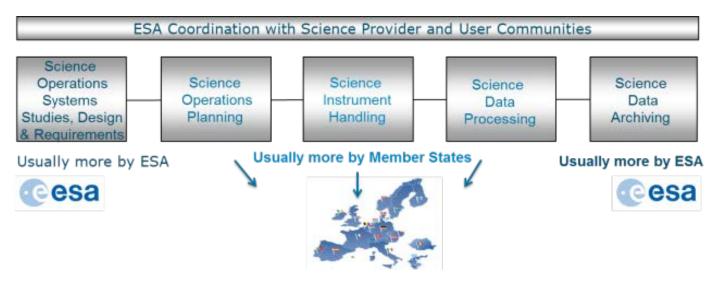


Background and context



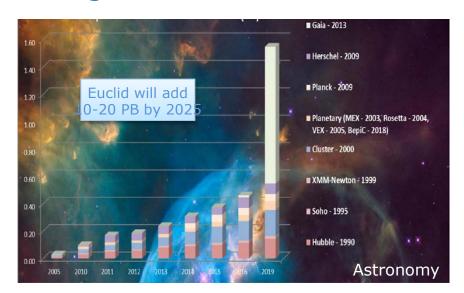


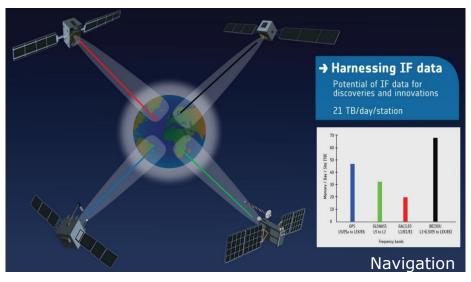
The **traditional Science Exploitation concept** is based on moving data and tools to the user, or to single user's organization, therefore transferred many times, replicated in many places, and with data exploitation taking place at users' premises.



Background and context







The fundamental principle of **SEPP's concept** is to move the user to the data and tools. Users access a **platform work environment** providing the data, tools, and resources required, as opposed to downloading, replicating, and exploiting data 'at home'.























Objectives: Preservation



- Long term preservation of Data Systems:
 - Mission data processing systems
 - Mission data analysis systems
- Enable easier access to standard legacy software
- Rescue the code and bring it to the (small) data
- On the fly instantiation of:
 - Full system environment
 - Predefined system environment (threads)



ESA UNCLASSIFIED - For Official Use

ESA | Slide 4

Objectives: Exploitation

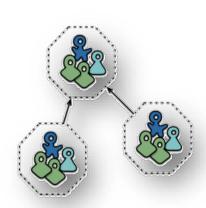


Exploitation platform for data

- Share your data (user storage space in the platform, VOSpace)
- Share your metadata (DB user space inside the archive, ...)
- Publish your data (VO protocols)

Exploitation platform for **software**

- Enable data processing where the data is
- Interactive data analysis Jupyter Notebook / Hub
- Generic pipeline development and testing environment
- User customized processing pipelines
- Share your code -> Science App Store

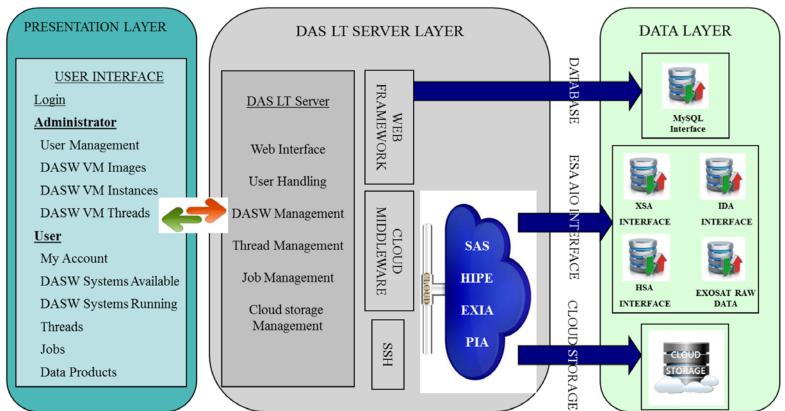


ESA UNCLASSIFIED - For Official Use

ESA | Slide 5

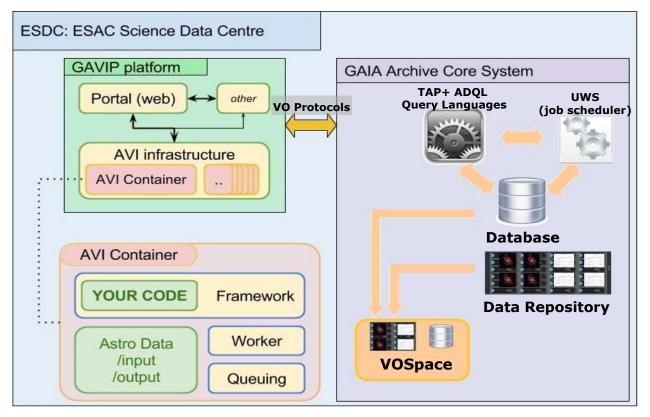
SCI related activities and systems: DASLT





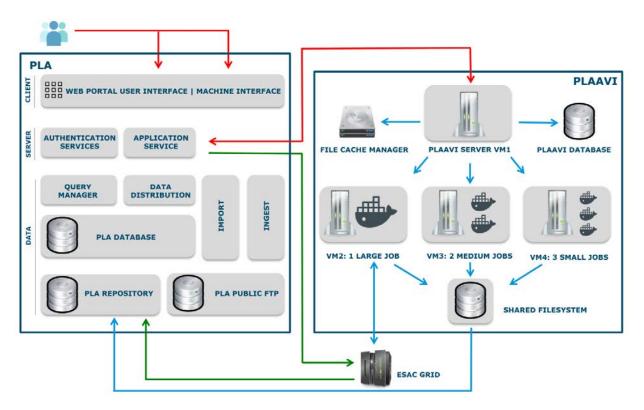
SCI related activities and systems: GAVIP





SCI related activities and systems: PLAAVI





SCI related activities and systems: Jupyter

































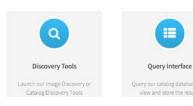


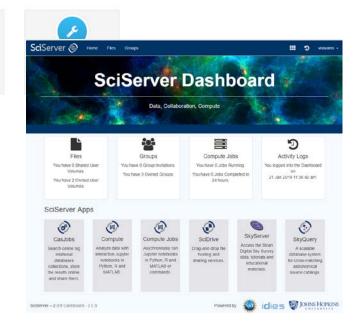
External Related activities and systems



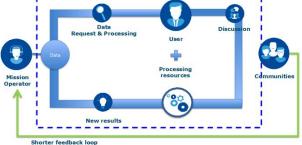
- SciServer
- NOAO Data Labs
- HST on AWS
- Thematic platform from EO
- LSST-Science Platform











EO (Exploitation) Platform

























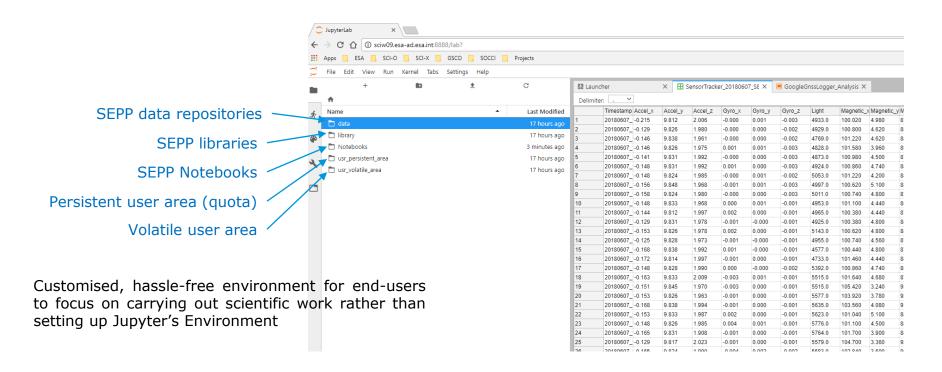




How does it look like?

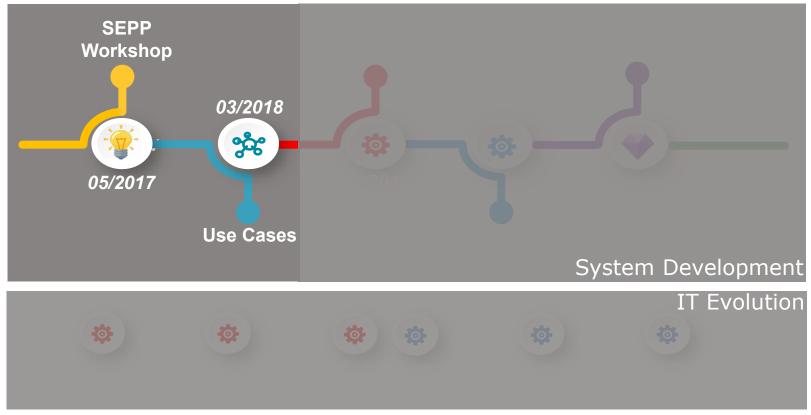


European Space Agency



The history so far





ESA UNCLASSIFIED - For Official Use

ESA | Slide 12

Multi-Mission Use Cases



5. Use Cases

5.1. SEPP-001: Planck Legacy Archive Collaborative Research Area (PLA-CRA)

5.1.1. Actors

- Primary actor: Members of the Planck Collaboration with knowledge of the Planck products that can access the full Planck repository and have privileges to run of system, with access to limited computing and storage resources, but who can request additional resources to the SEPP/ESDC.
- . Secondary actors: the general public, with or without prior knowledge of the Planck products and their complexity, willing to explore and analyze the Coucebi complex pipelines but on a limited range of products and with limited computing and storage resources until they become "trustable" users

5.1.2. Goal

SEPP System Several astronomical communities can benefit from the datasets of CMB experiments like Planck that cover a wide frequency astronomers that are not aware of the wide variety of products produced by the Planck Collaboration, or don't have the ces to handle and analyze these products

With this in mind, the goal is to provide the first collaborative research area for the astronomical community rave Background, millimeter and sub-millimeter science galactic and extragalactic science.

5.1.3. Scope

The Planck Legacy Archive Collaborative Research Area will provide the uses

- Collaborative tools: JupyterLab, R-Studio, Apache Zeppelin,
- defined in · High-level programming languages and interpreters: ctave and GDL (or Matlab and IDL, license permitting). Users can compile their own code.
- · Read-only access to the Planck repository or to
- · Libraries needed to handle Planck produc
- . Private storage areas to host the so
- · Public storage areas to host s
- Command line access ;
- F.A.Q and Example
- to learn how to explore the Planck repository and do simple processing of Planck data, modifying and

the users (VOspace, Dropbox, esabox, Google Drive ...)

sualization tools available

within the tools running in the environment (see ESASky, TOPCAT, Aladin ...) ort to the community for questions, doubts ... This should not require authentication.

5.1.4. Preconditions

- CAS to discriminate users and access rights to data (except for fully public functionality like the FAQ entry).
- 2. Ability to instantiate containers with the desired collaborative tool
- 3. Filesystem with access to Planck data
- 4. Code interpreters to read and execute scripts and notebooks
- 5. All the necessary libraries needed to manipulate Planck products
- 6. Filesystem where the scripts, code and results of the analysis can be saved.
- 7. Scratch area for intermediate processing of the products
- 8. Specialized visualization software not part of the collaborative tools like Aladin, Topcat, etc.

ESA | Slide 13







ESA UNCLASSIFIED - For Officia































s maps in Healpix format (cfitsio, Healpix) in various programming languages (Fortran, C, C++, Python and IDL).

ld be shared among a restricted group of people or the general public.

ne local virtual machine and, eventually, submit jobs to the cluster.







Early adopters, followers, observers





Early Adopters:

- Ready to adopt a SEPP scenario as soon as it is available
- First use cases to be implemented
- Planck, GSSC, JWST



Followers:

 Thought to adopt a SEPP scenario once positive feedback from Early Adopters is available



Observers:

 The need for a specific SEPP scenario is not clear today or has not explicitly been mentioned





ESA UNCLASSIFIED - For Official Use

ESA | Slide 14

SEPP scenarios and adoption



			Storage Area (Augitate) Web instantiation of legacy systems Available in 2019Q4 Available in 2020Q2 Availa						
			Storage Area	Interactive Analysis (Jupiter)	Web instantiation of legacy systems	Execution of project/user pipelines	Pipeline development framework	Interoperabiliity with VO	Store
			Available in 2019Q4	Available in 2019Q4	Available in 2019Q4	Available in 2020Q2	Available in 2020Q2	Available in 29	2020Q4
SEPP Use Cases	PLA-CRA	Planck Legacy Archive Collaborative Research Area	Early adopter	Early adopter	Observer	Follower	Follower	reculi	Observer
	BPC-IPSE	BepColombo Instrument Pipelines Scheduling & Execution	Follower	Follower	Follower	Follower	Follows		Observer
	EUC-CRA	Euclid Collaborative Research Area	Follower	Observer	Follower	Follower	proje	Early adopter	Observer
	XMM-SPCA	XMM-Newton legacy Science Processing Capability Area	Follower	Observer	Follower	Follows	O'	Observer	Observer
	GSSC-CRA	GSSC Collaborative Research Area	Early adopter	Early adopter	Observer	119HO	Early adopter	Observer	Observer
	PLT-PFIA	PLATO SOC Pipeline Framework Integration Area	Follower	Follower	Obser	O	Follower	Observer	Observer
	EXO-MARA	ESA Exoplanetary Mission Archive Research Area	Follower	Follower	nge	Observer	Observer	Observer	Observer
	JWST-WS	JWST Workspaces	Early adopter	Early	arver	Early adopter	Early adopter	Follower	Observer
	GAIA-IDE	Gaia Interactive Data Exploration	Follower	Way	Observer	Observer	Observer	Follower	Observer
	GAIA-SOP	Gaia python Script Offline Processing	Follow	arver	Observer	Follower	Follower	Follower	Observer
	GAIA- SVOP	Gaia Scientific Validation Offlip Process	on le	Observer	Observer	Follower	Follower	Follower	Observer
	Legacy missions- OTFI	Loos	erver	Observer	Early adopter	Observer	Observer	Observer	Observer
	ESDC-CRL	Cab	Follower	Follower	Observer	Follower	Follower	Follower	Observer
	INT-OSAP	Integral Offline Science Analysis Preservation	Observer	Observer	Observer	Observer	Observer	Observer	Observer

































Top Level scenarios



Workspace





user personalized storage (scratch, persistent and public mounted areas) and execution environment. It allows users to publish and share their assets.







frameworks to manipulate and analyze data interactively. Creation and sharing of documents that contain live code, visualization, etc

Pipeline





pipelines for data integration, transformation and analytics based on processing assets developed by SEPP actors





provides access to the platform's interface available to external systems and gathers all external dependencies integrated in the platform (example, VO protocols)

























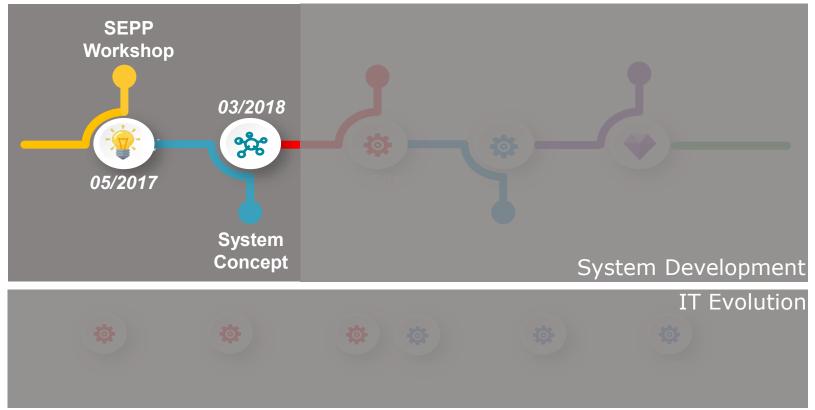






From now on



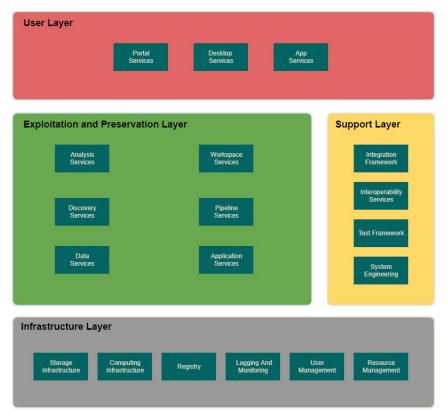


Reference Architecture



Reference Architecture is organised into layers resembling an N-tier architecture.

Layers designed to allow mission extensions of their System Domains.



































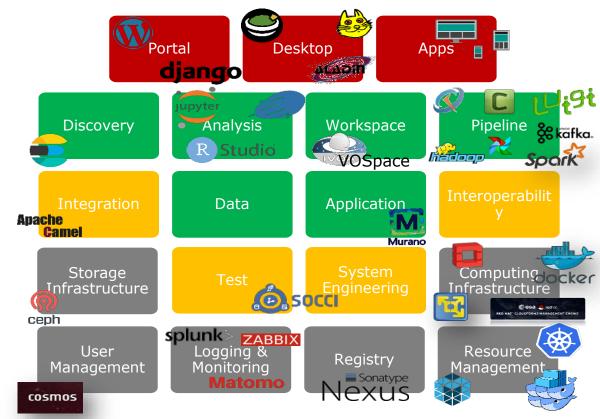
Technology Stack



PaaS and SaaS environments leveraging on mainstream technologies:

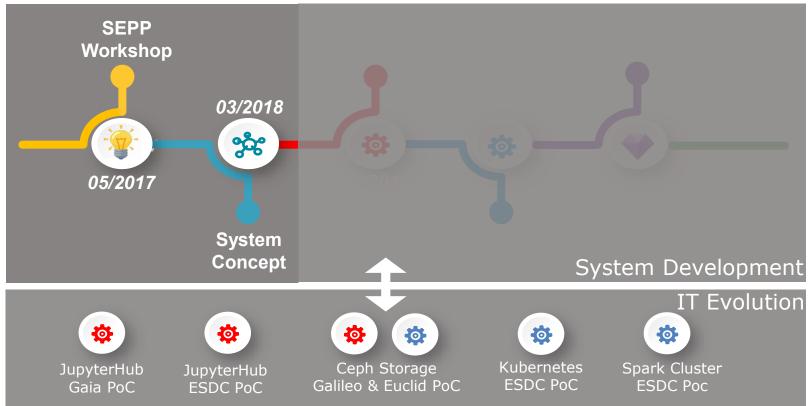
- Big data
- Cloud
- Virtualisation
- Containers

...



From now on























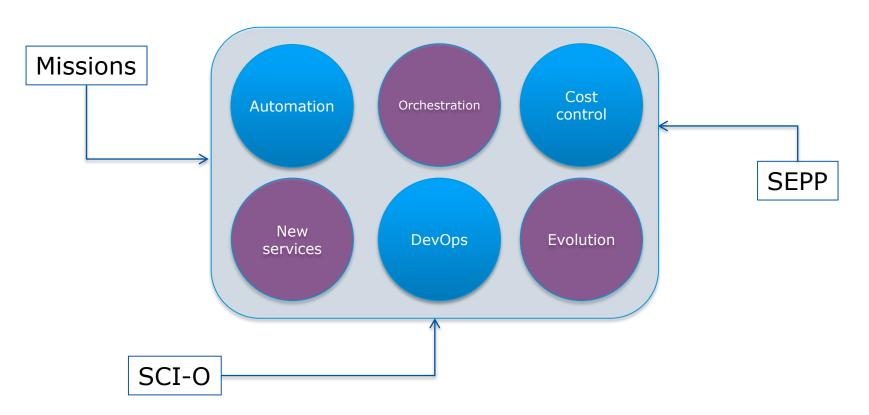






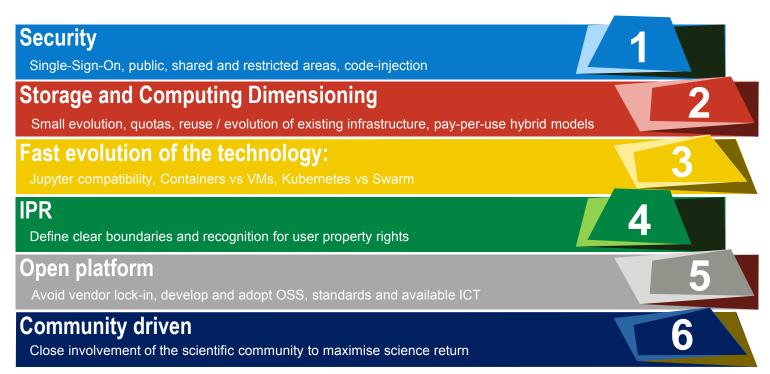
SCI-O IT infrastructure and SEPP





Challenges







Stakeholder Engagement



Meetings with stakeholders:

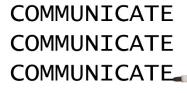
- Find the right periodicity: every 2 months?
- Keeping stakeholders aware of progress
- Capture stakeholders' feedback
- Individual meetings possible too

Platform demos:

- Available after the implementation of the first scenarios
- Mission oriented
- Available on-demand

SEPP project portal:

- Roadmap visibility
- Project documentation visibility
- Public portal





Reporting:

- At the stakeholders' meetings
- Executive reports on quarterly basis
- Periodic reporting through the SEPP project portal























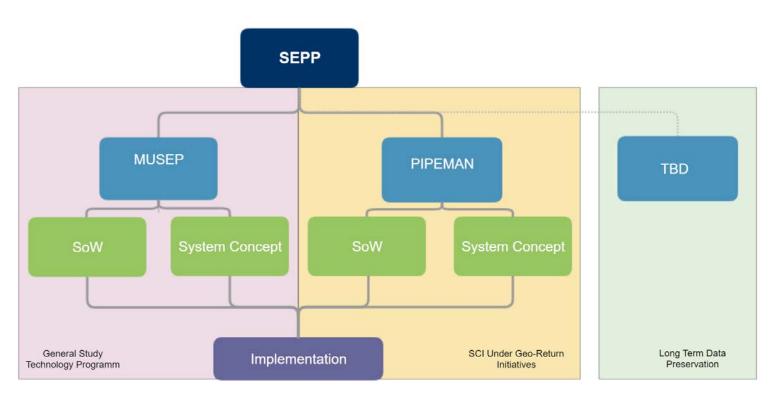






ESA Development Framework





























ESA Science Archives Long Term Strategy



SEPP-013: ESDC Collaborative Research Lab

Primary actors: astronomers, planetary scientists and heliophysicists with knowledge of one or more of the data products in the ESDC science archives and wanting to data mine, visualise and/or analyse data from one or a combination of missions.

Secondary actors: members of the general public, educators, students, etc. wanting to explore and analyse archival data in the ESDC Science archives.

Efficient long-term preservation of data, software and knowledge





Enabling maximum scientific exploitation of datasets

























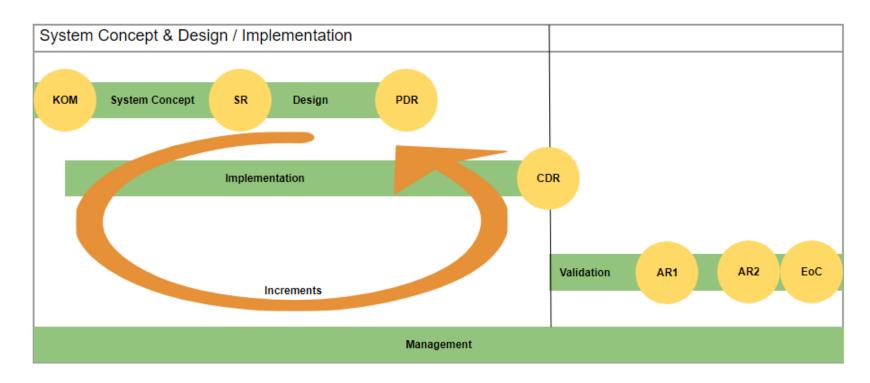






Projects Work Logic

























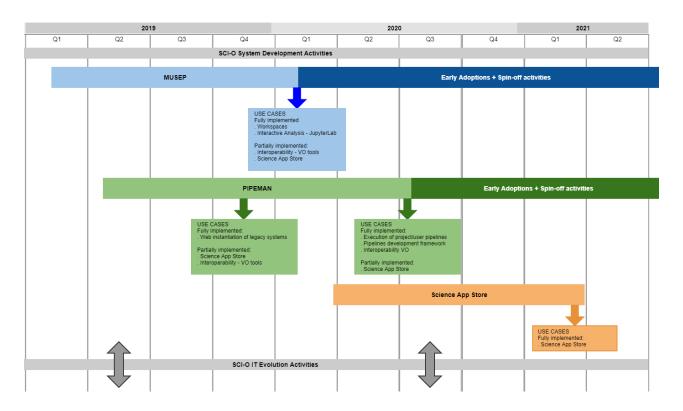






Next steps

























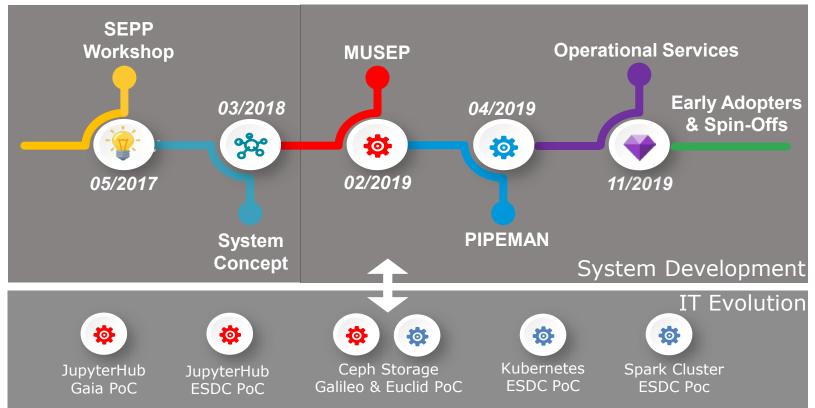






Roadmap summary





Conclusions



- SEPP's concept is to move the user to the data and tools
- SEPP covers both exploitation and preservation use cases
- SEPP uses knowledge from precursor activities in SCI and from external similar platforms
- SCI-O IT provides solutions to the missions and looks for synergies and commonalities with SEPP
- The implementation of scenarios and use cases will follow a **incremental approach**
- Stakeholder engagement is essential for the development of a successful platform



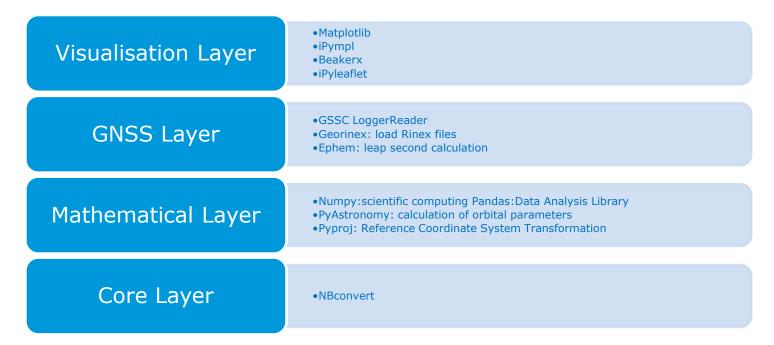
2// 014CF/22111FD 1 01 0111Clat 02C

LUM | Ullue Zu



GSSC Jupyter Notebook - Configuration

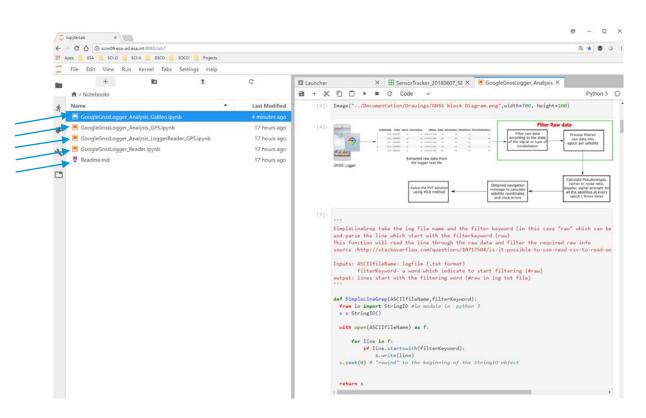




GSSC Jupyter Notebook - Repository



GSSC Notebooks currently available































GSSC Jupyter Notebook – Science Case



