THE EUCLID ARCHIVE SYSTEM:
A DATA-CENTRIC APPROACH TO BIG DATA

Sara Nieto on behalf of

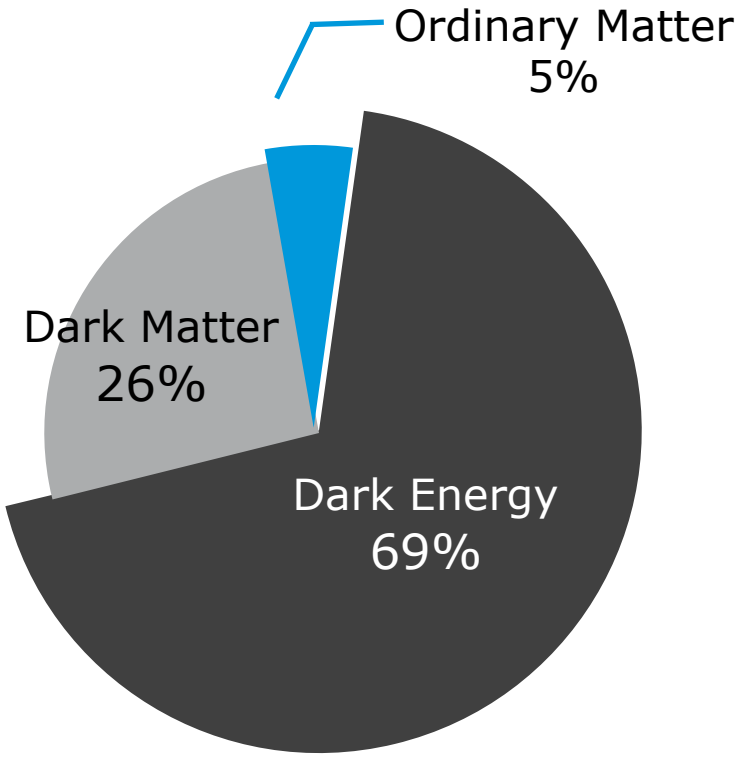
B.Altieri, G.Buenadicha, J. Salgado, P. de Teodoro
European Space Astronomy Center, European Space Agency, Spain
O.R. Williams, A.N.Belikov, D.Boxhoorn, B. Dröge, J.McFarland, A.Tsyganov, E.A. Valentijn
University of Groningen, Groningen, The Netherlands

Issue/Revision: 1.0
Reference: Euclid Archive System for ADASS @Trieste – 17 October 2016
Status: Issued
ESA UNCLASSIFIED - Releasable to the Public

Euclid Mission

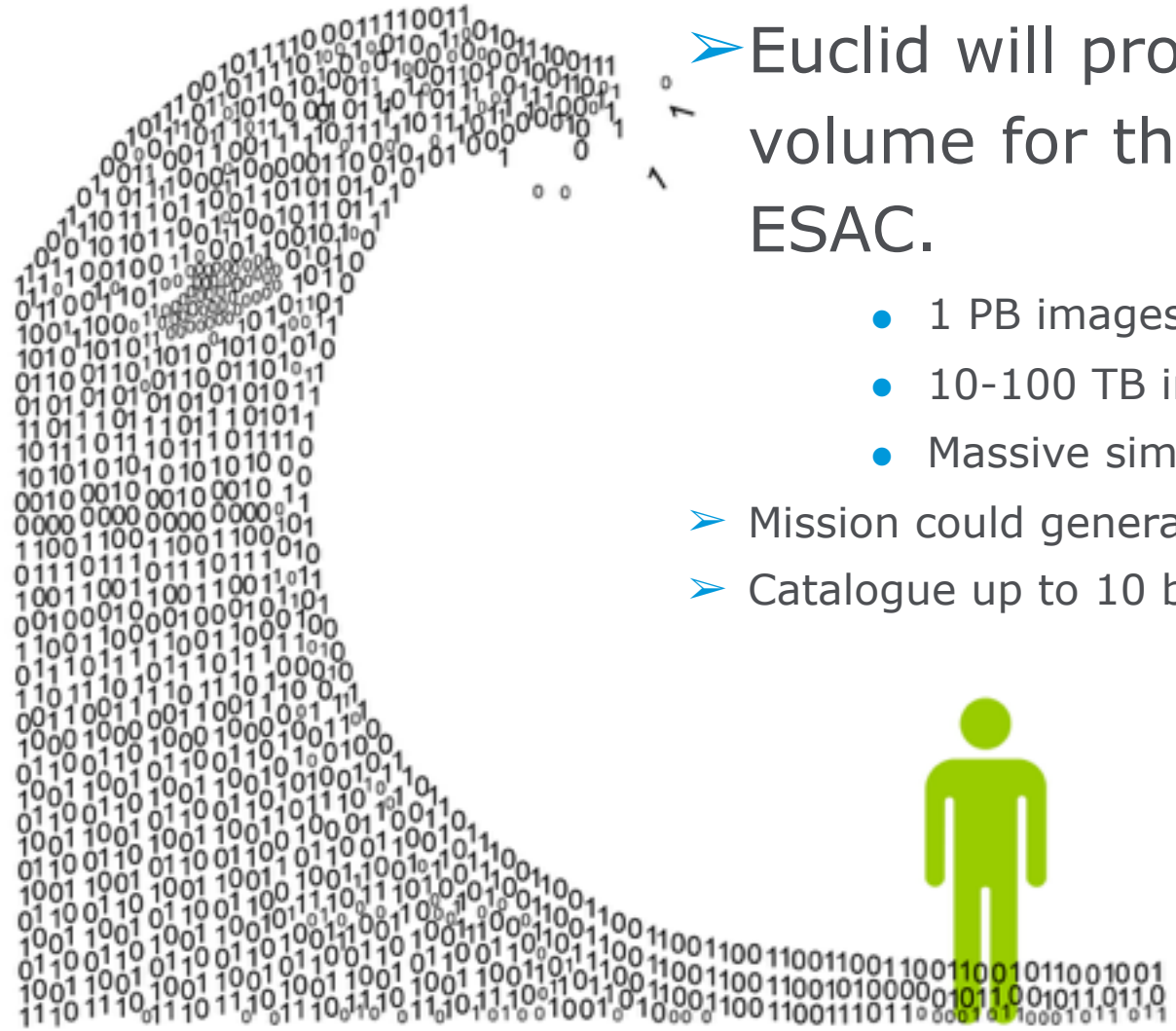
- 1.2m telescope, L2 orbit
- 5.5 years mission duration
- Map the sky in 1 optical band, 3 NIR bands and NIR slit-less spectroscopy
- Launch on Soyuz in Q4 2020

- **ESA** is responsible for the mission.
- The **Euclid Consortium** will supply ESA with the instruments and most of the SGS.
- Euclid Consortium & Other teams
  - 15 countries, 130 institutes, 1300 consortium members and 700 scientists



ESA UNCLASSIFIED - Releasable to the Public

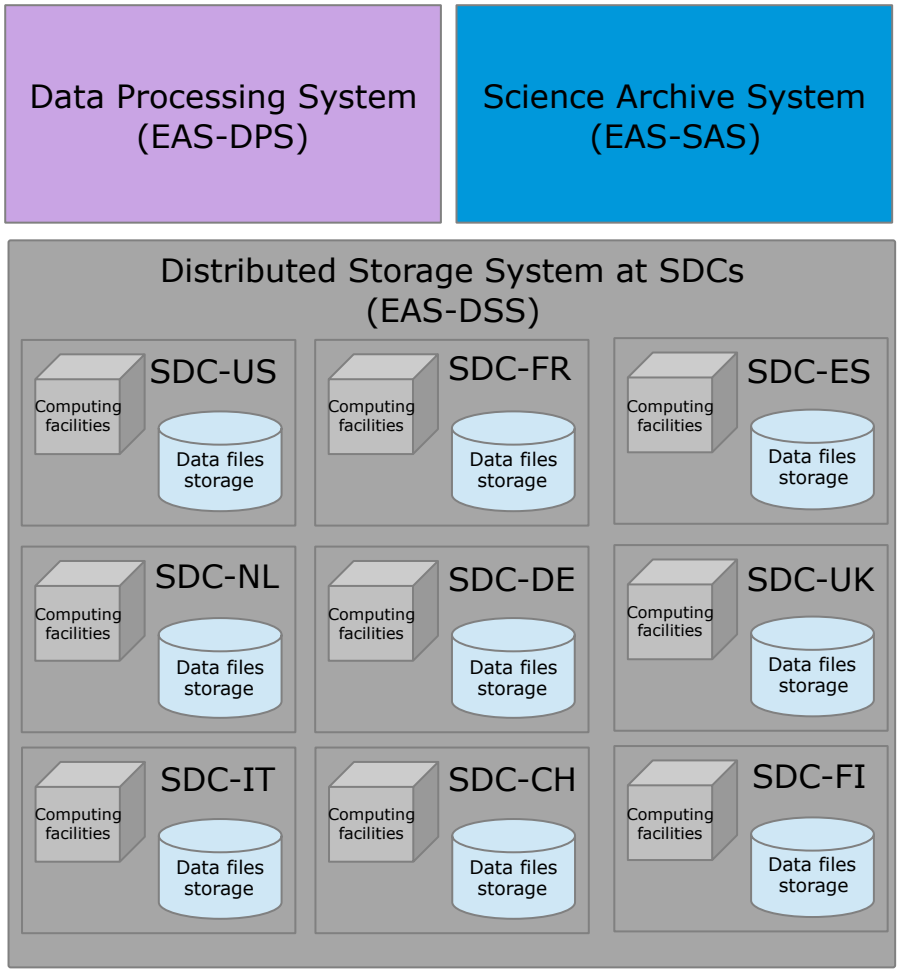
Euclid Data Challenge



- Euclid will produce the largest data volume for the Science Archives at ESAC.
  - 1 PB images
  - 10-100 TB images from ground obs.
  - Massive simulation efforts required
- Mission could generate 26 PB/year
- Catalogue up to 10 billions objects (tens of TBs)

ESA UNCLASSIFIED - Releasable to the Public

EAS Metadata Centric Approach

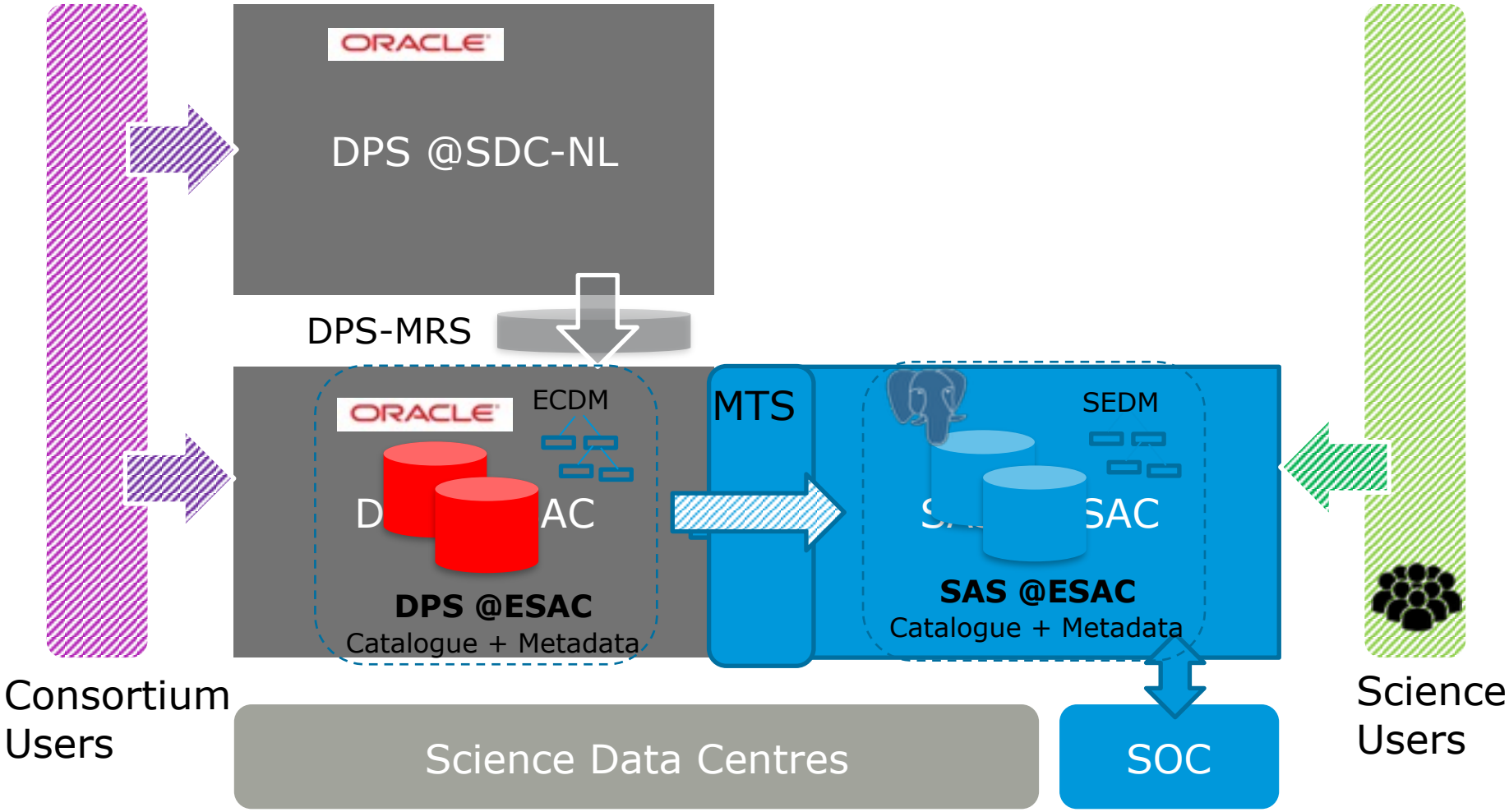


- EAS acts as interface between all ground system components
- Metadata is centrally stored
- EAS metadata contains all information except images
- EAS data distributed across SDCs
- EAS stores dependencies of data products
  - Avoid unnecessary reprocessing
  - Data provenance ensured
- SOC will also host a DSS

ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

EAS Decomposition



ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

Scientific Archive System

EAS-SAS @ESDC

The SAS is being built at the ESAC Science Data Centre (ESDC), which is responsible for the development and maintenance of the scientific archives for the Astronomy & Solar System missions of ESA.

Science Community

The SAS is focused on the needs of the scientific community and it will provide access to the most valuable scientific metadata from the EAS-DPS. **Euclid will produce the largest data volume for the Science Archives at ESAC.**

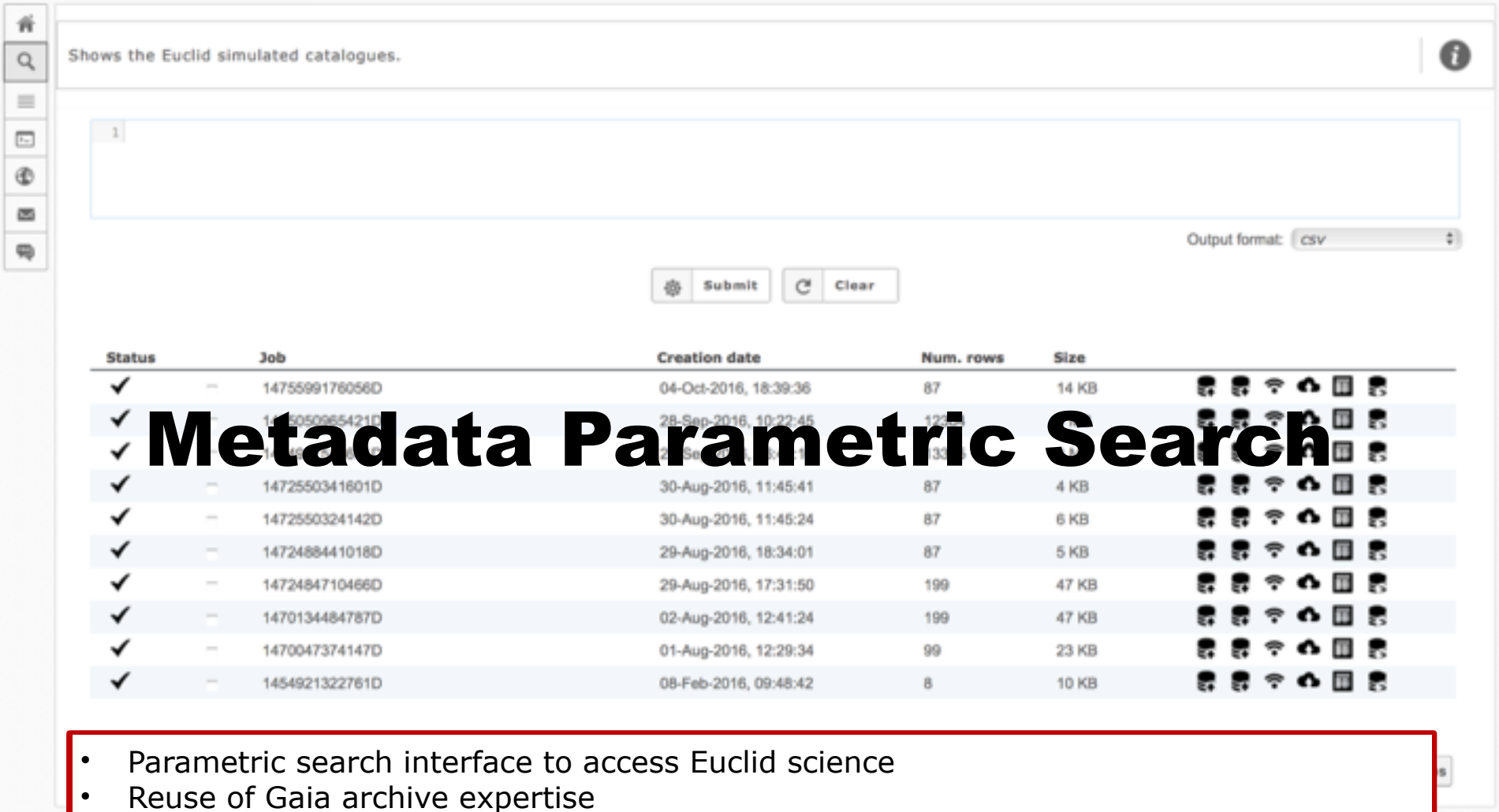
ESA UNCLASSIFIED - Releasable to the Public

Access to scientific data online

- ESDC latest GUI Look & Feel
- "Single Sign-On" Authentication based on ESA Cosmos
- Left side menu to access archive capabilities: search, maps visualization, explore query results and VOSpace sharing area.

ESA UNCLASSIFIED - Releasable to the Public

Euclid Science Archive System

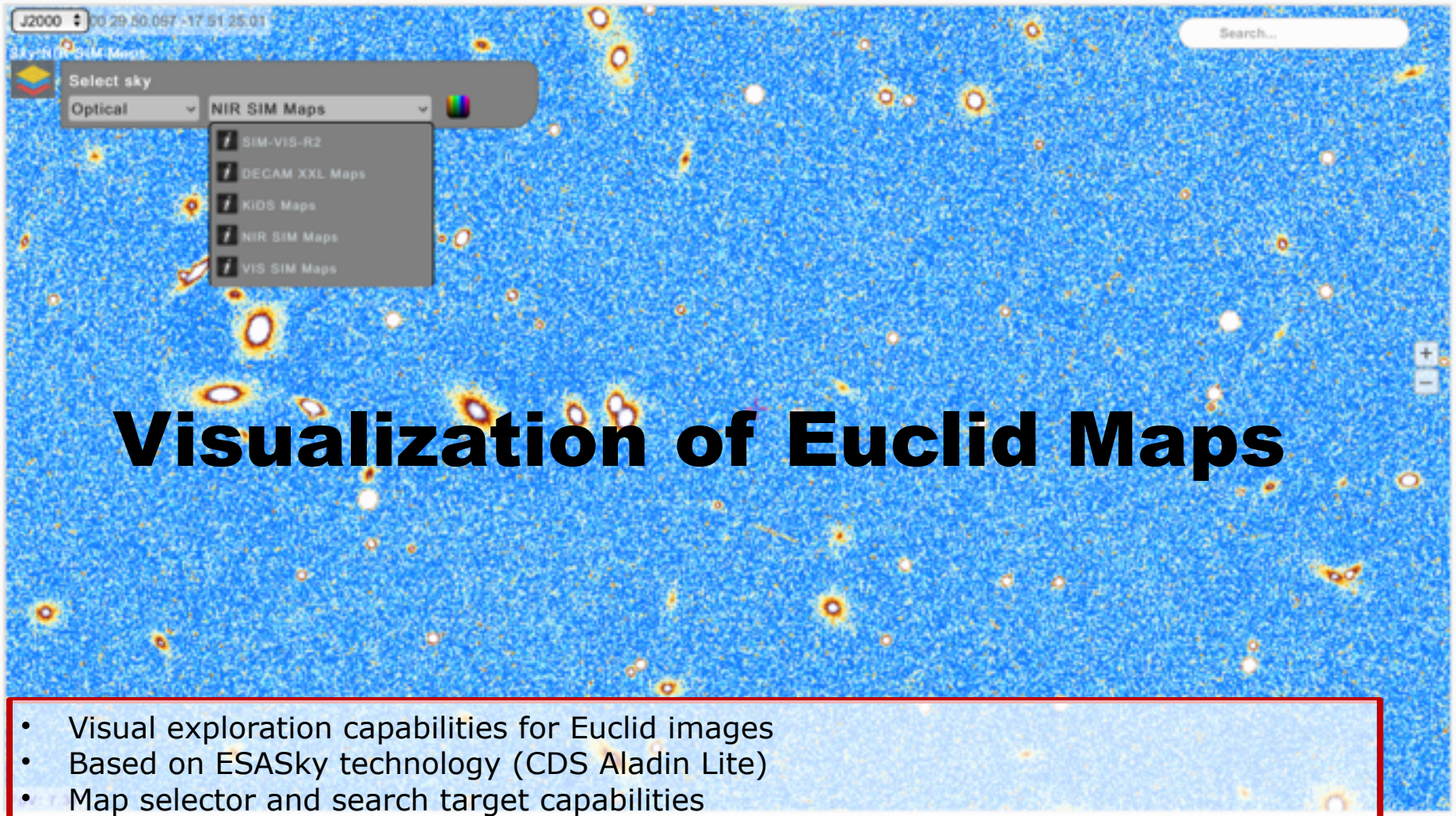


Metadata Parametric Search

- Parametric search interface to access Euclid science
- Reuse of Gaia archive expertise
- IVOA protocols: TAP+ Interface, UWS Job management
- Simulated and external catalogues

ESA UNCLASSIFIED - Releasable to the Public

Euclid Science Archive System



Visualization of Euclid Maps

- Visual exploration capabilities for Euclid images
- Based on ESASky technology (CDS Aladin Lite)
- Map selector and search target capabilities
- Future tools to work online: spectra tools, catalogues overlay, etc.

COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.

ESA UNCLASSIFIED - Releasable to the Public

Euclid Science Archive System

RESULTS Close All < Search #2 >

Search #2

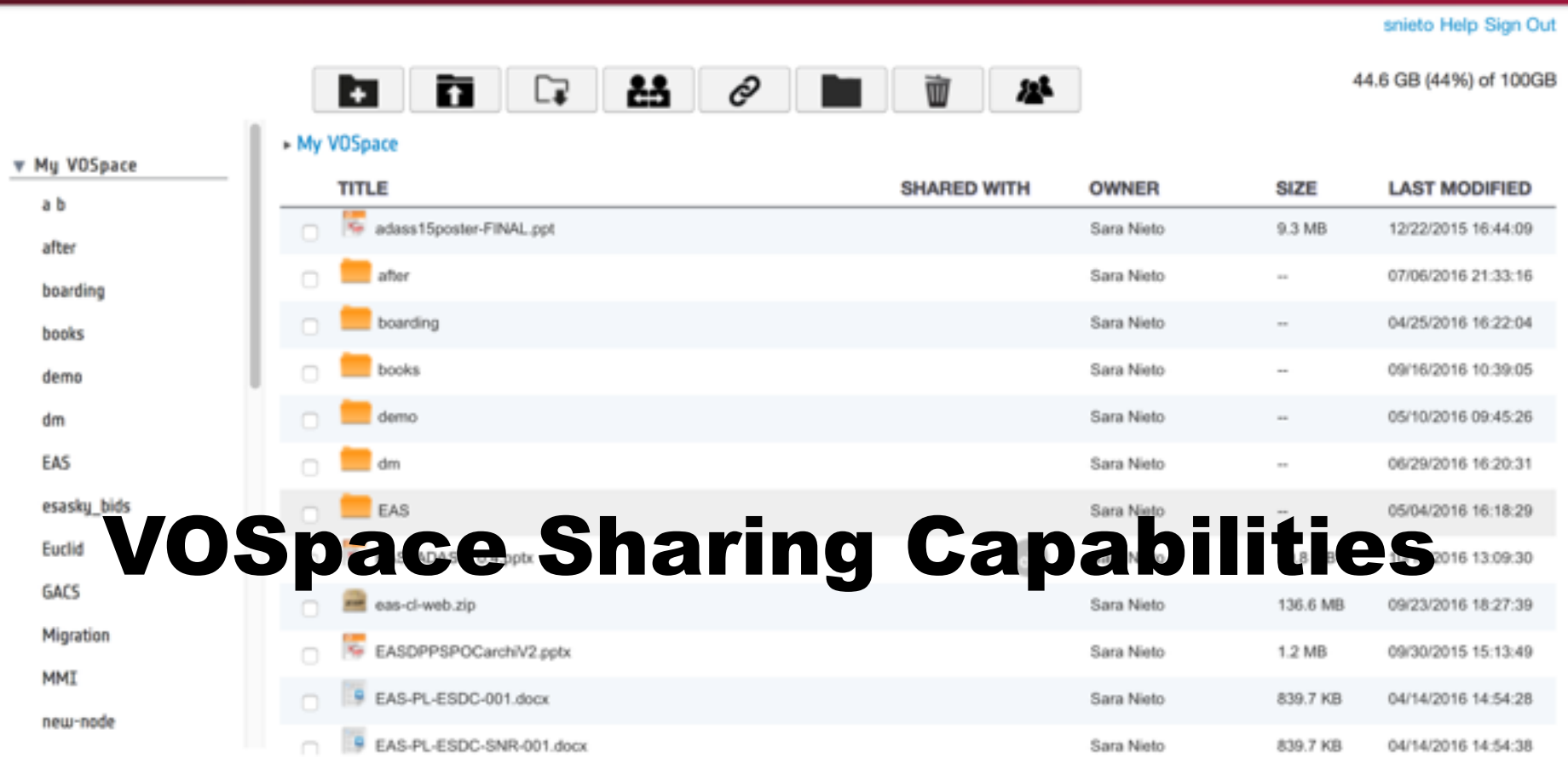
axis_number	file_name	file_status	first_type	image_dps_id	img_category	objectid	second_type	technique	thi
2	EUC_SIM_VIS_FIELD-34080-DITHER-3_20160425T134952.365Z_01.00.fits	PROPOSED	OBJECT	1	SCIENCE	318e4f57-058b-1c04-e053-b316a9c32cd4	SKY	IMAGE	WI
2	EUC_SIM_VIS_FIELD-11728-DITHER-1_20160426T045654.018Z_01.00.fits	PROPOSED	OBJECT	2	SCIENCE	318e9e40-ff79-25b5-e053-b316a9c3f174	SKY	IMAGE	WI
2	EUC_SIM_VIS_FIELD-11728-DITHER-3_20160426T013100.705Z_01.00.fits	PROPOSED	OBJECT	3	SCIENCE	318ee17a-aad5-2ea6-e053-b316a9c33fc5	SKY	IMAGE	WI
2	[illegible] 1_20160426T185417.576Z_01.00.fits	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] e053-b316a9c3db55	[illegible]	IMAGE	WI
2	EUC_SIM_VIS_FIELD-11728-DITHER-7_20160316T185439.110Z_01.00.fits	PROPOSED	BIAS	5	CALIB	31e1774c-e875-7ae6-e053-b316a9c37da6	OTHER	IMAGE	CA
2	EUC_SIM_VIS_FIELD-11728-DITHER-10_20160316T190827.605Z_01.00.fits	PROPOSED	BIAS	6	CALIB	31e1fdfe-9fdf-0af8-e053-b316a9c32b4a	OTHER	IMAGE	CA
2	EUC_SIM_VIS_FIELD-11728-DITHER-12_20160316T185844.638Z_01.00.fits	PROPOSED	BIAS	7	CALIB	31e257bd-799d-148d-e053-	OTHER	IMAGE	CA

1-20 of 87

Show query in ADQL form

Online Results Exploration

COPYRIGHT © EUROPEAN SPACE AGENCY. ALL RIGHTS RESERVED.



VOSpace Sharing Capabilities

- **http://vospace.esac.esa.int**
- Sharing capabilities through VOSpace Browser
- IVOA protocol for distributed storage
- Single Sign-on based on ESA Cosmos Authentication (LDAP

ESA UNCLASSIFIED - Releasable to the Public

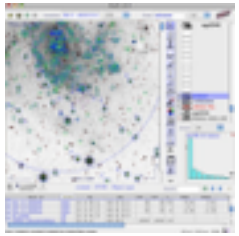
Euclid Scientific Archive: VO Interfaces

SAS will provide the tools and **VO interfaces** to enable the Software-to-Data Paradigm and "bring the software to the data" for the Euclid science.

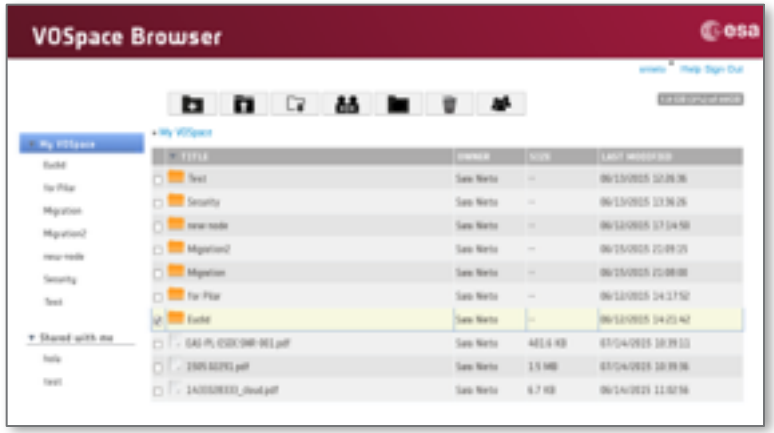
- **VO protocols** are part of the ESDC infrastructure
- **TAP+** parametric search for metadata en catalogues based on ADQL
- **Universal Worker Service** (UWS) to manage sync/async queries
- **VOSpace** to provide data sharing capabilities to the users community
- **SAMP** to interoperate with other astronomical analysis applications (Aladin, Topcat).

Browser

SAMP



Aladin, Topcat…



ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

Euclid Scientific Archive: Technology

- Backend is based on **Java**

- Web portal is based on **Google Web Toolkit** (JavaScript)

- Database is based on **PostgreSQL** and accessed through **TAP+ Interface**
  - Modules **pgSphere** and **Q3C** provide spherical data types, functions and operators to PostgreSQL

- **VO Interfaces (IVOA)**

- Access relies on **ESA LDAP** and CAS service for A&A control

ESA UNCLASSIFIED - Releasable to the Public

Scientific Archive Features

- Single access portal to Euclid Science
- Metadata Parametric Search
- X-match Analysis
- Online Results Exploration
- Level 2 Maps Visualization
- Catalogues Overlay on Euclid Maps
- Spectra Analysis
- Download Euclid Data
- Sharing Capabilities with Scientific Community
- Private storage area

  and others…

ESA UNCLASSIFIED - Releasable to the Public

Conclusions

- Euclid Archive System is a **collaborative effort** between **SDC-NL** and **ESAC**
- Euclid Archive System will face the **Big Data Challenge** and all its subsystems will be scaled up towards this target.
- The **Scientific Archive** is aimed to be the single access portal to the most valuable scientific metadata from the **EAS-DPS**.
- The **Scientific Archive** will provide the tools and interfaces to answer the scientific questions that Euclid is aimed to answer.

ESA UNCLASSIFIED - Releasable to the Public

Thanks for your attention

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

B. Altieri, G.Buenadicha, S. Nieto, J. Salgado, P. de Teodoro
European Space Astronomy Center, European Space Agency, Spain
O.R. Williams, A.N.Belikov, D.Boxhoorn, B. Dröge, J.McFarland, A.Tsyganov, E.A. Valentijn
University of Groningen, Groningen, The Netherlands

@ESAesdc

ESA UNCLASSIFIED - Releasable to the Public