

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Concrete case

Gilles Mathieu



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.

This material by the EOSC-Pillar is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



Biomedical Use Case within EOSC-Pillar



Work Package 6

Use-case #6 : Exploring reference data through existing computing services for the bioinformatics community



Agile FAIR Data for Environment and Earth System Communities
Earth Environment Sciences & Geosciences require a large panel and volume of data from satellite...

Integration of data repositories into EOSC based on communities approaches
The agriculture, food and environment research community faces many challenges common to all: data...

Software source code preservation, reference and access
Leveraging the experience of Software...

FAIR principles in data life-cycles for Humanities
This task aims to identify and develop use cases based on Social Sciences and Humanities (SSH) co...

Defining procedures and services to enforce data provenance for thematic communities and beyond
Due to data exploration complexity, provenance management is a key component in order to guarantee...

Exploring reference data through existing computing services for the bioinformatics community
Galaxy is a widely adopted workflow management systems for bioinformatics, aiming to make computa...

Suitable data formats for seismological big data provisioning via web services
Seismological web services have been designed some years ago with particular types of user and da...

Virtual definition of data sets according to RDA recommendations
At GEODON data centre is very difficult to offer big pre-assembled datasets to be downloaded, due...

Integrating heterogeneous data on cultural heritage
Heritage Sciences, Le...

EOSC-Pillar UC#6: some background

Inserm

La science pour la santé
From science to health



W
What

T6.6 : Investigate existing computing services for the bioinformatics community's reference data

W
Whom

Led by Inserm
Other partners: IBIOM, INFN; IFB

W
Where

France, Italy

W
When

July 2019 - December 2021

H
How

✓ Enhances existing national services in France and Italy: Galaxy , F2DS, D4science..

W
Why

- ✓ Different Galaxy deployments' reproducibility and consistency: ensure the same results regardless of the Galaxy instance used
- ✓ Make it simple to connect Galaxy to data sources.
- ✓ Personal health data protection: deployment in a private and secure environment

EOSC-Pillar UC#6: some background

UC#6 page on the EOSC-Pillar web site <https://eosc-pillar.eu/use-cases/exploring-reference-data-through-existing-computing-services-bioinformatics-community>

UC#6 “fact sheet”

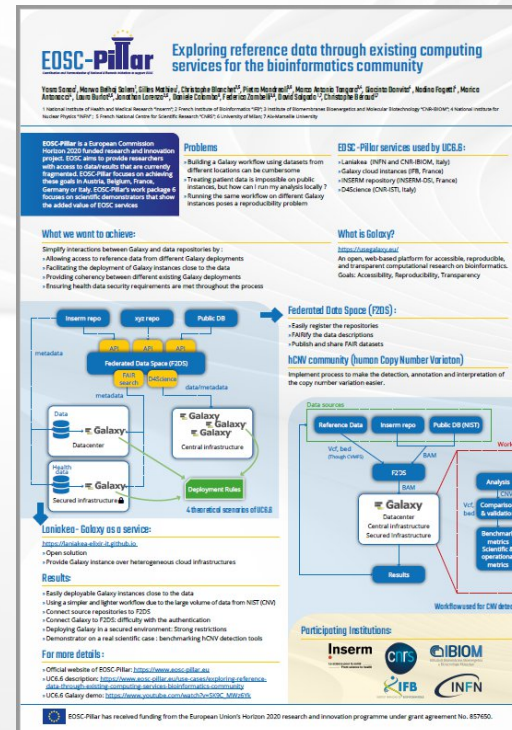
<https://doi.org/10.5281/zenodo.6726022>

UC#6 video demo

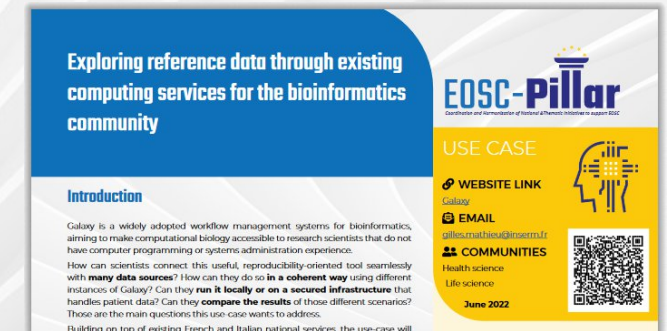
<https://youtu.be/WZey8XrCp1I>

UC#6 poster

<https://doi.org/10.5281/zenodo.7051283>



The poster is titled "EOSC-Pillar Exploring reference data through existing computing services for the bioinformatics community". It lists project partners: Yves Touff, Nicolas Ballez Odier, Gilles Malhotra, Christophe Bignon, Pierre Moreau, Marco Antonio Tavares, Sebastien Donati, Nicolas Fogel, Meissa Amrane, Lucie Boute, Jonathan Lecomte, Isabelle Colinet, Jean-Clément, David Soguel, Christophe Bignon. It describes the project's goal to explore reference data through existing computing services, focusing on interoperability and security. It lists problems like data silos and lack of interoperability, and solutions like federated data spaces and Galaxy instances. It also lists participating institutions: Inserm, CRIIS, IBIOM, FFB, and INFN.

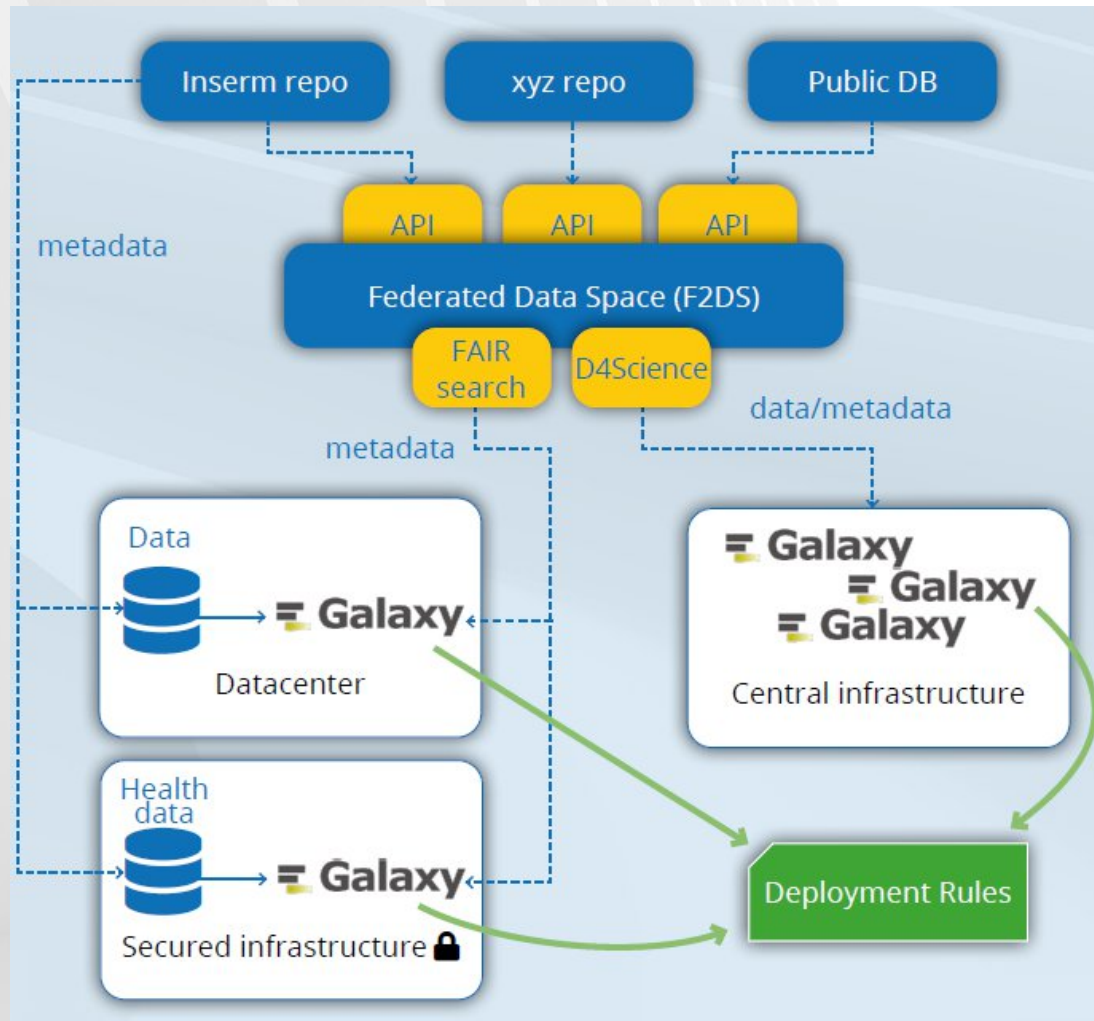


The poster is titled "Exploring reference data through existing computing services for the bioinformatics community". It includes a "USE CASE" section with a "WEBSITE LINK" (Galaxy), "EMAIL" (galaxy.matthieu@inserm.fr), and "COMMUNITIES" (Health science, Life science). It also features a QR code and the date "June 2022".



The thumbnail shows a computer monitor displaying a Galaxy workflow interface. The title is "Use case 6 - Reference data through existing computing services ...". It includes a "Watch Later" button and a "Share" button. At the bottom, it says "Watch on YouTube".

UC#6: what did we want to do?



4 theoretical scenarios

- 1 Use public Galaxy Instances
- 2 Deploy Galaxy locally
- 3 Deploy Galaxy on a secured infrastructure
- 4 Compare results of the 3 scenarios above

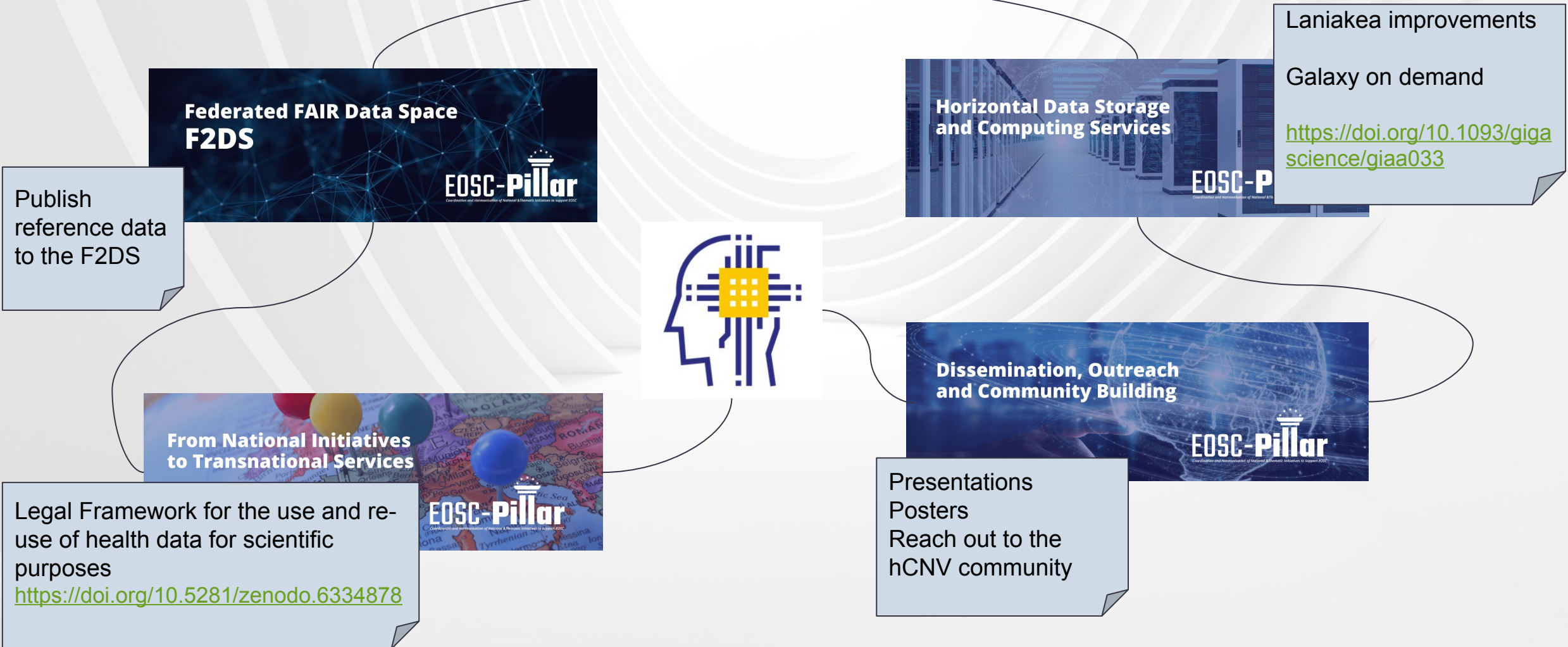
UC#6 challenges

- * Reproducibility and coherency of the various deployments
- * How to implement Galaxy in a private, secure environment with a data analysis workflow similar to that of its public equivalent
- * Integrating the service into a global authentication system
- * Make the service available to all members of the EOSC community

UC#6: what did we actually do?

- Implement and use better rules for Galaxy deployment
- Provide easily deployable Galaxy instances close to the data
- Publish and populate source/reference data to the F2DS
- Deploy Galaxy in a secured environment
- Provide a working demonstrator with a light workflow
- Analyse transnational health data restrictions
- Defined a full scale "real life" workflow with the hCNV community

UC#6: what we did - in the whole EOSC-Pillar picture



UC#6: How did that go?



Worked well

- * Collaborating internationally
- * Publishing data to F2DS
- * Enriching metadata in F2DS
- * Conceiving a Galaxy workflow
- * Using Laniakea@ReCAS
- * Deploying Galaxy locally
- * Deploying on a secured infra



Didn't work so well

- * Running workflow on huge files (memory problem)
- * Connecting F2DS to Galaxy (AAI problem)
- * Directly using F2DS fetch method from Galaxy (function missing)

How does this translate to “facing a new pandemic”?

- * Technical work
 - * Publishing data to a federated dataspace
 - * Technical aspects: Push vs pull model
 - * semantic aspects: how do we enrich? What is the benefit?
 - * Some of the analysis could be done in Galaxy
 - * Ex: virus sequencing data
 - * Designing global workflow
 - * Connect different tools: how do they communicate?
 - * Allow those tools to work in secured (health data) environments
- * Community awareness
 - * Spread the word that EOSC can help

A wider perspective: be preventive rather than curative

Proposing new/innovative solutions while in the middle of a crisis usually doesn't work...



A wider perspective: be preventive rather than curative

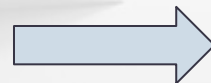
... but when the crisis is over, will to improve things decreases



What we have to bear in mind

A pandemic is not a marathon. You don't know the date, the cost, or what kind of training it will require.

The question is not whether or not we'll face a new pandemic. It's "when will it occur"



This might not be the best

Image credits: <https://keepcalms.com/>

Overcoming technical issues

Back to “what didn’t work”
in UC#6...

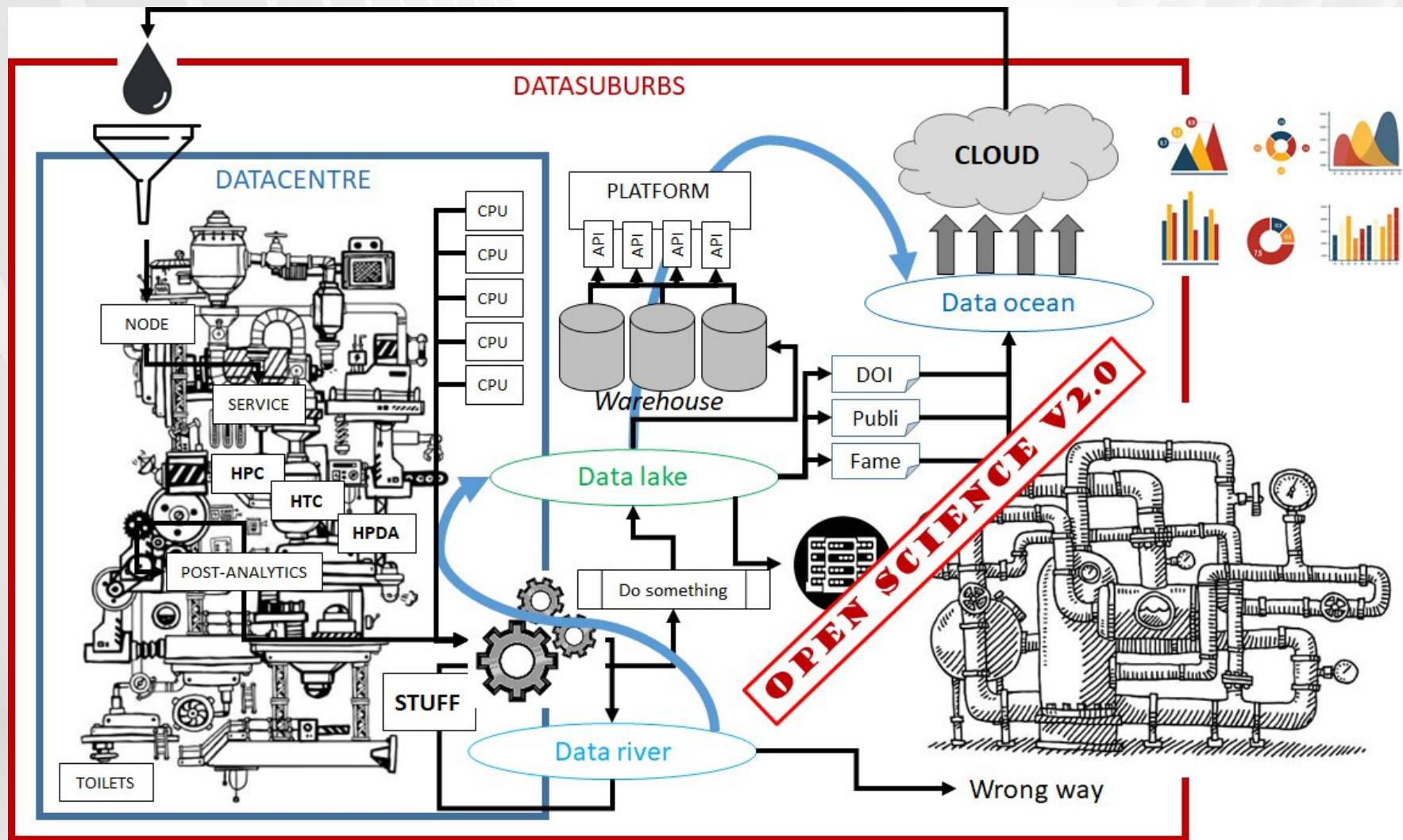
Didn’t work so well

- * Running workflow on huge files (memory problem)
- * Connecting F2DS to Galaxy (AAI problem)
- * Directly using F2DS fetch method from Galaxy (function missing)

- * **Technical resources issue**
- * **Technological issue**
- * **Functionality issue**

- Work on scalability/dimensioning
- Go beyond the “prototype” step
- Base architecture on users needs
- Think interoperability “by design”
- Improve reuse/sharing practices
- Use standards and converge
- Think long term

How EOSC is sometimes perceived



How we should build it



Reusing what we have is fine. But if we keep reusing bricks that are not compatible, building something coherent might prove difficult.



Final considerations



- * Technical issues are not the most blocking
 - * Act while we can
 - * Overcome local/personal interests
 - * Use projects to fulfill long-term goals...
 - * ... but use long-term goals to set up projects!

- * EOOSC is (and will be) what we make of it
 - * Get involved and put value in it
 - * Do our best to synchronise effort
 - * Get out of the comfort zone

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Thank you!

Get in touch with us!



www.eosc-pillar.eu



[@EoscPillar](https://twitter.com/EoscPillar)



[/company/eosc-pillar](https://www.linkedin.com/company/eosc-pillar)



EOSC-Pillar has received funding from the European Union's Horizon 2020 research and innovation Programme under Grant Agreement No. 857650.
This material by the EOSC-Pillar Consortium is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

