More than Bytes: Technical and Social Dimensions of Data Lifecycle Management at ESA's Space Science Data Archives

Rachana Bhatawdekar[1], Christophe Arviset[1], Bruno Merin[1], Guido de Marchi[1], Jos de Bruijne[1], Deborah Baines[1], Arnaud Masson[1], Mark Bentley[1], Hector Perez[2], and Ignacio Leon[2]

[1]European Space Agency, [2]Starion Group for ESA

ADASS XXXV, Gorlitz, Germany, 10/11/2025

ESAC Science Data Centre (ESDC)

ESAC Science Data Centre (ESDC)

European Space Astronomy Center, ESAC
Villanueva de la Canada, Madrid, Spain

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

ESDC Archives

- Central hub for ESA's space science missions — Astronomy, Planetary, Heliophysics and Human and Robotic Exploration
- Over 30 missions supported; from ISO (1998) → Euclid (2025)
- ~34,800 monthly users, 106 TB downloaded/month, 1.1 PB total archive size

The Planetary Science Archive

bepicolombo · cassini huygens · chandrayaan -1 · exomars · giotto · juice · mars express · rosetta · smart-1 · venus express

Astronomy Science Archives

cheops · euclid · esasky · exosat · gaia · herschel · hubble space telescope · integral · iso · jwst · lisa pathfinder · planck · xmm-newton

Future Archives

ariel · plato · smile · proba-3

Heliophysics Science Archives

cluster · double star · proba-2 · soho · solar orbiter · ulysses

Human and Robotic Exploration Science Archives*

HRE data archive · ISS- SolACES*

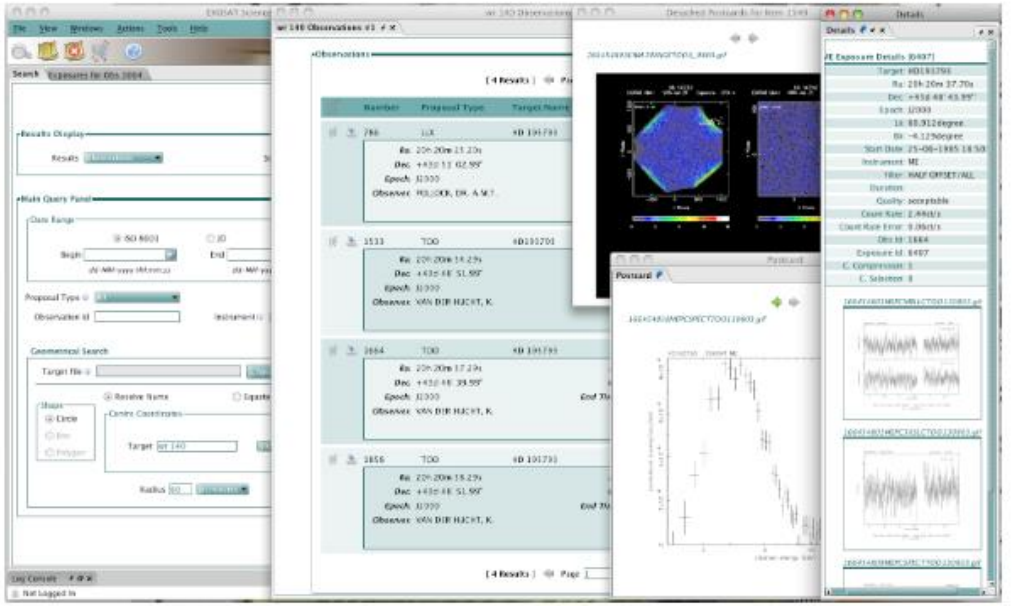
ESDC Statistics

Monthly Users (*)	Monthly Downloaded (*)	Archive Total Size
34 778	106.2 TB	1146.2 TB

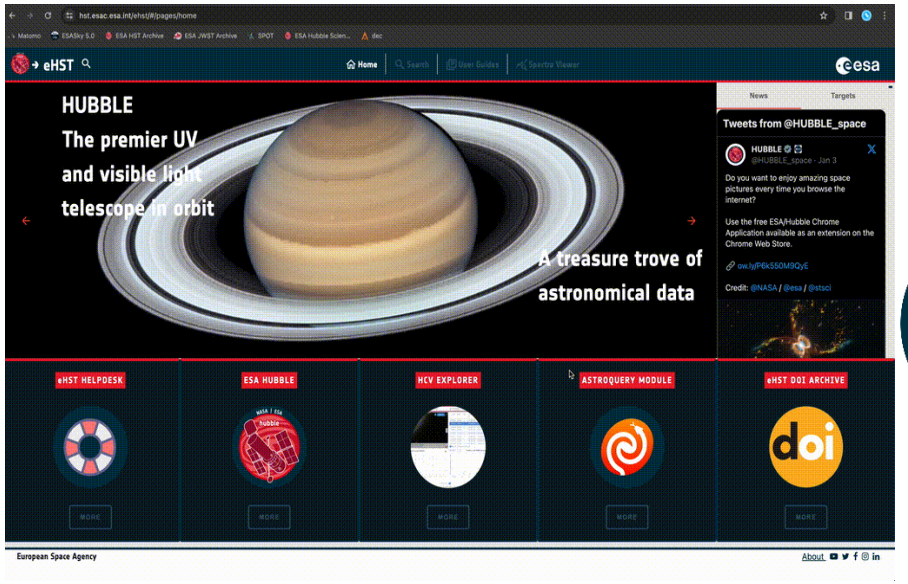
* Monthly averages in 2024

What *is* an archive really?

Archives: More than “Just a bunch of disks”



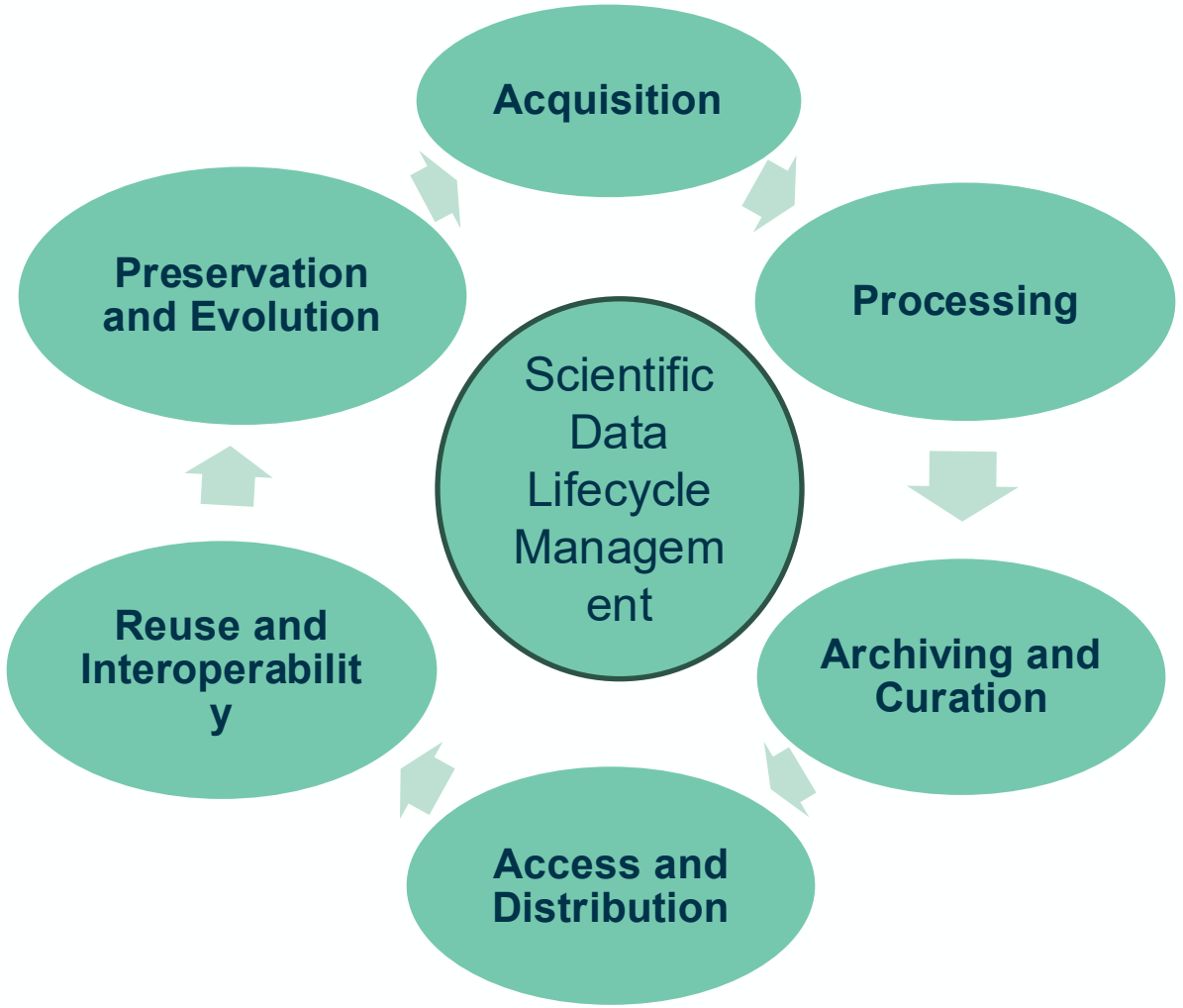
EXOSAT Science Archive



eHST Science Archive

- Host, curate, preserve and serve the data
- Ensure discoverability and interoperability (GUIs, APIs and VO protocols)
- Full data management systems (not just static storage)
- Provide tools for discovery, interfaces for access, and visualization services

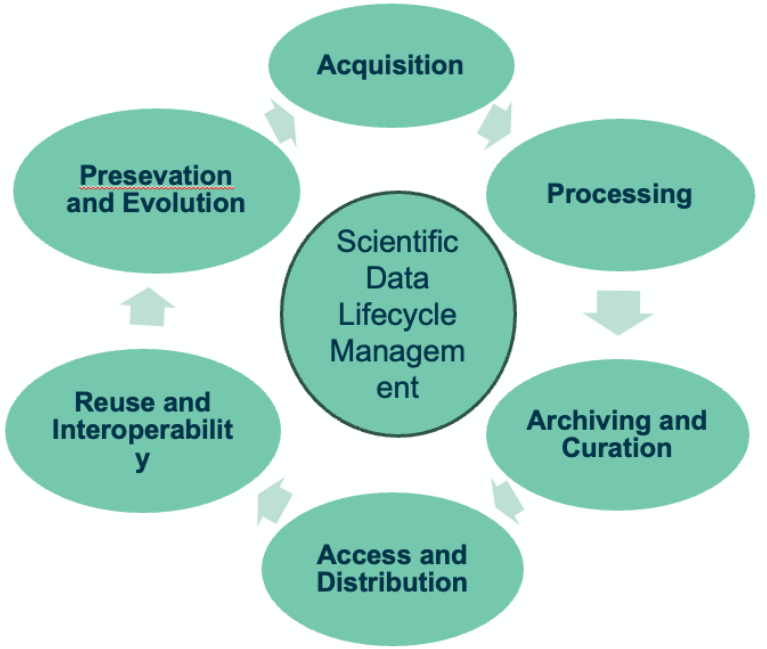
What is Data Lifecycle Management?



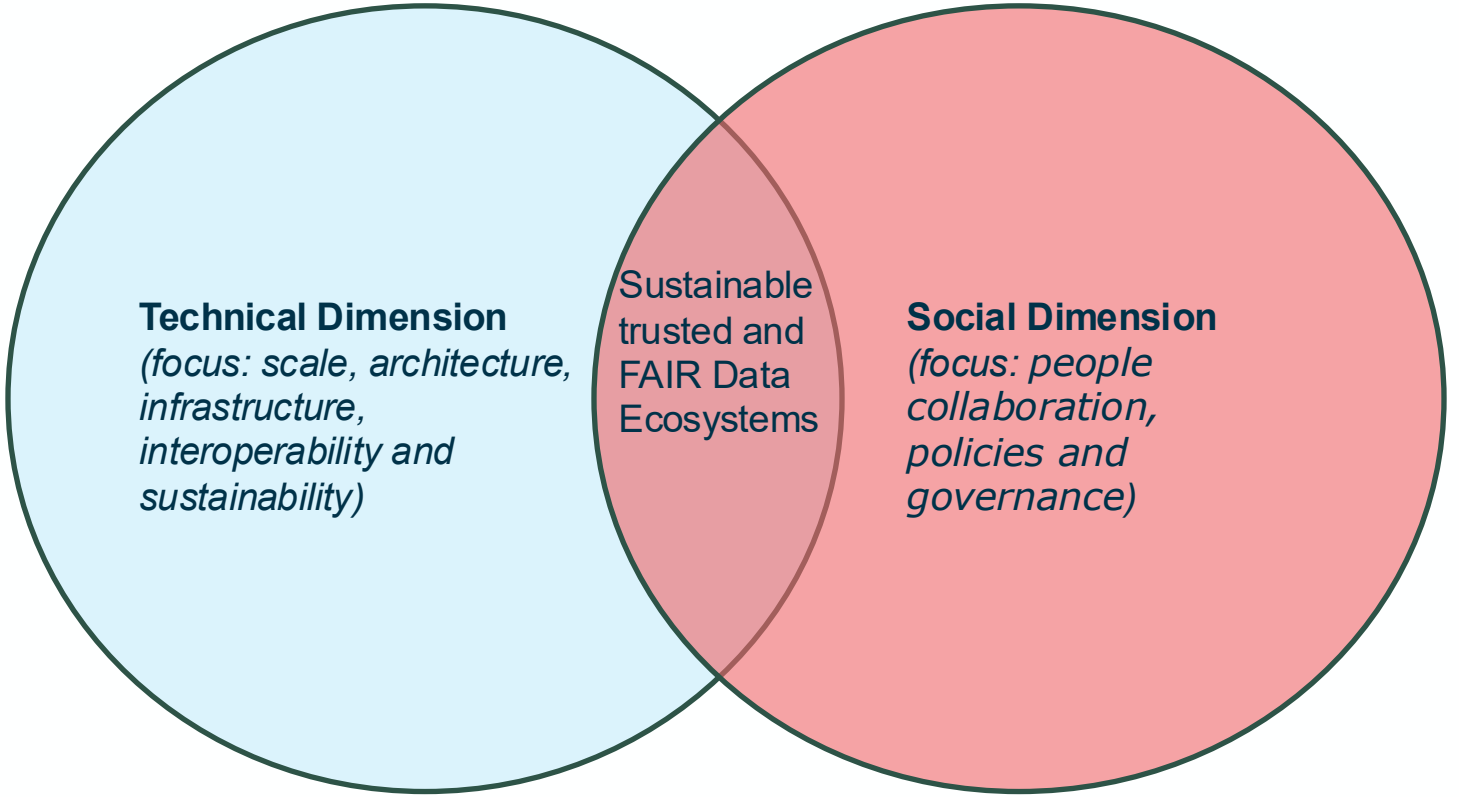
- Structured process: from data creation to long-term preservation and reuse
- Cradle-to-grave (and beyond) view: from orbital acquisition to decades after mission end
- Ensures data remain usable, discoverable, and interoperable over time
- For space missions, the data lifecycle management isn't just a technical process — it's a long-term responsibility to science

Why Data Lifecycle Management Matters?

- Space missions end — but their data must remain usable for decades
- Prevents data loss, obsolescence, and knowledge decay
- Ensures preservation, accessibility, and interoperability
- Enables cross-mission discoveries and new science long after launch
- Integrates **FAIR principles** (Findable, Accessible, Interoperable, Reusable)
- Supports sustainability and efficiency from design to legacy



Two inseparable dimensions of Data Lifecycle Management



Technical Dimensions

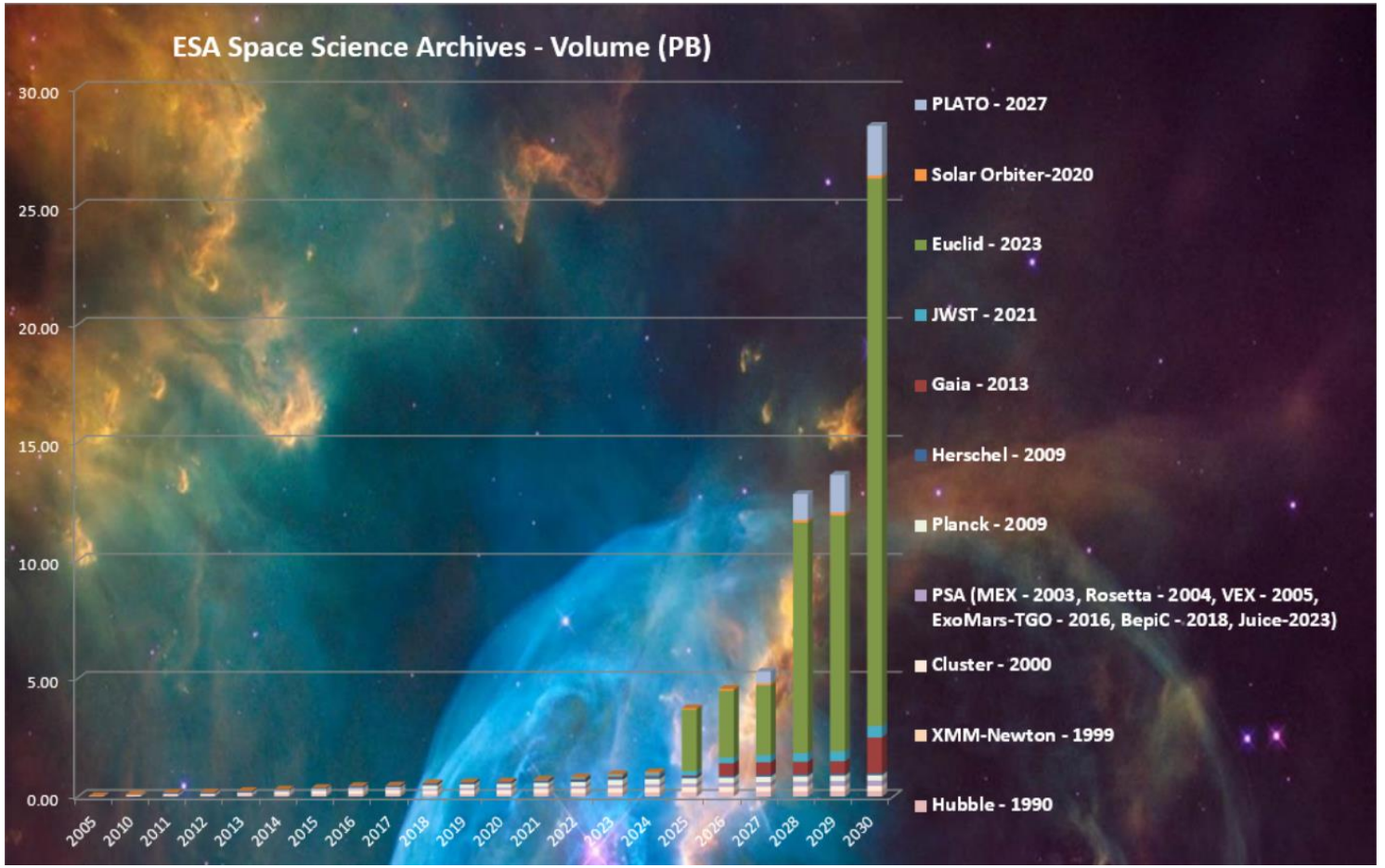
- Data growth and scalability (Managing petabyte-scale, multi-mission datasets)
- Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)
- Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)
- Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)

Technical Dimensions

- **Data growth and scalability (Managing petabyte-scale, multi-mission datasets)**
- Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)
- Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)
- Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)

Technical Dimensions

- **Data growth and scalability (Managing petabyte-scale, multi-mission datasets)**



New "big data" missions require a paradigm shift

*From "move the data to the user" towards "move the user to the data" => **Science Platforms***

Technical Dimensions

- **Data growth and scalability (Managing petabyte-scale, multi-mission datasets)**

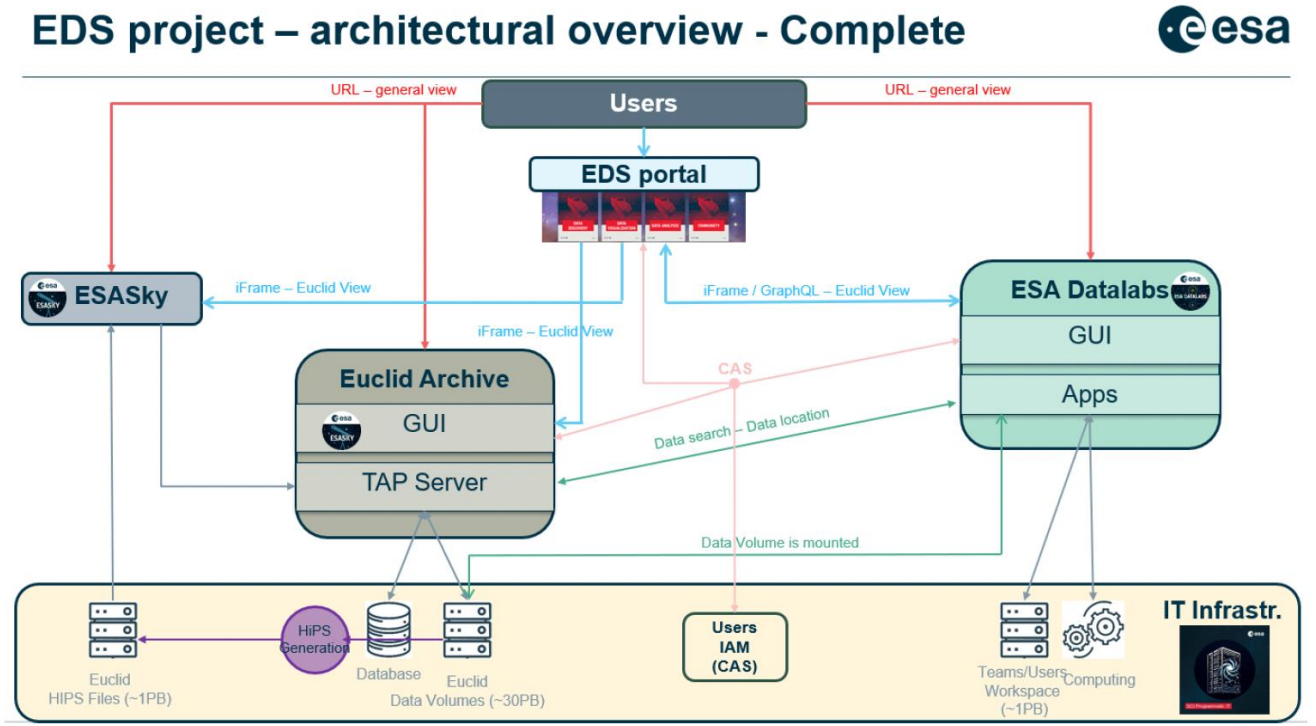


Figure 1: Euclid Data Space Architectural Overview

1. ESA Datalabs - our multi-mission science platform platform
2. Euclid Data Space (EDS) Portal – being developed for the Euclid mission

EDS version 1.0.0 released!
https://euclid.dataspace.esa.int

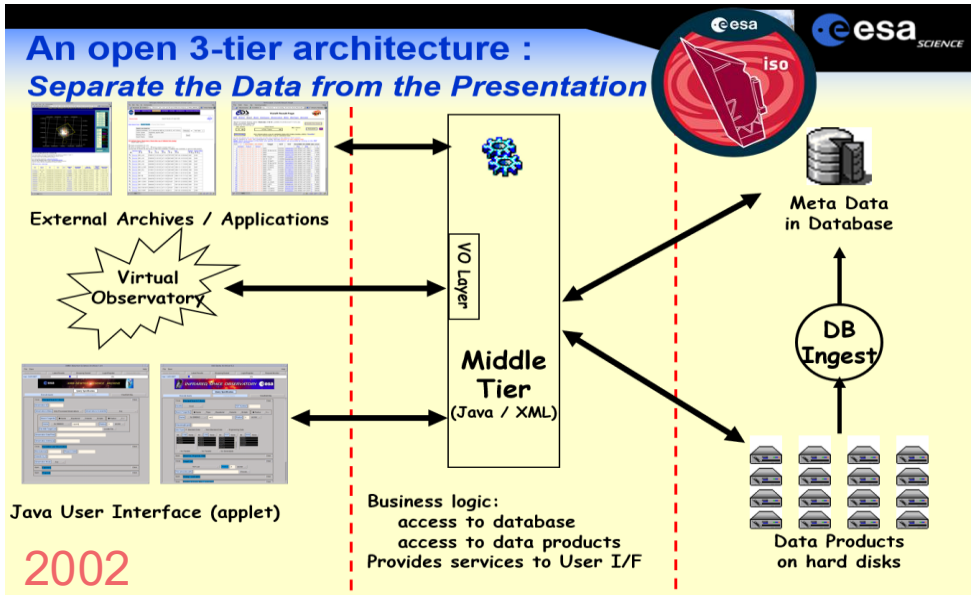
Sandor Kruk's talk ->
Nov 12 9AM

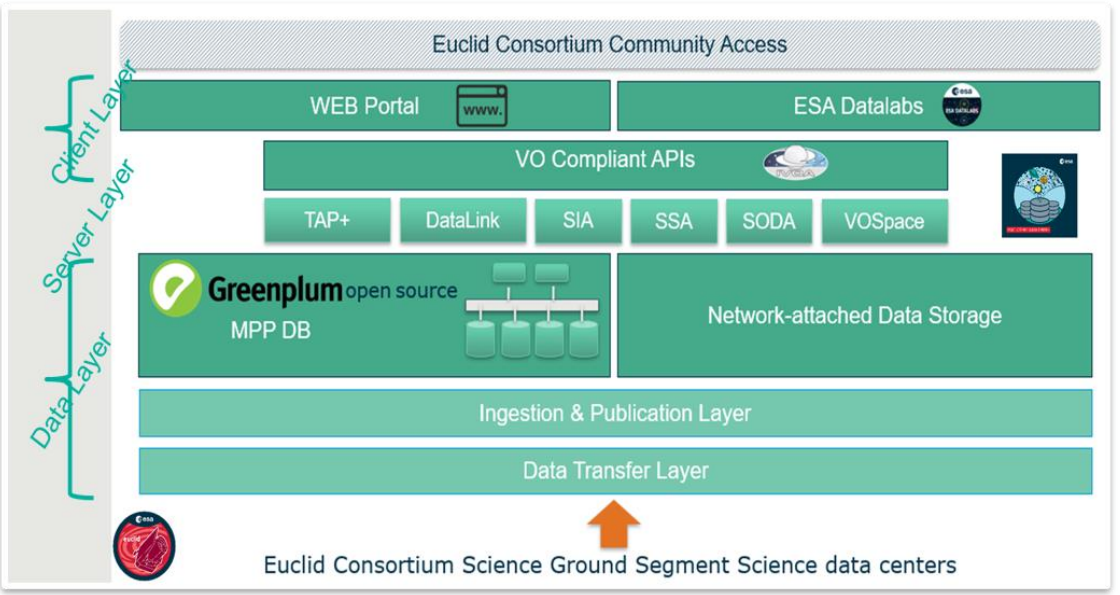
Technical Dimensions

- Data growth and scalability (Managing petabyte-scale, multi-mission datasets)
- **Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)**
- Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)
- Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)

Technical Dimensions

- **Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)**

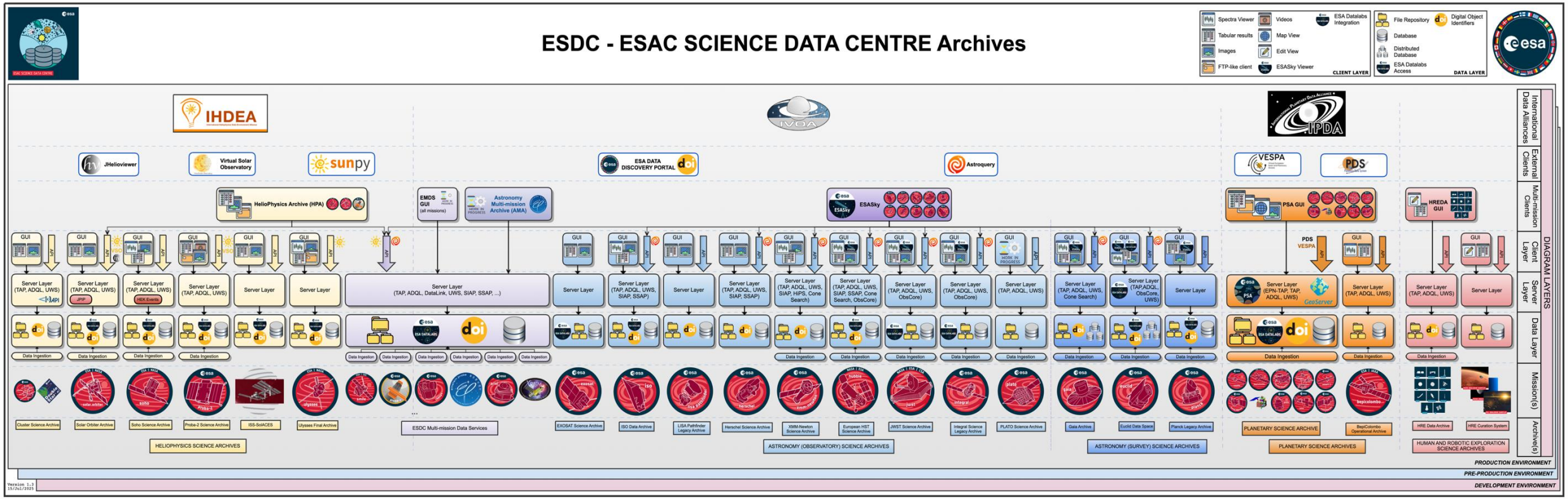




Stable 3 tier Architecture
Migration of individual layer to new technologies
From VO-layer towards VO-built-in
Easier interoperability (Astropy, ESASky, ESA Datalabs)

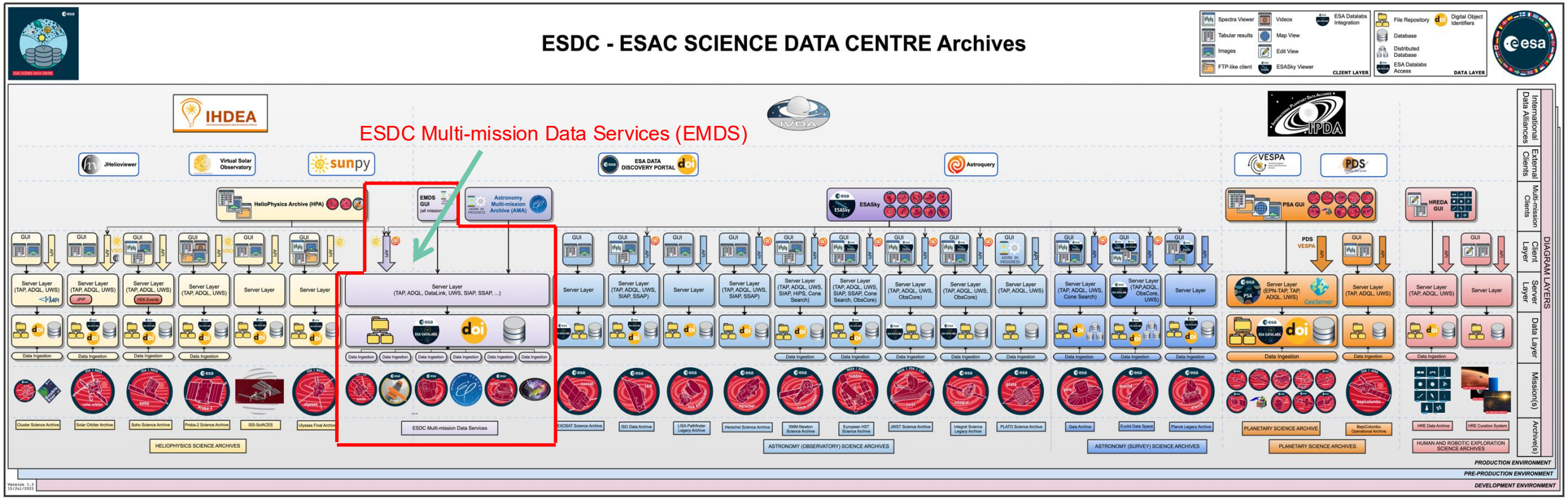
Technical Dimensions

- **Integrating diverse missions within a unified architecture**



Technical Dimensions

- **Integrating diverse missions within a unified architecture**



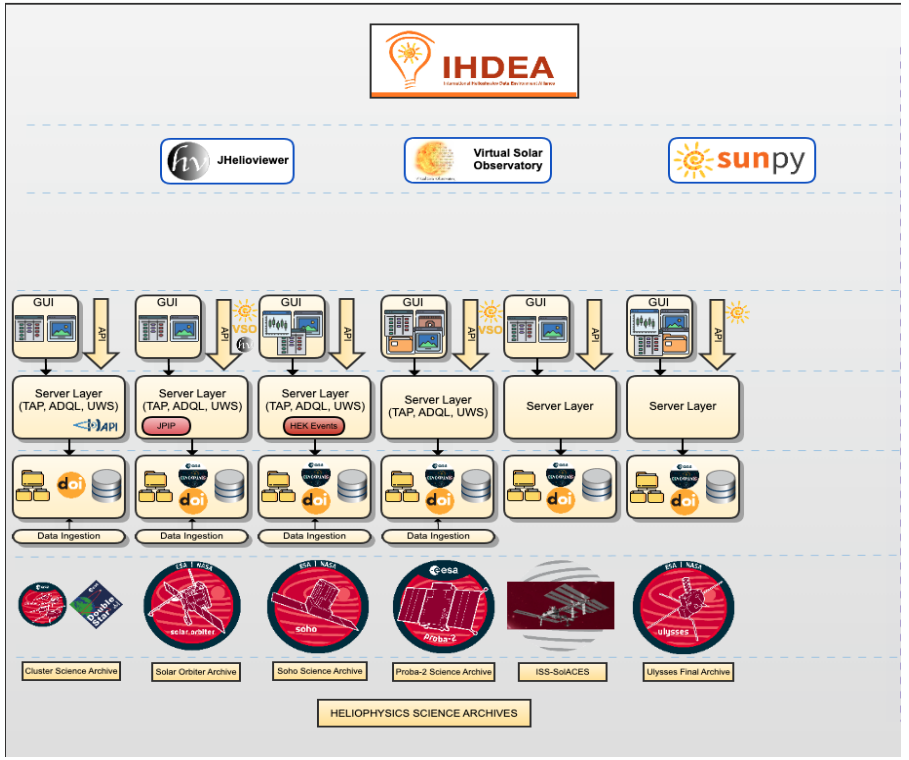
Technical Dimensions

- Data growth and scalability (Managing petabyte-scale, multi-mission datasets)
- Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)
- **Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)**
- Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)

Technical Dimensions

- **Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)**

- Standards live in the server layer (TAP, SIAP, SSAP, HAPI…)
- Enable cross-mission and international usability
- Support scientific cross-matches and multi-mission analyses
- Complement client-layer interfaces (GUIs, APIs, ESASky)

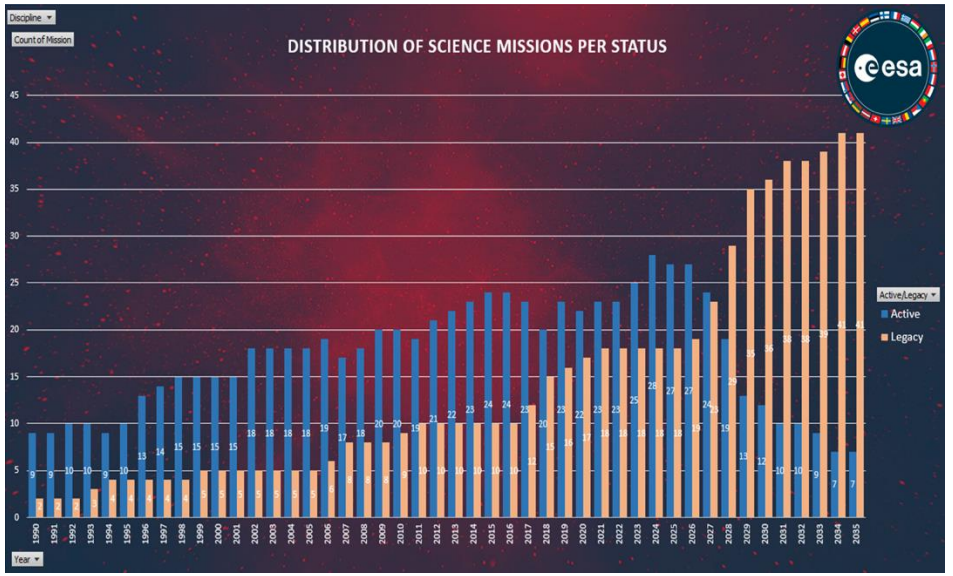


Technical Dimensions

- Data growth and scalability (Managing petabyte-scale, multi-mission datasets)
- Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)
- Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)
- **Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)**

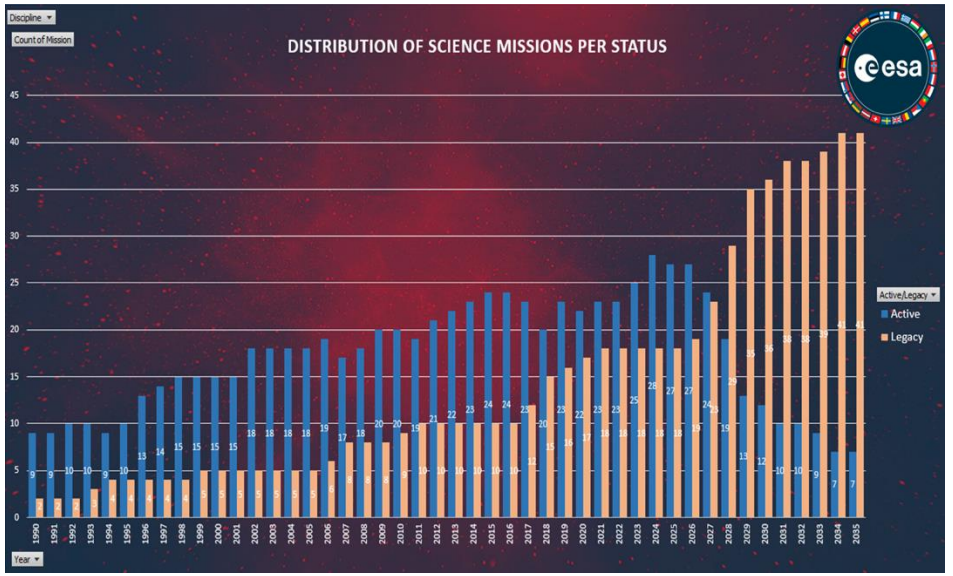
Technical Dimensions

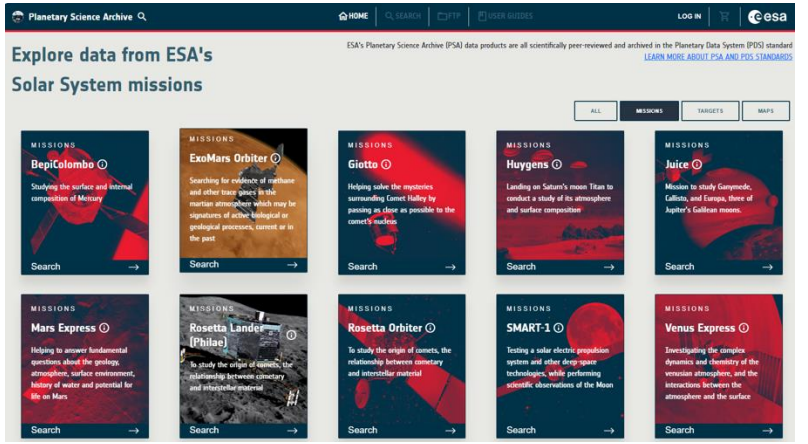
- **Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)**



Technical Dimensions

- **Multi-mission archives**

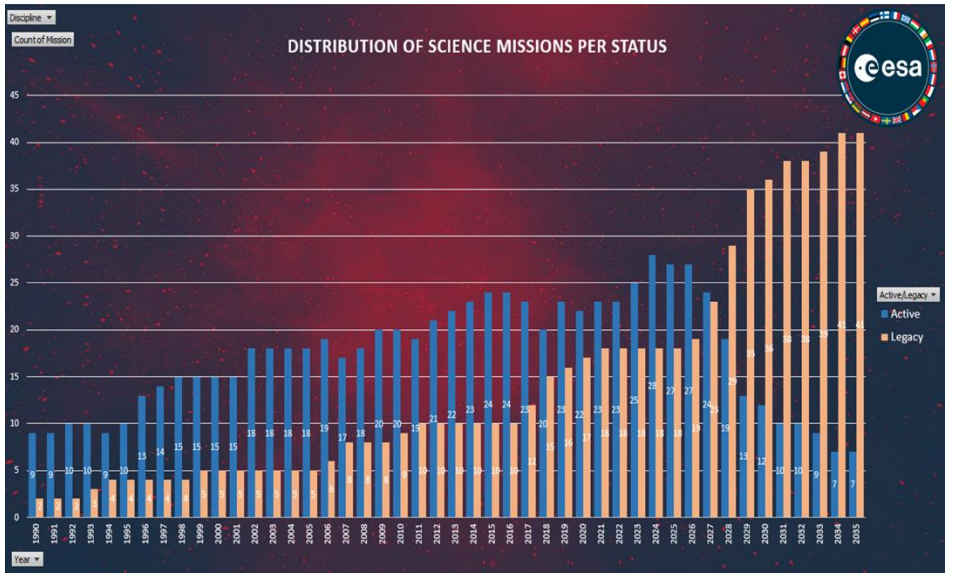


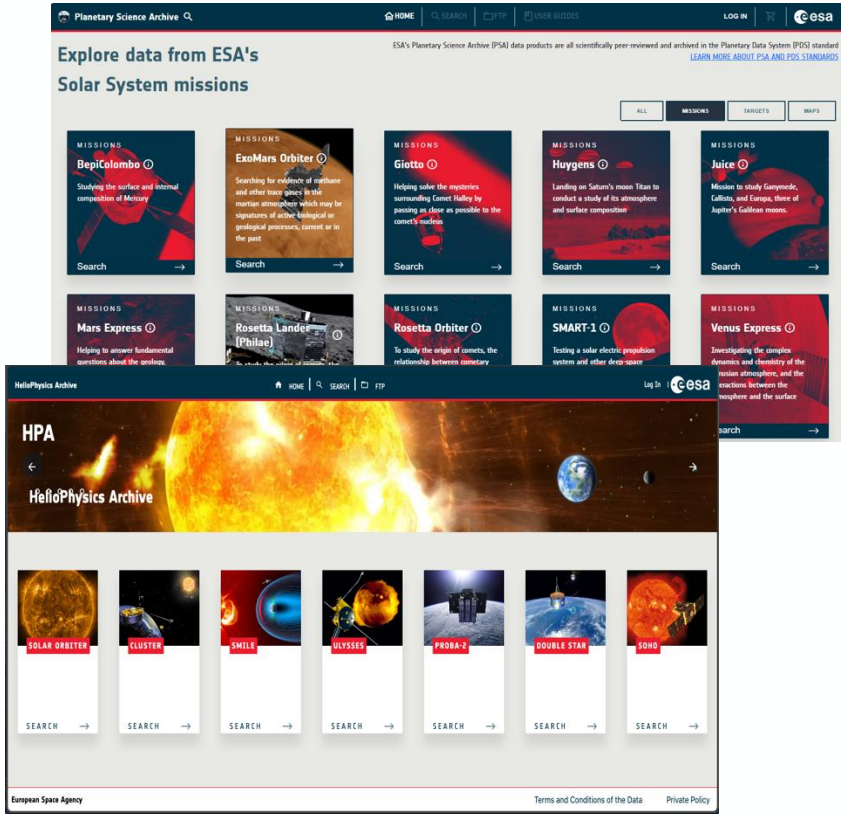


Planetary Science Archive (PSA) – https://psa.esa.int

Technical Dimensions

- **Multi-mission archives**

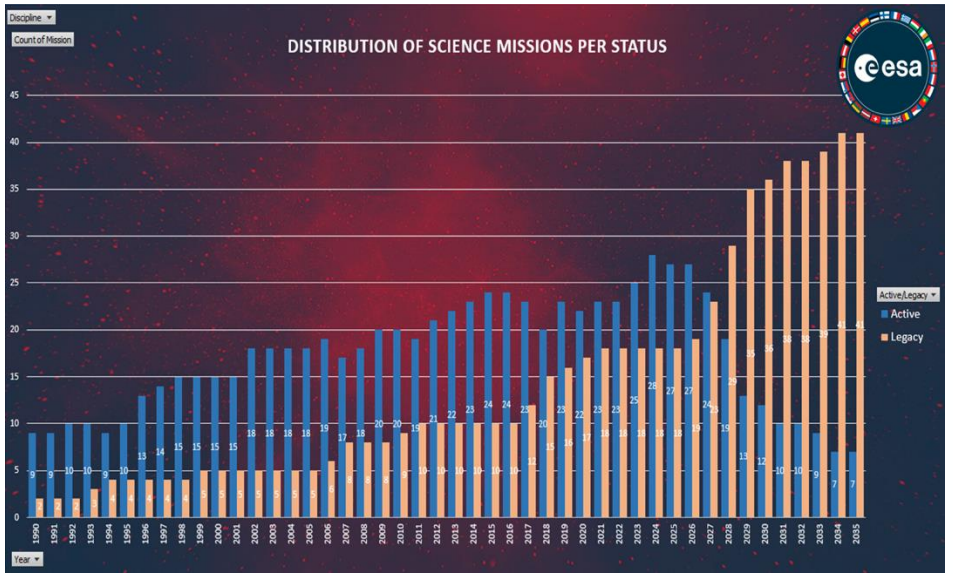


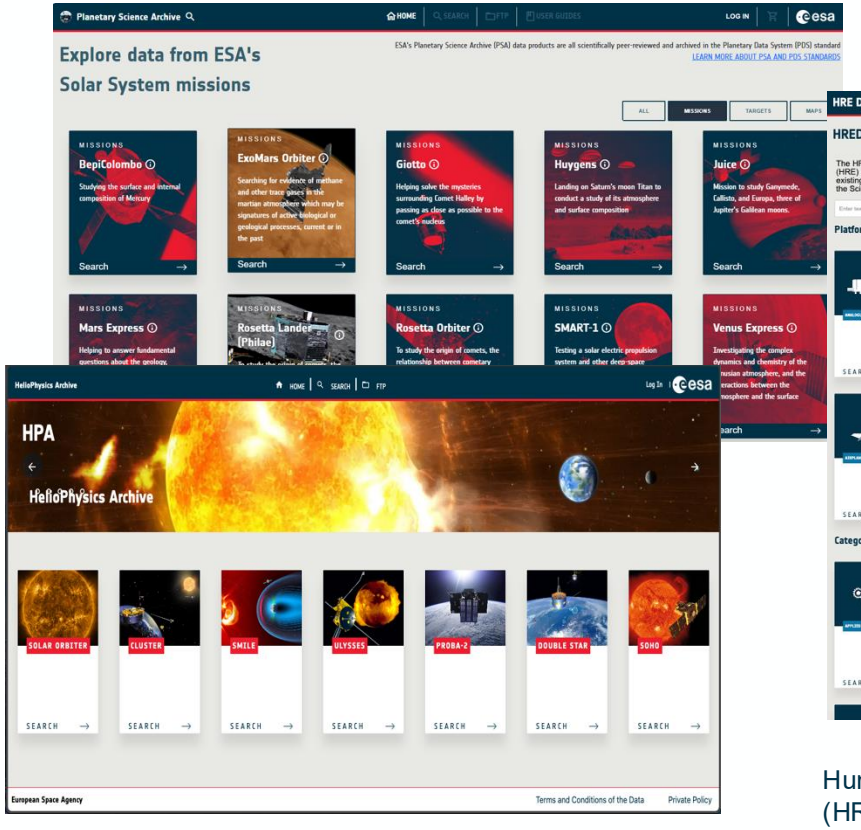


Heliophysics Archive (HPA) – https://hpa.esa.int

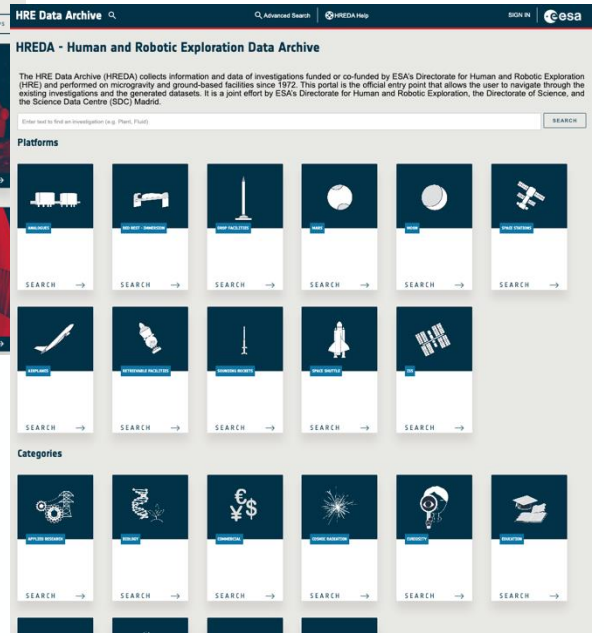
Technical Dimensions

- **Multi-mission archives**





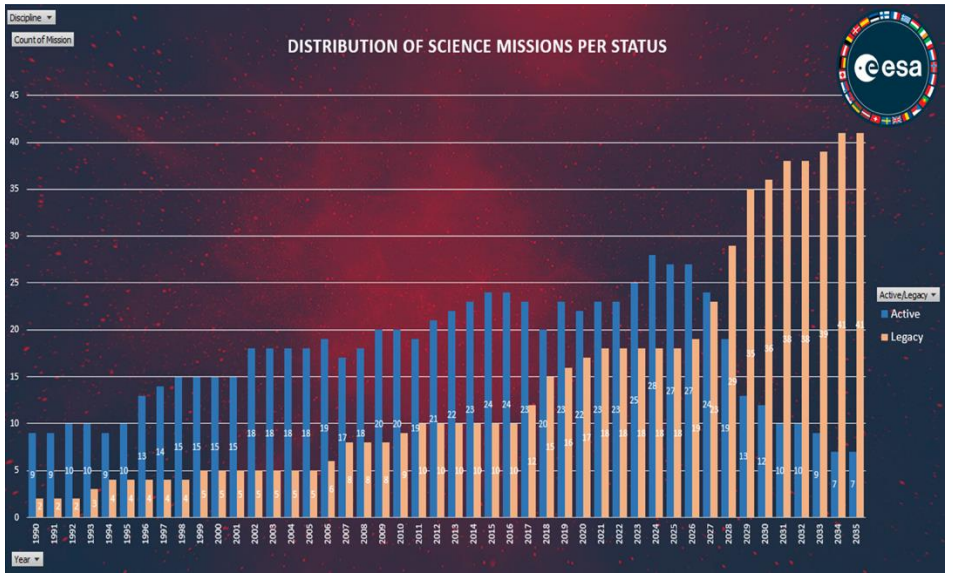
Heliophysics Archive (HPA) – https://hpa.esa.int

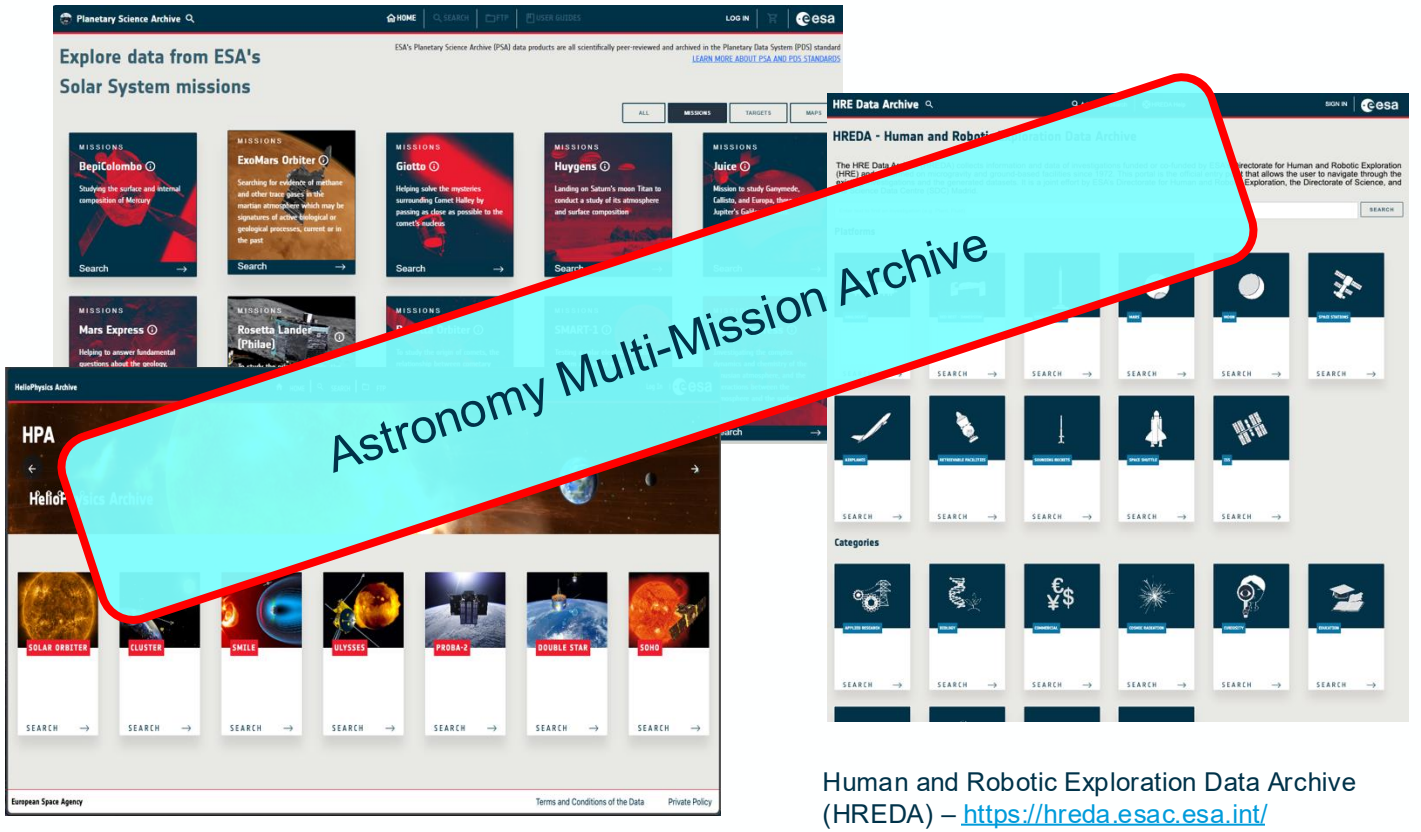


Human and Robotic Exploration Data Archive (HREDA) – https://hreda.esac.esa.int/

Technical Dimensions

- **Multi-mission archives**





Heliophysics Archive (HPA) – https://hpa.esa.int

Human and Robotic Exploration Data Archive (HREDA) – https://hreda.esac.esa.int/

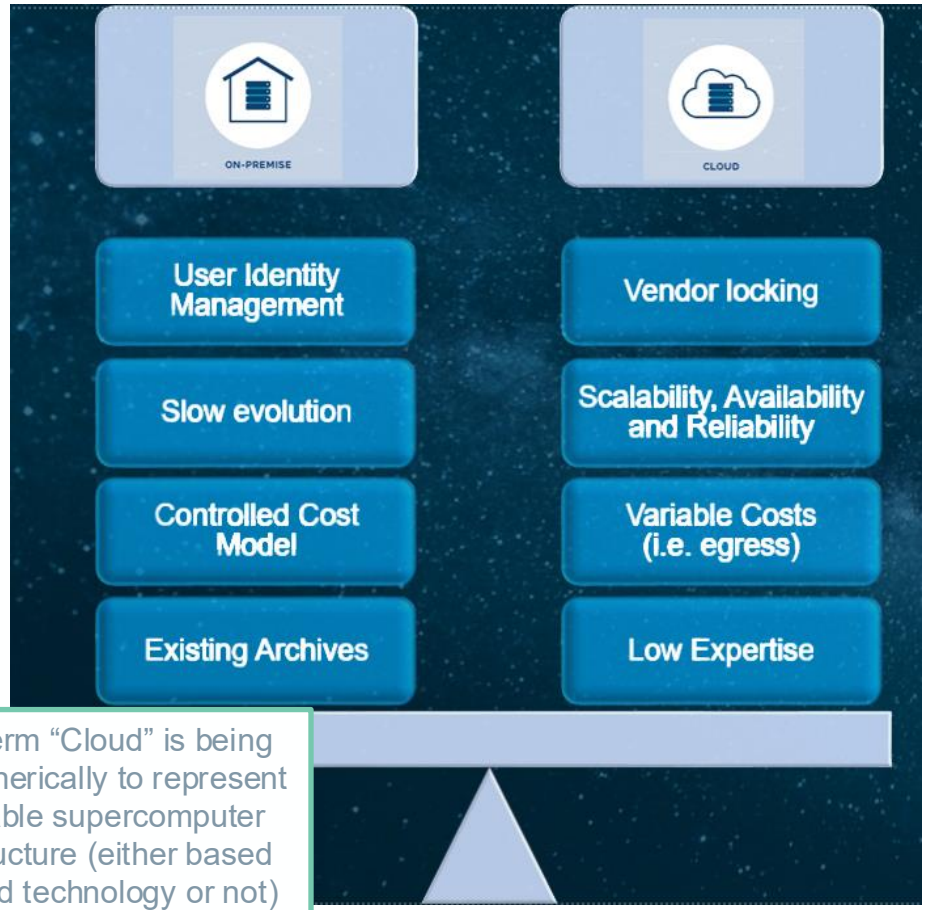
Technical Dimensions

- **Technology Evolution**

	1st Generation 1998-2006	2nd Generation 2006-2011	3rd Generation 2011-2019	4th Generation 2019-…
Client Layer	Java Applet AWT	Java Webstart Swing Jgoodies Infonode	Java GWT Vaadin	Javascript Angular UX design Link to Viz tools Link to ESA Datalabs
Server Layer	In-house Java layer	Open source Java server IVOA protocols layers on top		Open Source Java Server using IVOA protocols
Data Layer / Persistence	Sybase Oracle	PostgreSQL + pgSphere / postGIS Hibernate Ehcache		Greenplum
Data Ingestor	Java standalone application			Apache Camel

Technical Dimensions

- **Archives on premise or in the cloud**



* The term “Cloud” is being used generically to represent a scalable supercomputer infrastructure (either based on Cloud technology or not)

- Migrating all archives and EDL into the cloud needs to be thought through with care, i.e. proprietary and public data, identity management, colocation of storage and compute, …
  Requires expertise and planning to do it right

- Some clear advantages for immediate availability, scalability and reliability

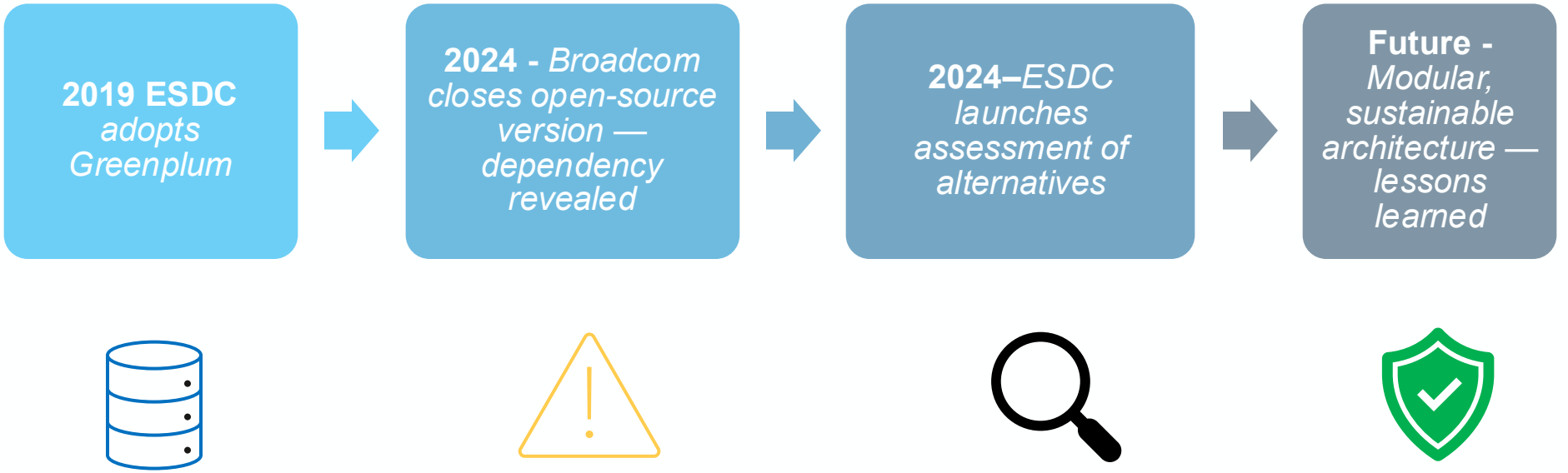
- Egress costs remain a question with respect to cost model. We want our archives to be used as much as possible, and we also need a controlled cost model

- ESA as a European Institution will go towards European cloud providers.

⇒ ESA Datalabs and Archives Cloud Study Project (EDACS)

Technical Dimensions

- **Responding to unplanned technology shifts**



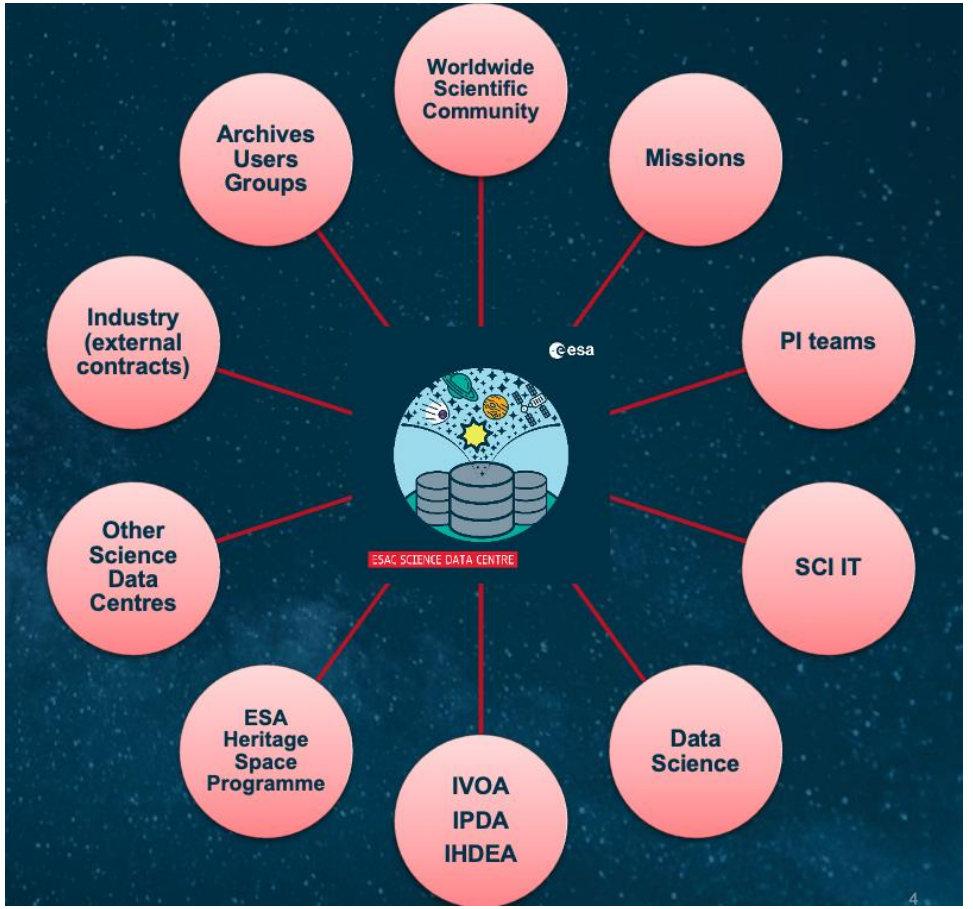
Technical sustainability is not only about selecting the right software, but it is also about ensuring that the ecosystem can adapt when external conditions change!

Technical Dimensions

- Data growth and scalability (Managing petabyte-scale, multi-mission datasets)
- Integrating diverse missions within a unified architecture (Scaling a common framework to support multiple scientific domains)
- Enabling interoperability and standards across missions (VO protocols, PDS4, HAPI— ensuring cross-mission and international usability)
- Evolving for the future while preserving the past (Building scalable, maintainable, and future-proof archives)

Social Dimensions

- **Cross-agency collaboration** – through international partnerships (IVOA, IPDA, IHDEA,) coordination with NASA, JAXA, CSA/NSSC, and partner centres like CDS, STScI/MAST, LINCC, SPACIOUS, and mission-specific collaborators (e.g. DZA for Gaia, IPAC for Euclid).
- **Policies and governance** – through policies and governance, ESA defines how its science-data can be accessed, cited, and shared responsibly
- **Knowledge continuity and expertise retention** – Preserving institutional know-how as missions transition into legacy phases
- **Empowering the scientific community** – Enabled through user-friendly interfaces, helpdesks, workshops, and the Archives User Group, ensuring user feedback directly shapes system evolution.
- **Open science and transparency** – Promoting accessible, standards-based, data sharing across international archives.



Take away messages

- **Data lifecycle management is essential:** Managing space mission data from acquisition to long-term preservation ensures usability decades into the future
- **Technical and social dimensions go hand in hand**: Scalable, interoperable systems and policies, governance, and collaboration together enable sustainable archives.
- **Archives are active enablers of science**: Beyond storing files, they provide discovery, access, and integration tools for cross-mission research.
- **Planning for the future is critical:** Platforms like EMDS, multi-mission archives, and technology evolution strategies help balance innovation with maintainability
- **Open standards enable interoperability:** By using protocols like TAP, HAPI.. ESDC ensures mission data can be consistently accessed, shared, and integrated across international archives
- **Lessons extend beyond ESA:** The approaches and experiences gained here can inform the design of future astronomical and interdisciplinary data infrastructures.