EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Development of a use case in bioinformatics within the EOSC-Pillar project

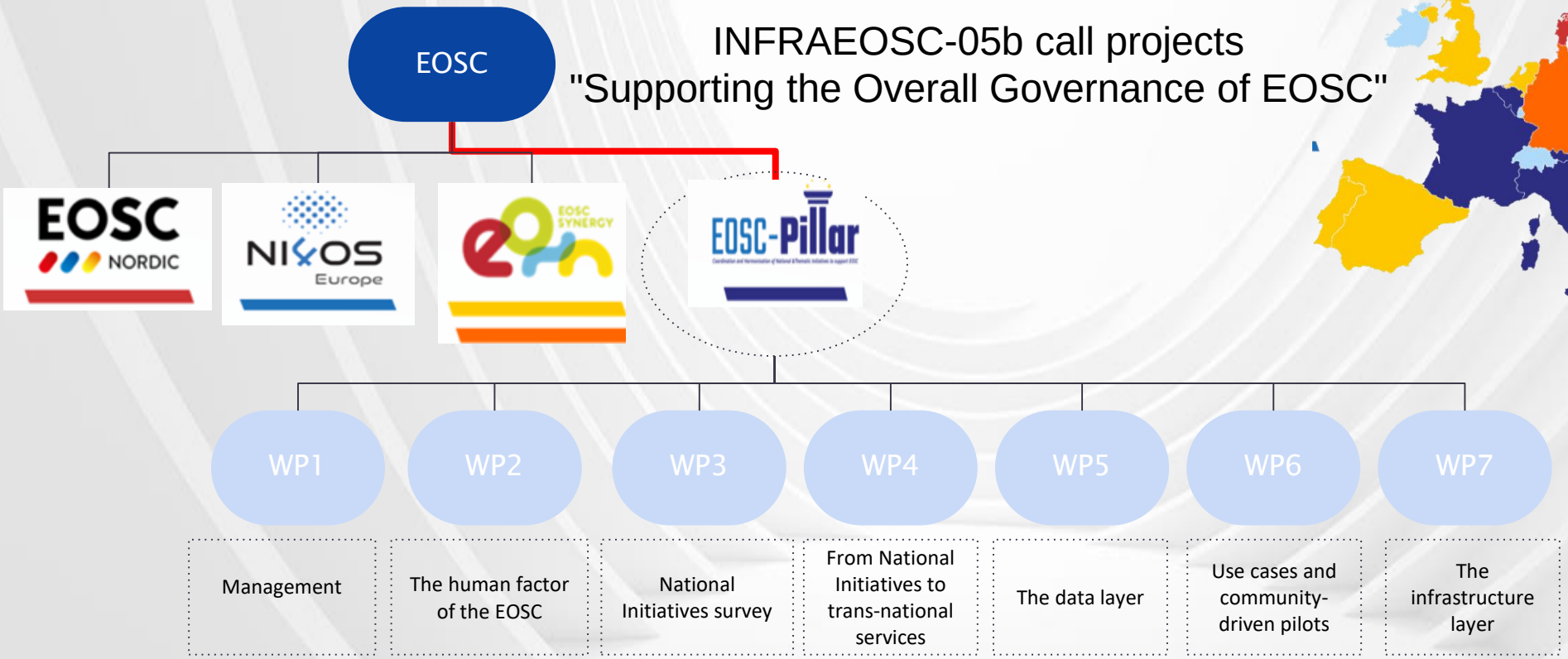
Marwa BELHAJ-SALEM, Gilles MATHIEU

INSERM DSI – SSDUN/DomIS

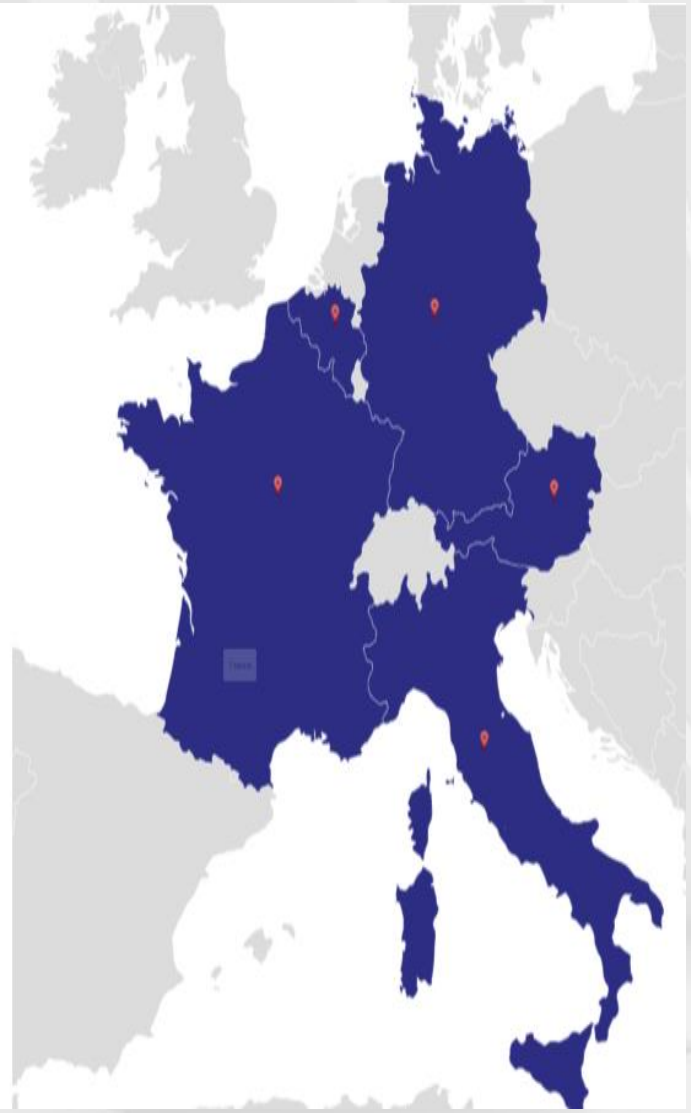
Dec. 2022

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

Position of EOSC-Pillar in EOSC



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

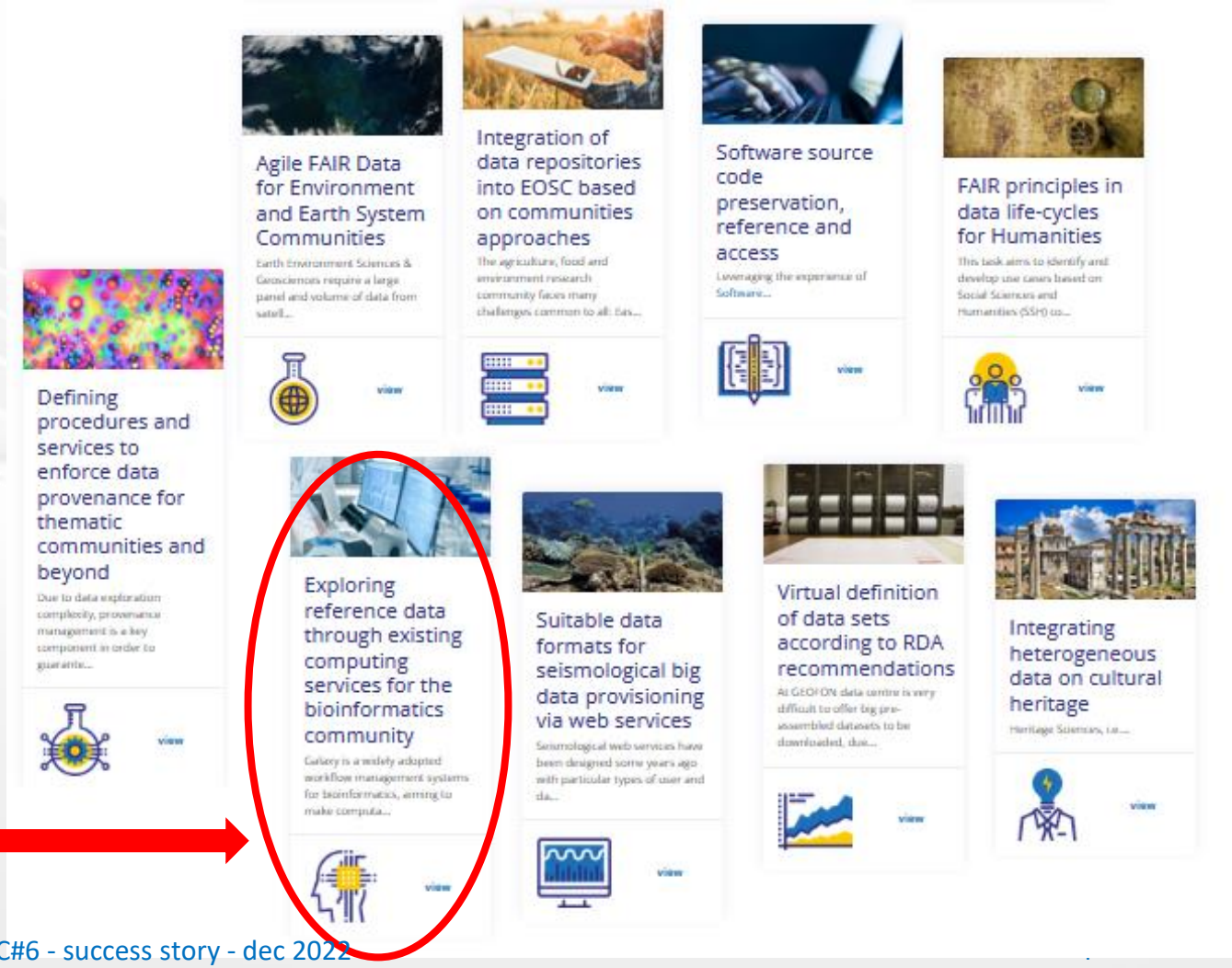
W Why

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

Biomedical Use Case within EOSC-Pillar

Work Package 6

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



Presentation of use case 6.6

What

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

Whom

Led by Inserm
Other partners: IBIOM, INFN; IFB

Where

France, Italy

When

July 2019 - December 2021

How

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

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

UC#6 challenges

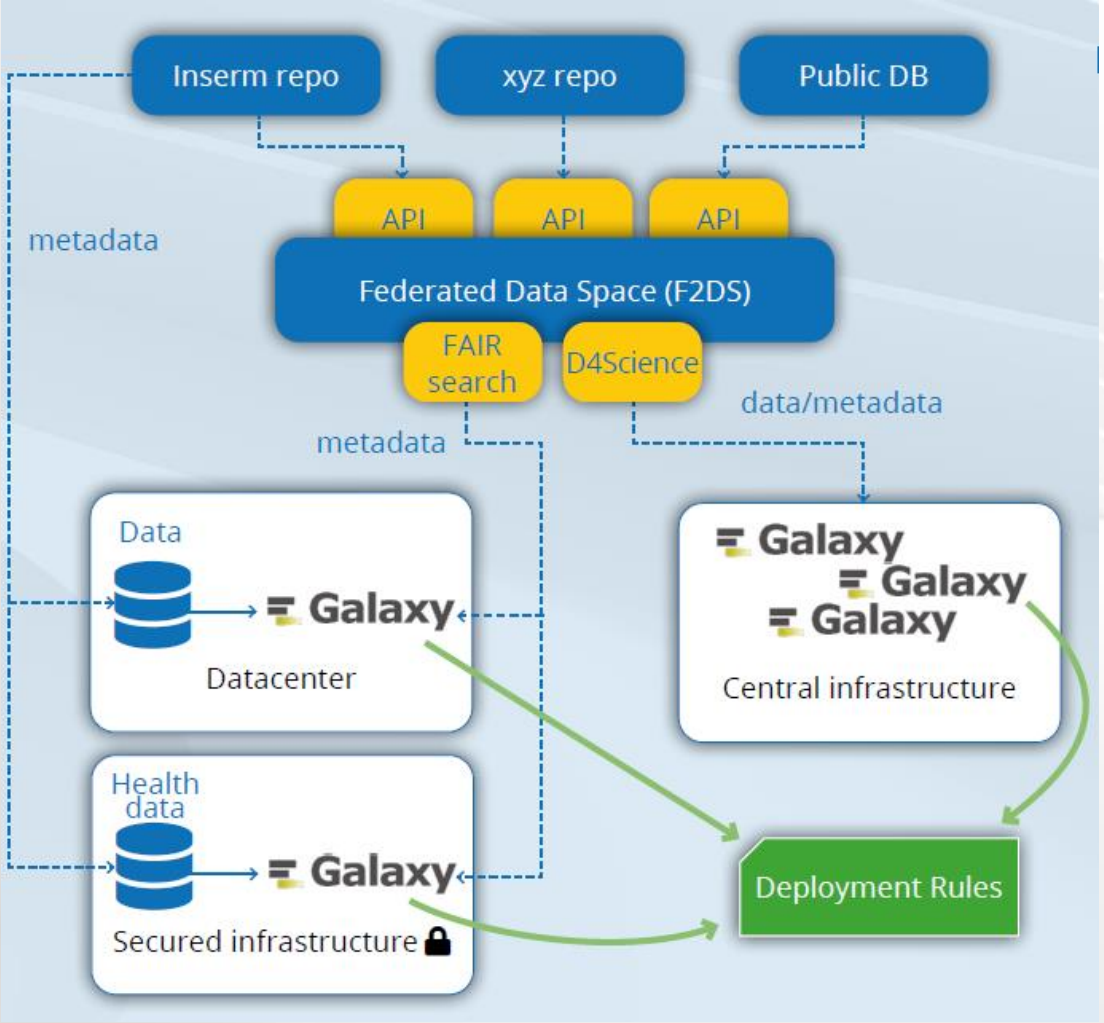
* Reproducibility and coherency of the various deployments

* 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 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: 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: How did that go?

Worked well

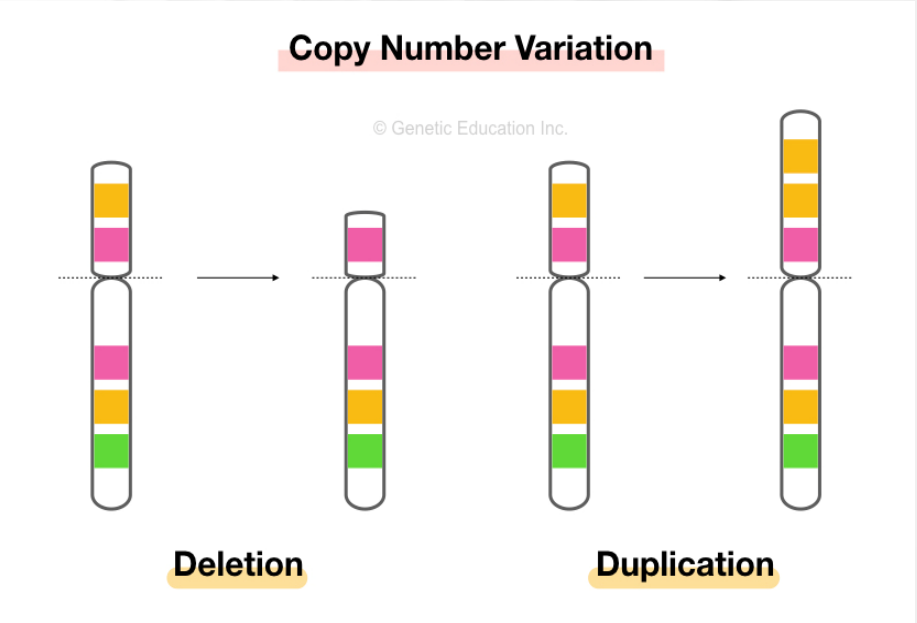
* Collaborating internationally
* Publishing data to F2DS
* Enriching metadata in F2DS
* Conceiving a Galaxy workflow
* Using Laniakea@ReCAS
* Deploying Galaxy locally
* Deploying on a secured infra

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)

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

More info

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”

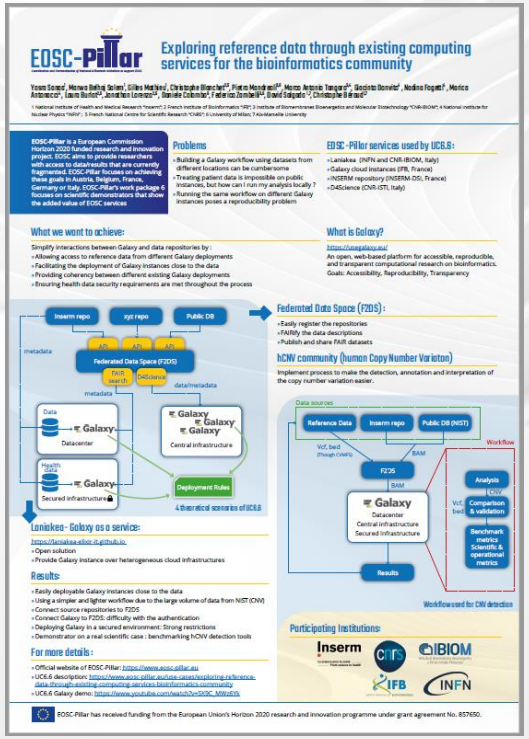
https://doi.org/10.5281/zenodo.6726022

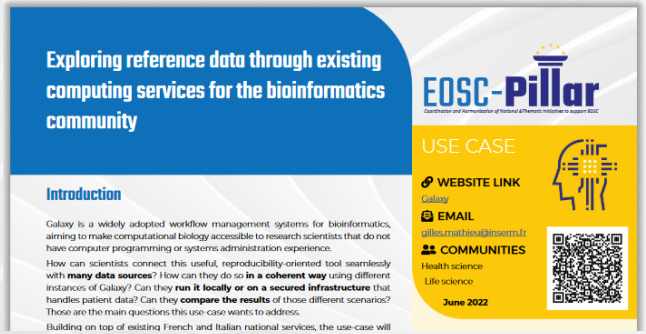
UC#6 video demo

https://youtu.be/WZey8XrCp1I

UC#6 poster

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





EOSC-Pillar

Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Thank you!

Get in touch with us!

www.eosc-pillar.eu

@EoscPillar

/company/eosc-pillar

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