Science Archives at the ESAC Science Data Centre

Christophe Arviset
Head of ESDC, ESAC

SCIOPS 2015 Workshop, ESO
24/11/2015

Issue/Revision: 1.0
Reference: ESAC Science Data Centre
Status: Issued
ESA UNCLASSIFIED - Releasable to the Public

→ ESA'S LEGACY OF SPACE OBSERVATORIES

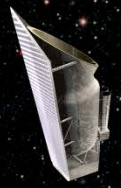
ESA has a long history of space science missions that have studied the skies across the full spectrum of light, charted the Milky Way Galaxy and probed the fabric of the Universe. Even after completing observations, the legacy of these missions lives on in the form of vast databanks, and paves the way for the next generation of space observatories.

hipparcos
The first astrometry satellite *(1989-1993)*

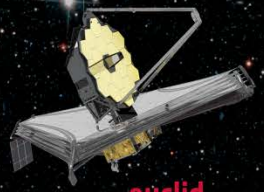
gaia
Surveying a billion stars *(2013-)*

iue
Analysing ultraviolet light from stars *(1978-1996)*

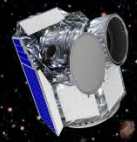
iso
Chemical analysis of celestial objects *(1995-1998)*



jwst
Observing the first light

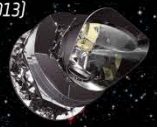


cheops
Sizing and first characterisation of exoplanets



herschel
Unveiling the cool and dusty Universe *(2009-2013)*

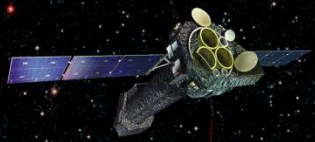
planck
Looking back at the dawn of time *(2009-2013)*



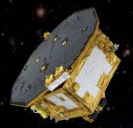
euclid
Exploring the dark Universe *(in development)*

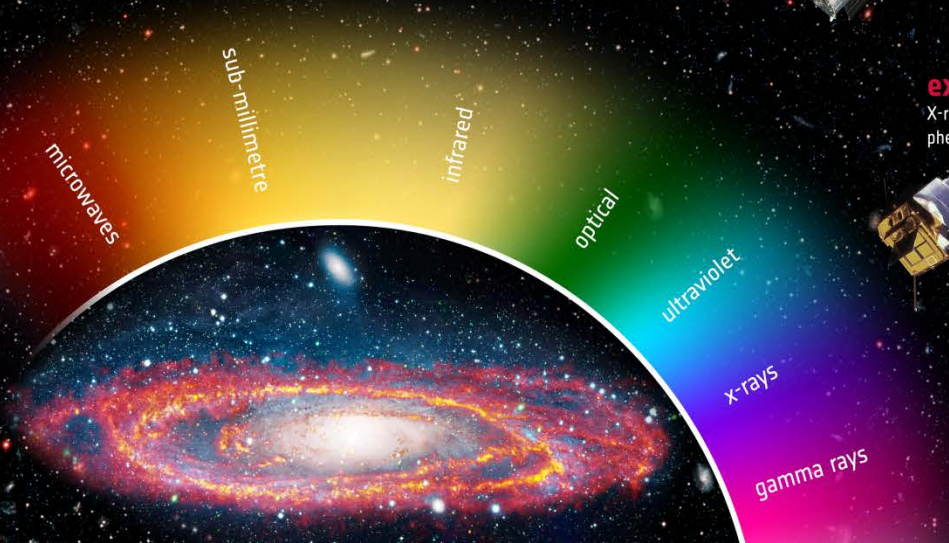
hst
Expanding the frontiers of the visible Universe *(1990-)*

xmm-newton
Seeing deeply into the hot and violent Universe *(1999-)*



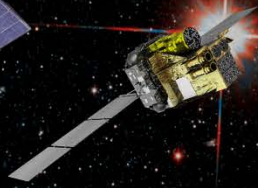
lisa pathfinder
Testing the technology for gravitational wave detection *(2015-)*





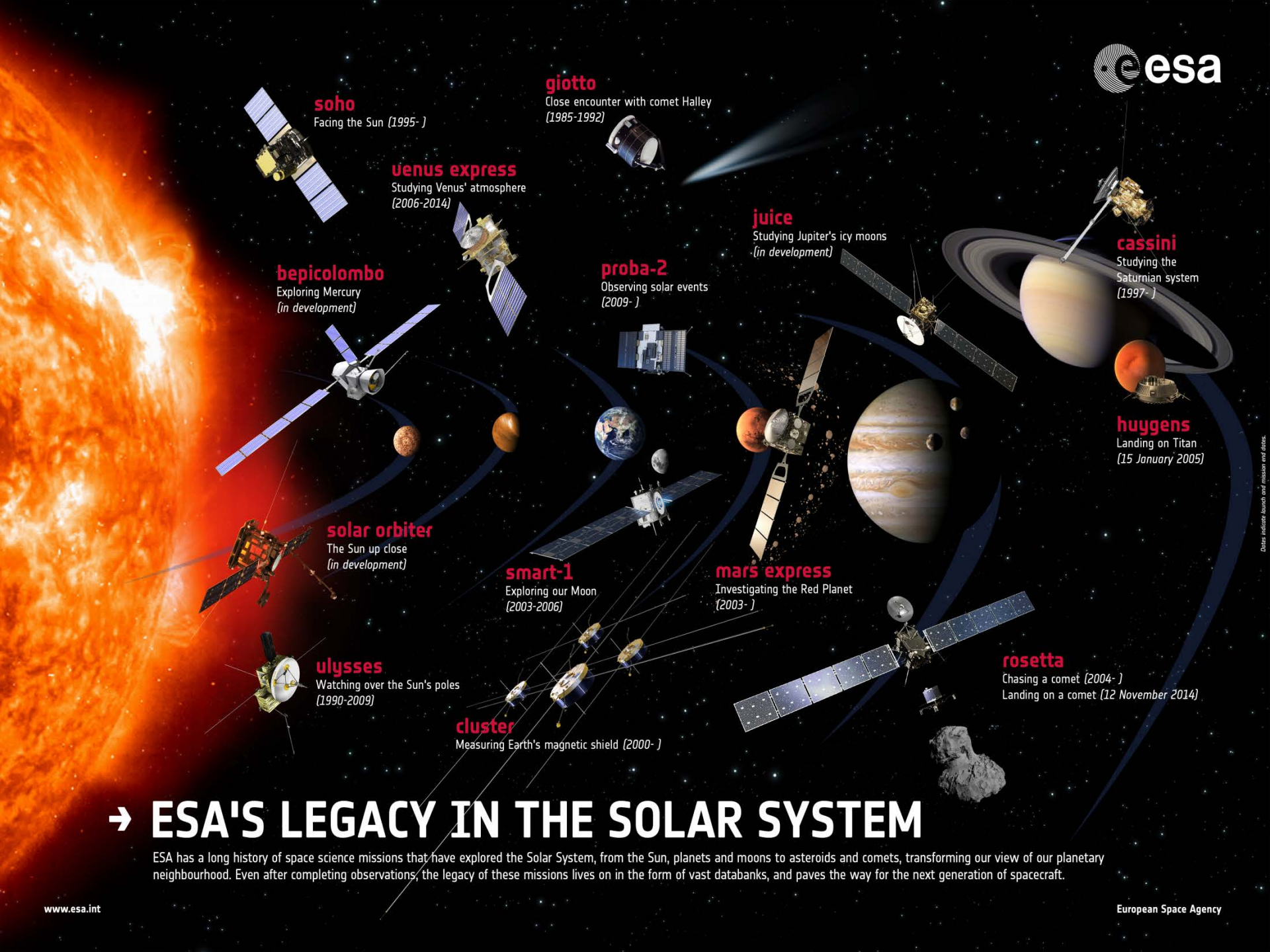
exosat
X-ray survey of high-energy phenomena *(1983-1986)*

integral
Seeking out the extremes of the Universe *(2002-)*



cos-b
Surveying the high energy Galaxy *(1975-1982)*

Dates indicate launch and mission end dates.



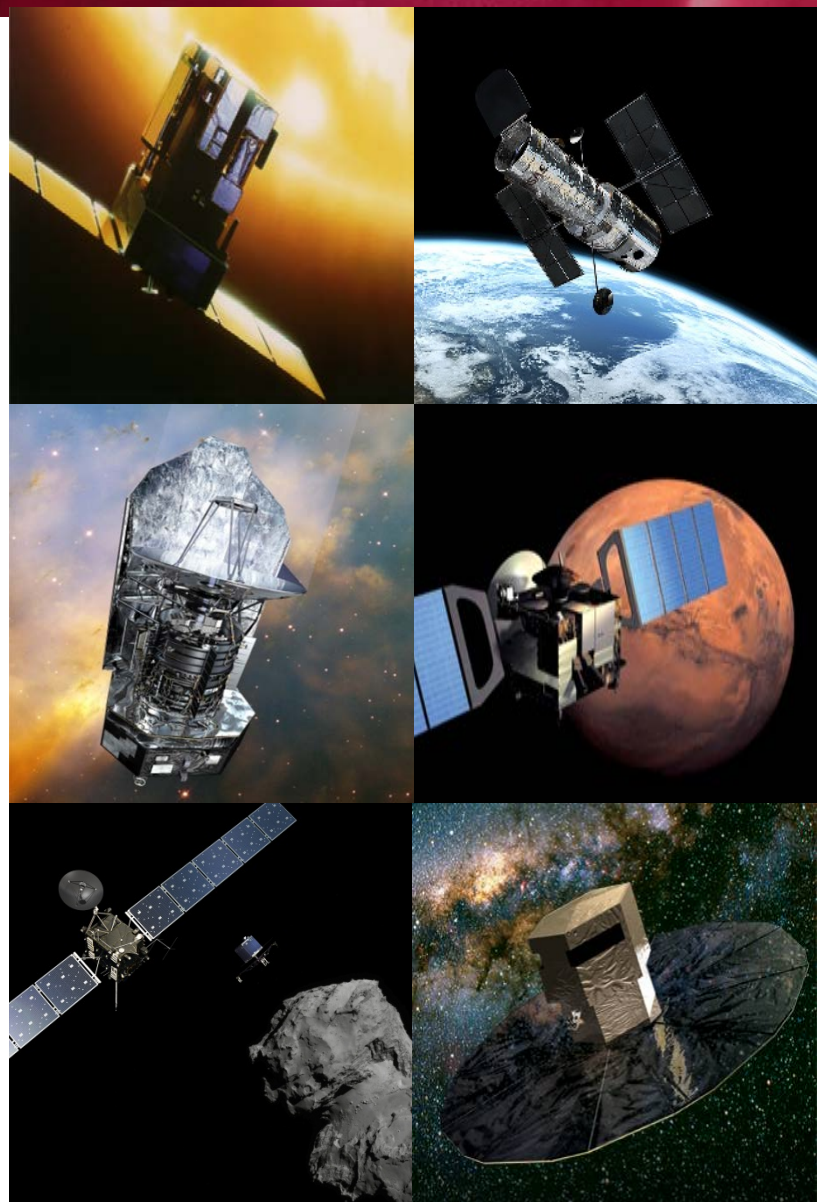
→ ESA'S LEGACY IN THE SOLAR SYSTEM

ESA has a long history of space science missions that have explored the Solar System, from the Sun, planets and moons to asteroids and comets, transforming our view of our planetary neighbourhood. Even after completing observations, the legacy of these missions lives on in the form of vast databanks, and paves the way for the next generation of spacecraft.

ESAC Science Data Centre

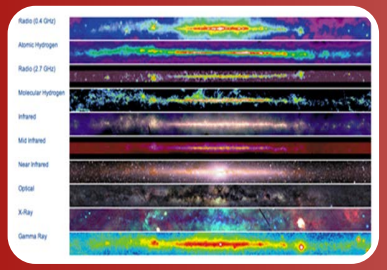
The Digital Library of the Universe

- 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



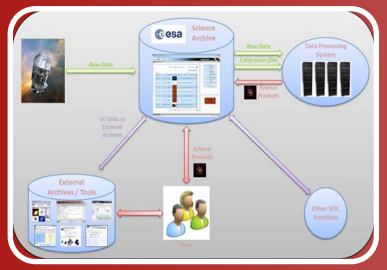
ESA UNCLASSIFIED - Releasable to the Public

ESAC Science Archives Strategy



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***

ESA UNCLASSIFIED - Releasable to the Public

ESDC organizational structure

H/ESDC

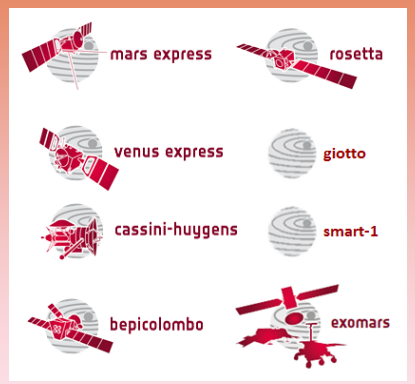
ESDC Head Scientist

Astronomy Archives Group

- Astro. Archives Science Lead
- Astro. Archives Technical Lead

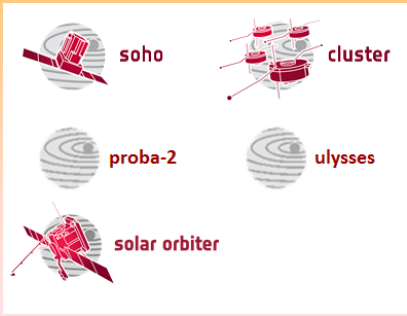
Planetary Archives Group

- Planetary Archives Science Lead
- Planetary Archives Technical Lead



Solar Helio Archives Group

- Solar Helio Archives Science Lead
- Solar Helio Archives Tech. Lead



Archive Software Engineers

Common software libraries and tools repository, full open source

Astronomy SOCs/SGSs

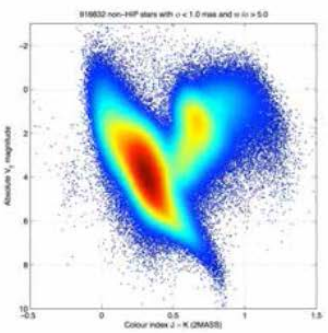
Planetary SOCs/SGSs

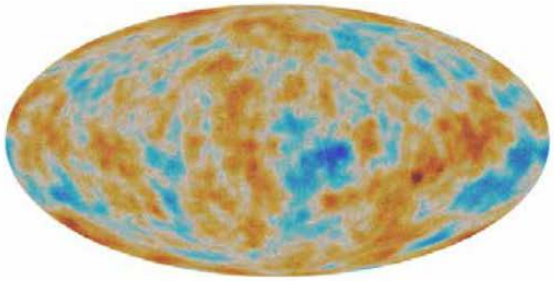
Solar Helio SOCs/SGSs

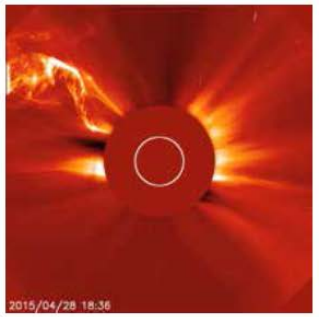
ESA UNCLASSIFIED - Releasable to the Public

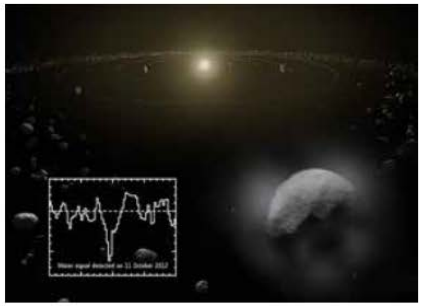
European Space Agency

Enable Maximum Science Exploitation







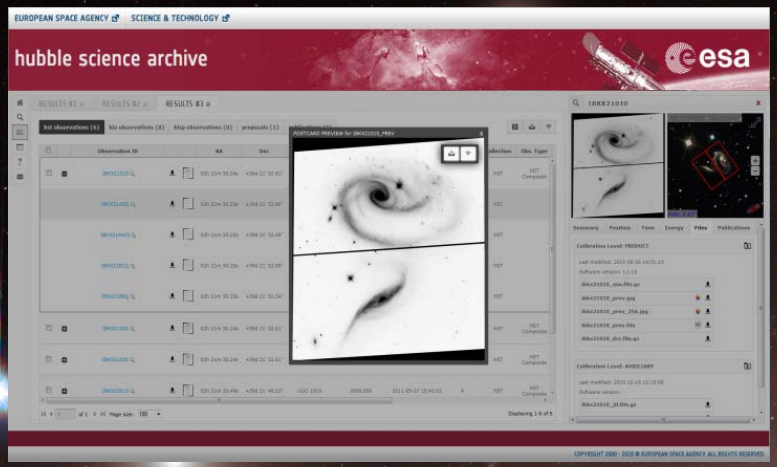


Scientists and Engineers working very closely together

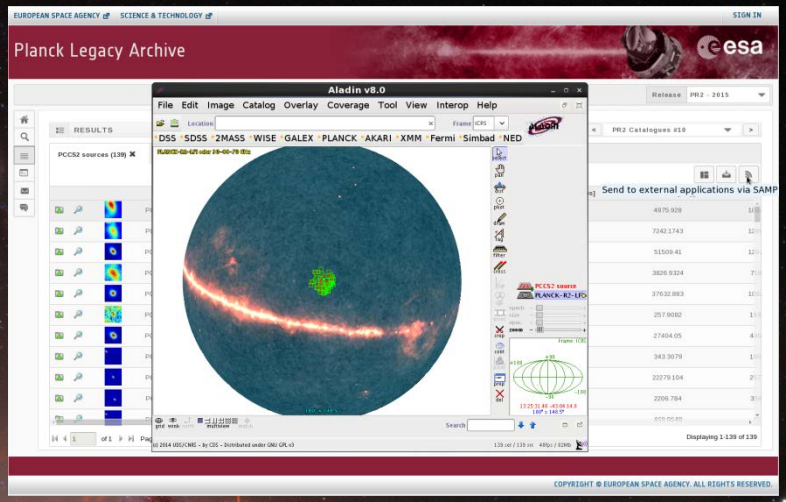
Science driven Archives

ESA UNCLASSIFIED - Releasable to the Public

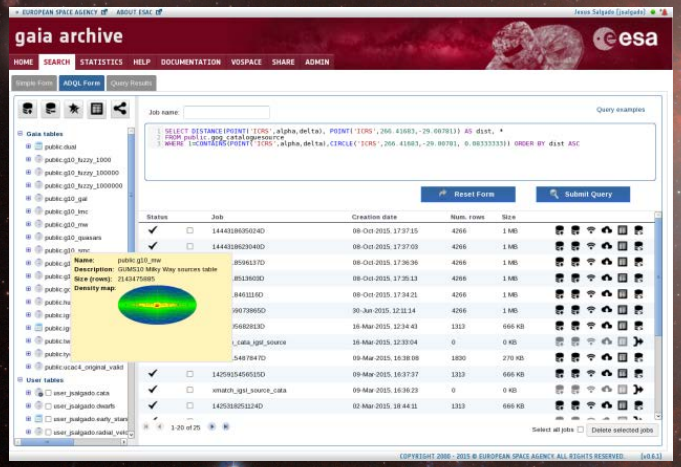
Individual Mission Science Exploitation



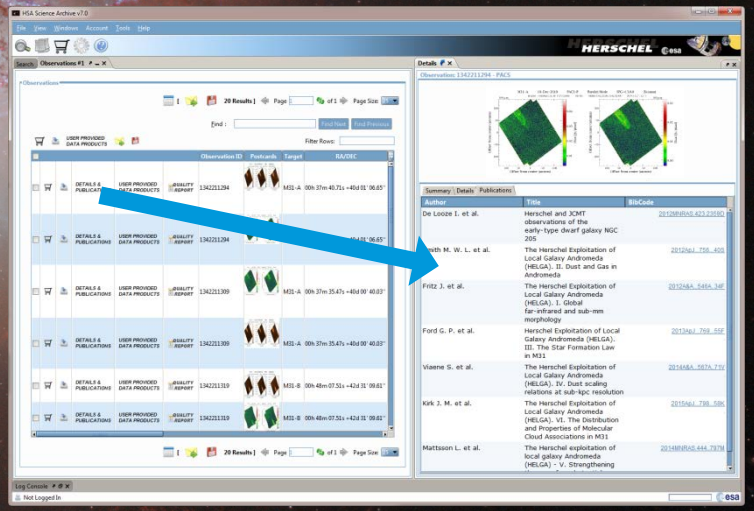
European HST Archive (*Baines*)



Planck Legacy Archive (*Dupac, Lopez*)

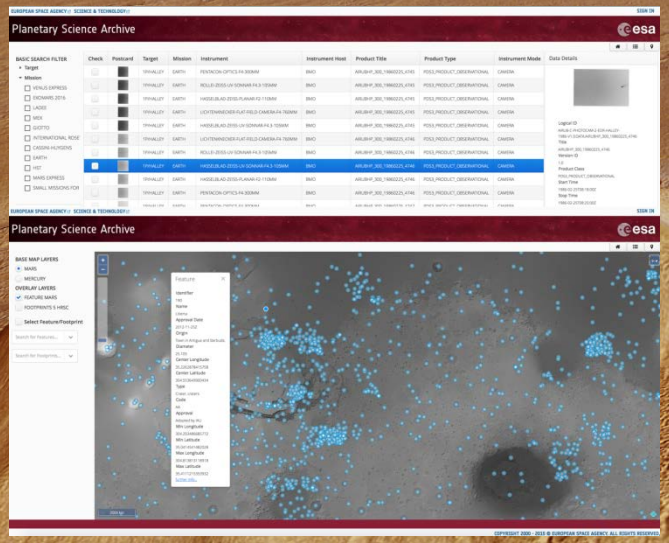


Gaia Archive consortium release (*Gonzalez*)

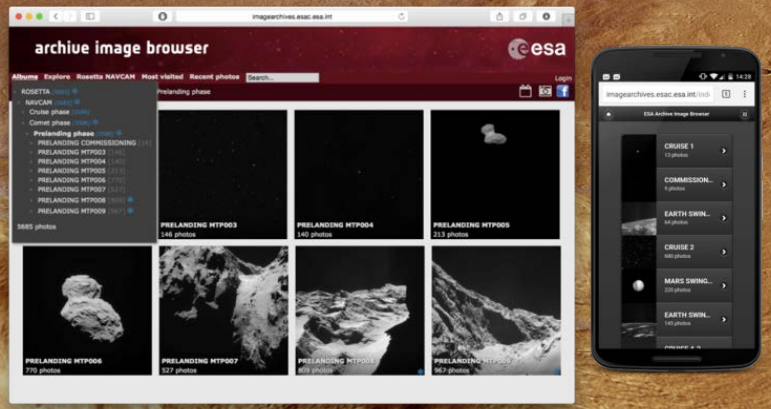


Herschel Science Archive (*Verdugo, Teyssier*)

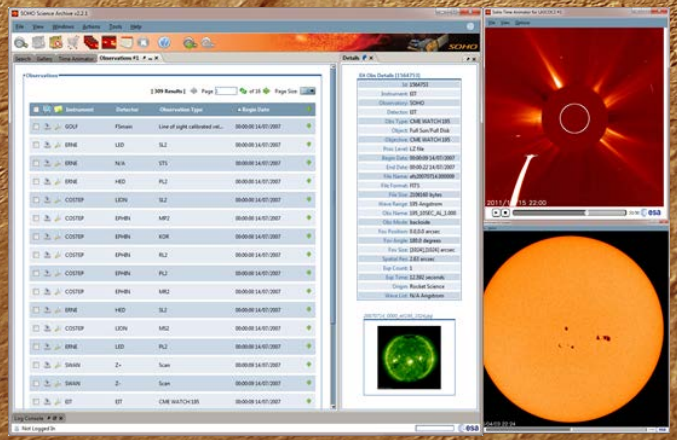
Individual Mission Science Exploitation



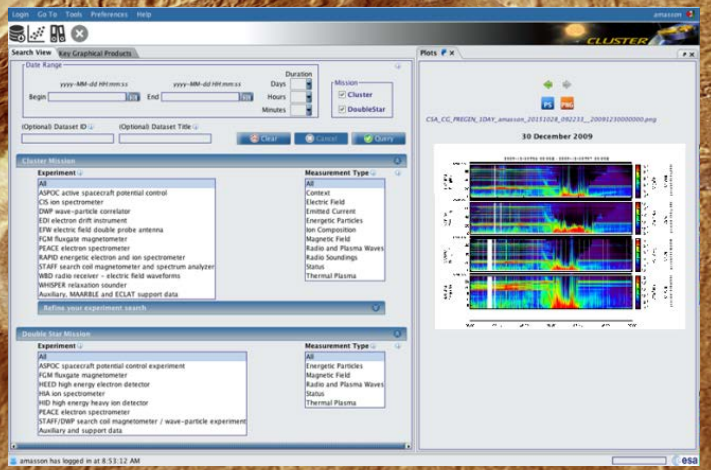
Planetary Science Archive (*Lim, Grotheer*)
BepiColombo Archive (*Martinez, Ortiz, Perez*)



Archive Image Browser (Rosetta)



Soho Science Archive
Including Proba-2 data



Cluster Science Archive,
Including Double Star data

Need for multi mission science

- Large and wide set of ESA astronomy archives
- Science is more and more multi wavelength

- As missions enter into post operations and legacy phase, need to “hide” projects jargons and complexity, and offer simpler science data exploration tools

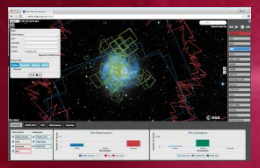
- Build on top of existing ESA archives
  - Re-use existing data infrastructure
  - Re-use of VO protocols (to make it open to other non ESA archives)

- ESASky for astronomy
- New PSA for planetary science

- Start thinking multi disciplinary (astronomy, planetary, solar heliospheric)

ESA UNCLASSIFIED - Releasable to the Public

ESA Sky (*Merin*)
Multi Mission Science Exploitation



esa

Multi Mission Interface (M

Bruno

ammi.esac.esa.int/mmi/

J2000 13 36 1.90 -30 03 48.3

m 83

Upload target list

Skies

XMM-Newton

HST

Herschel

Planck

Others P/DSS2/color

Selected: P/DSS2/color

Colour Map

Native | Greyscale | Rainbow | EOSB

Reverse

CygnusX-1

M31

M33

M83

M101

NGC 1365

WR85

WR89

WR110

DOBASHI 1444

eta Car

V709 CrA

SgrA

LMC

ACO 1656

IC 348

LMCX-3

PSR B1509-58

FoV: 55,94'

esa

Summary | 3XMM EPIC | HST | XMM-Newton | Herschel

Observations: XMM-Newton, HST, Herschel, ISO

Catalogues: 3XMM EPIC, XMM Slew, XMM OM

ESA Observations

Number of results: 10, 5, 0

X-Ray | Visible | Infrared

Wavelength

XMM-Newton HST ISO Herschel

ESA Catalogues

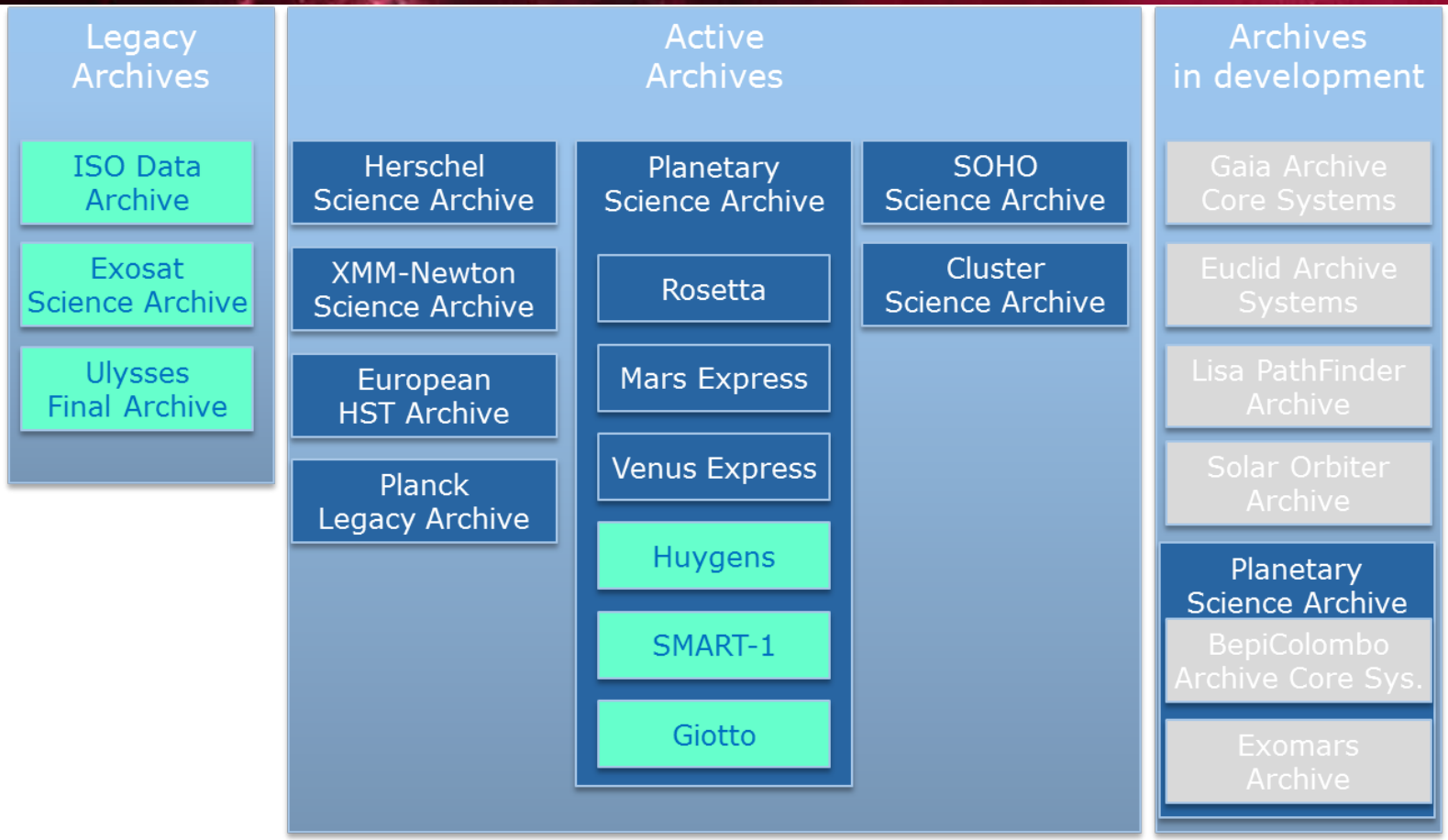
Number of results: 2000, 0

X-Ray | Visible | Infrared

Wavelength

XMM Slew 3XMM EPIC XMM OM

Long Term Preservation Strategy



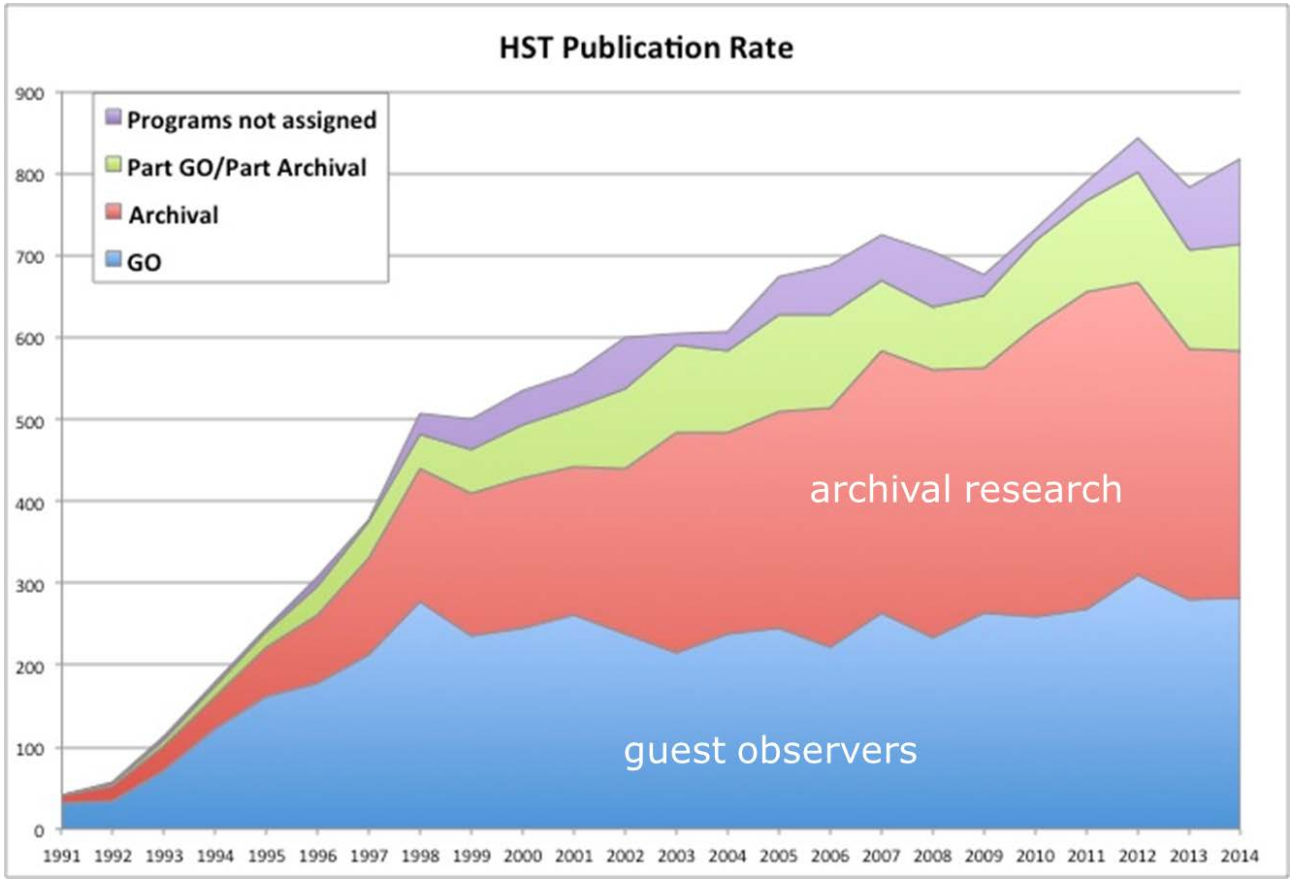
Preservation of data, software and knowledge

Regular Technology Evolution

ESA UNCLASSIFIED - Releasable to the Public

Archival Science through Publications

- Number of Publications is a usual metrics to measure the success of a space science mission
- HST produced more than 13.000 publications to date
- Number of publications coming from “archival” data (vs Guest Observers) has increased with time and now represents ~50% of HST annual publications
- Long Term Preservation of Data increases the science results from a mission !



ESA UNCLASSIFIED - Releasable to the Public

Long Term Preservation Strategy

- Consolidation of all ESA Space Science Archives at ESAC, with strong re-use across projects, ensuring easier and cheaper long term data preservation
  - Hardware infrastructure
  - Software architecture and code, including technology migration
  - Human technical and scientific expertise
  - Multi missions, multi instruments science exploitation

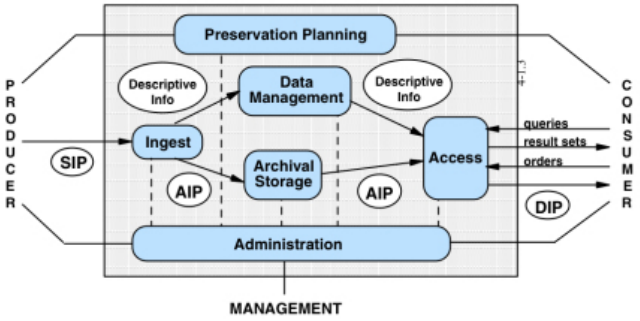
- Long term preservation of data processing capabilities (*Navarro*)
  - Preserve software coming from various places
  - Provide data processing capabilities as a “service”
  - Bring the “user software to the data” instead of the “data to the user”

- Sharing and preservation of knowledge, including international cooperation
  - IVOA, IPDA, within ESA (with EO), CCSDS, PV2015, …

ESA UNCLASSIFIED - Releasable to the Public

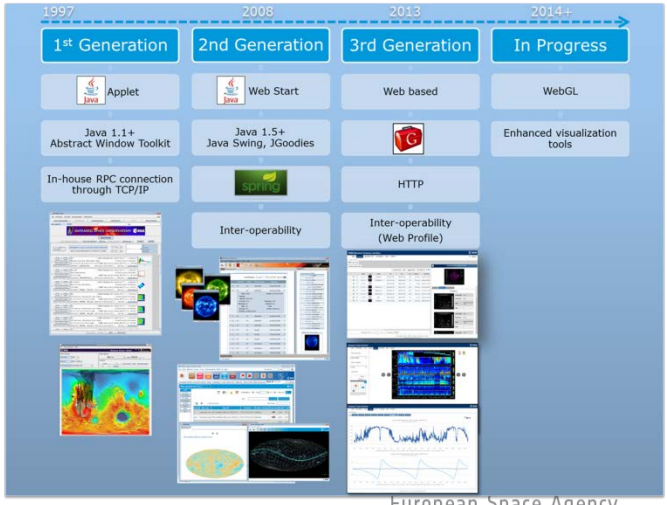
ESAC Archives Technologies

- Flexible and modular system architecture
  - OAIS standard (Open Archival Information System)
  - Interfaces with data providers, external tools and archives through VO protocols
  - Enabling building added value services on top of our archives



- Need to support all phases
  - Development, operations, post operations, and legacy (LTDP – long term data preservation)

- Technology monitoring and update at regular intervals
  - Java Rich Client to *Web Thin Layer*
  - Database: Commercial to *Open Source*
  - Web Services and Interoperability (*VO*)
  - "*Big Data*" challenges (Gaia, Euclid)



ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

Long Term Preservation
Upcoming Challenges

- More and more ESA missions entering Legacy phases
  - Venus Express, Planck in 2016
  - Herschel, LisaPF in 2018
- Preserve and improved science exploitation facilities
- Archive technology migration for some old archives (ISO, …)
- Setup Software preservation framework
- Define data persistent identifiers (ie DOI – Digital Object Identifiers)
- ESA LTDP+ programme being defined for 2017-2022

ESA UNCLASSIFIED - Releasable to the Public

International Collaboration and Interoperability amongst Archives

IVOA - International Virtual Observatory Alliance

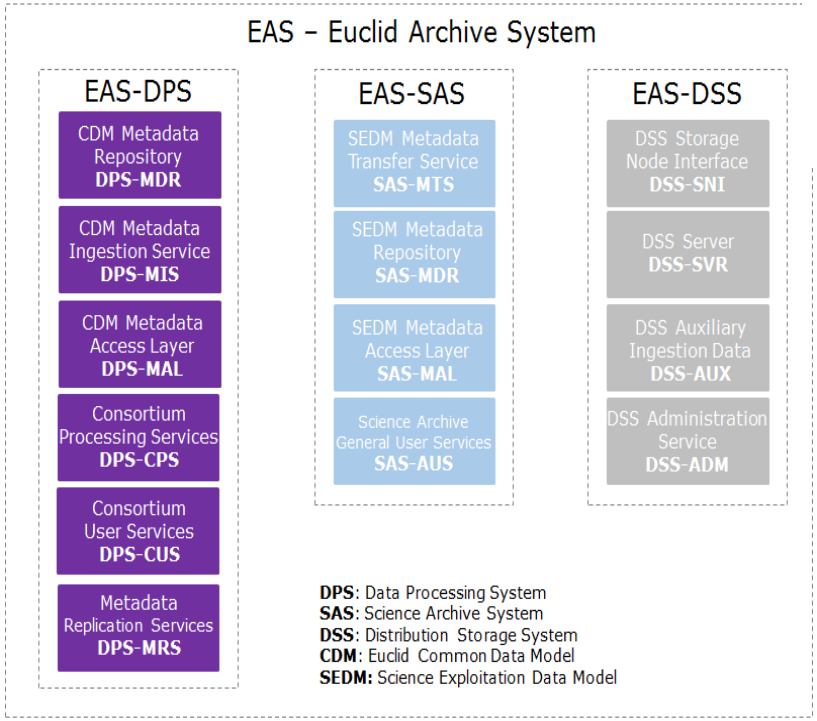
- Formed in 2002, 20 member projects
- Defines interoperability standards (VO framework) amongst astronomical (ground and space based) archives
- Working Groups and Interest Groups per technical domain (data access, data model, registry, applications, semantics, operations, …)
- http://www.ivoa.net/

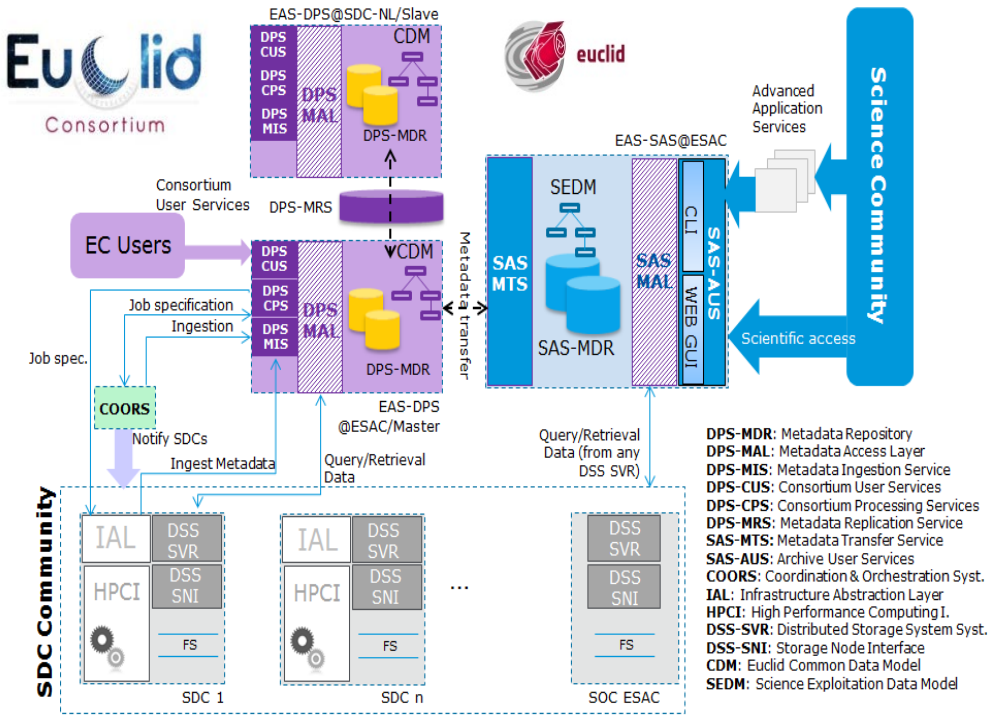
IPDA - International Planetary Data Alliance

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

ESA UNCLASSIFIED - Releasable to the Public

Archives Integration within Projects



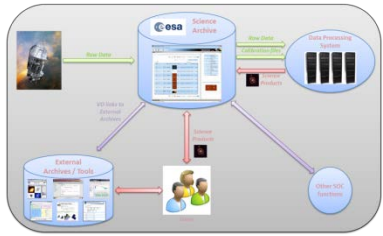


Archive is fully part of Science Operations, from the start

Strong collaboration with SGS and Consortium

ESA UNCLASSIFIED - Releasable to the Public

Archives full integrated in projects



- Archive is *not* an after thought at the end of the mission
- Archive is to be part of SciOps development from the beginning

- Archive is more and more a collaborative process with many groups involved
  - Within Archives Groups
  - With SOC/SGS
  - With Consortium
  - With external Community

- Need for more collaborative methods
  - Scrum methodology for daily interactions with SOC/SGS

- Open source development for collaboration with Consortium

ESA UNCLASSIFIED - Releasable to the Public

Standard Archives Architecture

Command line

http

xmm-newton science archive

Browser GUI

http

Database

SAMP

SIAP, SSAP, TAP

VO services

Data Repository

VO Apps

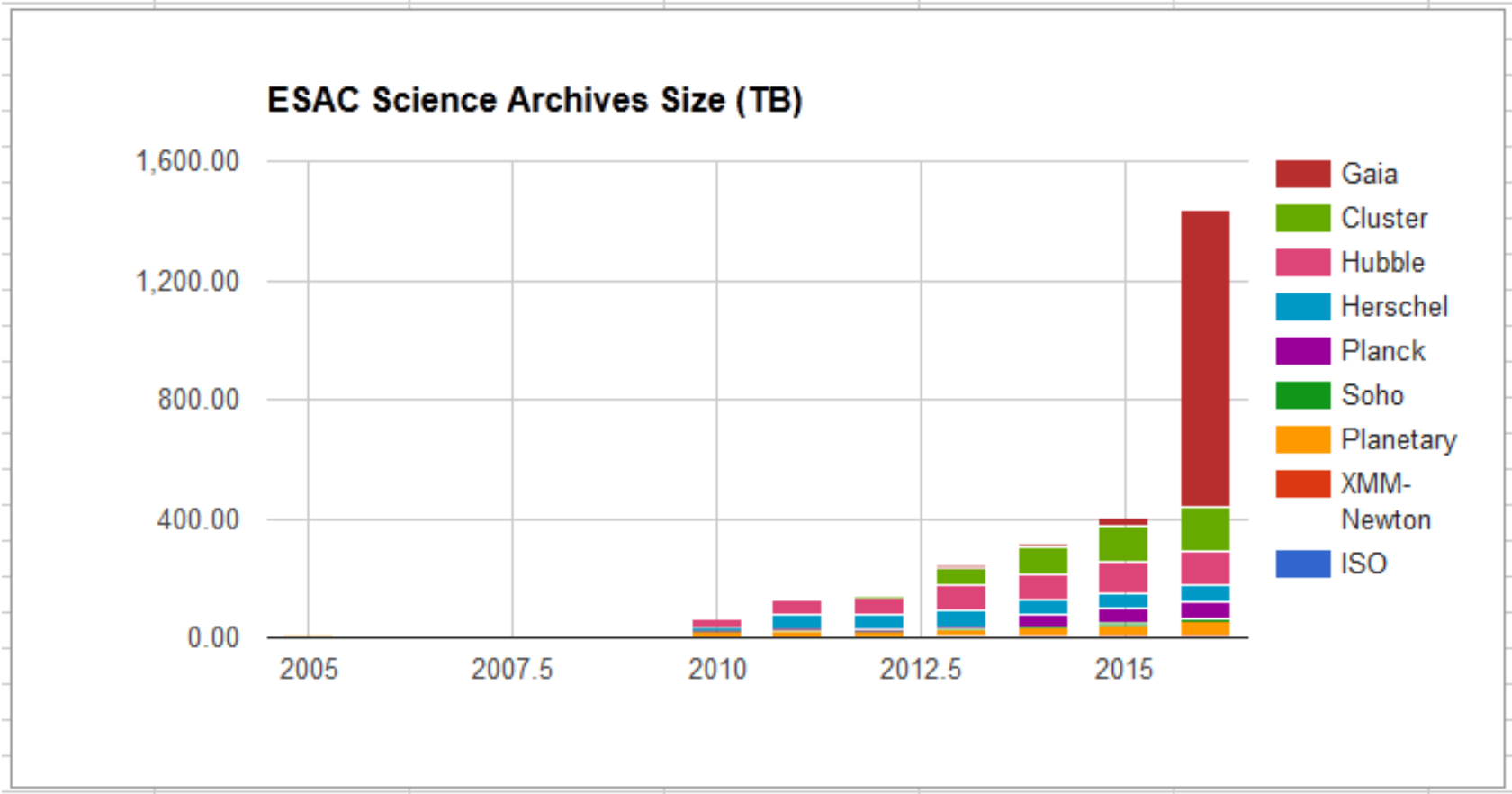
ftp

User Disk

ESA UNCLASSIFIED - Releasable to the Public

European Space Agency

ESAC Archives Volume Evolution



All data stored on hard disks and distributed through Internet

Euclid will add up to ~150 PBs by 2023

ESA UNCLASSIFIED - Releasable to the Public

Need for new paradigm

- 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

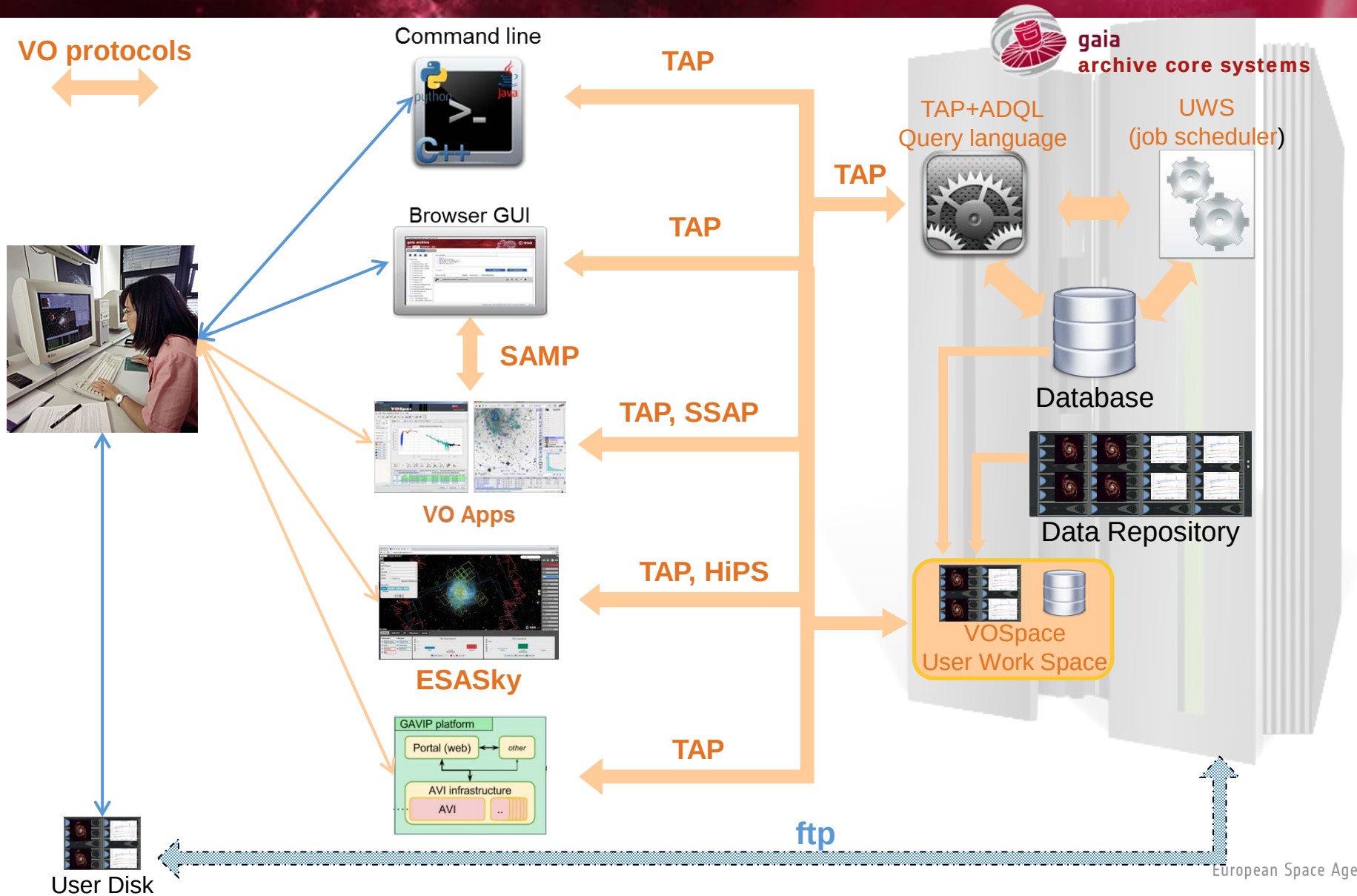
- User can not download all catalogue and all data
  - Need to have user workspaces *IN* the Archive
    - User database space, user disk space
  - User workspace shareable amongst various users

- Bring user code to the data (*OMullane*)
  - 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

ESA UNCLASSIFIED - Releasable to the Public

Gaia Archive Architecture



Mission and Archives metrics ?

- Mission success usually measured by number of scientific papers
- How to measure success of the archives ?
  - Number of active users / IP addresses
  - Volume of data being downloaded vs size of the archive
  - Ratio of data-out vs data-in ?
- Big differences between missions, between missions phases
- Interested to hear how other data centres report on archives usage…

ESA UNCLASSIFIED - Releasable to the Public

Science Archives

Fulfil various Science Operations functions

- Maximum science exploitation of data
- Ensure long term preservation of data, software and knowledge
- Support various functions of Science Ground Segments

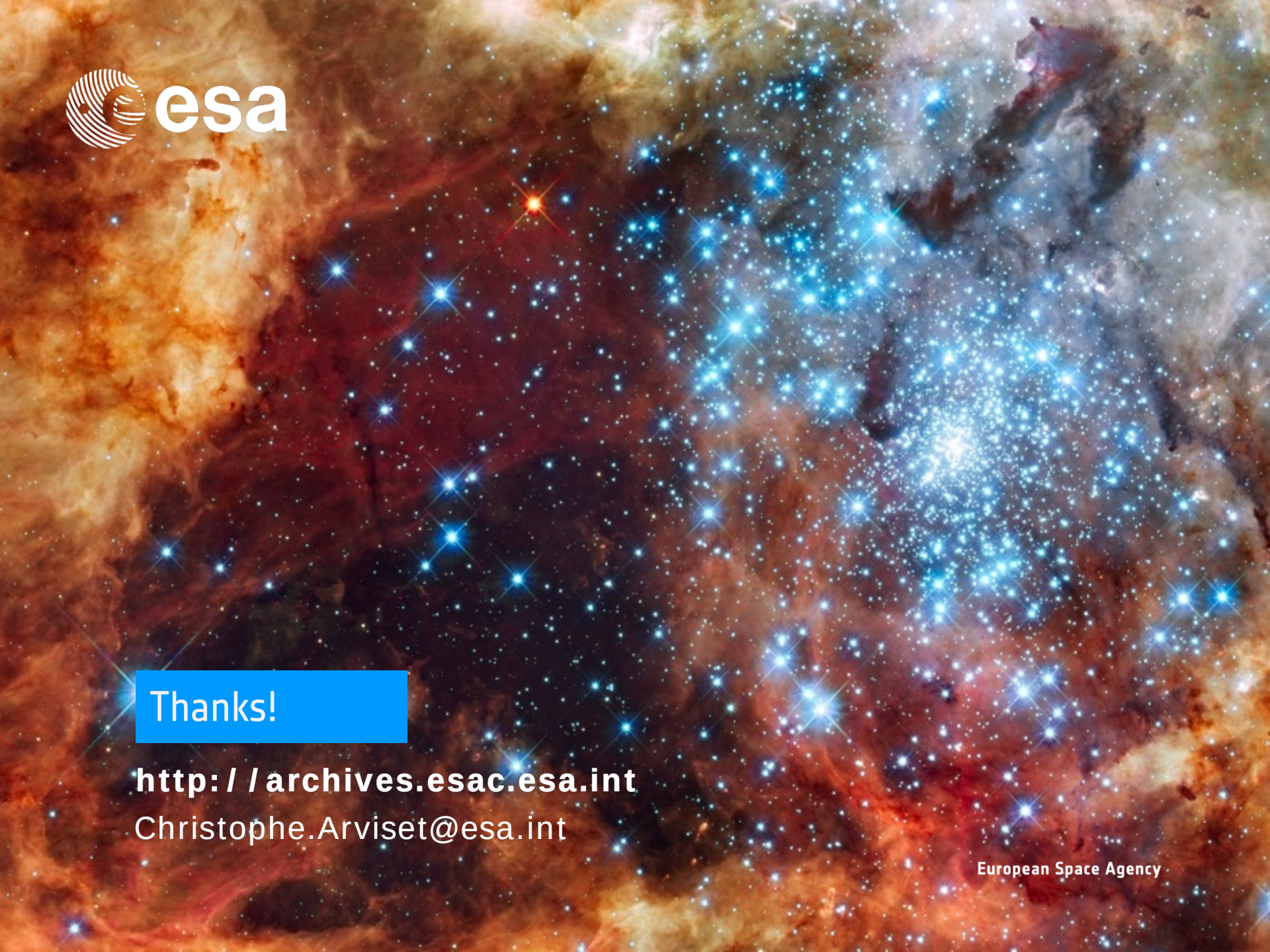
Scientists and Engineers working together

- Archives must be science driven
- Archives require strong IT expertise

Towards Archives 2.0

- Open data, open source, open archives

ESA UNCLASSIFIED - Releasable to the Public


esa
Thanks!
http://archives.esac.esa.int
Christophe.Arviset@esa.int
European Space Agency