Astroinformatics
Proceedings IAU Symposium No. 325, 2016
M. Brescia, S.G. Djorgovski, E. Feigelson,
G. Longo & S. Cavuoti, eds.

doi:10.1017/S1743921316012874

Data Management in the Euclid Science Archive System

P. de Teodoro[1], S. Nieto[2] and B. Altieri[3]

[1]ESAC Science Data Centre (ESDC), Villanueva de la Canada, Madrid, 28692, Spain
email: pteodoro@sciops.esa.int

[2]ESAC Science Data Centre (ESDC), Villanueva de la Canada, Madrid, 28692, Spain
email: snieto@sciops.esa.int

[3]ESAC Science Data Centre (ESDC), Villanueva de la Canada, Madrid, 28692, Spain
email: baltieri@sciops.esa.int

Abstract. Euclid is the ESA M2 mission and a milestone in the understanding of the geometry of the Universe. In total Euclid will produce up to 26 PB per year of observations. The Science Archive Systems (SAS) belongs to the Euclid Archive System (EAS) that sits in the core of the Euclid Science Ground Segment (SGS). The SAS is being built at the ESAC Science Data Centre (ESDC), which is responsible for the development and operations of the scientific archives for the Astronomy, Planetary and Heliophysics missions of ESA. The SAS is focused on the needs of the scientific community and is intended to provide access to the most valuable scientific metadata from the Euclid mission. In this paper we describe the architectural design of the system, implementation progress and the main challenges from the data management point of view in the building of the SAS.

Keywords. Euclid, Data Storage, Information System, Science Archives, European Space Agency

1. Introduction

In the era of Big Data in Astronomy missions, Euclid Laureijs *et al.* (2011) will play a key role to analyze the data coming form the Euclid satellite in order to better understand dark energy and dark matter by accurately measuring the acceleration of the universe. Euclid will probe the history of the expansion of the universe and the formation of cosmic structures by measuring the redshift of galaxies.

Around 175 PB will be generated from the Science Ground Segment Dowler *et al.* (2011). The output catalog will contain the description of around 10 billion objects with hundreds of attributes. Euclid will be launched on a Soyuz rocket from Kourou at the end of 2020.

To fulfill the processing and scientific requirements, the Euclid Archive System (EAS) is composed by three different subsystems: the Data Processing System (EAS-DPS), the Distributed Storage System (EAS-DSS) and the Scientific Archive System (EAS-SAS). The EAS follows a metadata centric approach, where DPS and SAS store and manage Euclid metadata. On one hand, DPS manages all metadata needed for processing while SAS is responsible to provide access to the public data releases transferred from DPS.

The aim of SAS is to provide the set of tools and interfaces to query the Euclid data and fulfill the requirements from the scientific community. SAS is being built by the ESAC Science Data Centre (ESDC), which is responsible for the development and maintenance of the scientific archives of ESA missions.

https://doi.org/10.1017/S1743921316012874 Published online by Cambridge University Press

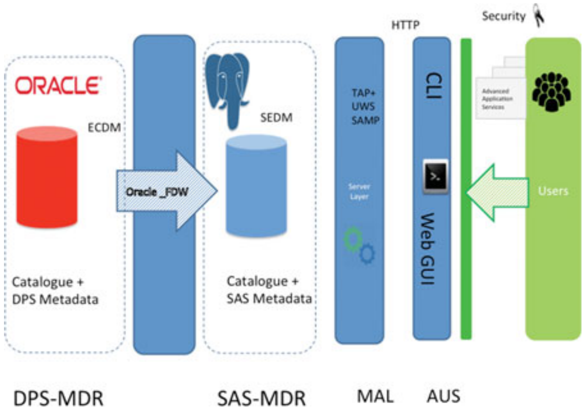


Figure 1. Data Transfer System from DPS to SAS and the public archive.

Finally, Euclid data is distributed across all SDC where DSS is a distributed set of interfaces that allow access to Euclid data in a seamless way. The Science Operations Centre (SOC) at ESA, will also host the full set of Euclid data in long term basis.

2. Data Transfer Between Systems

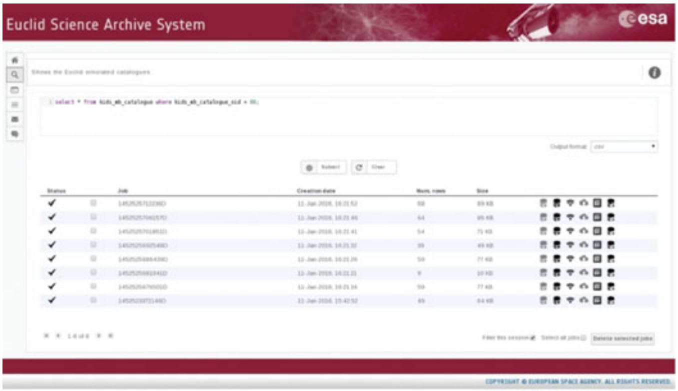
One of the critical subsystems of SAS is the mechanism in charge of transferring the metadata for the public releases. From the Data Processing System Metadata Repository (DPS-MDR) containing the metadata of the processed products, a Metadata Transfer Layer will allow to map the Euclid Common Data Model (ECDM) to the Scientific Exploitation Data Model (SEDM). These database systems can be from different providers. Currently the DPS-MDR is running in an Oracle database and the SAS data is stored in PostgreSQL. The mapping of data from the DPS into the SAS currently uses the PostgreSQL extension that provides Foreign Data Wrappers, which is a standardized way of handling access to remote objects from SQL databases for easy and efficient access allowing to map the data types between both different database providers in a seamlessly way. The mapping of both data models is a corner stone for the scientific archive. Currently in the processing of defining the scientific data model based on science use cases requested to the community using the information store in the Euclid Common Data Model, which is built by the Euclid Consortium. The SAS metadata plus some catalogues are stored into tables inside the database that will be finally queried. Figure 1 describes the elements involved in the transfer.

Later, the SAS data will be exposed through a common line interface or through the Web GUI of the ESA archives to the users. To allow to expose this data the TAP+ protocol Dowler *et al.* (2011), the UWS (Universal Worker Service) defined by the IVOA as a web-service which allows execution of one or several jobs in an asynchronous manner and SAMP as a messaging protocol that enables astronomy software tools to interoperate and communicate. There is also an authentication process for the users to be able to access privileged data. The portal makes use of the common look and feel and structure used by the 3rd generation of scientific archives.

3. Euclid ESA Archive

The Euclid ESA archive follows the latest generation of archives being developed by the ESAC Space Data Centre and its experience building scientific archives. The SAS builds

https://doi.org/10.1017/S1743921316012874 Published online by Cambridge University Press


Euclid Science Archive System
esa
Shows the Euclid simulated catalogues
Output format
csv
Submit
Clear
Filter this session
Select all jobs
Delete selected jobs
COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.

```
1 select * from kids_wb_catalogue where kids_wb_catalogue_oid = 88;
```

Status	Job	Creation date	Num. rows	Size
✓	[illegible]	11-Jan-2016, 16:21:52	68	83 KB
✓	[illegible]	11-Jan-2016, 16:21:46	64	85 KB
✓	[illegible]	11-Jan-2016, 16:21:41	54	71 KB
✓	[illegible]	11-Jan-2016, 16:21:32	39	49 KB
✓	[illegible]	11-Jan-2016, 16:21:26	59	77 KB
✓	[illegible]	11-Jan-2016, 16:21:21	9	10 KB
✓	[illegible]	11-Jan-2016, 16:21:16	59	77 KB
✓	[illegible]	11-Jan-2016, 15:42:52	49	64 KB

Figure 2. TAP Query interface.

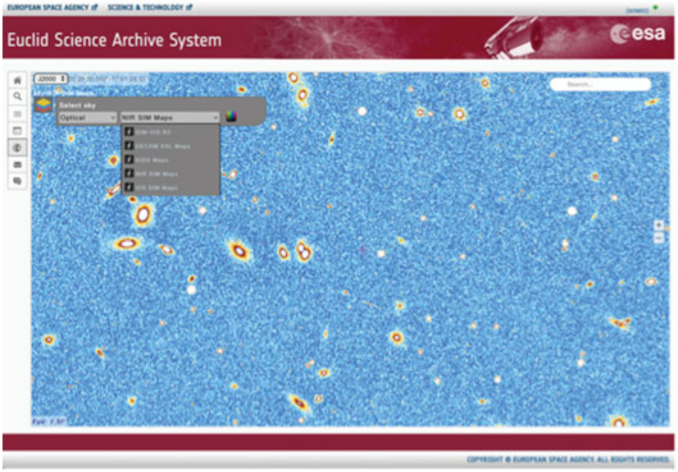
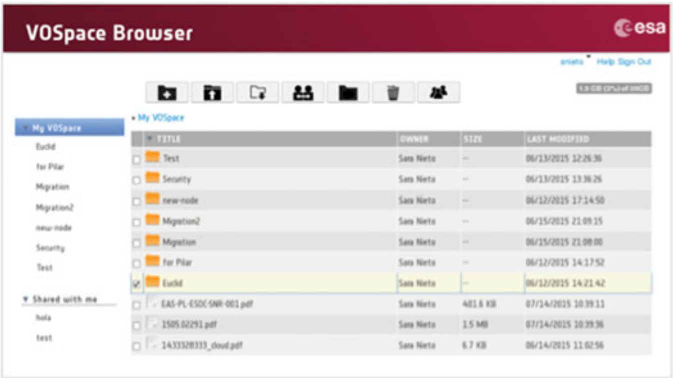


Figure 3. Euclid visualization interface.

on top of a three-tier modular architecture (client, server and data layer). This Euclid Scientific Archive is currently on development, building it on simulated data using a Java backend and a GWT front-end. The Euclid ESA Archive portal will show a Home page from where it will be possible to navigate to the top features. Also it offers Single Sign-On Authentication based on ESA Cosmos. On the left side, there is a menu showing different archive capabilities: search, maps visualization, explores query results and VOSpace user area. The Search link will navigate to a page where it will be possible to write queries in ADQL (Astronomy Data Query Language), the results will run asynchronously in will be shown in a query panel. See Figure 2.

In the Results panel, the results are displayed and it will be possible to download the data files associated with a search. This option will connect with the internal storage where the specific files will be located. The size of the SAS compared with the Euclid DPS metadata will be about a 20%. The scientific requirements on the SAS cover three main areas: parametric search for metadata and catalogues, data retrieval and the visualization of images and spectra, and more. Inherited from the experience on the ESDC ESASky a visualization tool is available allowing the exploration of the astronomical resources using a useful and intuitive web interface, see Figure 3. In the future it will show the maps created with Euclid data for analysis and exploration in an interactive way. In the

https://doi.org/10.1017/S1743921316012874 Published online by Cambridge University Press



• My VOSpace

	TITLE	OWNER	SIZE	LAST MODIFIED
□	Test	Sara Nieto	--	06/13/2015 12:26:36
□	Security	Sara Nieto	--	06/13/2015 13:36:26
□	new-node	Sara Nieto	--	06/12/2015 17:14:50
□	Migration2	Sara Nieto	--	06/15/2015 21:09:15
□	Migration	Sara Nieto	--	06/15/2015 21:08:00
□	for Pilar	Sara Nieto	--	06/12/2015 14:17:52
☑	Euclid	Sara Nieto	--	06/12/2015 14:21:42
□	EAS-PL-ESDC-SNR-001.pdf	Sara Nieto	401.6 KB	07/14/2015 10:39:11
□	1505.02291.pdf	Sara Nieto	1.5 MB	07/14/2015 10:39:36
□	1433328333_cloud.pdf	Sara Nieto	6.7 KB	06/14/2015 11:02:56

Figure 4. VOSpace Browser interface

current prototype it shows some external catalogs such as KiDS de Jong *et al.* (2013), DES and Euclid simulations that are processed using the Hierarchical Progressive Survey (CDS Library) creating HIPS maps.

4. ESA VOSpace: Astronomical data sharing and dissemination

VOSpace (Graham *et al.* 2015), the IVOA protocol for distributed data storage, will be available in the Euclid Science Archive System providing a storage abstraction layer and sharing capabilities transparent for the user on top of an independent file-sharing service available to the scientific community for access and data sharing. VOSpace, in Figure 4, allows the users to request science data from the archives of ESA missions and store them locally. To keep data locally is not an affordable option for most of the Euclid users, as it was not for Gaia users, for that reason VOSpace will provide to the users a private storage area directly connected to the archives of ESA missions saving local resources.

Users can select their own VOSpace area to store the query results. Once in VOSpace, users can access, transfer and share the data results with the rest of the science community. As science does not end with a query, any other tool or protocol can access VOSpace through the REST interface. This will allow the users to compose their own data pipelines and share the results with other colleagues.

5. Future Development

In the following years prior to the start of the mission, the Scientific Data Model will evolve and many challenges will be ahead among them the scalability of the system regarding space and resources. We plan as well to create a system which will support a dynamic data model: accommodating changes in ECDM without re-creation of the metadata storage scheme and migrating the data between different versions of the data model.

References

Laureijs, R. & others 2011, *Euclid Definition Study Report*
Arviset, C. & others 2016, *Big Data Challenges and New Paradigm for the Gaia Archive*
Dowler, P. & others 2011, *IVOA Recommendation: Table Access Protocol Version 1.0*
Graham, M. & others 2015, *IVOA recommendation: VOSpace specification v2.0*
de Jong, J. T. A. & others 2013, *The Kilo-Degree Survey*

https://doi.org/10.1017/S1743921316012874 Published online by Cambridge University Press