

Maximizing data accessibility at the ESAC Science Data Centre

J. González-Núñez, C. Arviset, J. Salgado
ESAC Science Data Center (ESDC)

Issue/Revision: 1.0

Reference: Asteriscs EDP ESDC Archives

Status: Issued

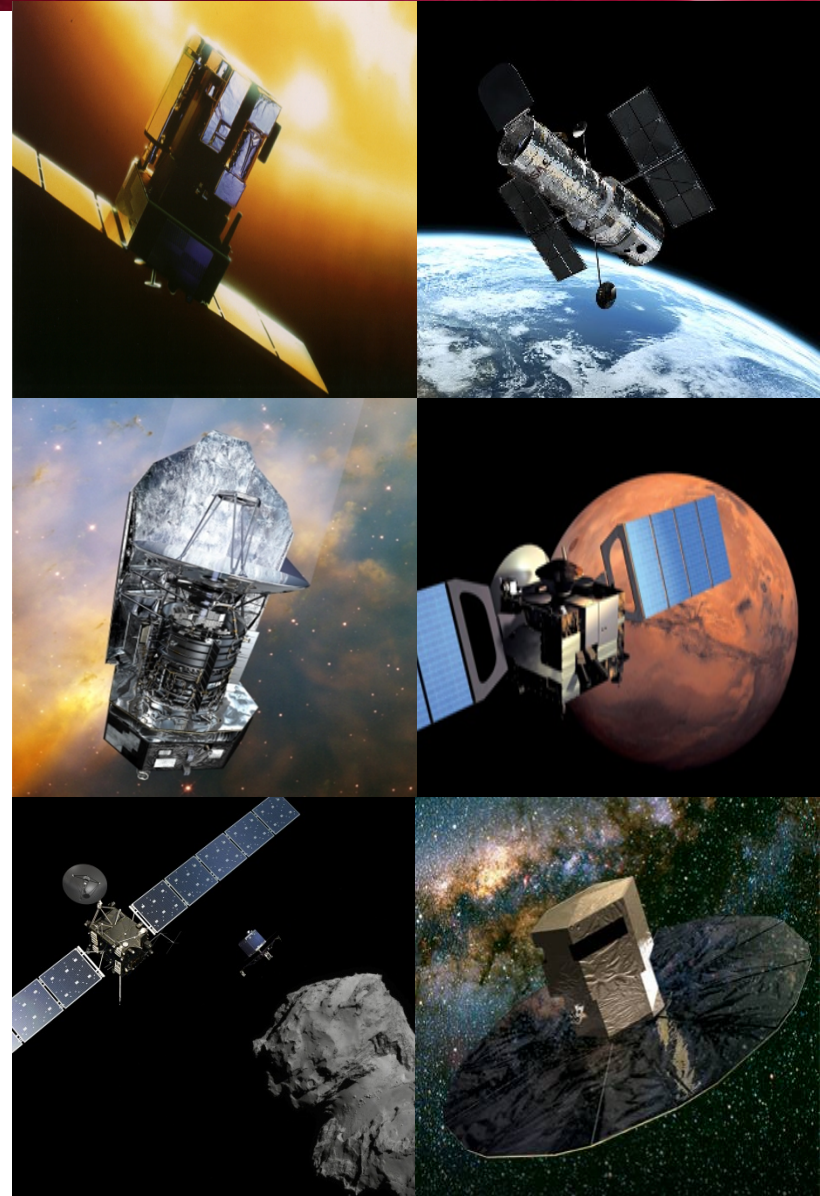
ESA UNCLASSIFIED - Releasable to the Public

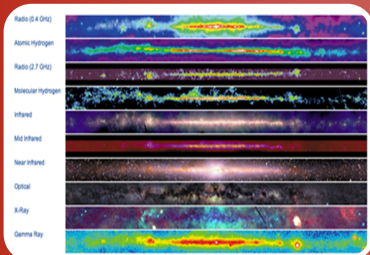
ESAC Science Data Centre

The Digital Library of the Universe



- Large set of science archives co-located at ESAC are a major research asset for community :
 - Astronomy, Planetary, Solar Heliospheric
- Different types of data:
 - Raw data, calibrated processed data, high level data products, ...
 - All data public and available on-line after a short proprietary period
- Need to be kept readily available for future users and novel uses by various types of users:
 - Scientific Community (public access)
 - PI team and observers (controlled access)
 - Science Operations Team (privileged access)
- Archive Strategy Plan for 5-20+ years

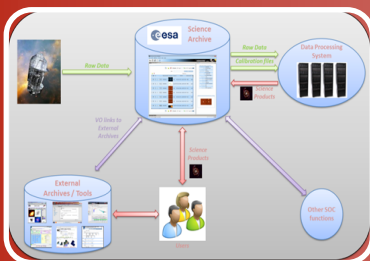




Enable maximum **scientific exploitation** of data sets



Enable efficient **long-term preservation** of data, software and knowledge, using modern technology



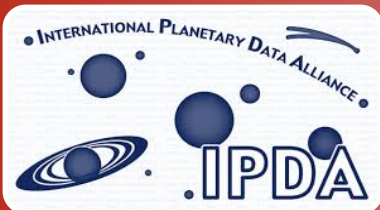
Enable cost-effective archive production by **integration in, and across, projects**

International Collaboration and Interoperability amongst Archives



IVOA - International Virtual Observatory Alliance

- Formed in 2002, 20 member projects
- Defines interoperability standards (VO framework) amongst astronomical (ground and space based) archives
- Working Groups and Interest Groups per technical domain (data access, data model, registry, applications, semantics, operations, ...)
- <http://www.ivoa.net/>

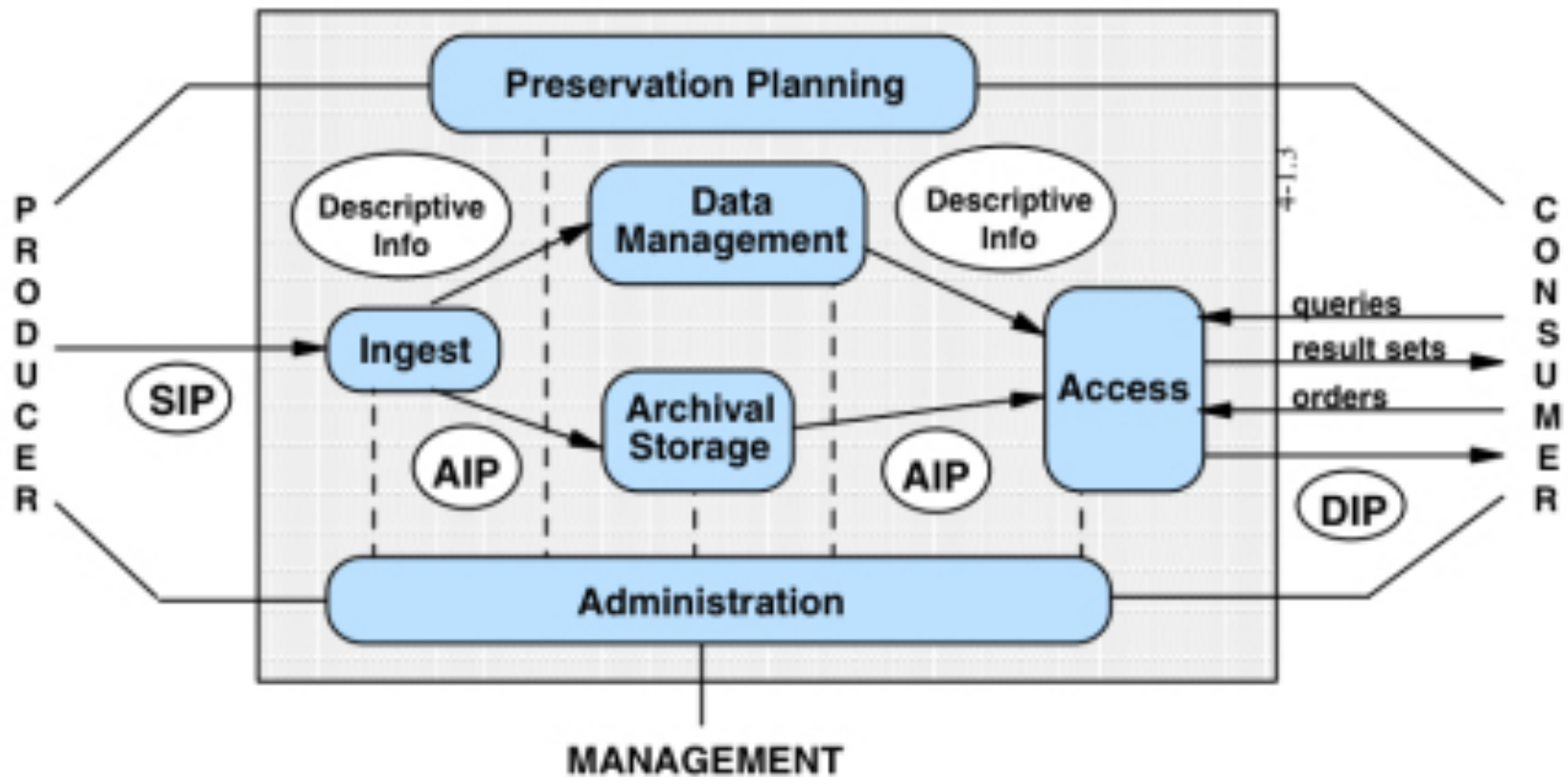


IPDA - International Planetary Data Alliance

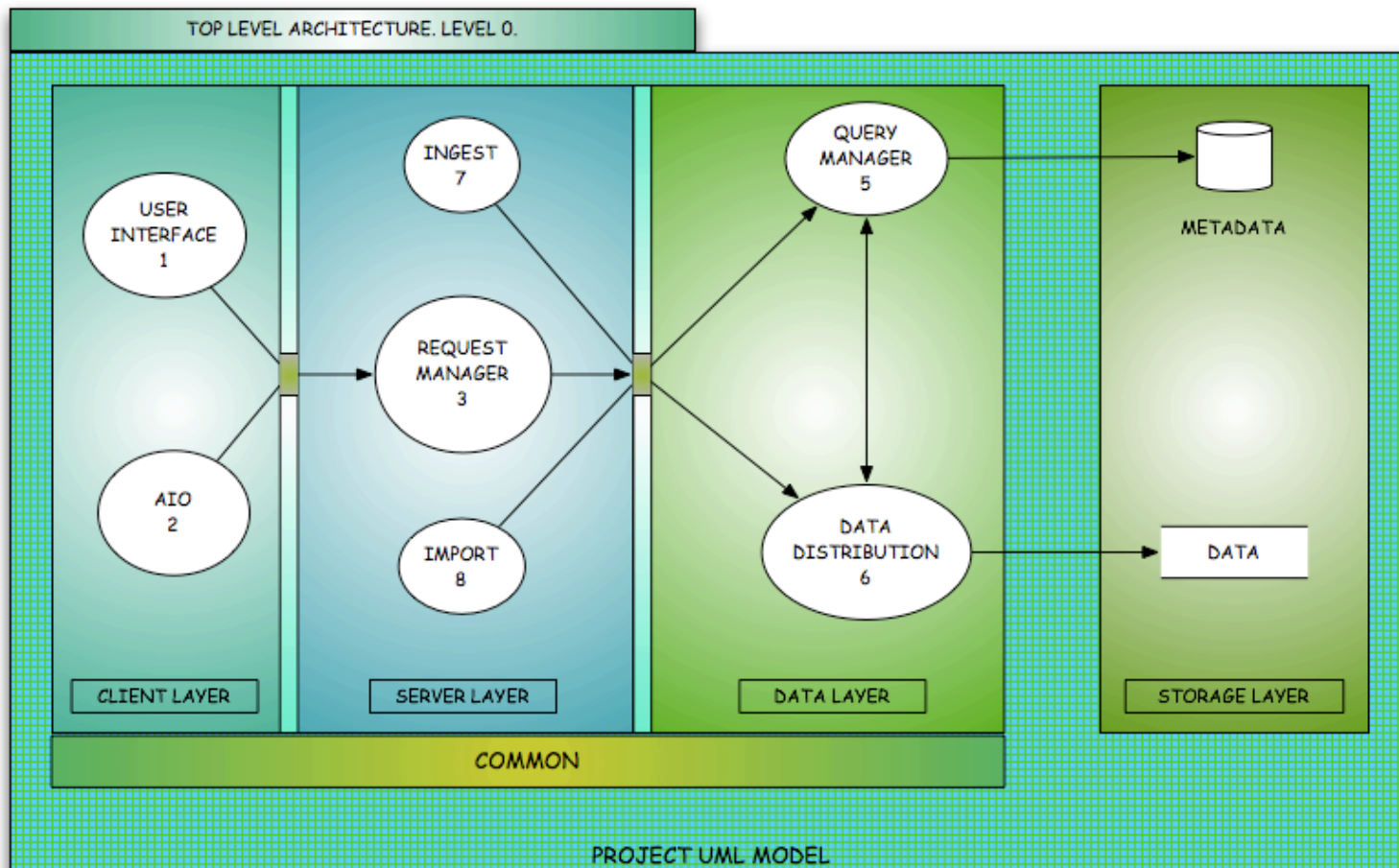
- Formed in 2004, 12 space agencies
- Defines archiving guidelines for planetary data
- Defines interoperability standards amongst planetary archives
- <http://planetarydata.org/>

Data Architecture @ ESDC archives

OAIS to ABSI

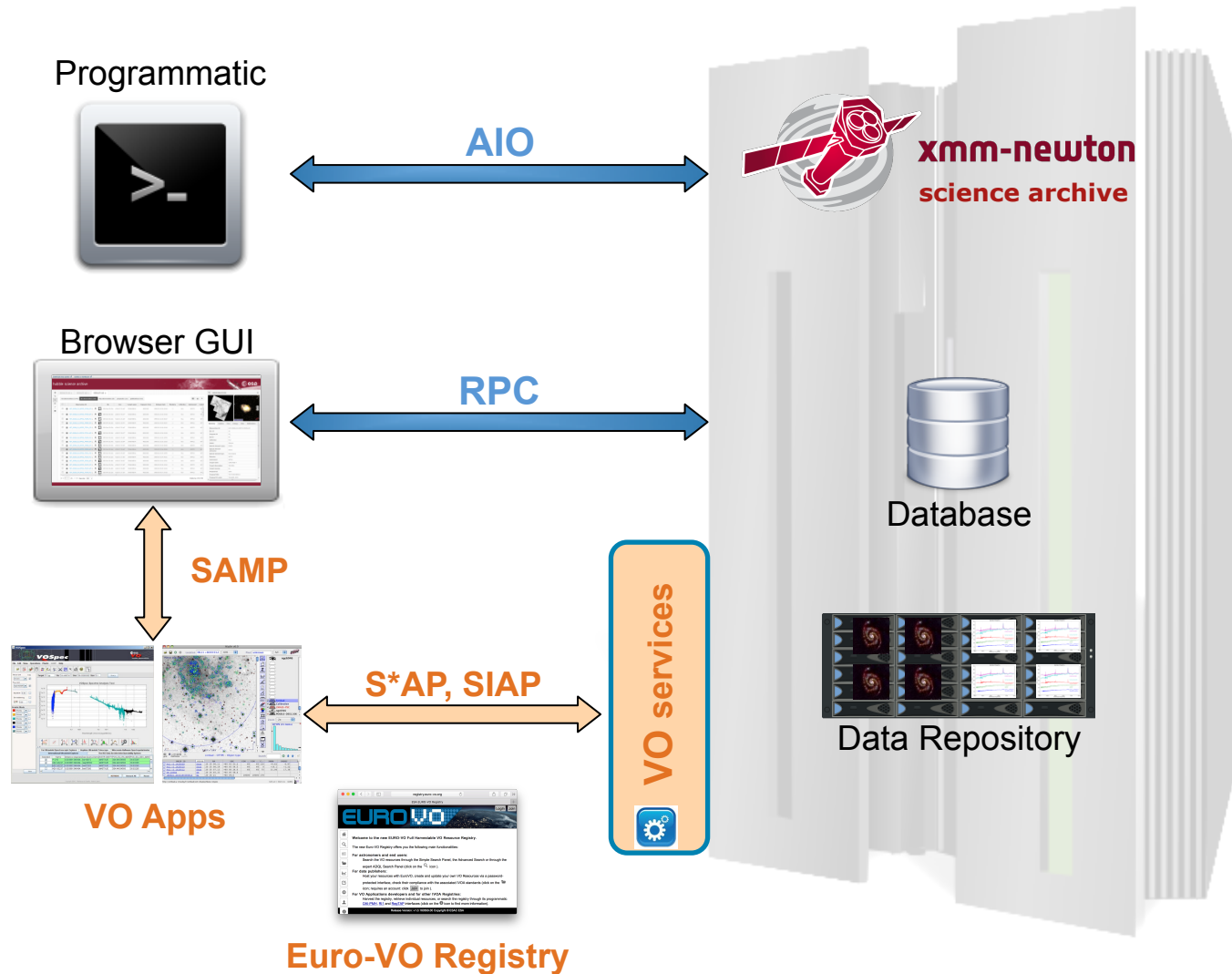


OAIS (2003)



Archives Building System Infrastructure (2006)

ABSI based architecture



1. SOC integrated archives:

- a. Archive as a science operations building block
- b. Data processing metadata & intermediate products

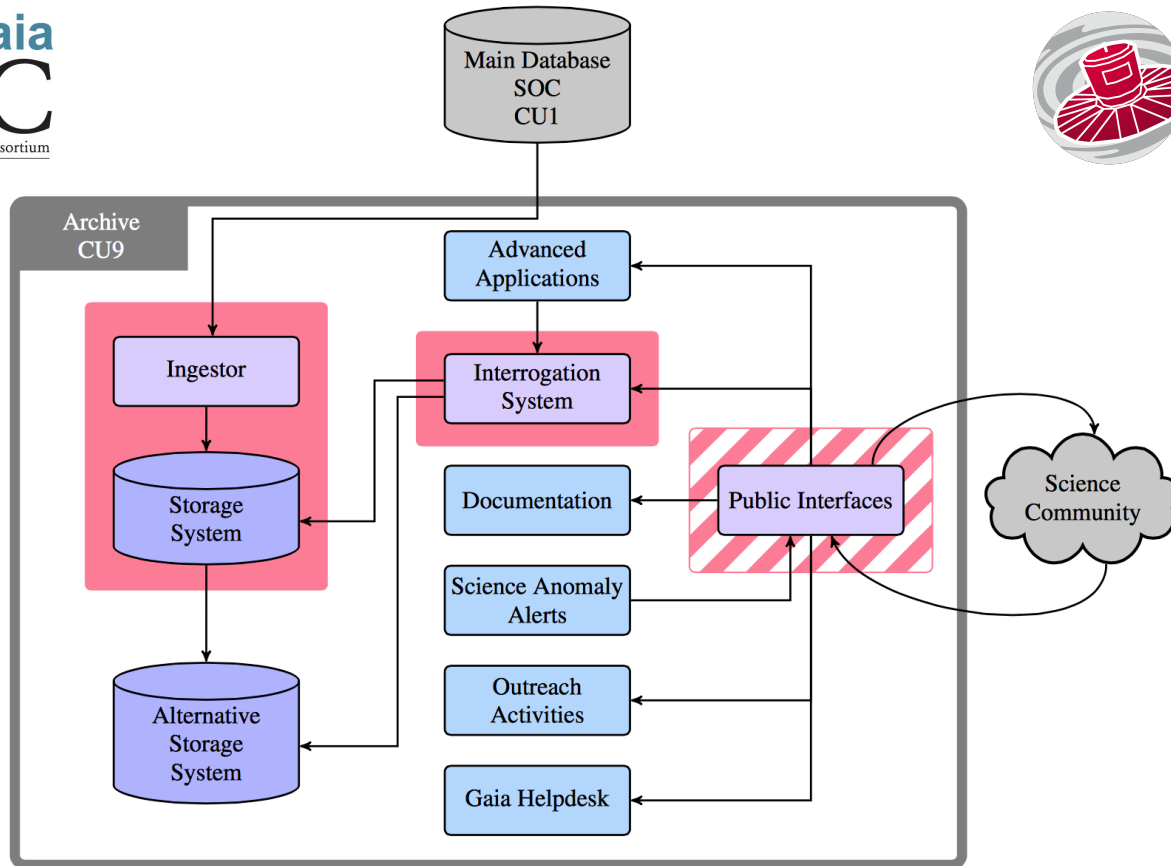
2. Archive consortiums (DPAC, Euclid Consortium)

- a. Integration with WP-developed modules:
Visualization, DM, Xmatch, Validation etc.

3. Large data volumes:

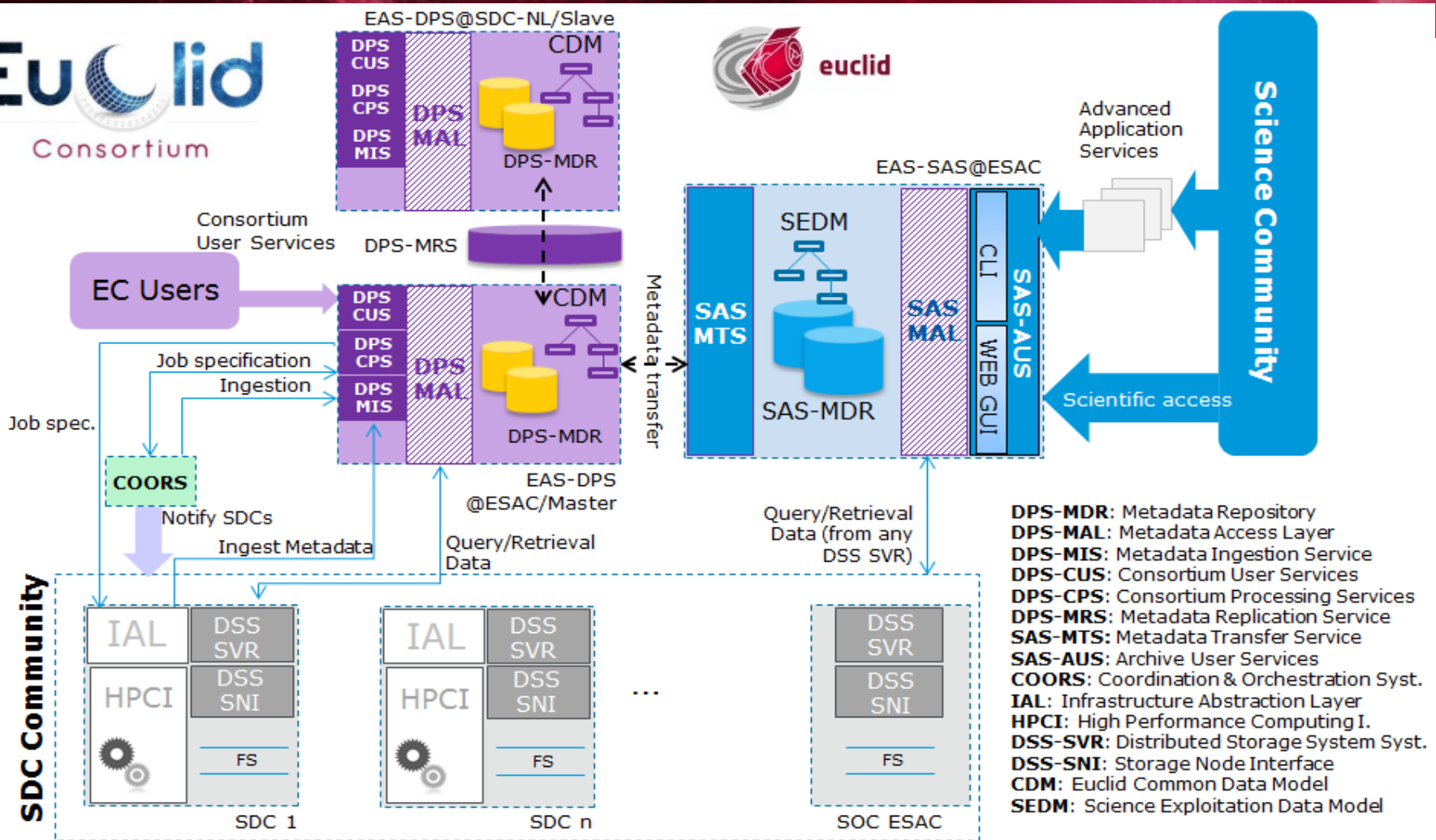
- a. Distributed data storage
- b. 'Big Data' analysis techniques

Archives Integration within Consortium



Archive is fully part of Science Operations, from the start
Strong collaboration with SGS and Consortium

Archives Integration within Consortium



Strong collaboration with SGS and Consortium

Statistics up to June 5

Nominal Mission Data

Type of Data	Amount	
Science telemetry	17 TB	
Astrometry transits	22.5×10^9	225×10^9 images
Photometry transits	22.5×10^9	45×10^9 images
Spectroscopy transits	1.5×10^9	4.5×10^9 spectra
Main Database	44TB	

- $\approx 30GB/day$ – $> \approx 100TB$ total
- with products can be 1PB total data by mission end

The Gaia Archive



gaia.esac.esa.int

Gaia Archive Core Systems Web Interface (***PROTOTYPE***)

EUROPEAN SPACE AGENCY ABOUT ESAC Juan Gonzalez (jgonza01)

gaia archive

HOME SEARCH STATISTICS VISUALIZATION HELP DOCUMENTATION VOSPACE SHARE ADMIN

Simple Form **ADQL Form** Query Results

Job name:

Query examples

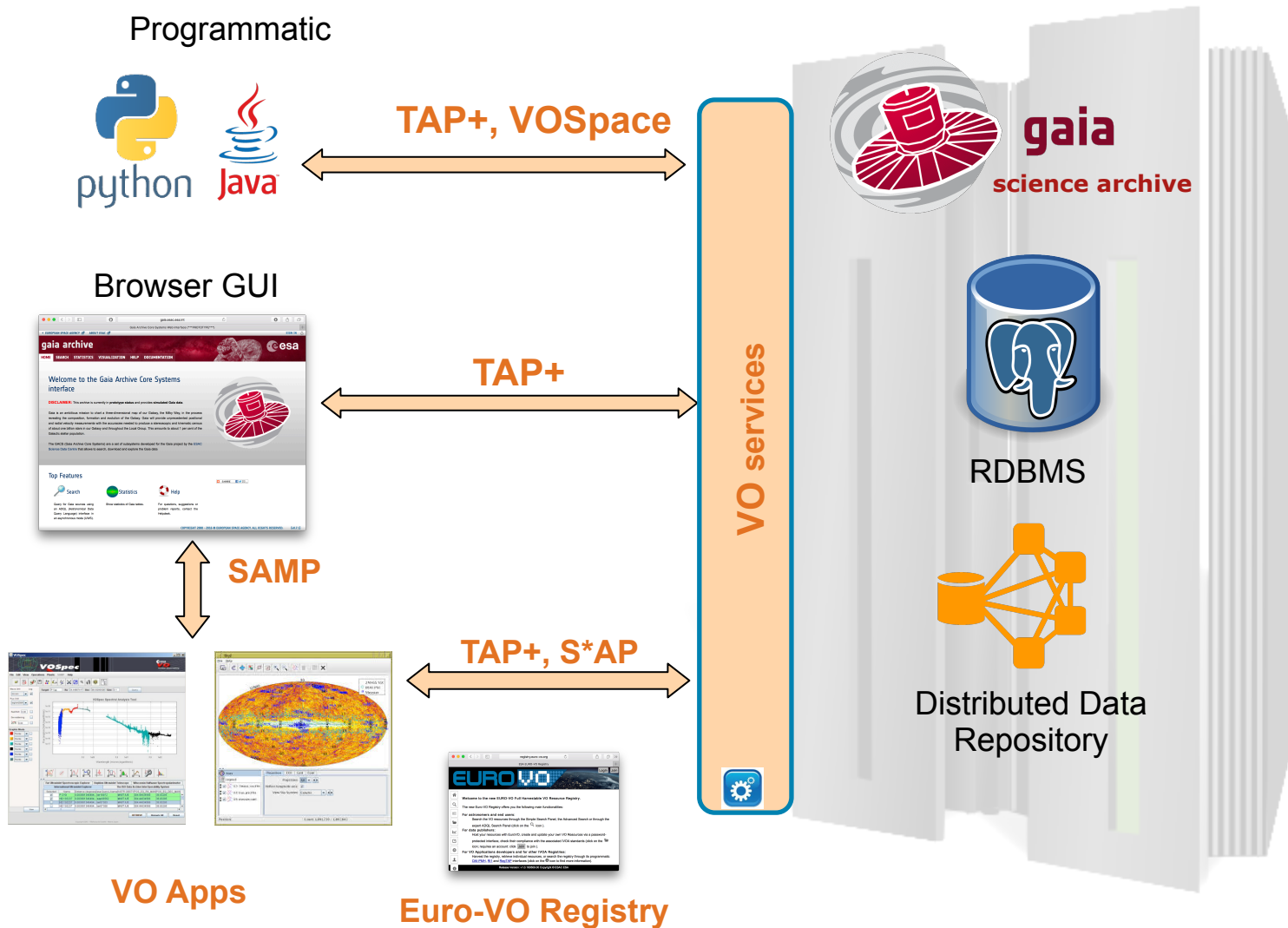
```
1 SELECT DISTANCE(POINT('ICRS',ra,dec), POINT('ICRS',266.41683,-29.00781)) AS dist, *
2 FROM public.gaia_source
3 WHERE 1=CONTAINS(POINT('ICRS',ra,dec),CIRCLE('ICRS',266.41683,-29.00781, 0.08333333)) ORDER BY dist ASC
```

Status	Job	Creation date	Num. rows	Size	
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1465561988522D"/>	10-Jun-2016, 14:33:08		0 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1464016796714D"/>	23-May-2016, 17:19:56		0 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1464016558821D"/>	23-May-2016, 17:15:58		0 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1462381049252D"/>	04-May-2016, 18:57:29	695	440 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1459261694854D"/>	29-Mar-2016, 16:28:14	100	72 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1459172451097D"/>	28-Mar-2016, 15:40:51	117096	81 MB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="xmatch_hipparcos_newreduction_tycho2"/>	28-Mar-2016, 15:40:34		0 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1456926874367D"/>	02-Mar-2016, 14:54:34	4	13 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>
<input checked="" type="checkbox"/>	<input type="checkbox"/> <input type="text" value="1456926865564D"/>	02-Mar-2016, 14:54:25	2	12 KB	<input type="button" value="i"/> <input type="button" value="🗨"/> <input type="button" value="🗨"/> <input type="button" value="📶"/>

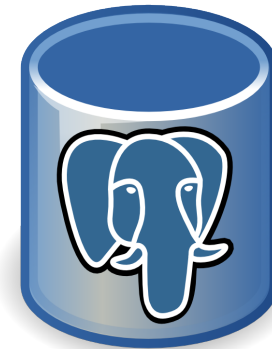
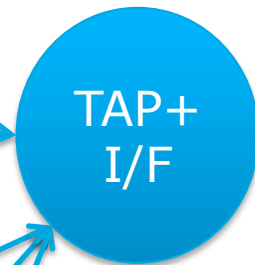
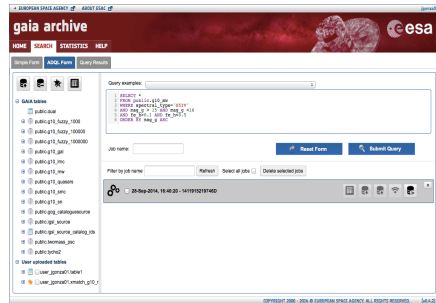
1-20 of 37 Select all jobs

COPYRIGHT 2000 - 2016 © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED. (v0.7.1)

Gaia Archive Current Architecture



Extending VO protocols: TAP +



Command line tools

External Apps

Data Validation

Public area

Restricted area

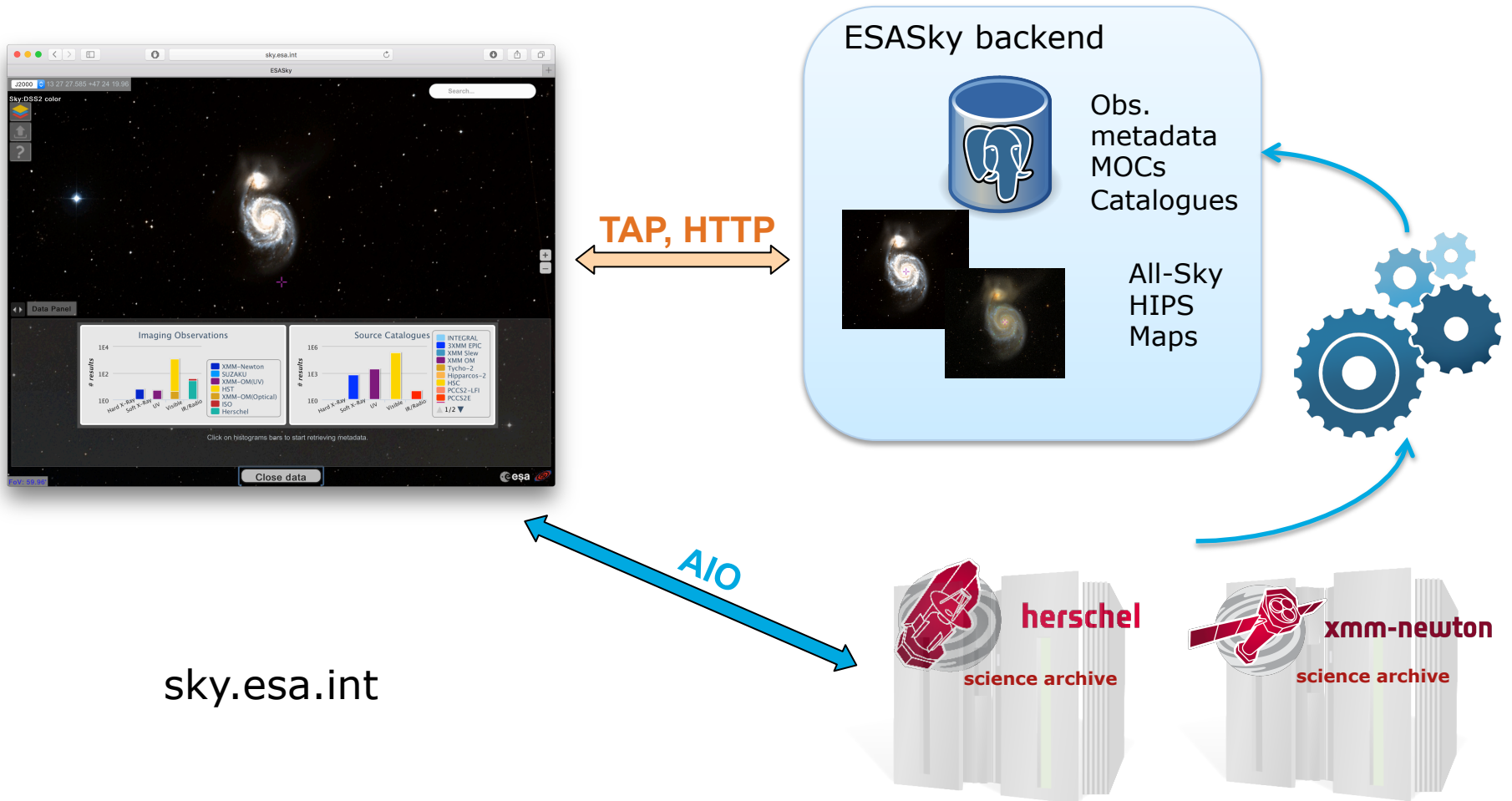
User Space

- Publicly released catalogues

- Catalogues during validation or proprietary exploitation

- User-uploaded data

ESASky : Multi-mission visualization



ASTERICS at ESDC archives

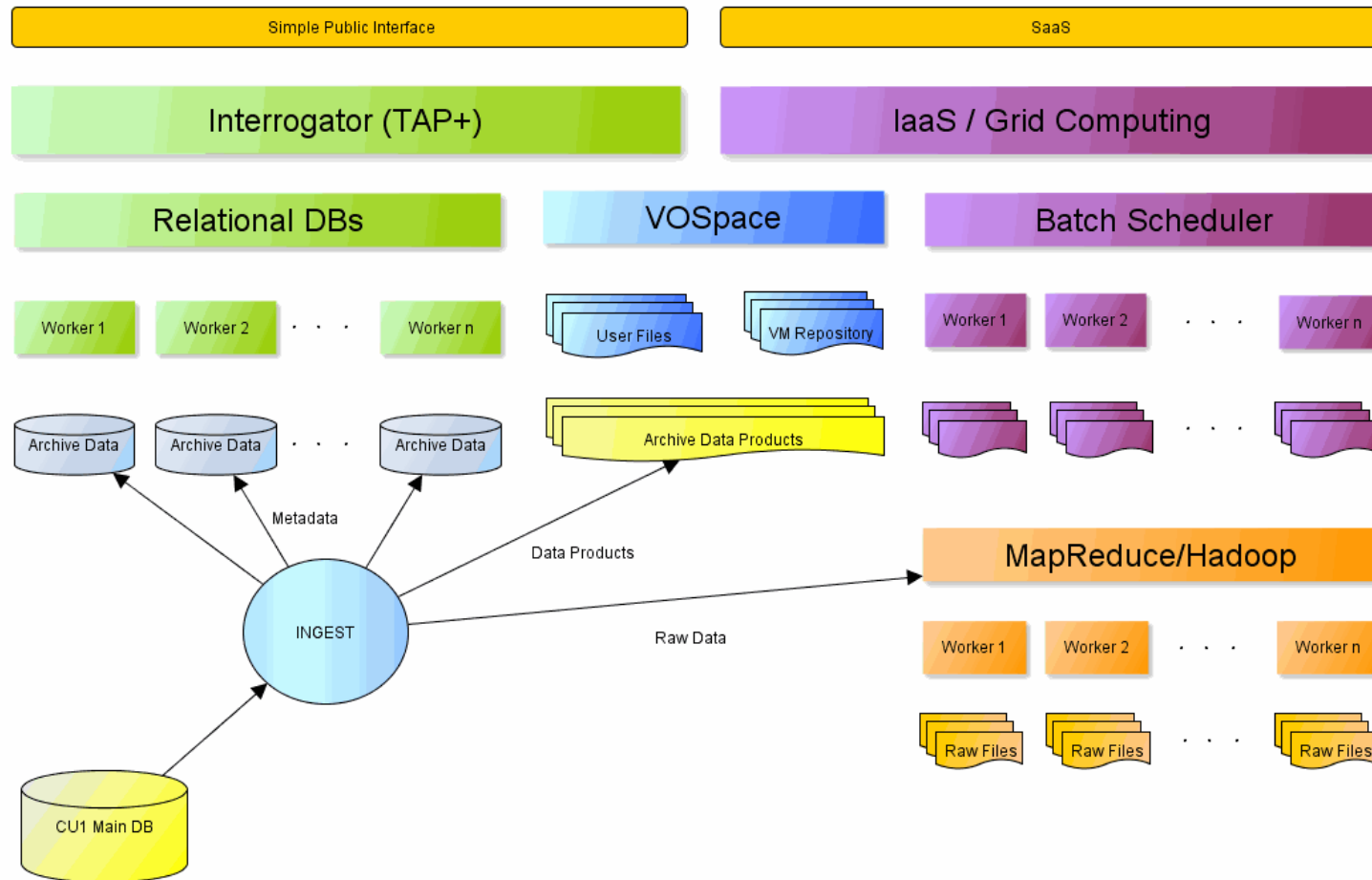
1. ADQL Auto-complete

- Provide suggestions in the Gaia Archive ADQL Query editor
- Stelios Voutsinas (WFAU, University of Edinburgh) - GENIUS - ASTERICS

2. pgSphere development

- Extension maintenance, configuration management
- Implementation of Gaia CU9 requested features
- Markus Nullmeier (ARI, Universität Heidelberg) - ASTERICS

Challenges



Gaia Archive Preparation (2011)

1. “Bringing the code to the data”

- a. GAVIP project (Docker)
- b. TAP+: user areas, user defined functions

2. Cloud Services

- a. ESAC Cloud

3. Large scale processing (Spark, Hadoop):

- a. Data Mining Work Package
- b. ‘Grand Challenges’
- c. Prototype simple use cases

Thanks!

<http://www.cosmos.esa.int/web/esdc>