

Space Science Data Handling at the ESAC Science Data Centre

Dr. Bruno Merín (ESA, Head of the ESAC Science Data Centre, SCI-SAA)

ESA_Lab “Space Data Management Workshop”, Venice, 8th of May of 2023

bruno.merin@esa.int

- Space science data in ESA's Scientific missions
 - (science-driven) data management
 - International standardisation
- AI in space science data management
 - Citizen science projects
 - Virtual assistants
- The future of space science

Space science data management in Europe today

ESA Coordination with Science Provider and User Communities

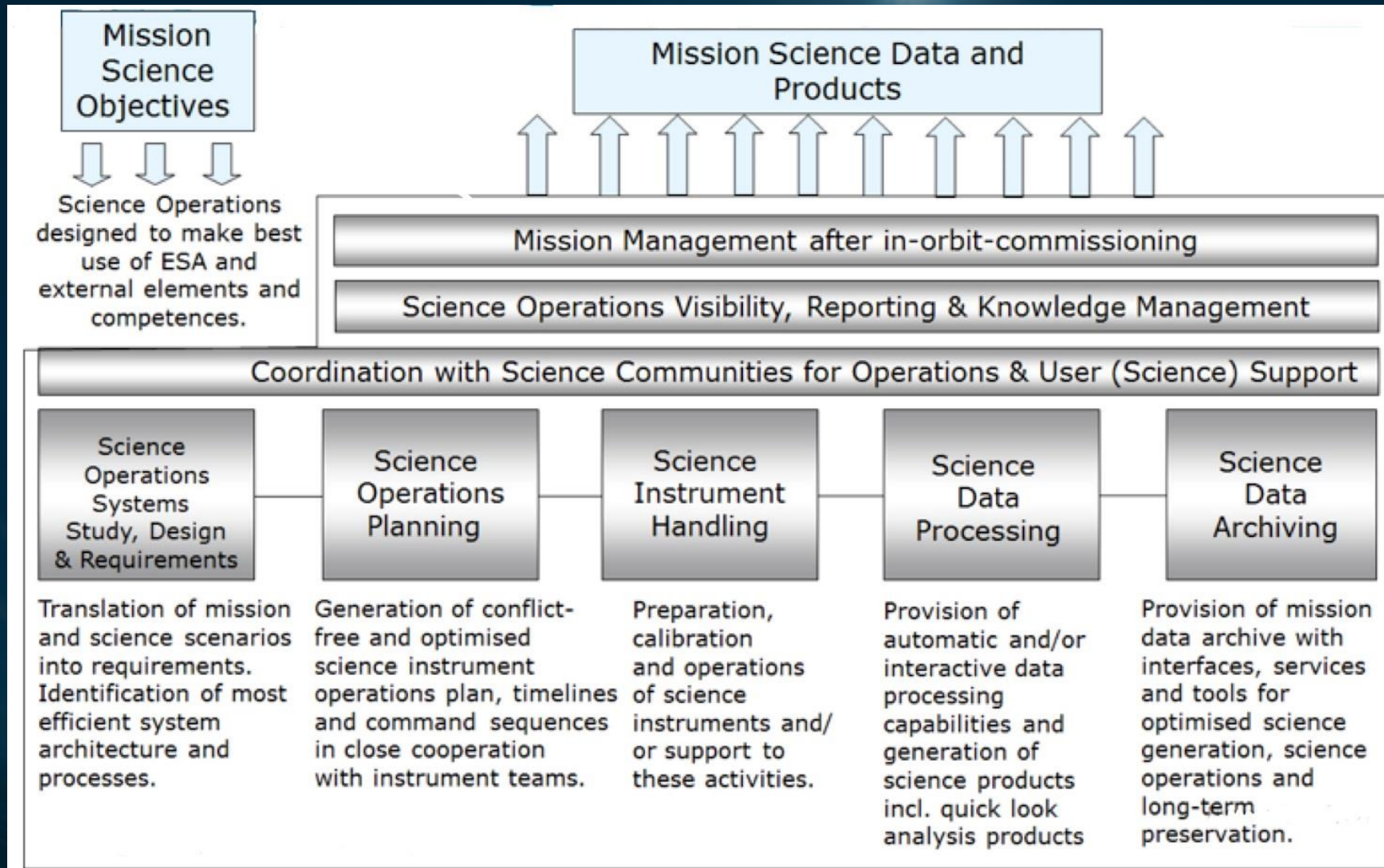


Usually more by ESA

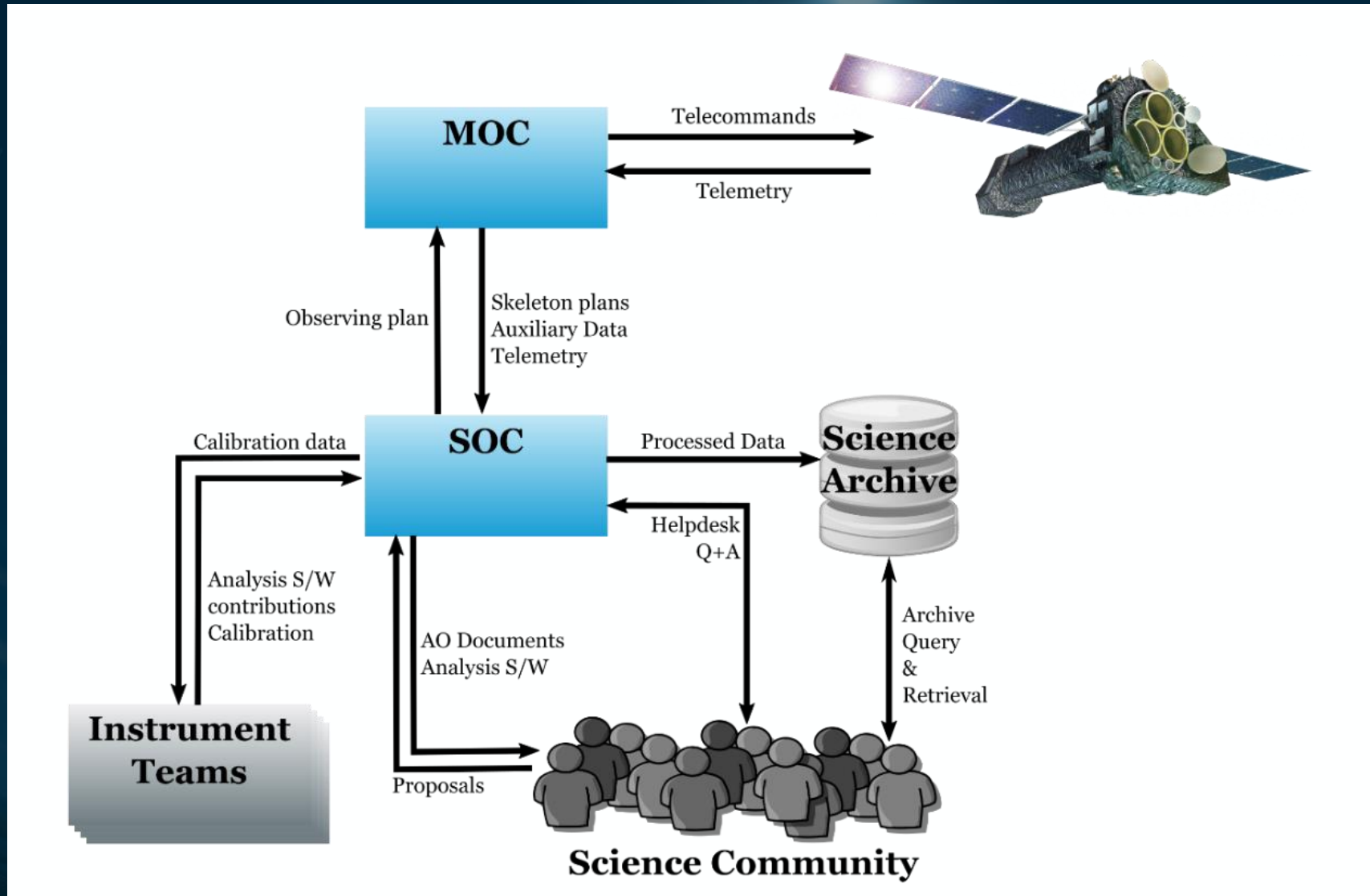
Usually more by Member States

Usually more by ESA

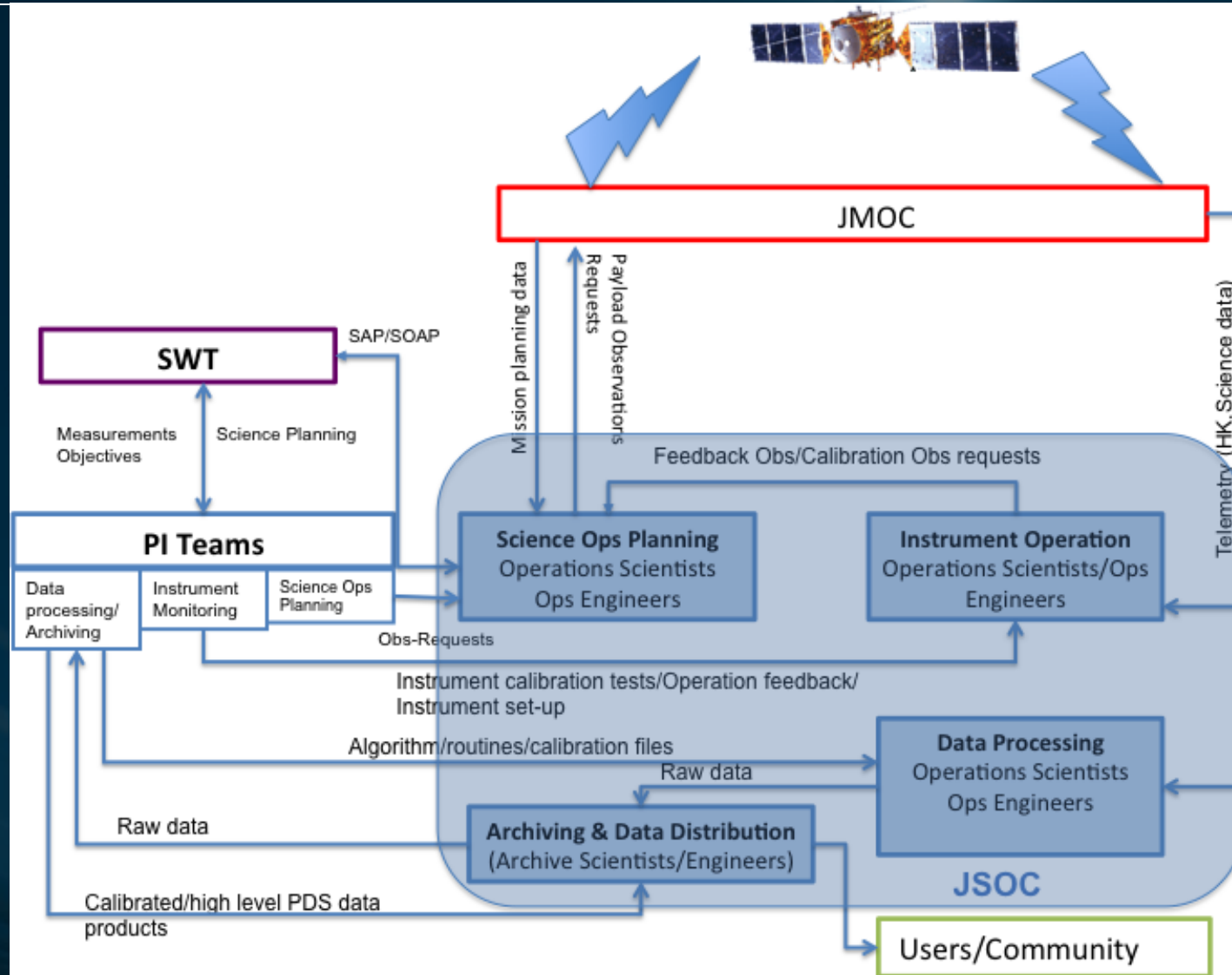




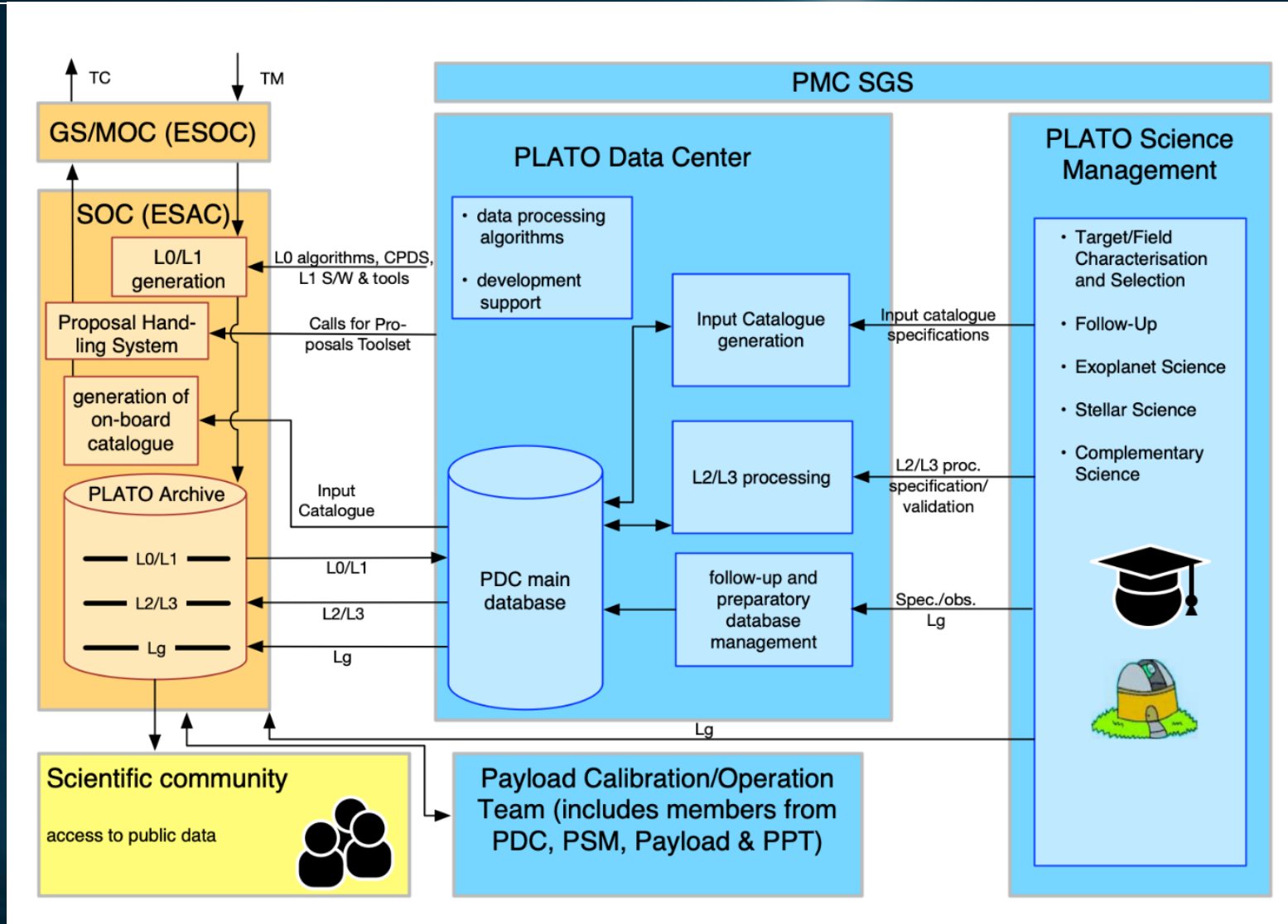
Example of data flow for a space science mission



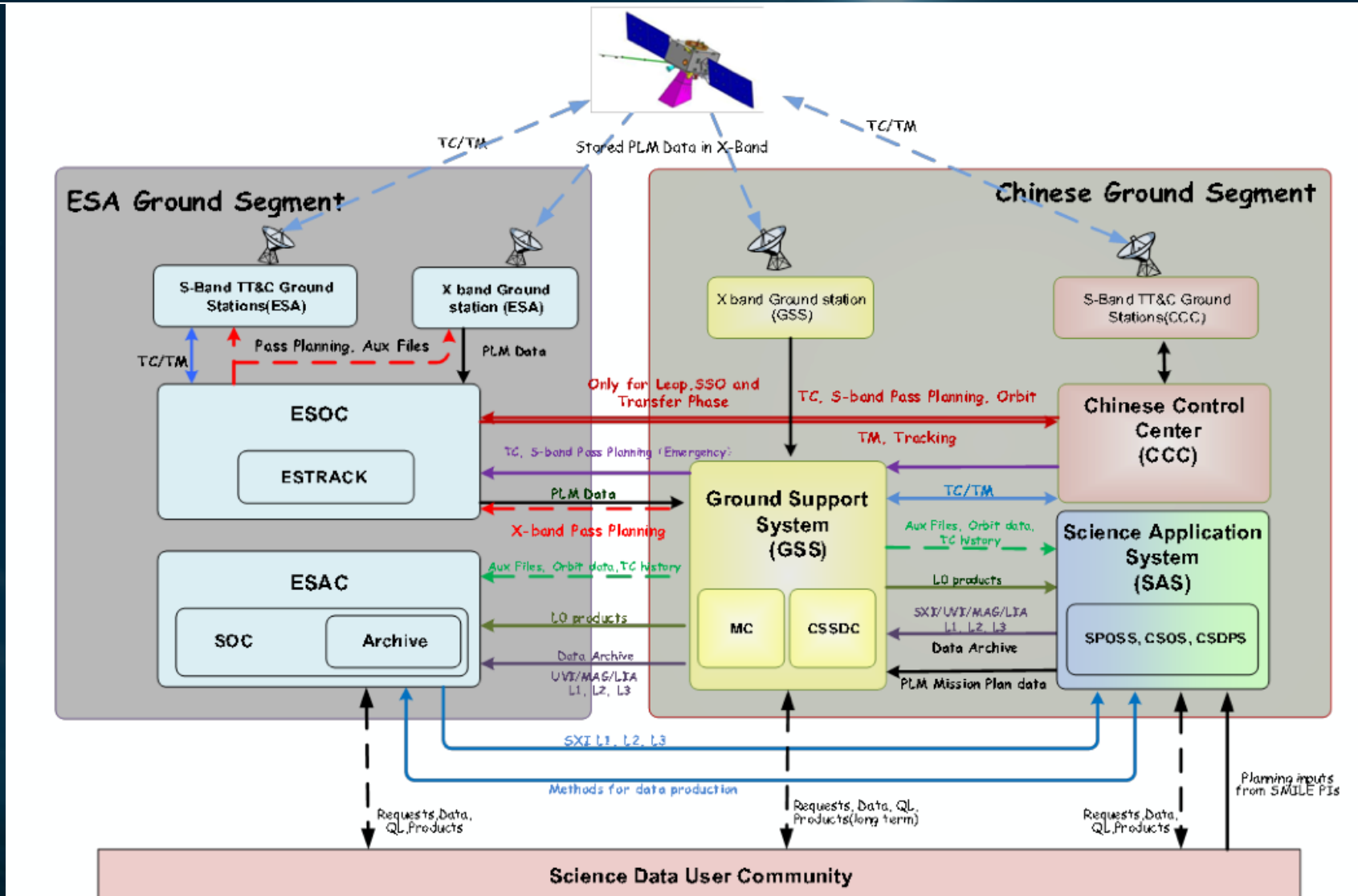
Example data flow for a Planetary mission : JUICE



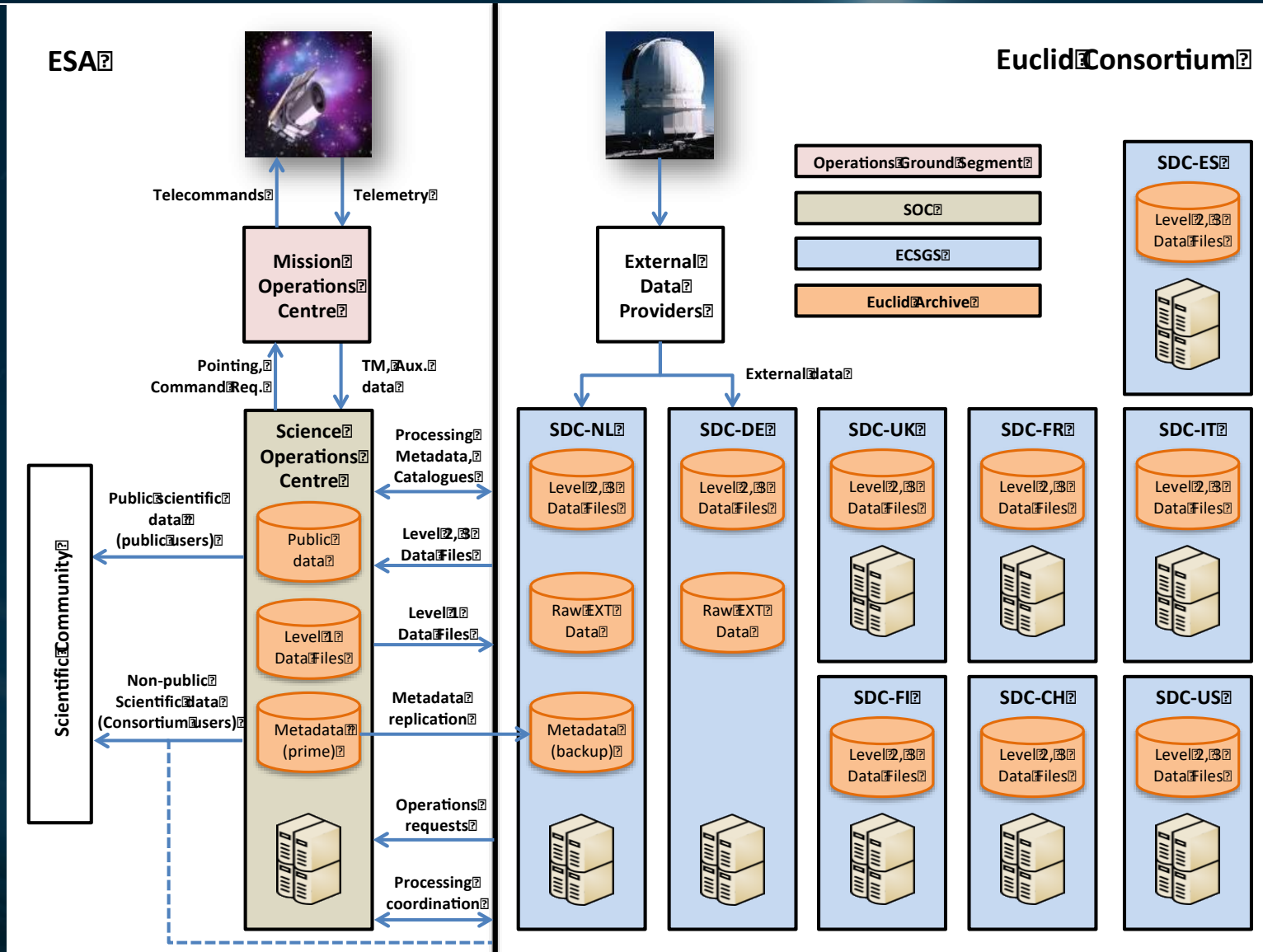
Example data flow for an Astronomy survey : PLATO



Example data flow for a collaborative mission : SMILE



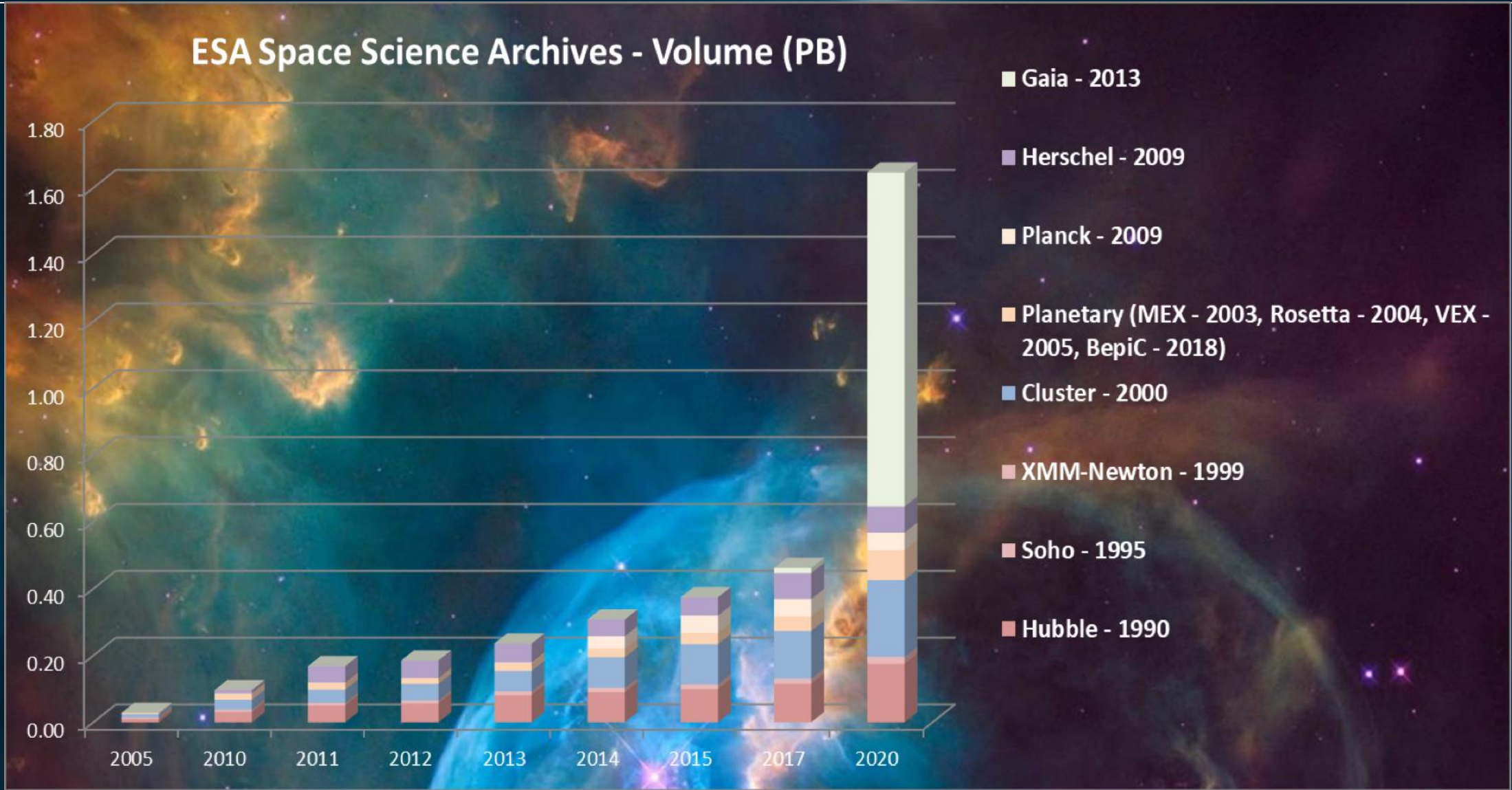
Example data flow for a Astronomy mission : Euclid



Science is about revisiting the data with new knowledge

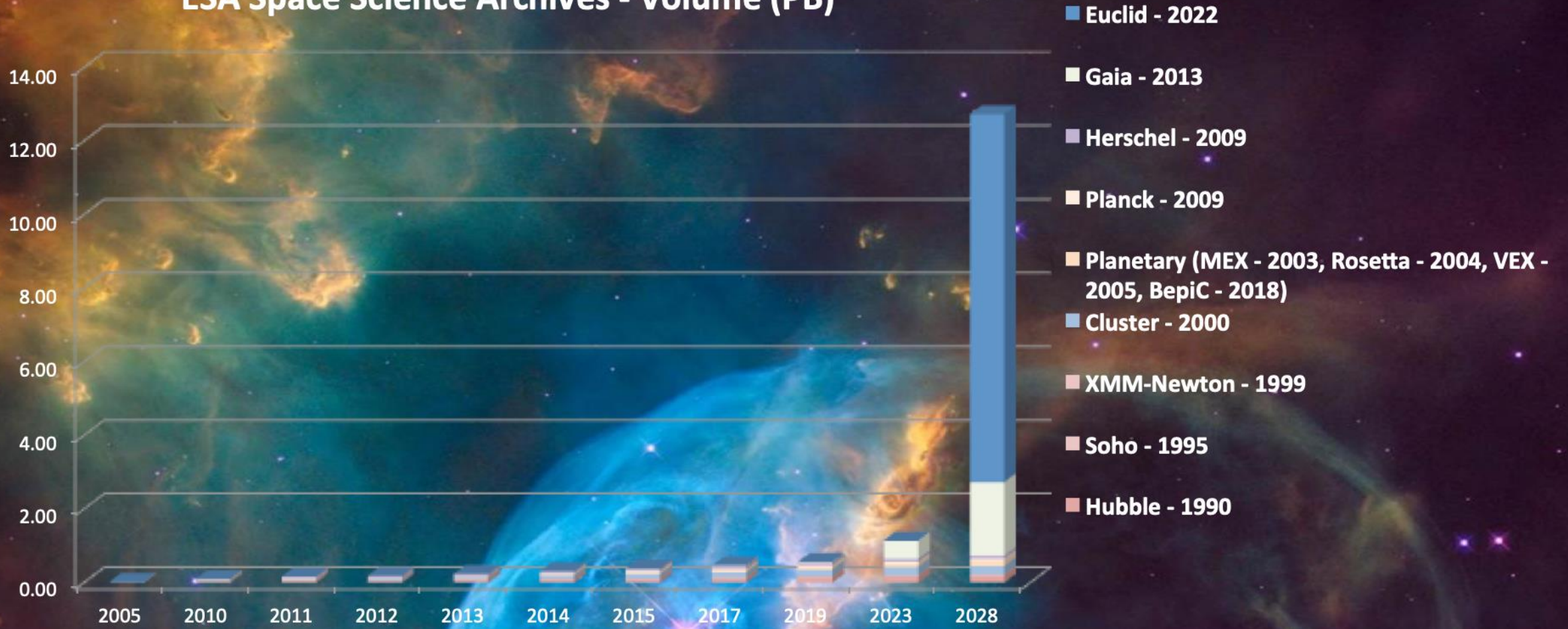
re-search

Complex data products grow in size and complexity

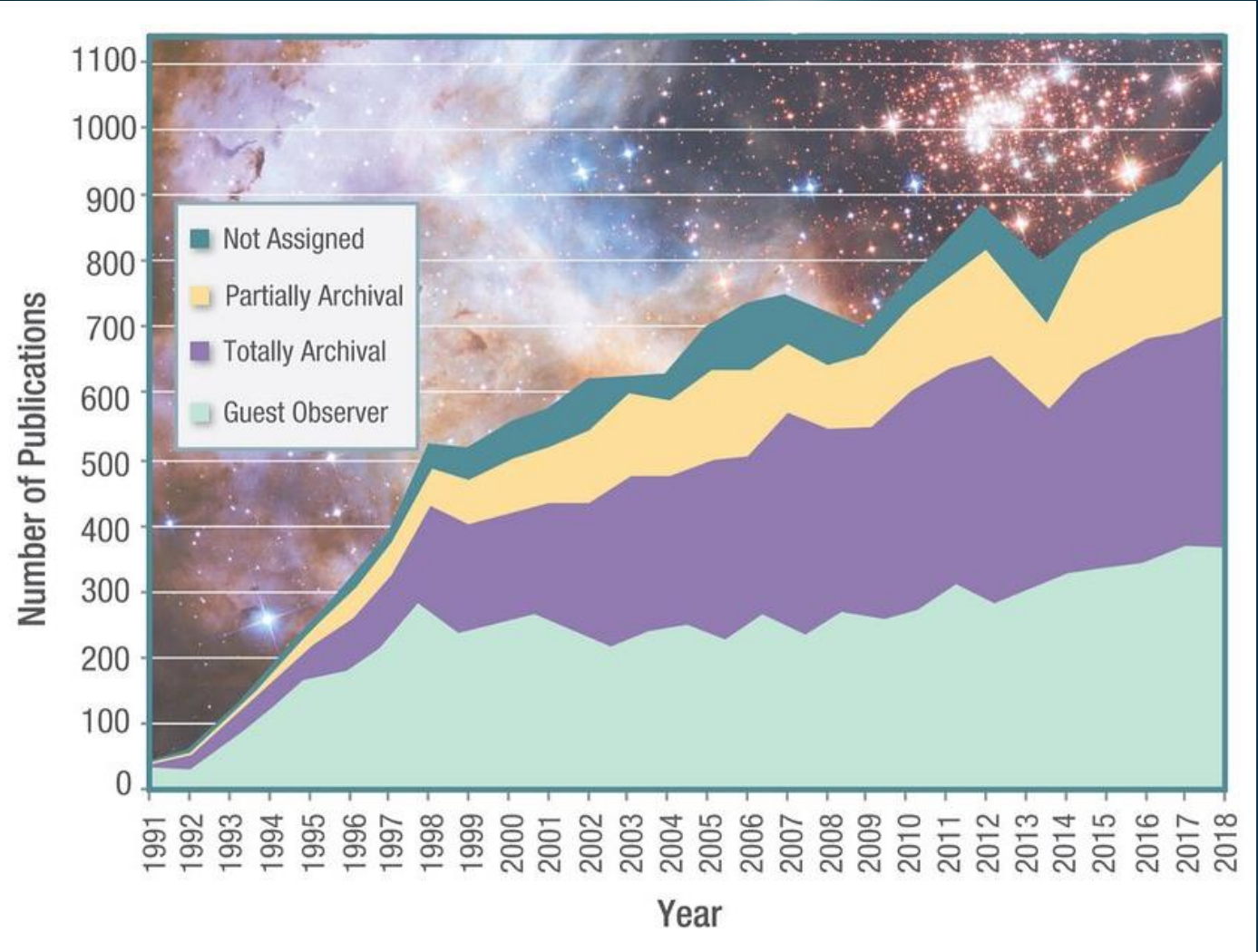


.. And they are getting very large very soon..

ESA Space Science Archives - Volume (PB)



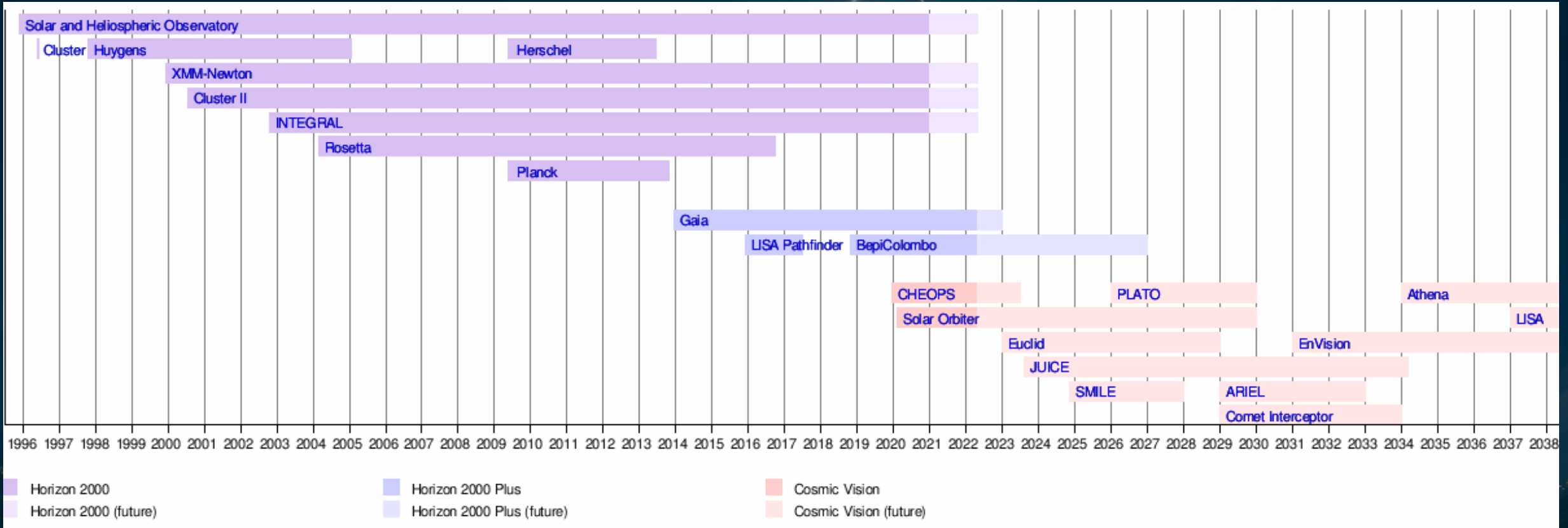
Astronomy is becoming a data-driven field



<http://archive.stsci.edu/hst/bibliography/pubstat.html>



Hosting science data for decades



24 missions developed, operated and curated for over four decades

International collaboration in standardisation



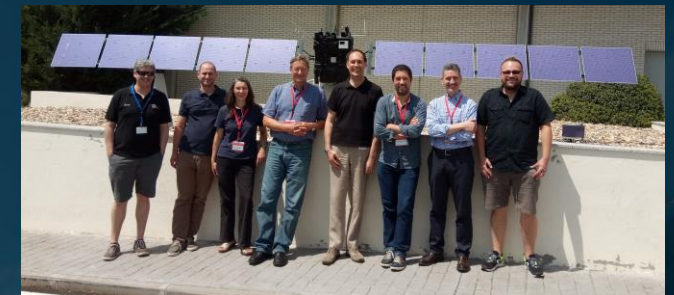
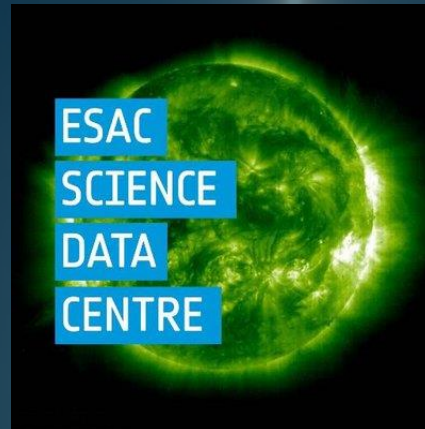
International Planetary Data Alliance



International Heliophysics Data Environment Alliance

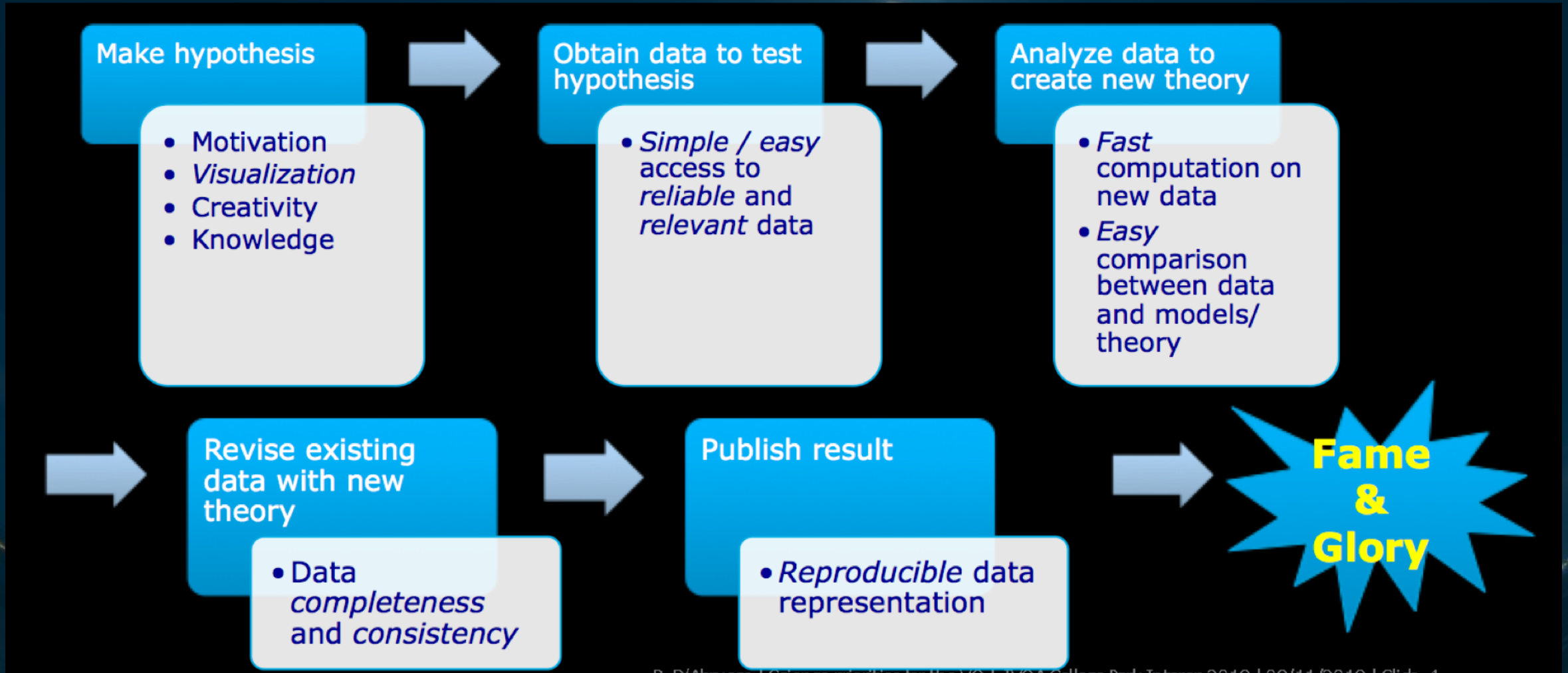


International Virtual Observatory Alliance



Astronomy, Heliophysics and Planetary Archive Users Groups

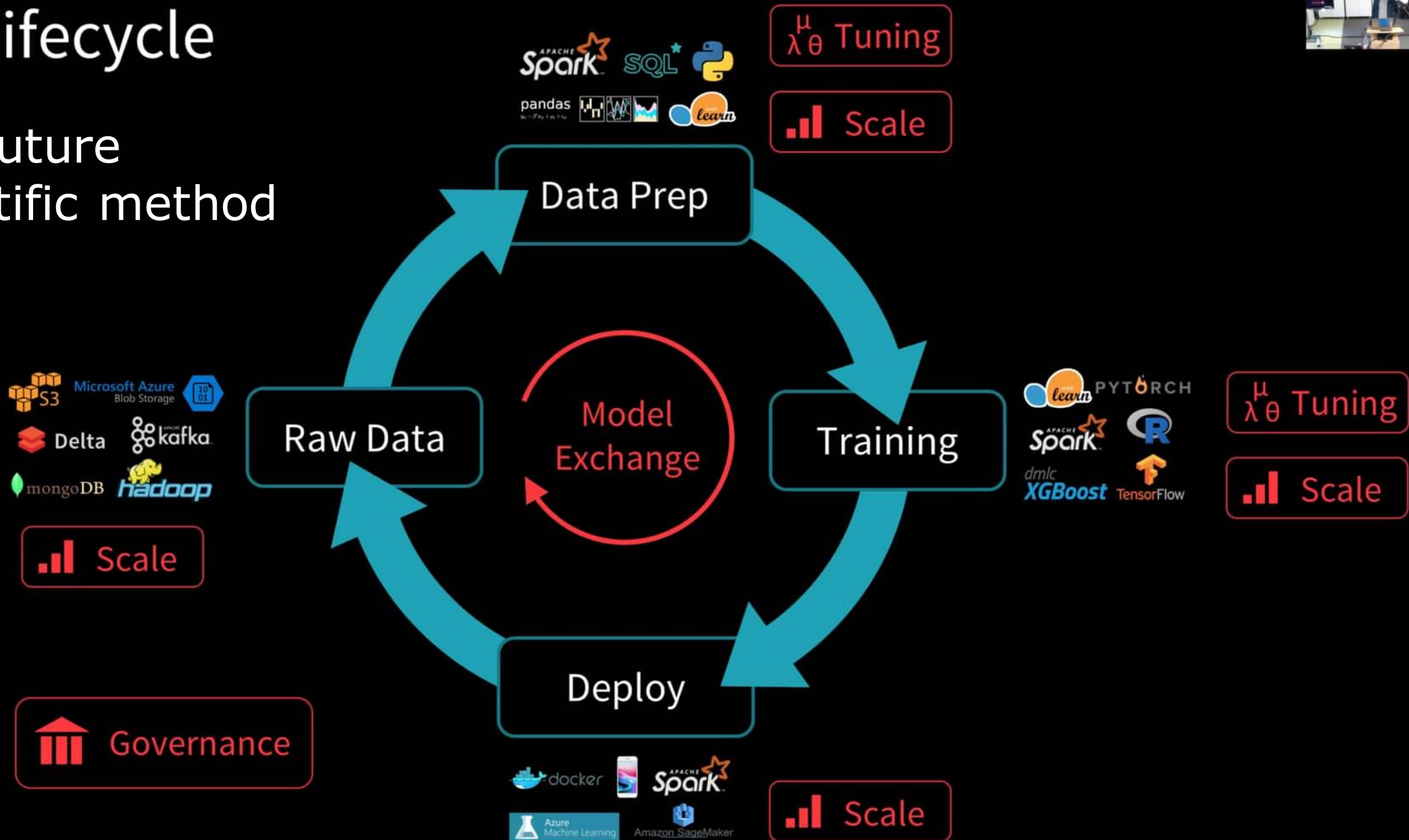
The traditional scientific method



P. D'Abusco | Science priorities for the VO L VOA College Park, Intern 2018, 1.08/11/2018 | Slide 4

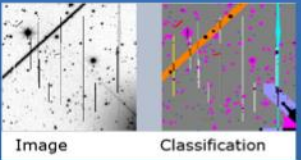
ML Lifecycle

The future scientific method



How AI can fit in our data chain

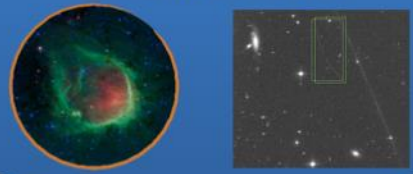
Deep learning classifiers of image contents for data-driven searches or quality control: images with asteroid trails or cosmic rays



NLP to automate linking of data to scientific papers. Automatic generation of statistics on Papers on our missions

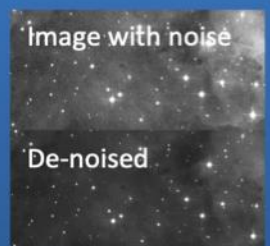


Deep learning classifiers of image contents for data-driven searches: e.g. search images with Interstellar medium bubbles



ESAC Science Data Centre Value Chain

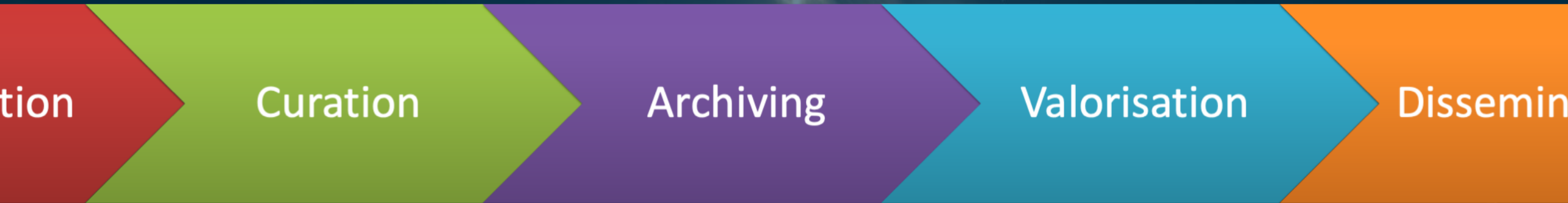
Noise reduction in astronomical images with neural networks, gaining a 50 % in S/N



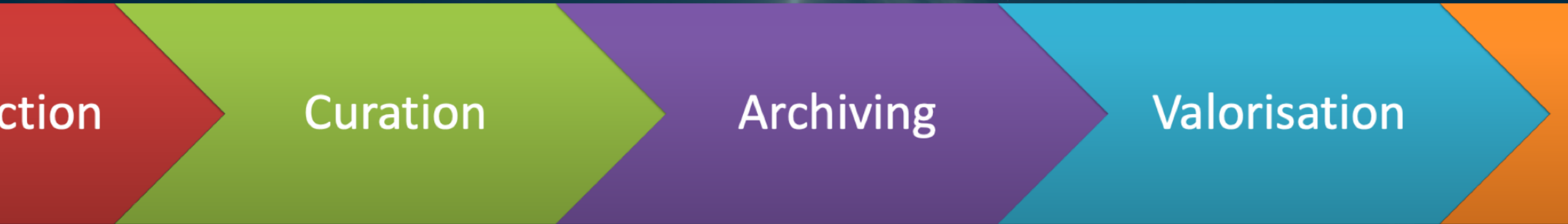
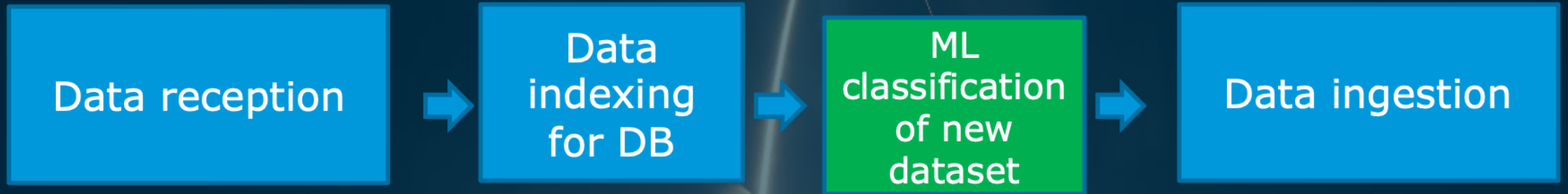
Virtual Assistants to support users on our Web Portals (e.g. Cosmos, ESASky)



AI / ML to enhance science data



AI / ML to enhance science data





TASK

TUTORIAL

Is there an asteroid trail visible in the images?

Yes

No

Impossible to tell

NEED SOME HELP WITH THIS TASK?

Done & Talk

Done



FIELD GUIDE

More than 11,000 volunteers classified 19 years of HST images finding a few thousand new trails of asteroids in just a few months!



Trained Google Auto-ML with the labels from volunteers



← HST_labels_70_15_15 [LABEL STATS](#) [EXPORT DATA](#)

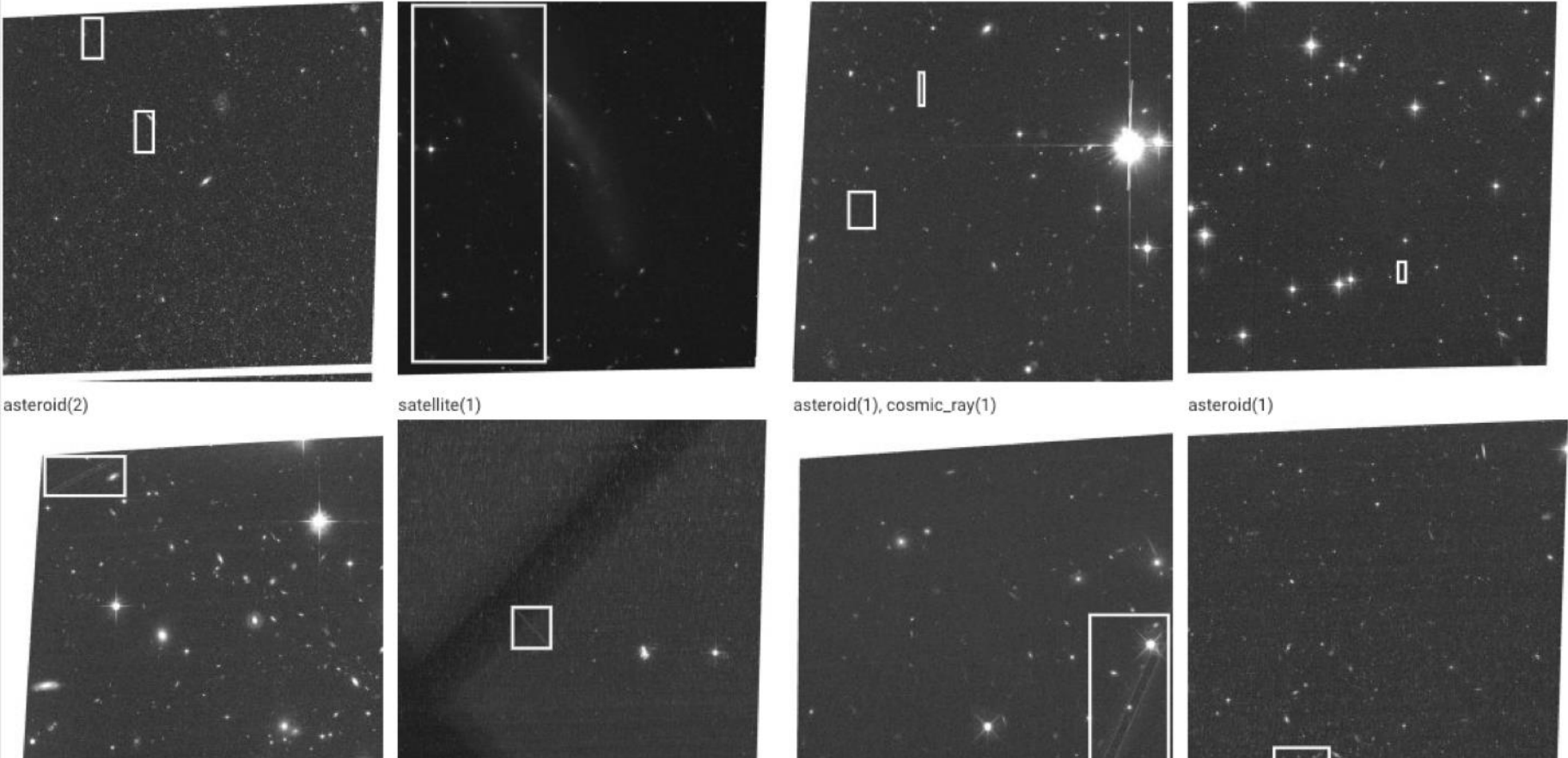
IMPORT **IMAGES** TRAIN EVALUATE TEST & USE Object detection

All images	4,556	<input type="checkbox"/> Filter images
Labelled	4,537	<input type="checkbox"/> Select all
Unlabelled	19	

Filter labels

- asteroid 1,302
- cosmic_ray 1,331
- gravitational_lens_arc 492
- satellite 1,613

[ADD NEW LABEL](#)



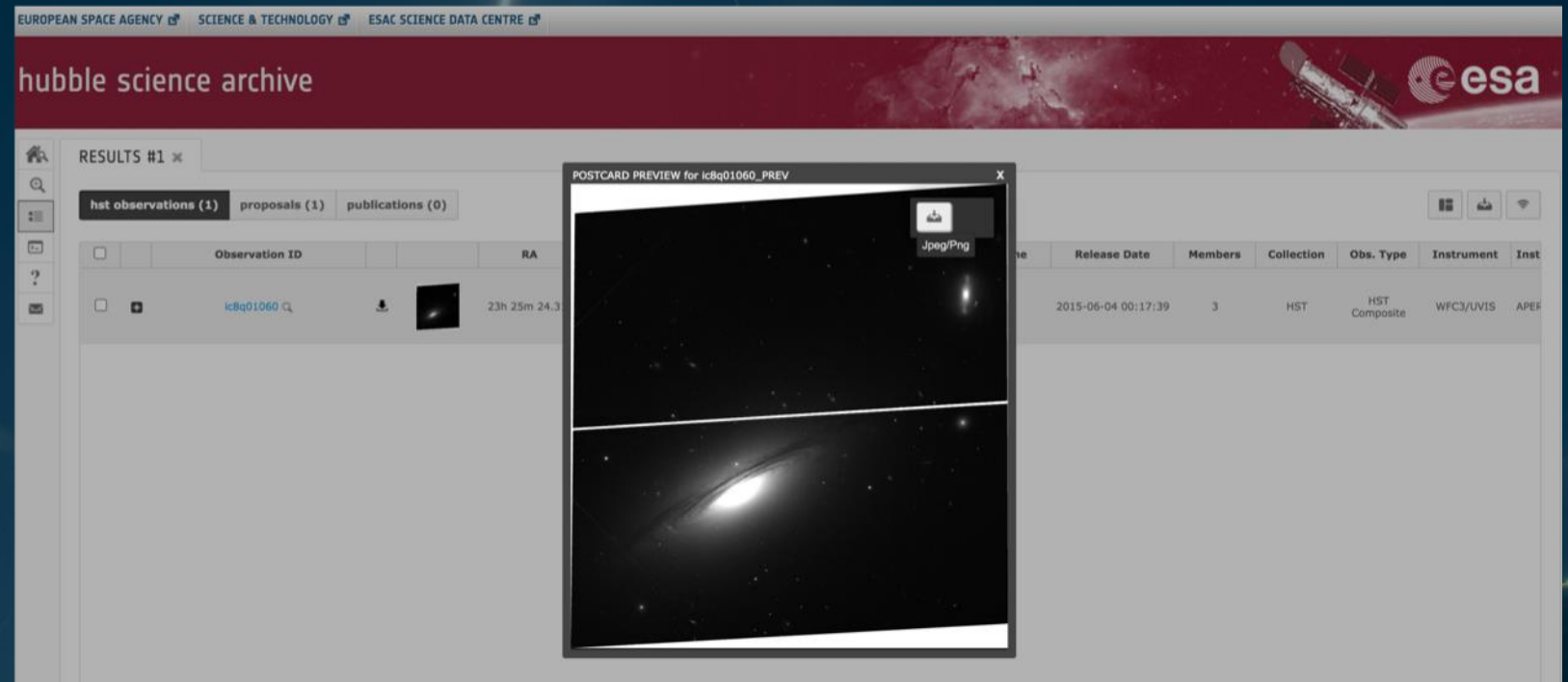
asteroid(2) satellite(1) asteroid(1), cosmic_ray(1) asteroid(1)



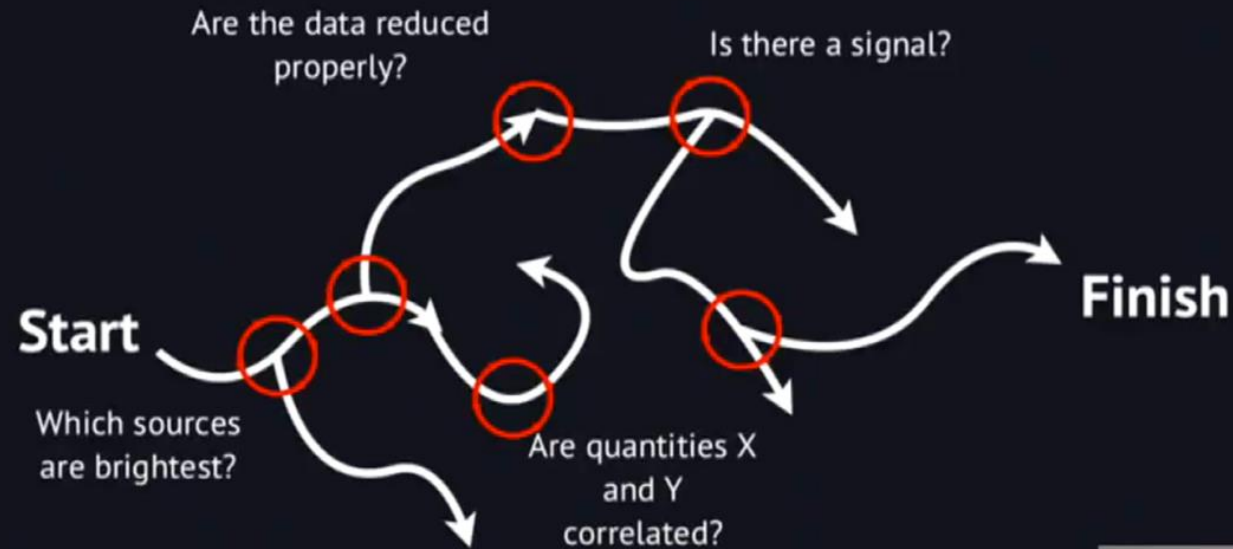


Classifying the eHST archive

- 40,151 “composite” HST ACS/WFC and WFC3/UVIS images (x4 cutouts = 160k cutouts)
- 122 computer Node Hours (~ €350)
- Run on multiple nodes for ~3 hours
- Batch classification on Google Cloud: ~7 hours (38 Node Hours, €60)

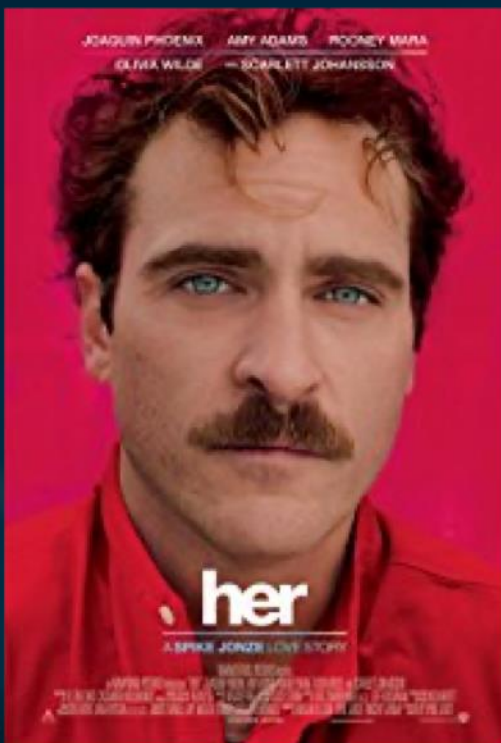


How will the future look like?



Chris Beaumont - Hackable User Interfaces and The Future of Data Analysis in Astronomy <https://www.youtube.com/watch?v=SES7DMwq9Lc>

The most advanced ultimate data query system should enable a dialogue with the user, like in the movie "her" (2013)



ESASky already features a virtual assistant !

http://sky.esa.int

J2000 14 29 42.908 -62 40 46.16 FoV: 6.0' X 4.4' 2MASS color JHK

Sci. Mode En

Search...

Proxima Centauri b

Proxima Centauri b (also called Proxima b or Alpha Centauri Cb) is an exoplanet orbiting within the habitable zone of the closest star to the Sun—the red dwarf star Proxima Centauri, which is in a triple star system. It is located about 4.2 light-years from Earth in the constellation of Centaurus, making it the closest known exoplanet to the Solar System. Best seen in the near-infrared 2MASS sky. [Read more](#)

Target List

Closest exoplanetary systems

- Proxima Centauri b
- Epsilon Eridani
- Ross 128
- Groombridge 34
- YZ Ceti
- Tau Ceti
- Luyten star

can you show me all the young stars observed in X-rays?

3 minutes ago

Which target from this list interests you? Choose by specifying number or target ID!

1. JWST_EROs
2. SpiralGalaxies
3. PeculiarGalaxies
4. InteractingGalaxies
5. GalaxyClusters
6. BrightNebulae
7. DarkNebulae
8. GlobularClusters
9. OpenClusters
10. StarFormationRegions
11. SupernovaRemnants
12. SupermassiveBlackHoles
13. BrownDwarfs
14. BrownDwarfsInMultipleSystems
15. ClosestExoplanetarySystems
16. CESAR_ISM
17. CESAR_Galaxies
18. CESAR_Colours

3 minutes ago

9

2 minutes ago

The player can be used to start the navigation typing "play", to stop it with "pause", to change to the next available option with "next", and to go back to the previous option with "previous". These commands apply to the skies panel and to the target list panel.

Type your message

3 767 12 981 201 771 27 548

Search...



ML MWP


<input type="checkbox"/>		Name	id	wid	hei	score
<input checked="" type="checkbox"/>		2965	2965	0.0469026696833339	0.0452581568469874	0.562308788299561
<input checked="" type="checkbox"/>		2966	2966	0.0295482488183723	0.0285908031789202	0.999996662139893
<input checked="" type="checkbox"/>		2967	2967	0.0176068844646977	0.0170042056114692	0.996415853500366
<input checked="" type="checkbox"/>		2968	2968	0.0248743126943793	0.0237259600980728	0.932490587234497
<input checked="" type="checkbox"/>		2969	2969	0.0234461656885685	0.0225580393023321	0.911825120449066
<input checked="" type="checkbox"/>		2970	2970	0.0554963909035564	0.053729375508177	0.99999761581421
<input checked="" type="checkbox"/>		2971	2971	0.038421063194		0.9569488525

3574 Observation(s) in the chosen area

THE EUROPEAN SPACE AGENCY

ESA Datalabs [0.3.0/BETA]

ESA Datalabs is available as a private beta release, public access is planned for Q3 2022. If you wish to join our private beta programme, please [submit your request here](#).



«YOU CAN EITHER MOVE YOUR QUESTIONS OR THE DATA. [...] OFTEN IT TURNS OUT TO BE MORE EFFICIENT TO MOVE THE QUESTIONS THAN TO MOVE THE DATA.»

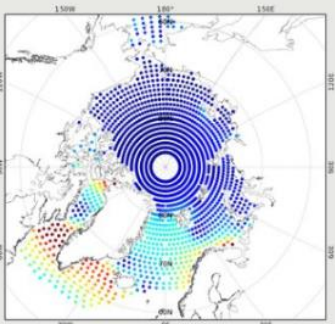
Jim Gray, eScience: A Transformed Scientific Method

BRING YOUR QUESTIONS TO THE DATA

There is a new paradigm, opening completely new opportunities for discovery – a data-intensive approach to science. In many domains, we have entered what could be called the golden age of surveys, with several large-scale projects, spanning decades, between finished, ongoing, and planned activities. ESA is responsible, or is a major partner, in several of these initiatives.

There is, however, a new profound change: data has become a major technological challenge. Increases by multiple orders of magnitude in dataset size means that transferring data to a scientist is often unfeasible.

ESA datalabs gives you a privileged position; bring your code directly to ESA's infrastructure – there is a great set of tools and programming languages are flexible – and execute it with direct access to ESA's archives.



Thanks for your attention !

<http://archives.esac.esa.int>

<http://sky.esa.int>

<http://datalabs.esa.int>

bruno.merin@esa.int