client is only the one that the client could need for visualization at the selected zoom level. That implies different resolution orders and a tesellation of the sky following *HEALPix* [5] indexes. In contrast with other kind of map projections used by other applications like, e.g. Google Sky, HEALPix makes use of

 doi: 10.2788/854791 

equiareal elements and good projections on the map poles, which is particularly important in astronomy.

A Table Access Protocol complaint server, TAP, [6] is used to discover observational data from ESA astronomical missions by using geometrical queries on accurate footprints. An accurate footprint of an observational based mission is considered the spatial coverage on the sky where photons have been accumulated. It is particularly important to have this kind of accurate footprint for the observations so a search for data for a certain astronomical object does not provide false positives in the results. This is why, for most of the missions, the footprints have been also generated for the ESA Sky project using only the best possible description by, e.g., ignoring sections of the sky where there is no real science data, although that spatial region could be inside the instrument coverage projection on the sky.

Most of the ESA catalogues are also provided through this TAP interface and connected to the ESA Sky graphical interface to enable the implementation of complex science use cases. Also, as all this metadata are available through a standard service, other TAP clients can make use of this service to implement, e.g. access to command line scripts.

As a last step, direct links to the full data are present to redirect the users to the different project archives. By doing that, a user can navigate from quite general all-sky maps in different wavelengths to the ultimate science data from the project archives.

The ESA Sky user interface makes use of Google Web Toolkit (GWT) technology [7] that allows a quite configurable and powerful interface for the users and, at the same time, maintainable code from the technical point of view. GWT components are written in the Java language (although, only a subset of the java classes that can be mapped to javascript are supported) and then converted to javascript after compilation. This is why a number of Integrated Development Environments (IDEs) are available to code, debug and compile GWT, allowing the development of very complex projects that can be evolved in an easy and controlled way.

## 3. ESA SKY GRAPHICAL USER INTERFACE

In order to provide the best possible HiPS for the different ESA missions, the generation of these maps have been coordinated between ESDC members for products generation, ESA mission members for selection of the best image products, filtering and cut-outs and CDS for providing the provides to the user HiPS generator library. ESDC has implemented a GWT wrapper on top of Aladin Lite to provide a better integration in a more complex GWT User Interface. ESA Sky makes use of the powerful CDS Aladin Lite component to display All-Sky mosaics following the HiPS format. Aladin Lite is the simplified javascript version of the well-known java application Aladin from CDS and allows an easy integration in any page/archive. In this particular case, the integration has implied the creation of a wrapper around this tool to have a GWT compliant component that can be reused, not only within ESA Sky but, also inside any of the ESA astronomical archives.

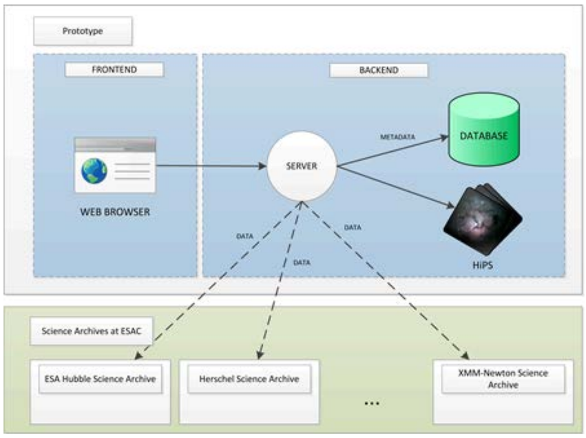


**Figure 1: ESA Sky Architecture**

The link to the ESA project archives is done by over-imposed detailed footprints that are shown under user requests. This is performed through the TAP service, which when invoked returns for different areas of the sky not only the detailed footprints but also tabular metadata with characterization information for the different observations. These metadata are shown in a tabular display for further access. The mechanism used for the access of observational metadata is based on the IVOA ObsCore TAP standard [8]. These metadata contain a link to a quick look image in graphical format and a link to access the data for this specific observational item. Both links are pointing to the original project archive so there is no duplication of scientific data at the ESA Sky level.

The same concept is applied to source catalogues with the representation of the positions as points in the graphical interface and tabular metadata with possible links to source related products like, e.g. finding charts.


doi: 10.2788/854791

A search on the interface provides to the user with an estimation of the number of observations per mission and sources per catalogue on the displayed area of the sky. In order to provide a fast response to this massive query, both at observational level and at source catalogues level, a *HEALPix* density calculation has been pre-calculated for the full sky so the number of observations and sources displayed is then a very fast sum of pre-computed density numbers over the visible region of the sky.
For the visualization, the HEALPix Multi-Order Coverage maps (MOCs) [9] are displayed for the observational results, when the area of the sky shown is very large. This is particularly useful when the mission spatial coverage is only a small fraction of the full sky as it helps to identify easily the regions of the sky where the mission has observations.

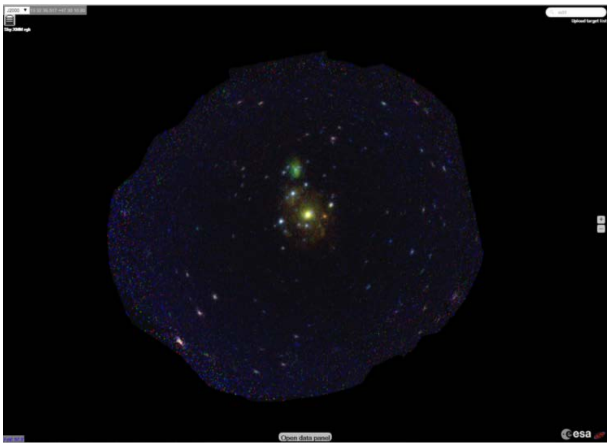


**Figure 2: Fraction of the improved XMM-Newton HiPS map, created by XMM SOC and ESDC (P. Rodriguez, E. Racero et al)**

The combination of MOCs, fast counts for all ESA observational data and catalogues and the presentation of results per spectral range provides to the user a quite accurate estimation of the amount of data available per wavelength range for all ESA astronomical missions in only one step. This is a rather innovative way to access big amounts of data by accessing summary histograms. This approach can be used not only in astronomy and for this particular use case, but also from a general point of view when the amount of data is so huge that a tabular representation in the first go could not provide a clear order of magnitude of the magnitude of the data found.

## 4. DATA IN ESA SKY

Currently, ESA Sky includes access to the data from the following missions:

- XMM-Newton images: HiPS all sky maps generated at ESAC with the support of XMM SOC. Access to observational data through detailed footprints and to data connecting to the XSA (XMM-Newton Science Archive)
- XMM-Newton catalogues: Slew catalogue, 3XMM, OM
- Herschel images: HiPS all sky maps generated at ESAC with the support of Herschel SOC. Access to observational data through detailed footprints and to data connecting to the HSA (Herschel Science Archive)
- HST images: HiPS all sky maps generated at ESAC. Access to observational data through detailed footprints and to data connecting to the eHST (European HST archive at ESAC)
- HST catalogues: Hubble Source Catalogue
- ISO images: HiPS all sky maps generated at ESAC
- Planck maps: HiPS all sky maps generated at ESAC
- Planck catalogues: PCCS2 (Second Planck Catalogue of Compact Sources), PGCC2 (Second Planck Catalogue of Galactic Cold Clumps) and PSZ2 (Second Planck Sunyaev-Zeldovich catalogue)
- Integral images: HiPS all sky maps generated at ESAC with the support of Integral SOC
- Integral catalogues: INTEGRAL General Reference Catalogue

Other major source catalogues are available for exploration and the access to the first Gaia catalogue will be also available as soon as the catalogue is made public (expected for mid 2016).

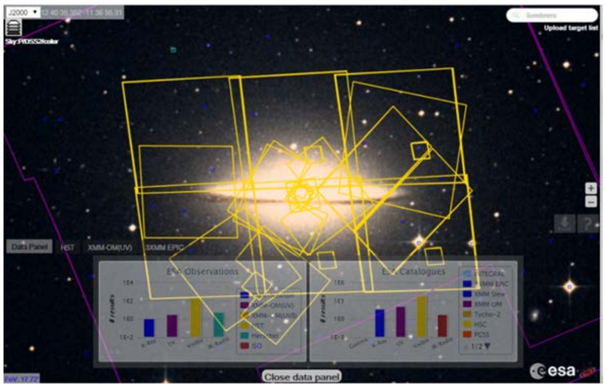


**Figure 3: ESA Sky User Interface**

## 5. INCLUSION OF NON-ESA DATA


doi: 10.2788/854791

Due to the interest from different projects and from the community, inclusion of other missions, including non-ESA ones, is also foreseen. By following some simple recipes provided by ESDC, data from other data centres can be made accessible on the ESA Sky interface. The mechanism to publish data into ESA Sky implies for the different data centres:

- Generation of HiPS maps: Taking all the images provided by the project as input, a HiPS structure is generated for exploration. This procedure is not only technical but, also, a good knowledge of the data facilitates the improvement of parameters and cuts to have the best possible all sky map. The generation of HiPS maps allows the publication of data not only inside ESA Sky, but also inside other VO applications
- Access to observational footprints: For the observational data, detailed footprints describing the spatial coverage of the observations should be generated. These footprints will be represented as an STC-S polygons [10]. Different techniques to generate these footprints are described in the ESDC how-to pages
- A table, containing general observational metadata should be generated by the data provider, fulfilling a common data model, similar to IVOA ObsCore DM
- The previous table could be exposed to the public by a TAP server by the data provider or directly provided to the ESA Sky team for ingestion inside the application data base
- The observational table should contain a link to the data located in the data provider side, so users will finally access the data from the data provider

Following these steps, data can be show in the ESA Sky interface and users can access project data directly from the data provider nodes.

Further information at:
http://www.cosmos.esa.int/web/esdc/esasky-contributing

## 6. FUTURE PLANS

Currently, ESA Sky has been focused on the access to images from different missions and source catalogues. A future enhancement will allow the discovery of spectral data in a similar way to the discovery of images. Spectral slits will be displayed in the interface for the different missions and a dedicated user interface module will allow the visualization and retrieval of spectral data. In the long term, there are plans to provide support for time domain data by the inclusion of time lines and time dependent HiPS maps.

At the same time, there are plans to include more missions into the ESA Sky interface in the short term and the connection to other external services.

Also, access to external VO applications will be obtained by connecting the ESA Sky interface to them using the IVOA SAMP protocol and access to other ESA archives by HTML5 interactions.

## 7. ACKNOWLEDGEMENTS

We acknowledge the support of the CDS team members, Pierre Fernique and Thomas Boch on the Aladin Lite component and HiPS generation. From ESA Science Operations Centers, we acknowledge Pedro Rodriguez (XMM-Newton SOC), Guillaume Belanger (Integral SOC) and Marcos Lopez-Caniego (Planck SOC) for their support on the creation of the HiPS maps for these missions.

Also we acknowledge some of the former ESA Sky team members: Iñaki Ortiz, Ignacio León, Andy Pollock and Michael Rosa.

## 8. REFERENCES

[1] Arviset et al, *Long-term Strategy for ESA Science Archives*, 2014ASPC..485..159A, ASP conference series, vol. 485, 2014, p.159
[2] Richardson, L., Ruby, S. *RESTful web service*, ISBN 978-0-596-52926-0 (2007)
[3] Boch, T.; Fernique, P., *Aladin Lite: Embed your Sky in the Browser*, *2014ASPC..485..277B*, ASP conference series, vol. 485, 2014, p.277
[4] Calabretta M.R., Boudewijn F.R., *Mapping on the HEALPix grid*, Mon. Not. R. Astron. Soc. 381, 865–872 (2007)
[5] Fernique, P.; Boch, T.; Pineau, F.; Oberto, *Create and Publish a Hierarchical Progressive Survey (HiPS)*, 2014ASPC..485..281F, ASP conference series, vol. 485, 2014, p.281
[6] Dowler P. et al, *IVOA Recommendation: Table Access Protocol Version 1.0*, 2011arXiv1110.0497D
[7] Dewsbury, R., *Google Web Toolkit Applications*, ISBN 978-0-321-50196-7, Prentice Hall, (2007)
[8] Louys M. et al, IVOA Recommendation: Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0, arXiv:1111.1758 (2011)
[9] Boch T, et al, MOC - HEALPix Multi-Order Coverage map, http://www.ivoa.net/twiki/bin/view/IVOA/IvoaApplications (2014)
[10] Rots, A., STC-S: Space-Time Coordinate (STC) Metadata Linear String Implementation, IVOA, (2007)