Gaia and Euclid

And their user requirements for VO

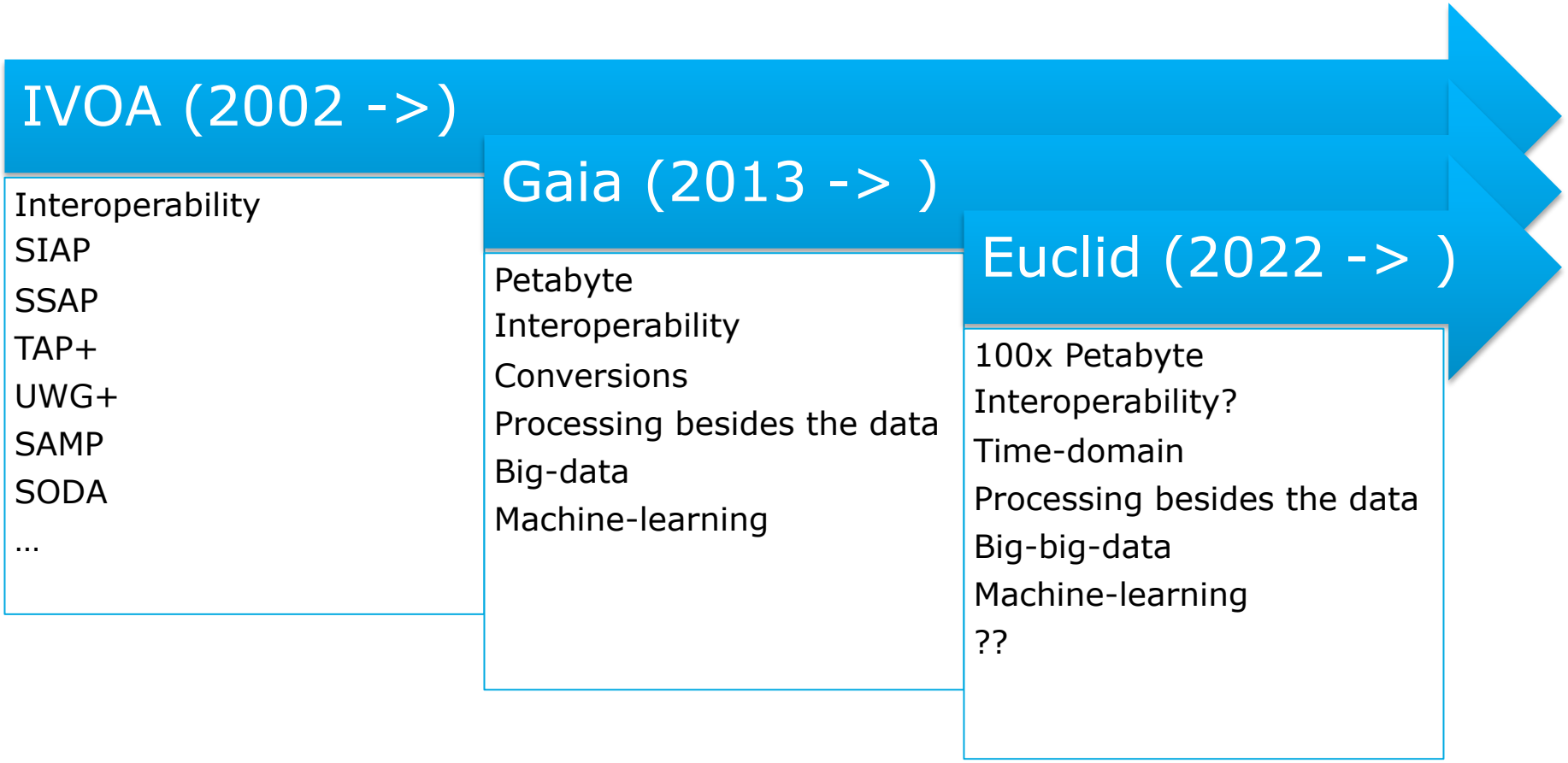
Bruno Merín (ESAC Science Data Centre),
Alcione Mora (Gaia Archive Scientist),
Bruno Altieri (Euclid Archive Scientist),
Deborah Baines, Jesús Salgado, Juan Gonzalez and Christophe Arviset (ESAC Science Data Centre)

Issue/Revision: 1.0
Reference: Gaia + Euclid + VO
Status: Issued

Outline

1. Motivation
2. Gaia
3. Euclid
4. Common user requirements for the IVOA

Motivation: cross-fertilization



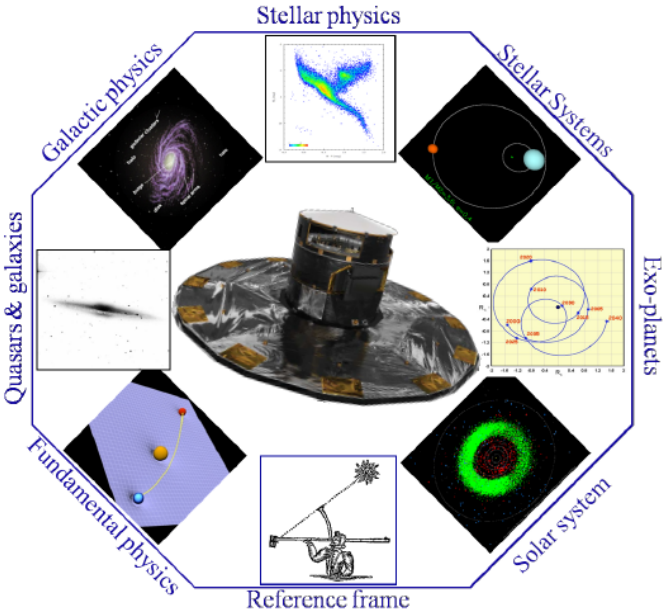
Gaia: mission concept

Gaia scanning the sky

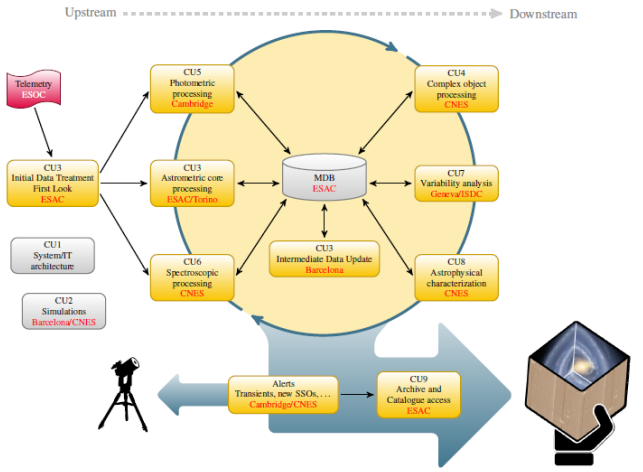
https://youtu.be/BnFyzZGWuYs

Gaia Satellite and Data Overview

- ESA Corner Stone mission, launched 19/12/2013
- Stereoscopic Census of the Galaxy over 5 years
  - 1-2 billions sources with unprecedented accuracy
  - 100TB downlink
  - Up to 1PB calibrated data
    - 10^{11} telescope transit
    - 10^{12} astrometric observations
    - 150 x 10^{6} Spectra

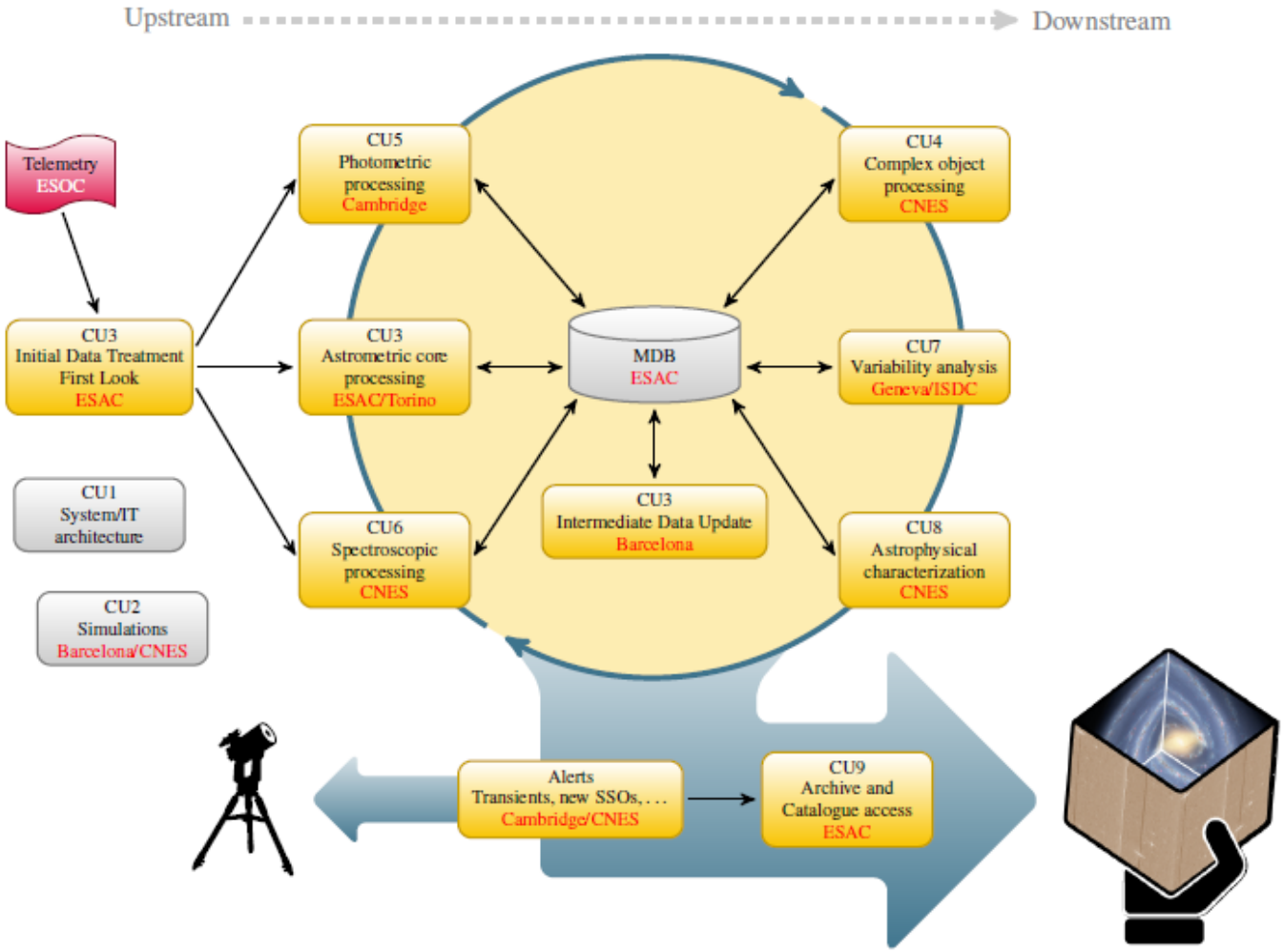


- Big data processing challenge as well !
  - (outside the scope of this presentation)



- 1st public release of Gaia catalogue in summer 2016
- ~1 new release per year
- Final catalogue ~2022

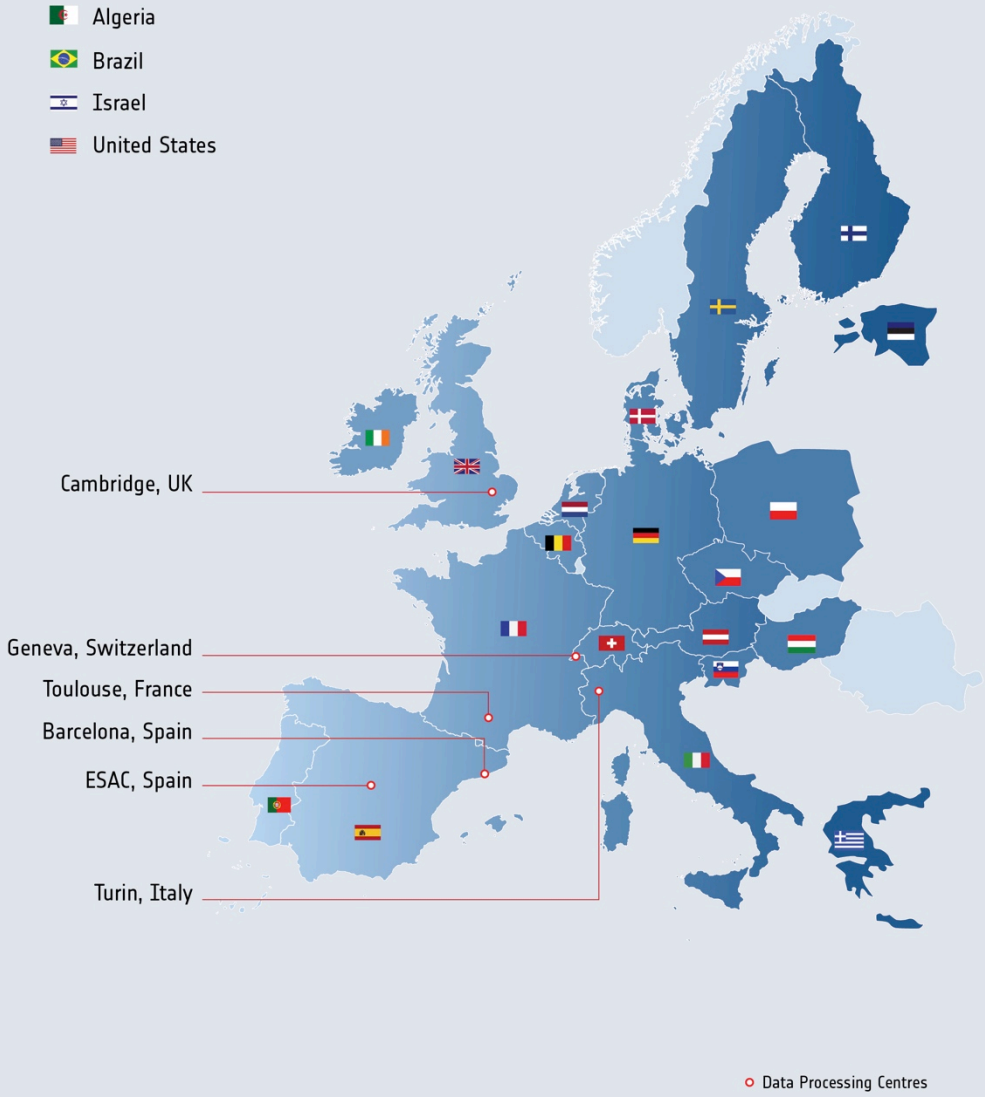
Gaia Data Processing and Analysis Consortium DPAC



Gaia Satellite and Dat

- ESA Corner Stone mission, laun
- Stereoscopic Census of the Gala
  - 1-2 billions sources wit accuracy
  - 100TB downlink
  - Up to 1PB calibrated da
    - 10^{11} telescope tra
    - 10^{12} astrometric
    - 150 x 10^{6} Spectra
- Big data processing challenge a
  - (outside the scope of this pres
- 1st public release of Gaia catalo
- ~1 new release per year
- Final catalogue ~2022

Gaia Data Processing and Analysis Consortium, DPAC, as in 2015



Gaia is definitely a major astronomy "Big Data" project

Volume

- 1PB of data in total, not really "big data"

Velocity

- Massively complex data processing challenges, 10^{20} FLOP

Variety

- Source catalogue, spectras, telescope transits

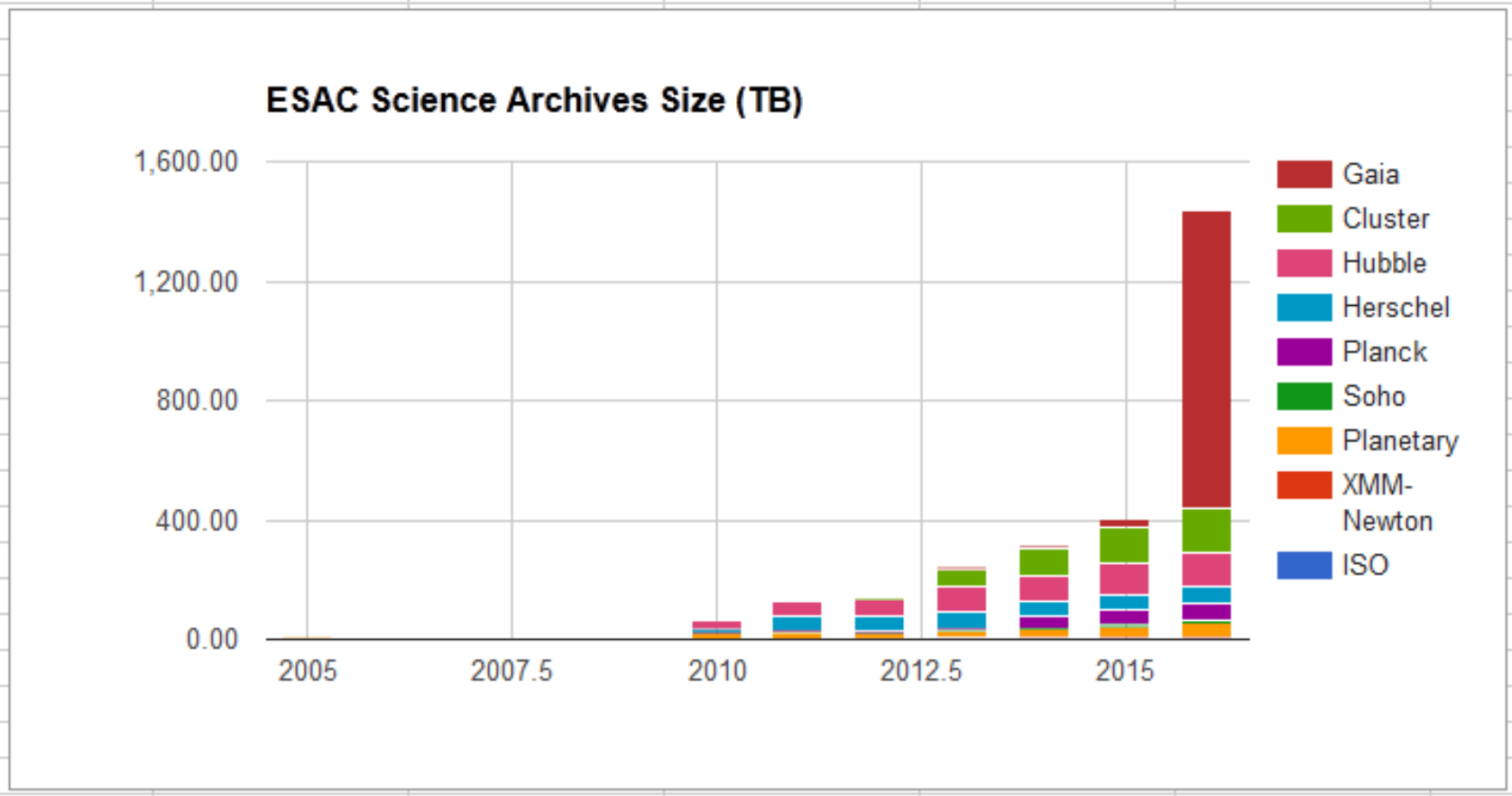
Veracity

- Astrometry, photometry and spectroscopy with high quality

Value

- Believed to revolutionize astronomy
- Most consistent, complete, and accurate astrometric dataset to date

ESAC Archives Volume evolution



All data stored on hard disks and distributed through Internet

Euclid will add up to ~150 PBs by 2023

Gaia top priority use-cases

- Three top priority science use-cases, from Gaia Archive Preparation:
  - "I'm completely new to Gaia. Tell me all about what is contained in the Gaia Archive, and give me some clear examples illustrating how it might be useful to my science. "
  - "I would like to be able to select objects based on any set of the variables provided in the Gaia catalogue position, parallax, astrophysical parameters, proper motion uncertainties etc. These selections should not be limited to simple "axis-parallel" cuts or cone cuts, but permit a broader array of functions/functional dependencies. "
  - I want the RVS spectra of my favourite source(s)
- From http://great.ast.cam.ac.uk/Greatwiki/GaiaDataAccess/GdaScenariosFeedback

IVOA Specifications used in Gaia Archive

ADQL: Astronomical Data Query Language

Language used to query data

TAP+: Table Access Protocol

A protocol to access tables that contain the data

UWS+: Universal Worker Service Pattern

A jobs scheduler/handler to manage data queries

(Some extensions required to fulfil science needs)

VOSpace: Interface to distributed storage

A virtual storage system (a VO "dropbox++")

SAMP: Simple Application Messaging Protocol

A protocol for applications to inter connect amongst them

Need for new paradigm: Archive 2.0

- New ways required to access the Gaia catalogue and associated data
  - Powerful query mechanism, asynchronicity of results
  - One "query interface" for all archive services and VO services

- User can not download all catalogue and all data
  - Need to have user workspaces *IN* the Archive
    - User database space, user disk space
  - User workspace shareable amongst various users

- Bring user code to the data
  - Part of the user workspace in the archive
  - Share code with other users

The user works with the data WHERE the data is : Archive 2.0 concept

Other requirements

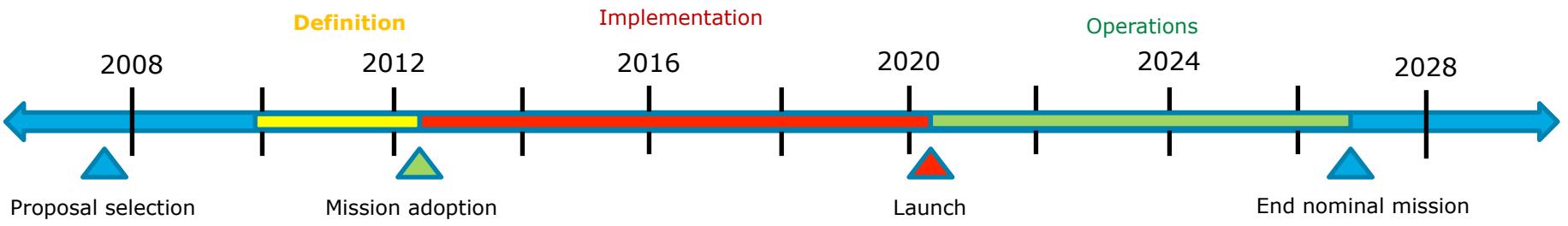
- TAP+ / UWS+: user authentication needed for scientific validation prior to data releases (DONE!)
- Interoperability between different archives (CDS/Vizier, MAST, IPAC) vs local replication for performance reasons
- ADQL box redefinition -> Geodetics? Only fixed values?
- ADQL coordinate system and epoch conversions (for large tables)
- ADQL/TAP unit conversions
  - Gaia Flux not equal Flux
  - different magnitude systems (AB/Vega)
  - not direct translation from parallax to distance
  - Filter conversion (G, BP, RP to e.g. ugriz/UBVRI)
- Time-series for transit data (70 measurements times 1.e9 sources)

Euclid: mission concept

- **Goal**: To study the Dark Energy and Dark Matter in the high-redshift universe via a large-scale massive imaging and spectroscopic survey of all the extragalactic sky (~ half of the total sky) with very high imaging and spectroscopic quality.

- Launch date: 2022, then six years of survey, yearly Data Releases with huge mosaics in optical + near-IR, near-IR spectra, catalogs and cosmological parameters

- Archive prototype already being developed at Consortium and ESA



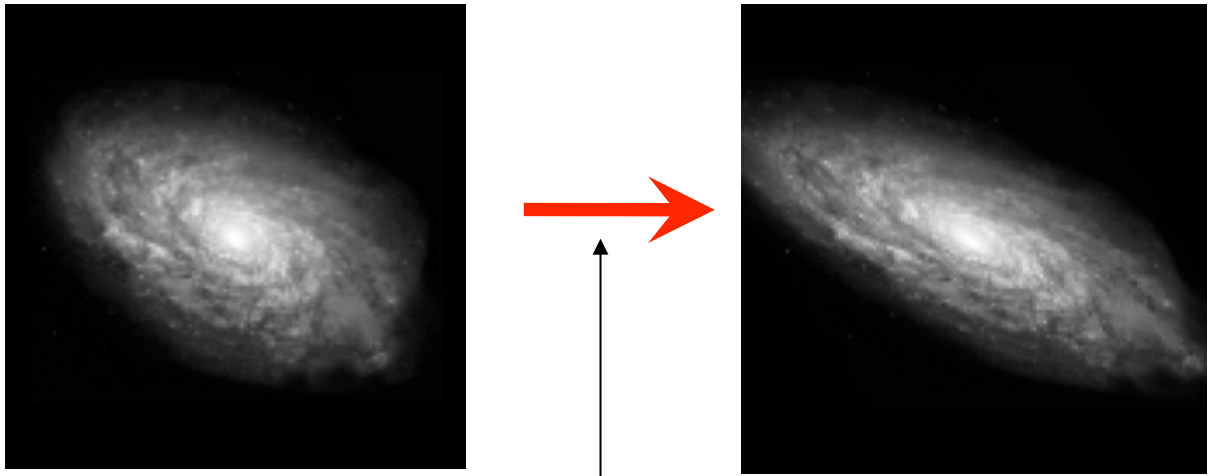
Euclid: mission concept

Exploring the dark side

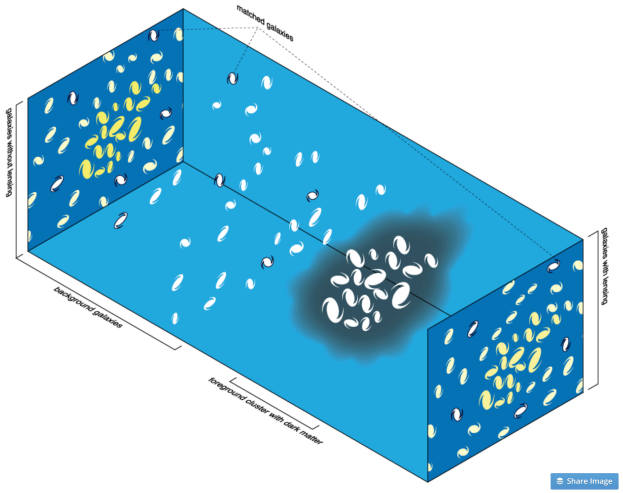
https://youtu.be/wQfjeJDuBh0

Weak Lensing

The weak lensing distortion is simply a (very small) change in ellipticity and position angle of a galaxy

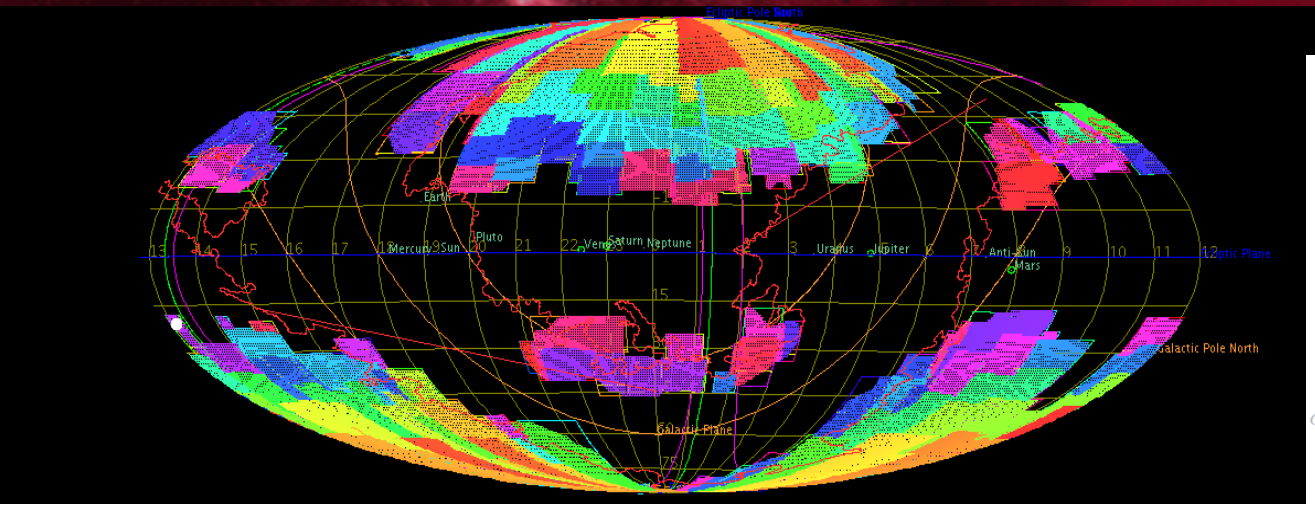


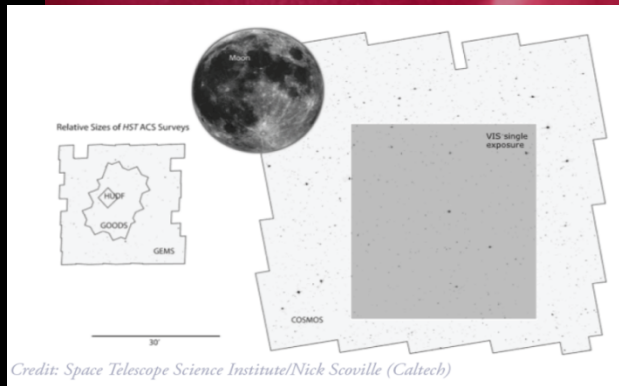
Dark Matter



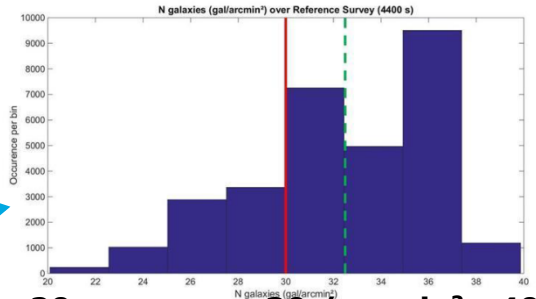
Euclid WL galaxies are ≥0.3 arcsec

EUCLID – Reference Sky Survey & Science return





Credit: Space Telescope Science Institute/Nick Scoville (Caltech)



20 **30 /arcmin²** **40**

- Ecliptic Coordinates, Mollweide Projection. Different colours indicate different survey years – total 6 years
- Avoidance regions: Ecliptic plane (zodiacal light) and low (<30 deg) galactic latitudes (extinction, star density)
- Number of resolved galaxies/arcmin^2 per VIS field (mean=32/arcmin^2)

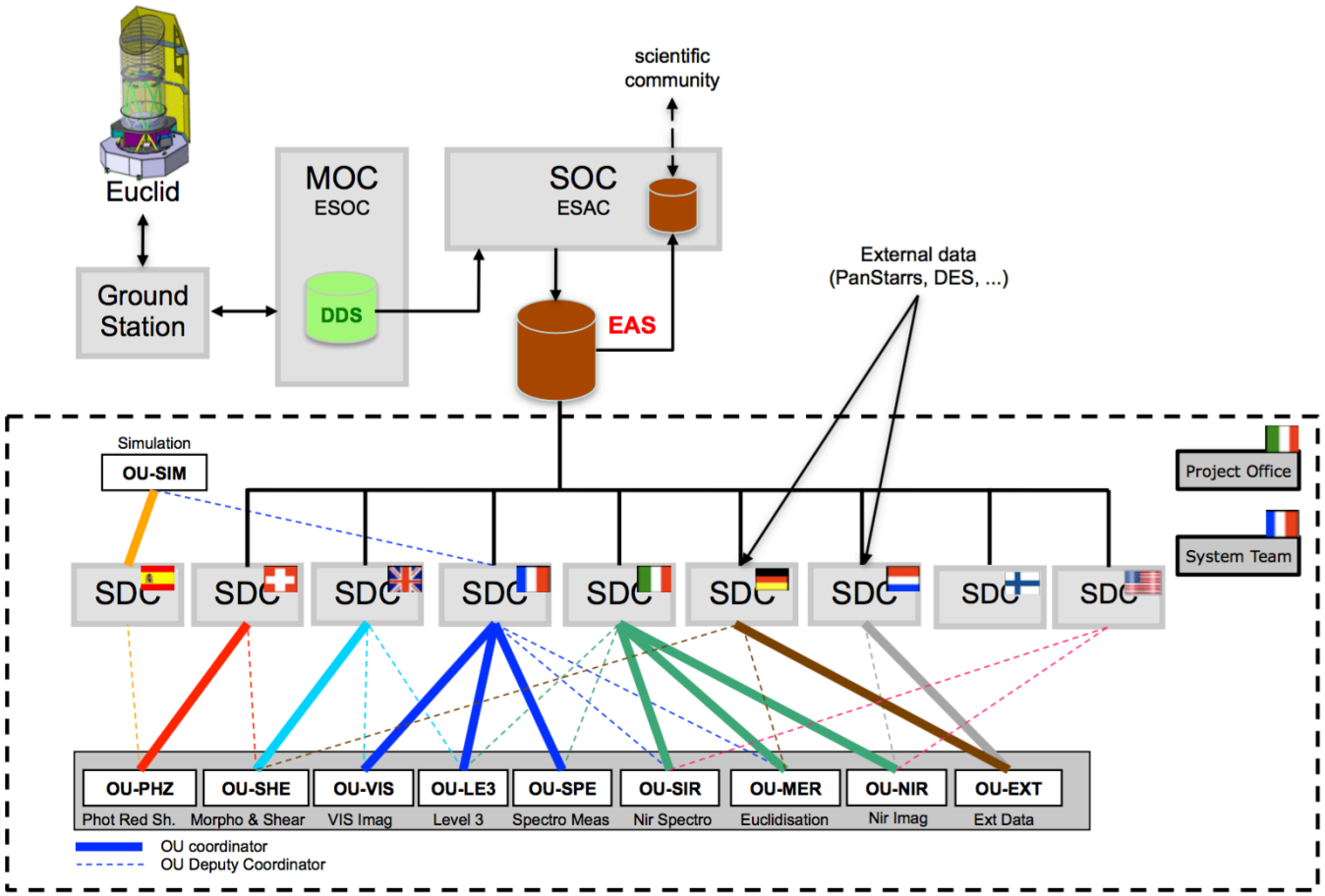
→ Euclid will survey >36% of sky down to AB >24 mag (~500TB raw data)

- **Excellent image quality (~0.15 arcsec FWHM) in visual band (550-900 nm): > 1.5 billion resolved galaxies with *z*<2**
- **Near-Infrared Imaging Photometry (Y, J, H bands) of these galaxies and H_{alpha} Spectroscopy (1.25-1.85 micron) of >25 million galaxies**
- **The Euclid Consortium will perform the scientific data processing in 9 Euclid data centres**

Euclid Consortium – Science Ground Segment

EC-SGS functions

- The Instrument Operation Teams (IOTs): maintenance of the instruments, production of weekly instrument reports.
- The Science Data Centres (SDCs), host the IOTs and run Processing Functions (pipelines) produced by Organisational Units (OUs);
- SDC's use Level 1 data from SOC and produce Level 2 and 3. Reprocess external data: Level E.



euclid European Space Agency

Euclid Consortium – Science Ground Segment

euclid European Space Agency

Euclid top-priority use-cases

- From Euclid Science Archive Use-cases document:

  - Query large data mosaics and extract cut-outs per frequency to inspect individual sources and/or extractions and be able to compare them with ancillary multi-wavelength data (SDSS, DECCam, HST, JWST, etc).
  - Non-positional searches: identify sub-samples that satisfy some complex set of conditions in the flux, color, shape and/or redshift spaces for the different Data Releases.
  - Time-domain searches: find fast moving objects (asteroids, solar system objects), or photometrically variable objects

All this should be done on a very data-intensive & distributed context.

Euclid is the Hubble Deep Field but for all (extragalactic) sky

Zooming into GOODS-S from all-sky with ESASky

https://youtu.be/zt-HtyPT5z4

IVOA Specifications used in Euclid Archive

- **ADQL: Astronomical Data Query Language**
  - Language used to query data
- **TAP+: Table Access Protocol**
  - A protocol to access tables that contai
- **UWS+: Universal Worker Service Pattern**
  - A jobs scheduler/handler to manage data queries
- **VOSpace: Interface to distributed storage**
  - A virtual storage system (a VO "dropbox++")
- **SAMP: Simple Application Messaging Protocol**
  - A protocol for applications to inter connect amongst them
- **HiPS: Hierarchical Progresive Surveys**
  - A format to visualize very large amounts of data in a light way

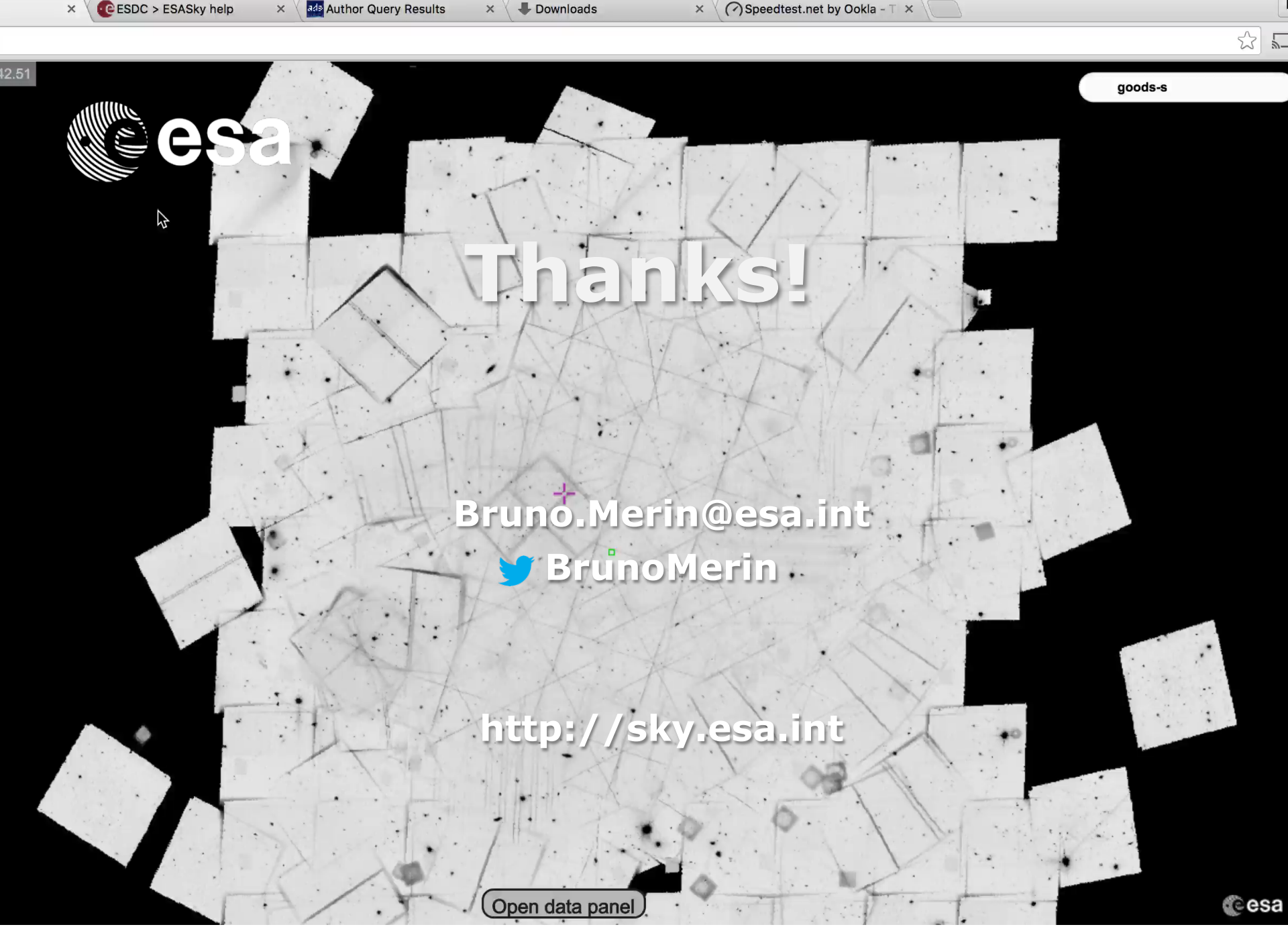
(Some extensions required to fulfil science needs)

Final words

- Both the Gaia and Euclid missions/projects are customers and contributors to the IVOA standards, protocols and formats already thanks to their co-location at the ESAC Science Data Centre with VO specialists

- Many future applications require IVOA standards and protocols to handle very large amounts of data and allow easy execution of machine learning applications in very distributed and complex international organizations

- The IVOA should continue to be responsive to unexpected changes in the user requirements from big data providers to keep them engaged and participating


ESDC > ESASky help
Author Query Results
Downloads
Speedtest.net by Ookla
goods-s
esa
Thanks!
Bruno.Merin@esa.int
BrunoMerin
http://sky.esa.int
Open data panel
esa