

EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Feedback on the development of a use case in bioinformatics within the EOSC-Pillar project

Marwa BELHAJ-SALEM, Gilles MATHIEU

INSERM DSI – SSDUN/Domis
October 2022



Planning of the presentation

- What are EOSC and EOSC-Pillar projects
- Galaxy and UC6 challenges
- The 4 theoretical scenarios
- Services used : Laniakea: Galaxy, F2DS
- hCNV's example
- Exchange of impressions: are the tools really useful, what is missing, is EOSC relevant, etc.

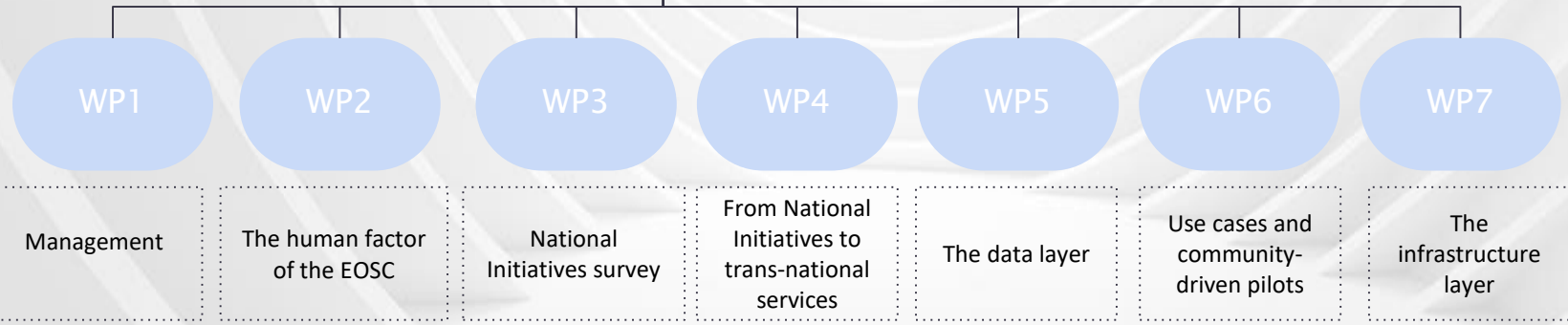
What's EOSC

- The European Open Science Cloud (EOSC) – <https://eosc-portal.eu/>
- Initiative of the European Commission started in 2015 and it is a contribution of Member States and partners
- 53 projects funded under H2020 with 81.5 million € over 2021 and 2022 for the first calls under Horizon Europe
- Based on combining existing research data infrastructures and Virtual environment
- Enables researchers to find, access and reuse research data, services and tools (FAIR Principles)

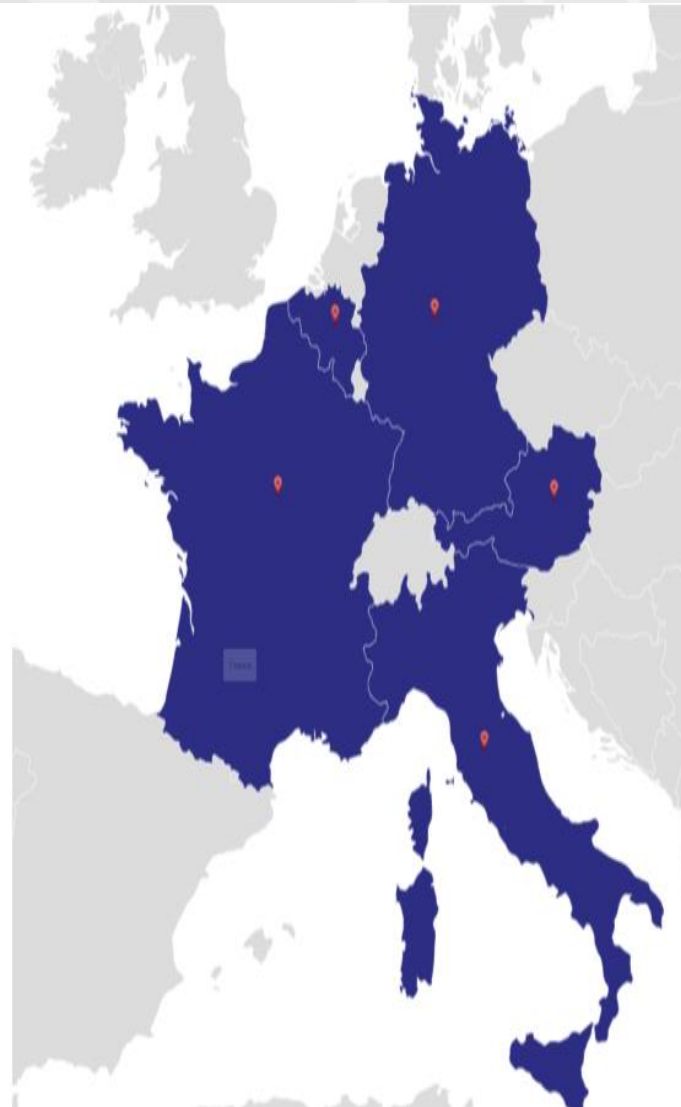
Position of EOOSC-Pillar in EOOSC



INFRAEOOSC-05b call projects
"Supporting the Overall Governance of EOOSC"



What's EOSC-Pillar



W

what

EOSC-Pillar – Coordination and harmonization of national initiatives, infrastructures and data services in Central and Western Europe

W

whom

A consortium of 18 partners

W

Where

France, Belgium, Germany, Italy, Austria

W

When

July 2019 - December 2022

H

How

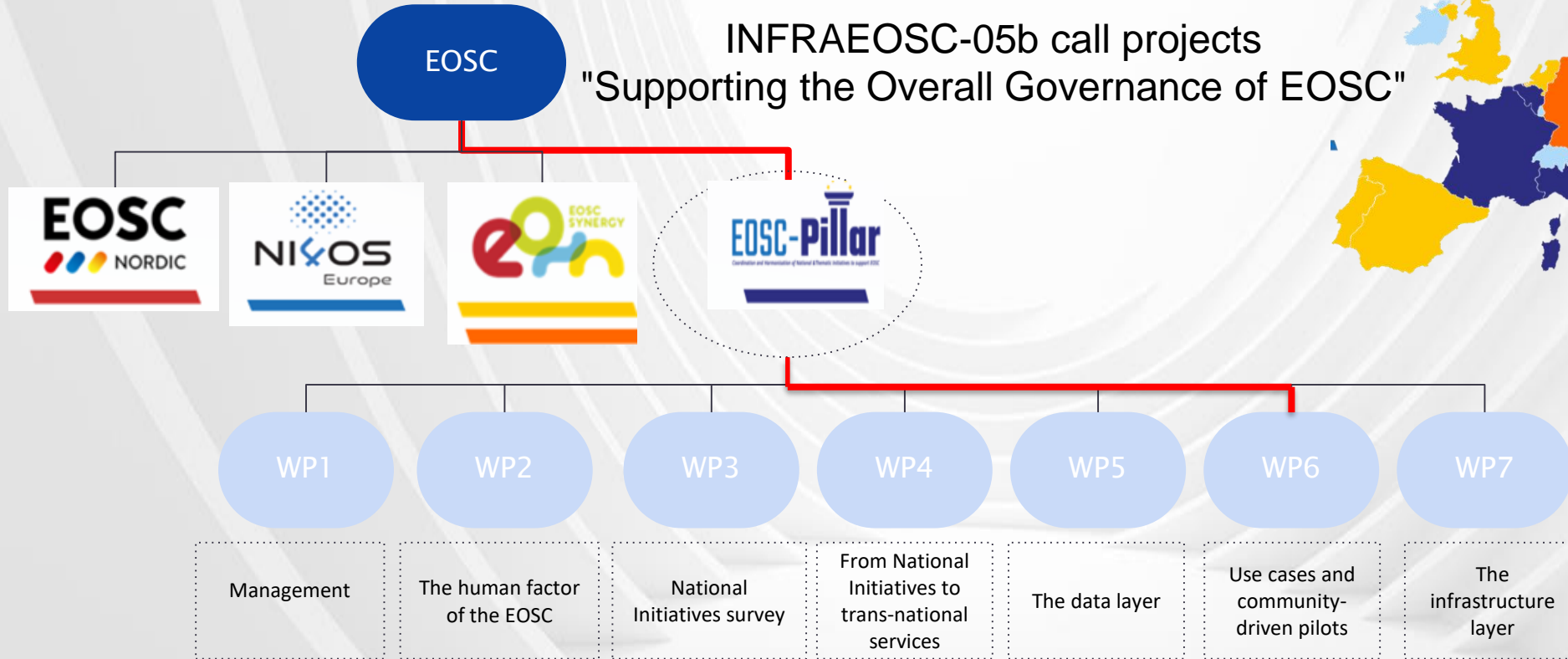
- ✓ Coordinates with regional EOSC projects
- ✓ Supports EOSC implementation by building on national and thematic initiatives developed by research communities.
- ✓ Coordinates data infrastructures and services by bringing together leaders of national initiatives.
- ✓ Various services for management, analysis, and storage

W

Why

- ✓ Better European-level exchange of research data (different fields)
- ✓ Data reuse across borders and scientific disciplines through the federation of existing infrastructures and services

Use Case 6's position in EOSC-pillar



Presentation of use case 6.6



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..

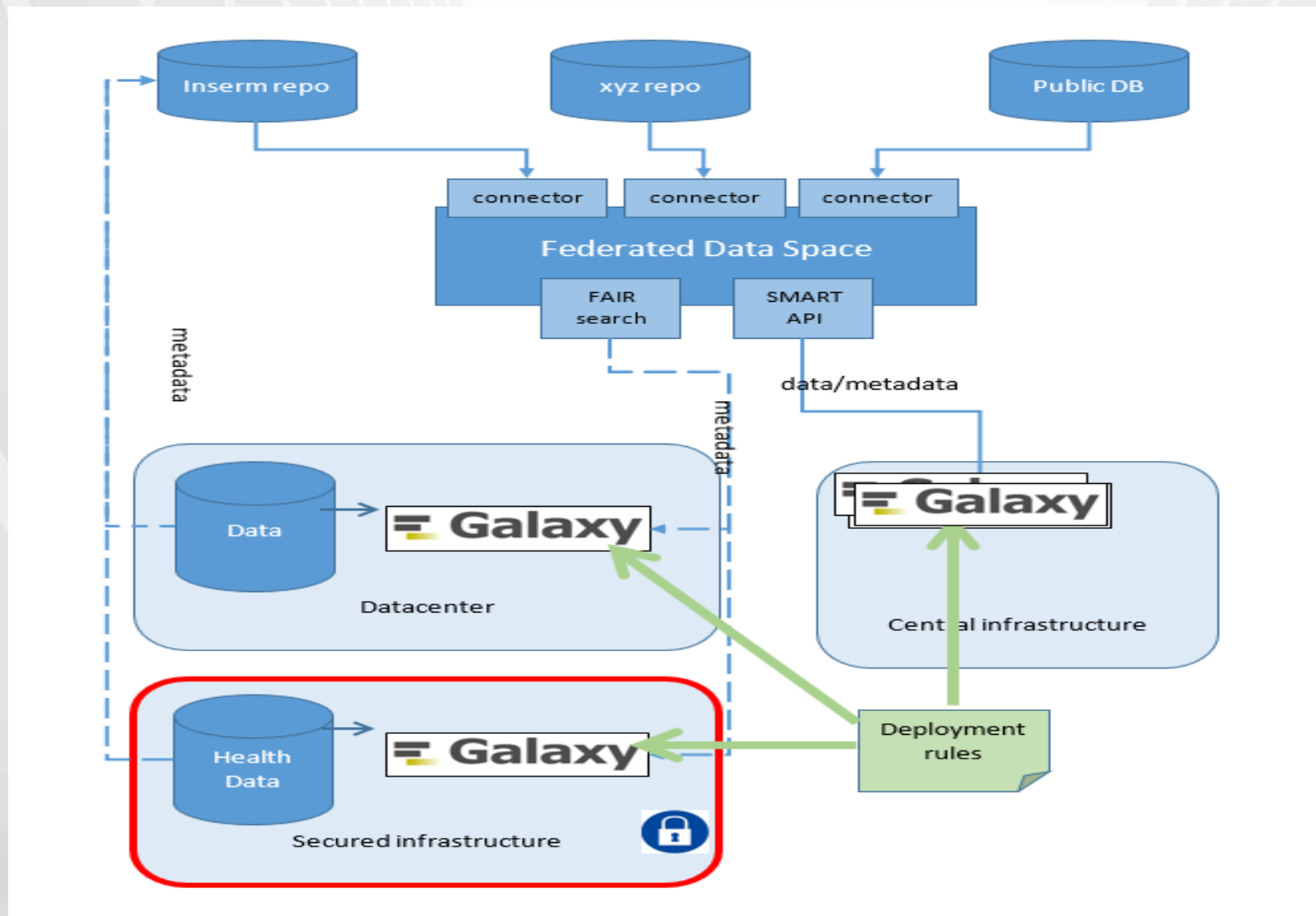
W
Why

- ✓ 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

Challenges of UC 6

- The 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

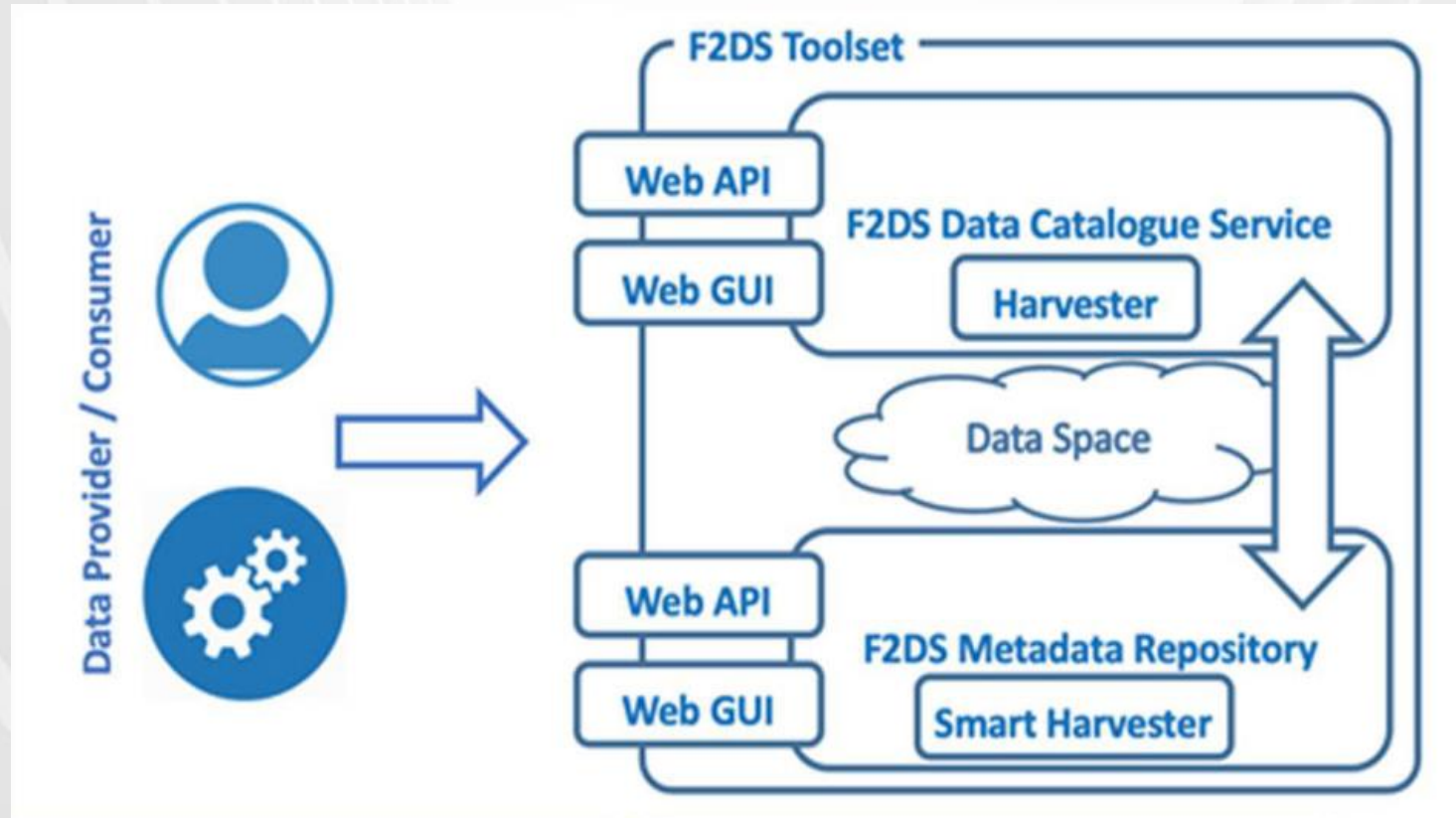
The 4 theoretical scenarios



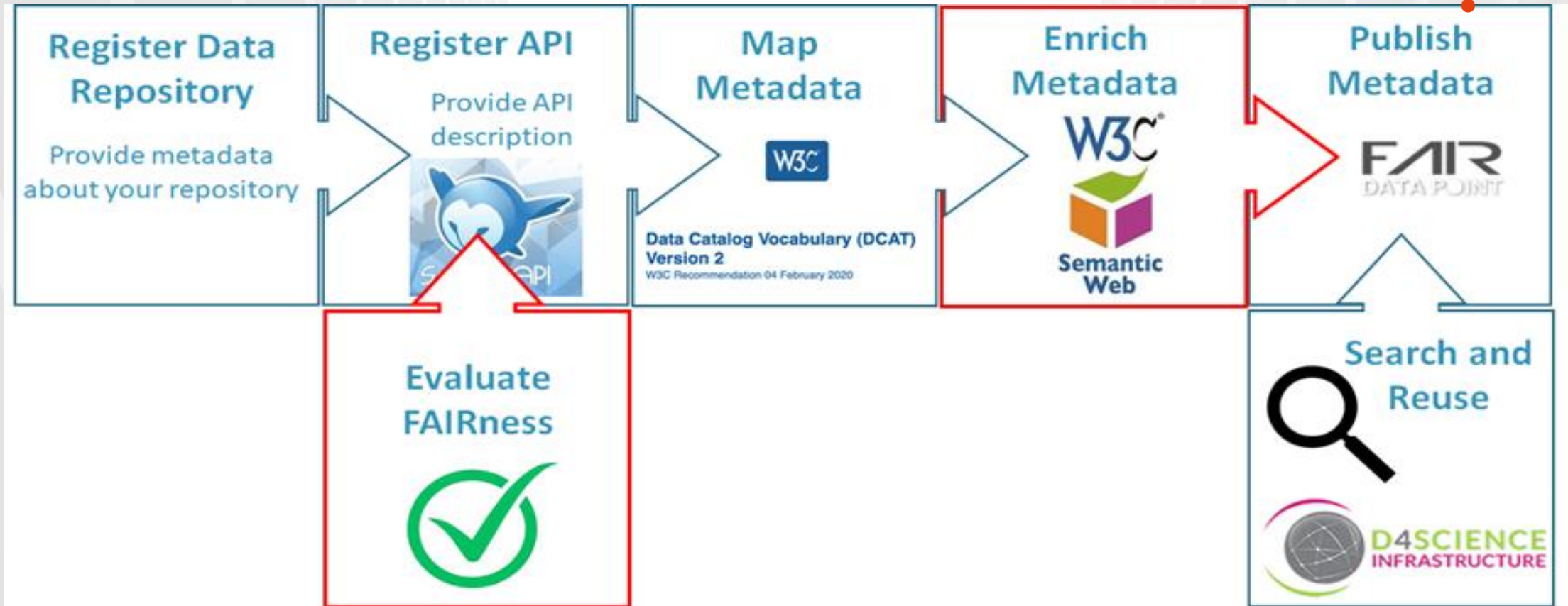
Galaxy

- Galaxy is a well-known bioinformatics workflow management system <https://galaxyproject.eu>
- Galaxy aims to make computational biology accessible to research scientists who do not have computer programming or systems administration experience
- **Laniakea – Galaxy** as a service provided by INFN and CNR-IBIOM. Laniakea is a software framework that facilitates the provisioning of on-demand Galaxy instances as a cloud service over e-infrastructures <https://doi.org/10.1093/gigascience/giaa033>
- Galaxy instances deployed by IFB in France

The Federated FAIR Data Space (F2DS) aims to provide tools to improve the overall FAIRness of datasets



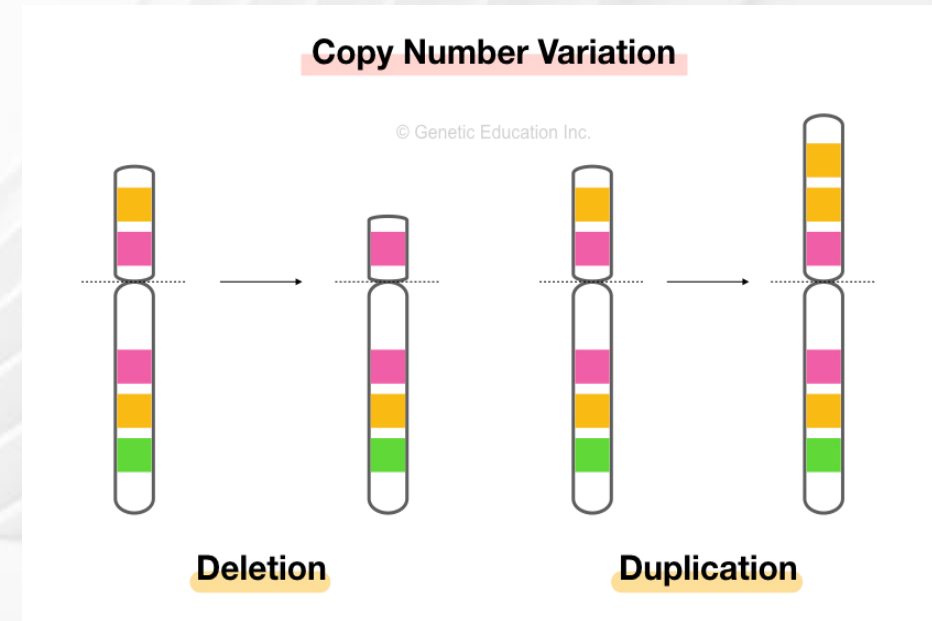
F2DS Overall Architecture



the Metadata Repository and the Data Catalogue Service

A concrete application: hCNV tool benchmarking (1)

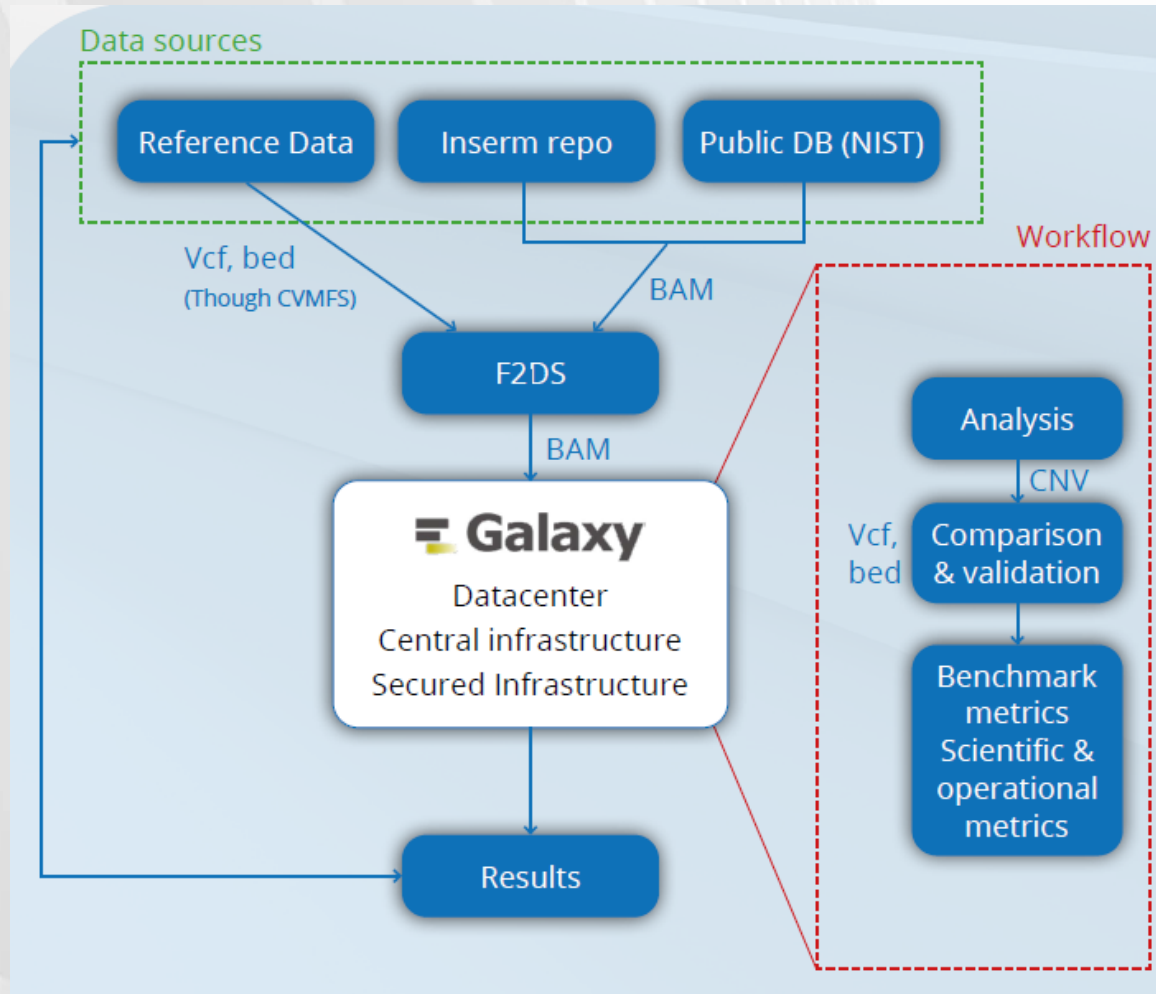
- hCNV: Human Copy Number Variation
 - Genome modification during mitosis
 - Duplication : a gene is copied twice or more
 - Deletion: a gene is not copied
 - Play a role in some diseases
- hCNV detection
 - Represents a major challenge
 - Needs tools: “CNV callers”



A concrete application: hCNV tool benchmarking (2)

- Reference hCNV data
 - produced by NIST (National Institute of Standards and Technology, US)
 - <https://www.nist.gov/programs-projects/genome-bottle>
- Benchmarking of hCNV callers
 - Run hCNV caller on a reference sample
 - Compare results with the NIST “gold standard”
- One of the tasks of the hCNV Elixir community
 - <https://elixir-europe.org/communities/hcnv>

A concrete application: hCNV tool benchmarking (3)



- Select 2 samples within the set of NIST reference data
- Put 1 in INSERM repository, and reference it in the F2DS
- Reference the 2nd sample in the F2DS directly pointing to its location on NIST database
- Reference in the F2DS the output files (VCF format) that corresponds to the NIST gold standard
- Run the analysis workflows on Laniakea
- Run the comparison workflow to compare obtained results on Lanikea with Gold Standard results
- Publish the result of the comparison (statistic table) in Inserm repository, and reference it in the F2DS

Technical limitation

- Related to Galaxy

Laniakea Galaxy instances comes at most with 32GB of RAM

memory space limitation

- Related to F2DS

F2DS does not offer a direct link to the raw data of Galaxy using the 'fetch function'

- Authentication problem

Difficulties linking between D4science and Galaxy: they have different AAI

General impression: is this worth it?

- Was EOSC-Pillar useful?
 - From an Inserm point of view, yes
 - It helped us thinking about our services with a wider perspective
- EOSC is (and will be) what we make of it! We need to:
 - Get involved and put value in it
 - Do our best to synchronise effort, not only internationally
 - Stop adopting the “wait and see” attitude

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/)

