Use Cases for the ESAC Science Exploitation and Preservation Platform (SEPP)

Christophe Arviset, Vicente Navarro, Ruben Alvarez
ESA Operations Department, ESAC, Spain

BIDS 2019, Munich, 21/02/2019

@carviset

→ COSMIC OBSERVERS

esa

CONCEPTS

IN DEVELOPMENT

OPERATIONAL

LEGACY


spica


theseus


webb
(2021)


ariel
(2028)


euclid
(2022)


cheops
(2019)


plato
(2026)


xrism
(2021)


einstein probe
(2022)


athena
(2031)


lisa
(2034)


hubble
(1990–)


gaia
(2013–)


xmm-newton
(1999–)


integral
(2002–)

microwaves

sub-millimetre

infrared

optical

ultraviolet

x-rays

gamma rays

gravitational waves


planck
(2009–2013)


herschel
(2009–2013)


iso
(1995–1998)


akari
(2006–2011)


hipparcos
(1989–1993)


corot
(2006–2014)


iue
(1978–1996)


exosat
(1983–1986)


hitomi
(2016)


suzaku
(2005–2015)


cos-b
(1975–1982)

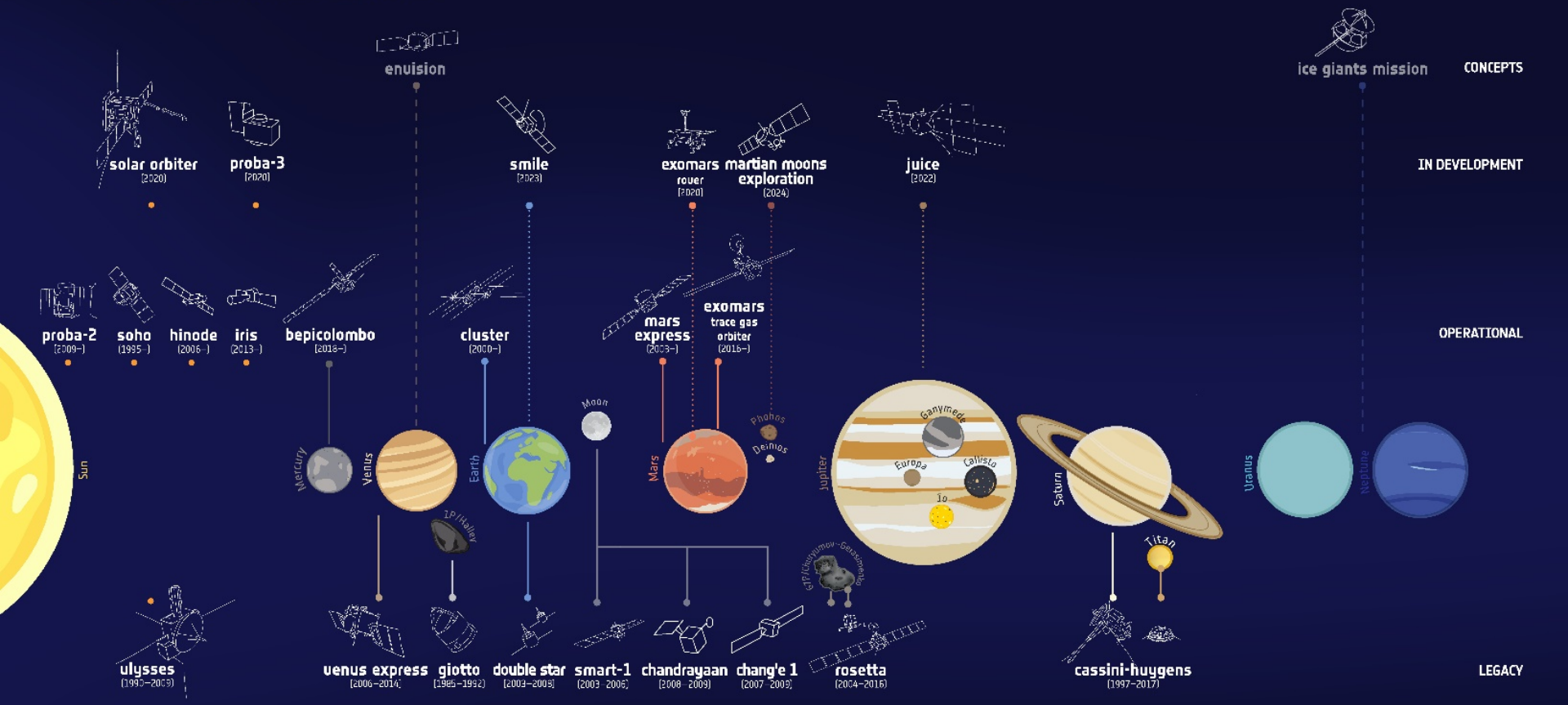

lisa pathfinder
(2015–2017)


microscope
(2016–2018)

#Space19plus

Space19

→ SOLAR SYSTEM EXPLORERS

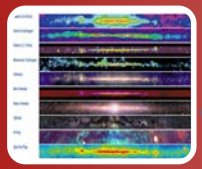


#Space19plus

Space19

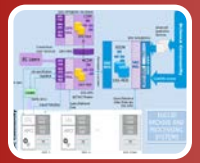
ESA Space Science Data Archives

archives.esac.esa.int



Enable maximum ***scientific exploitation*** of data sets

Enable efficient ***long-term preservation*** of data, softw… knowledge, using moder…



Enabl… …ve archive production by ***in… …n in, and across, proje…***

Science Exploitation and Preservation Platform

ESAC SCIENCE DATA CENTRE

Astronomy Science Archives

- esasky
- exosat
- gaia
- herschel
- hubble space telescope
- iso
- lisa pathfinder
- planck
- xmm-newton

Heliophysics Science Archives

- cluster
- double star
- ISS-SolACES*
- proba-2
- soho
- ulysses

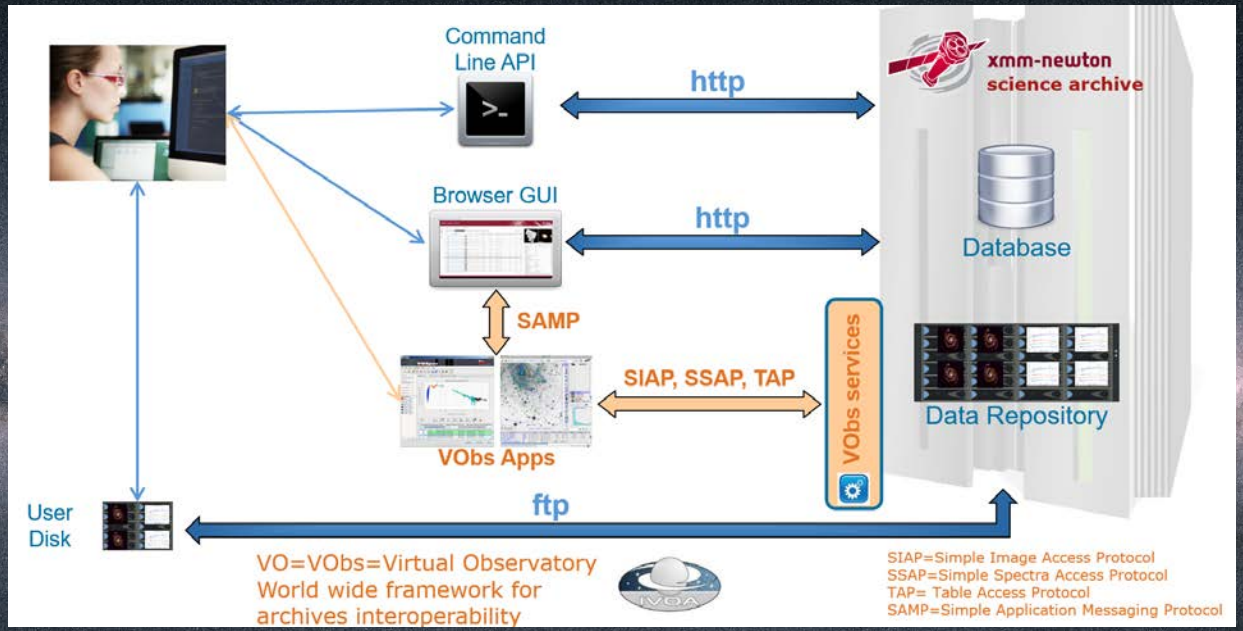
The Planetary Science Archive

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

Future Archives

- euclid
- juice
- solar orbiter

Background and context



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

→ HOW MANY STARS WILL THERE BE IN THE SECOND GAIA DATA RELEASE?

esa

gaia

position & brightness on the sky

1 692 919 135

surface temperature

161 497 595

red colour

1 383 551 713

blue colour

1 381 964 755

parallax and proper motion

1 331 909 727

radius & luminosity

76 956 778

amount of dust along the line of sight

87 733 672

radial velocity

7 224 631

14 099
Solar System objects

550 737
variable sources

www.esa.int

The second data release of ESA's Gaia mission is scheduled for publication on 25 April 2018.

European Space Agency

New Archive usage paradigm

Data exploitation from science community of mission data brings challenges for us

Execution of user's code at server side

Move code to the data

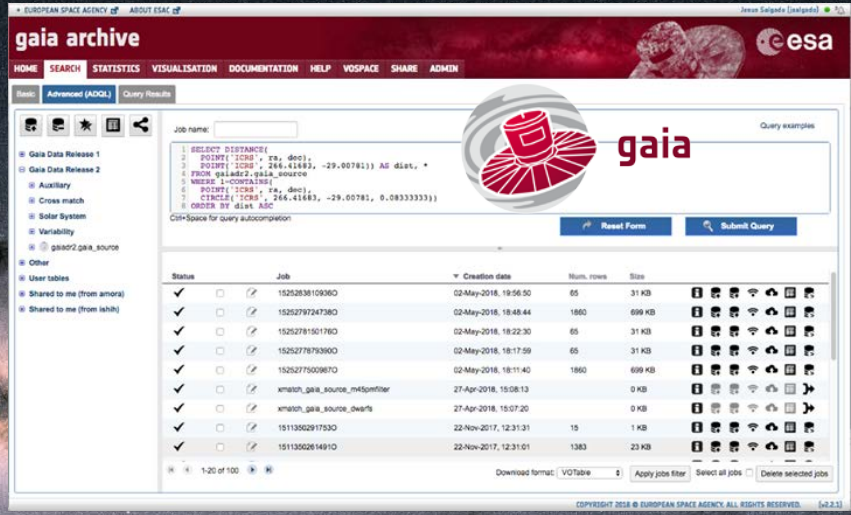
Sharing data (collaborative)

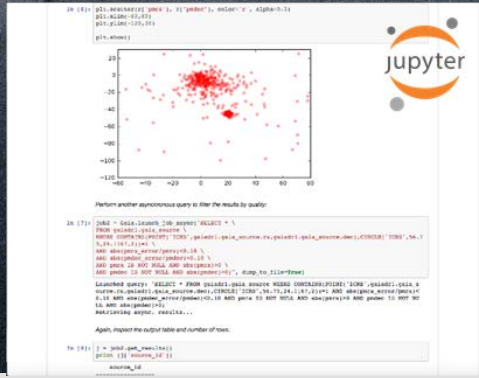
State of the art technologies of big data mining

Visualization techniques

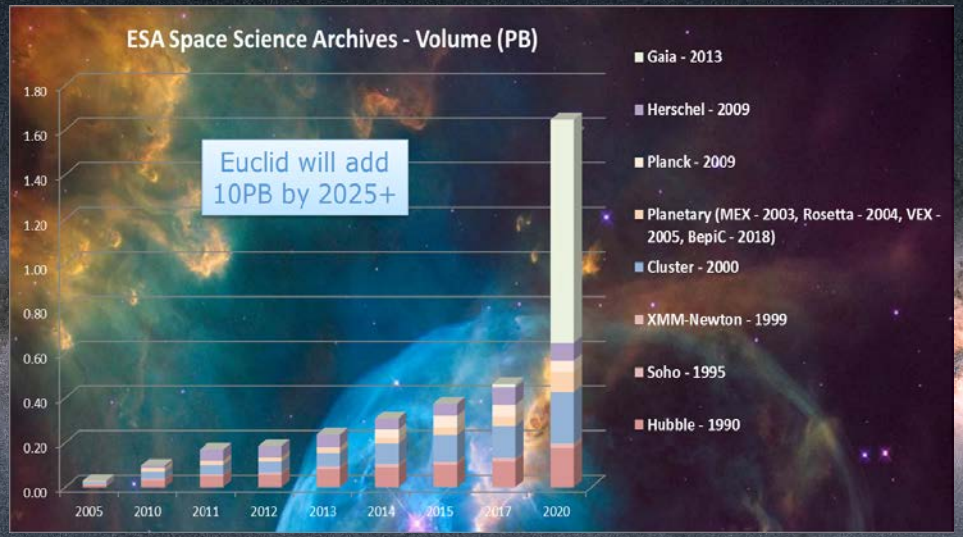
For the community:

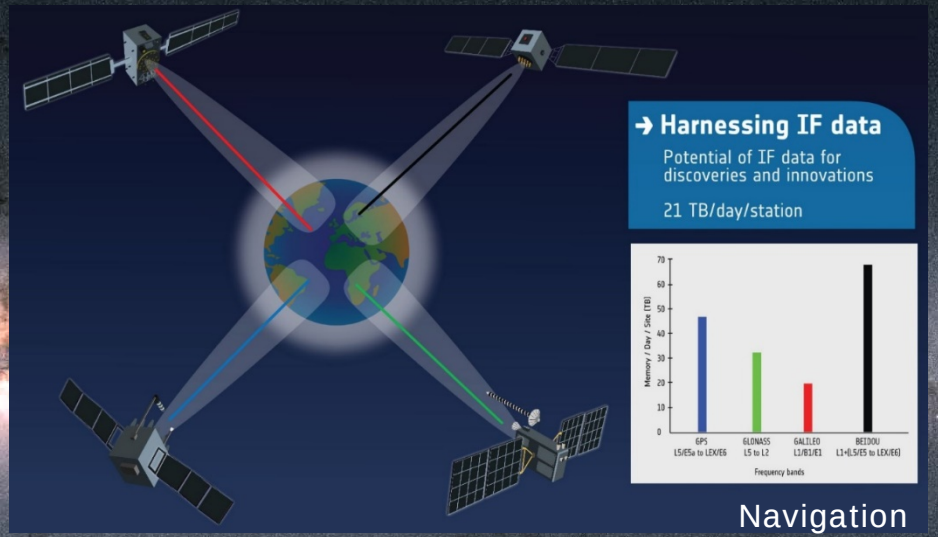
New way of working for scientists





Background and context





The fundamental principle of **SEPP's concept** is to move the user to the data and tools.

Users access a **science platform** providing the data, tools, and resources required, as opposed to downloading, replicating, and exploiting data 'at home'.

Top Level scenarios

Data Exploitation

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

Collaborative Research env.

VOSpace

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

Pipeline Develop. env.

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

Software Preservation

provides on-the-fly instantiation to legacy software, through full environment or predefined processing threads

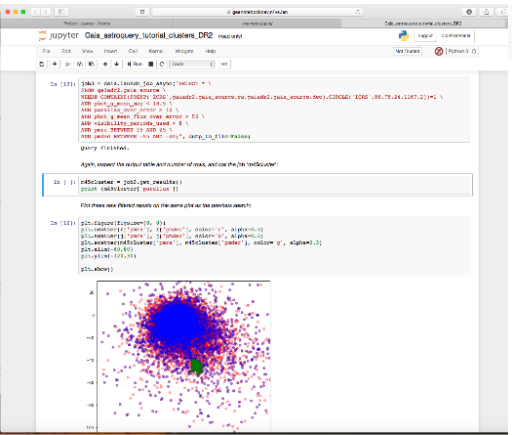
Data Exploitation

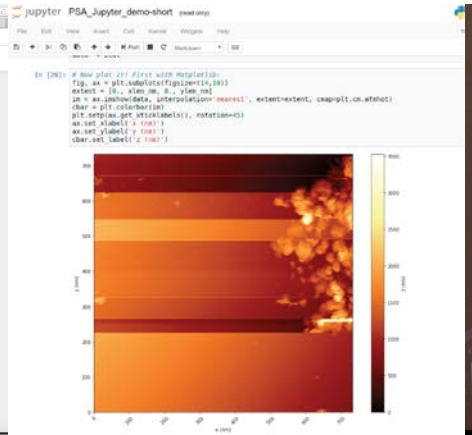
Data Exploitation

Enable data processing where the data is, ie the Archives

Provide generic pipeline development and testing environment

Interactive data analysis Jupyter Notebook / Hub





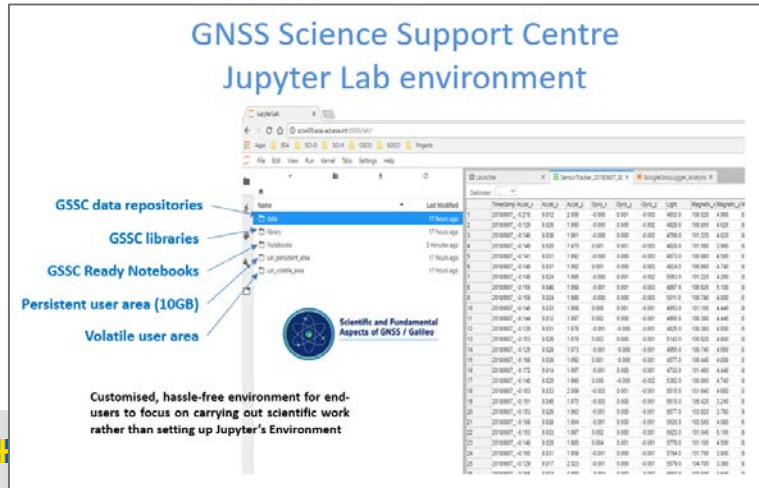
Collaborative Research Environment

Collaborative Research env.
VOSpace

Share your data (user storage space in the platform, VOSpace)

Share your metadata (DB user space inside the archive, …)

Publish your data through standard data protocols (ie VO)



Data Pipeline

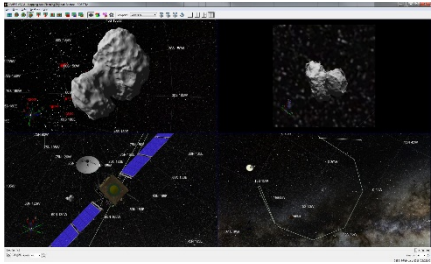
Pipeline Develop. env.

Luigi kafka

Develop your own customized processing pipelines

Share your code through a *Science App Store*

Legacy Software Preservation

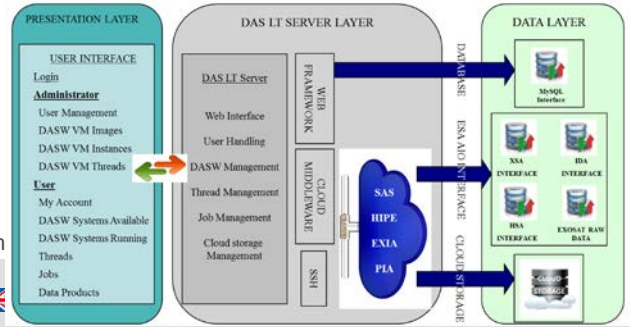


Software from Legacy Missions

- Mission Planning System (ie Rosetta on the 67P comet)
- Could be re-instantiated in years for future missions (ie JUICE)

Data Processing and Analysis Software

- Enable easier access to legacy software
  - On-the-fly instantiation of full system / predefined processing threads
- Reproducibility of data processing in the future (FAI**R** principle)
- Rescue the code and bring it to the (small) data



Early adopters, followers, observers

Early Adopters:

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

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

SEPP scenarios and adoption

			SEPP Scenarios						
			Storage Area Available in 2019Q4	Interactive Analysis (Jupiter) Available in 2019Q4	Web instantiation of legacy systems Available in 2019Q4	Execution of project/user pipelines Available in 2020Q2	Pipeline development framework Available in 2020Q2	Interoperability with VO Available in 2020Q2	Science App Store Available in 2020Q4
SEPP Use Cases	PLA-CRA	Planck Legacy Archive Collaborative Research Area	Early adopter	Early adopter	Observer	Follower	Follower	Early adopter	Observer
	BPC-IPSE	BepColombo Instrument Pipelines Scheduling & Execution	Follower	Follower	Follower	Follower	Follower	Observer	Observer
	EUC-CRA	Euclid Collaborative Research Area	Follower	Observer	Follower	Follower	Follower	Early adopter	Observer
	XMM-SPCA	XMM-Newton legacy Science Processing Capability Area	Follower	Observer	Follower	Follower	Follower	Observer	Observer
	GSSC-CRA	GSSC Collaborative Research Area	Early adopter	Early adopter	Observer	Early adopter	Early adopter	Observer	Observer
	PLT-PFIA	PLATO SOC Pipeline Framework Integration Area	Follower	Follower	Observer	Follower	Follower	Observer	Observer
	EXO-MARA	ESA Exoplanetary Mission Archive Research Area	Follower	Follower	Observer	Observer	Observer	Observer	Observer
	JWST-WS	JWST Workspaces	Follower	Follower	Observer	Follower	Observer	Follower	Observer
	GAIA-IDE	Gaia Interactive Data Exploration	Early adopter	Early adopter	Observer	Observer	Observer	Early adopter	Observer
	GAIA-SOP	Gaia python Script Offline Processing	Early adopter	Early adopter	Observer	Follower	Follower	Early adopter	Observer
	GAIA-SVOP	Gaia Scientific Validation Offline Processing	Early adopter	Early adopter	Observer	Follower	Follower	Early adopter	Observer
	Legacy missions-OTFI	Legacy Missions On The Fly Instantiation	Observer	Observer	Early adopter	Observer	Observer	Observer	Observer
	ESDC-CRL	ESDC Collaborative Research Lab	Follower	Follower	Observer	Follower	Follower	Follower	Observer
	INT-OSAP	Integral Offline Science Analysis Preservation	Early Adopter	Early Adopter	Early Adopter	Early Adopter	Early Adopter	Early Adopter	Early Adopter

Top Level scenarios

Data Exploitation

jupyter

RStudio

Scientific and Fundamental Aspects of GNSS / Galileo

euclid

hubble space telescope

planck

esasky

gaia

Collaborative Research env.

IVOA

VOSpace

S3

euclid

jwst

gaia

planck

Scientific and Fundamental Aspects of GNSS / Galileo

Pipeline Develop. env.

Luigi

kafka

bepicolombo

euclid

Scientific and Fundamental Aspects of GNSS / Galileo

Software Preservation

docker

exosat

integral

herschel

rosetta

lisa pathfinder

xmm-newton

iso

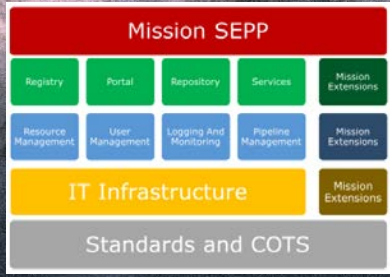
Conclusion

New missions call for a paradigm shift for ***science data exploitation***

- From “bring the data to the user” to “bring the user to the data”
- Closer interaction between archives and data processing services

Legacy missions call for data and software ***long term preservation***

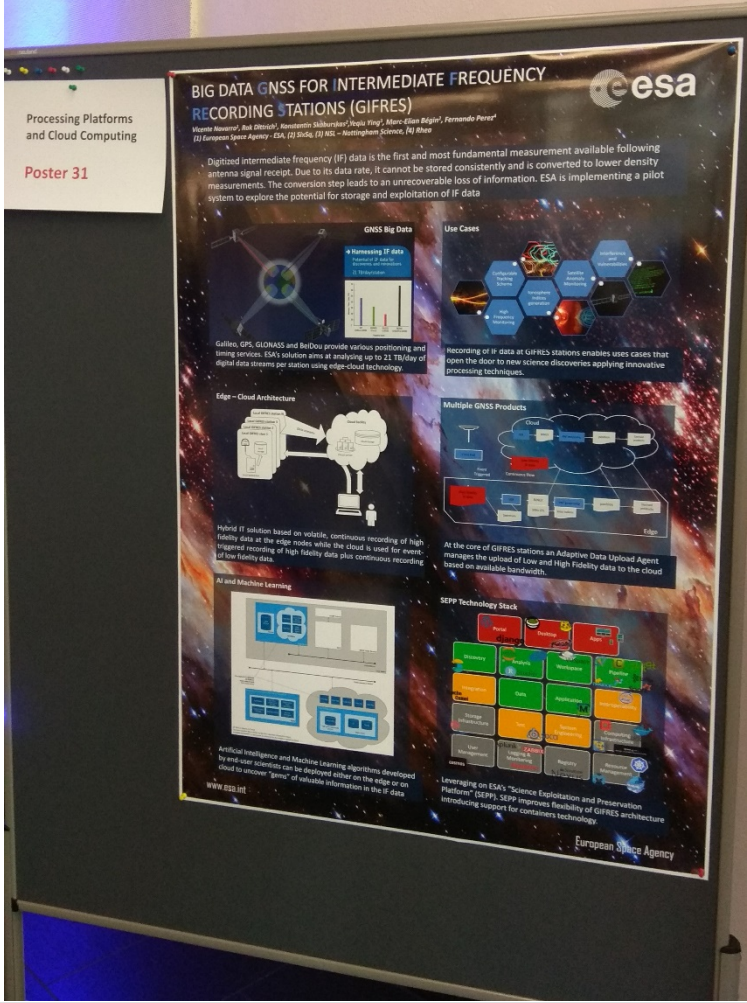
New scientists call for ***collaborative research environment***



Science Exploitation and Preservation Platform

@carviset


Thank You
Asante
Merci
Obrigado
Multumesc
Kiitos
Spasibo
Juspaxar
Arigato
Grazie
Dank Je
Matondo
Vinaka
Welalin
Salamat
Kia Ora
Matur Nuwun
Maake
Chokrane
Mochchakkeram
Terma Kasih
Raibh Maith Agat
Ua Tsaug Rau Koj
Nirringrazzjak
Cam on ban
Dankon
Mamana
Obrigado


Processing Platforms and Cloud Computing
Poster 31
BIG DATA GNSS FOR INTERMEDIATE FREQUENCY RECORDING STATIONS (GIFRES)
esa
Digitized intermediate frequency (IF) data is the first and most fundamental measurement available following antenna signal receipt. Due to its data rate, it cannot be stored consistently and is converted to lower density measurements. The conversion step leads to an unrecoverable loss of information. ESA is implementing a pilot system to explore the potential for storage and exploitation of IF data
GNSS Big Data
Harnessing IF data
Use Cases
Galileo, GPS, GLONASS and BeiDou provide various positioning and timing services. ESA's solution aims at analysing up to 21 TB/day of digital data streams per station using edge-cloud technology
Recording of IF data at GIFRES stations enables uses cases that open the door to new science discoveries applying innovative processing techniques.
Edge – Cloud Architecture
Multiple GNSS Products
Hybrid IT solution based on volatile, continuous recording of high fidelity data at the edge nodes while the cloud is used for event-triggered recording of high fidelity data plus continuous recording of low fidelity data
At the core of GIFRES stations an Adaptive Data Upload Agent manages the upload of Low and High Fidelity data to the cloud based on available bandwidth.
AI and Machine Learning
SEPP Technology Stack
Artificial Intelligence and Machine Learning algorithms developed by end-user scientists can be deployed either on the edge or on cloud to uncover "gems" of valuable information in the IF data
Leveraging on ESA's "Science Exploitation and Preservation Platform" (SEPP). SEPP improves flexibility of GIFRES architecture introducing support for containers technology
www.esa.int
European Space Agency