EOSC-Pillar
Coordination and Harmonisation of National & Thematic Initiatives to support EOSC

Exploring reference data through existing computing services for the bioinformatics community

Yosra Sanaa[1], Gilles Mathieu[1], Christophe Blanchet[2,5], Pietro Mandreoli[3,6], Marco Antonio Tangaro[3,4], Giacinto Donvito[4], Nadina Fogetti[4], Marica Antonacci[4], Laura Burlot[2,5], Jonathan Lorenzo[2,5], Daniele Colombo[6], Federico Zambelli[3,6], David Salgado[1,7]

1 National Institute of Health and Medical Research "Inserm"; 2 French Institute of Bioinformatics "IFB"; 3 Institute of Biomembranes Bioenergetics and Molecular Biotechnology "CNR-IBIOM"; 4 National Institute for Nuclear Physics "INFN" ; 5 French National Centre for Scientific Research "CNRS"; 6 University of Milan; 7 Aix-Marseille University

EOSC-Pillar is a European Commission Horizon 2020 funded research and innovation project. EOSC aims to provide researchers with access to data/results that are currently fragmented. EOSC-Pillar focuses on achieving these goals in Austria, Belgium, France, Germany or Italy. EOSC-Pillar's work package 6 focuses on scientific demonstrators that show the added value of EOSC services

Problems

» Building a Galaxy workflow using datasets from different locations can be cumbersome
» Treating patient data is impossible on public instances, but how can I run my analysis locally ?
» Running the same workflow on different Galaxy instances poses a reproducibility problem

EOSC-Pillar services used by UC6.6:

» Laniakea (INFN and CNR-IBIOM, Italy)
» Galaxy cloud instances (IFB, France)
» INSERM repository (INSERM-DSI, France)
» D4Science (CNR-ISTI, Italy)

What we want to achieve:

Simplify interactions between Galaxy and data repositories by :
» Allowing access to reference data from different Galaxy deployments
» Facilitating the deployment of Galaxy instances close to the data
» Providing coherency between different existing Galaxy deployments
» Ensuring health data security requirements are met throughout the process

What is Galaxy?

https://usegalaxy.eu/
An open, web-based platform for accessible, reproducible, and transparent computational research on bioinformatics.
Goals: Accessibility, Reproducibility, Transparency

4 theoretical scenarios of UC6.6
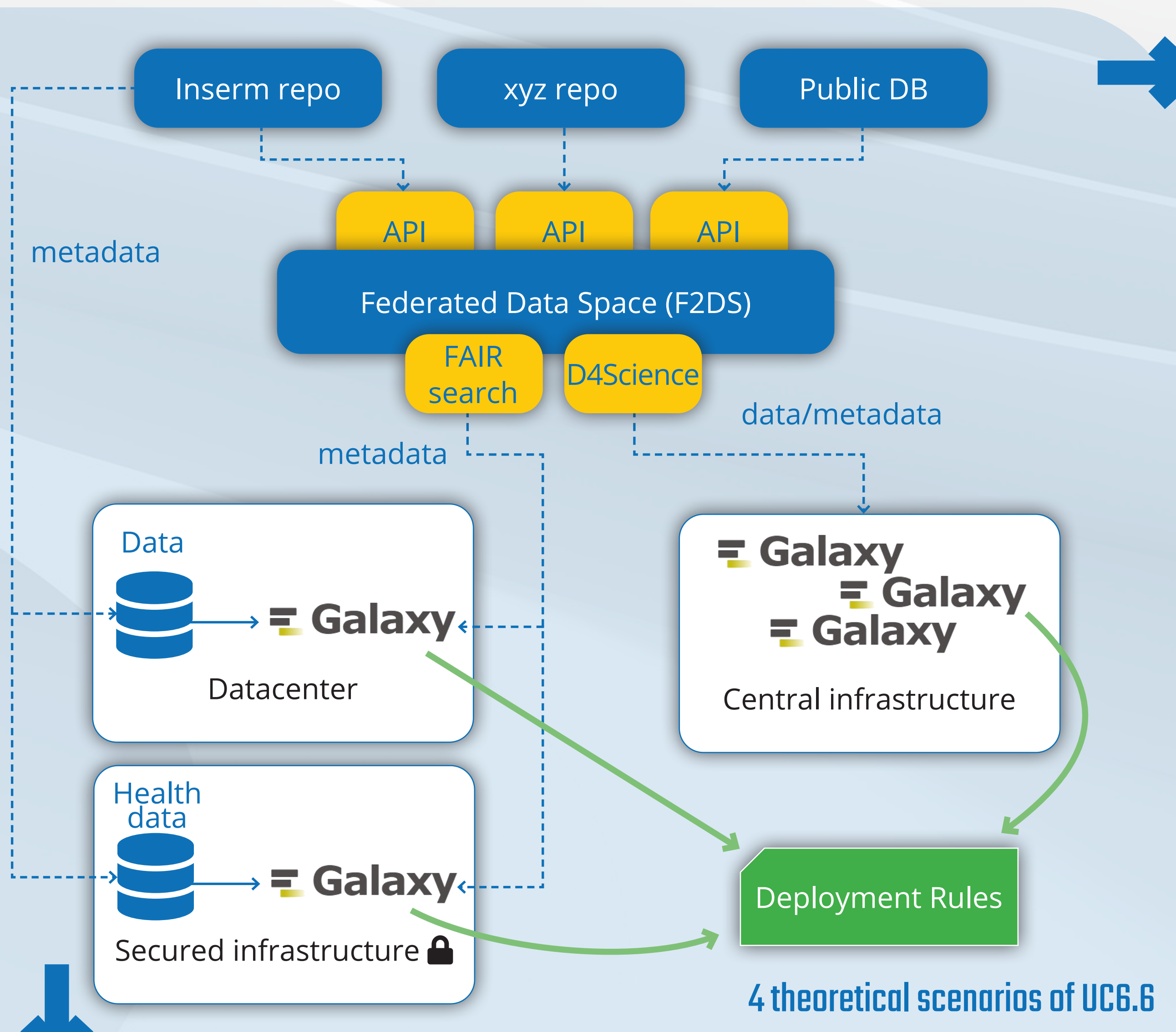
Federated Data Space (F2DS) :

» Easily register the repositories
» FAIRify the data descriptions
» Publish and share FAIR datasets

hCNV community (human Copy Number Variaton)

Implement process to make the detection, annotation and interpretation of the copy number variation easier.

Workflow used for CVN detection
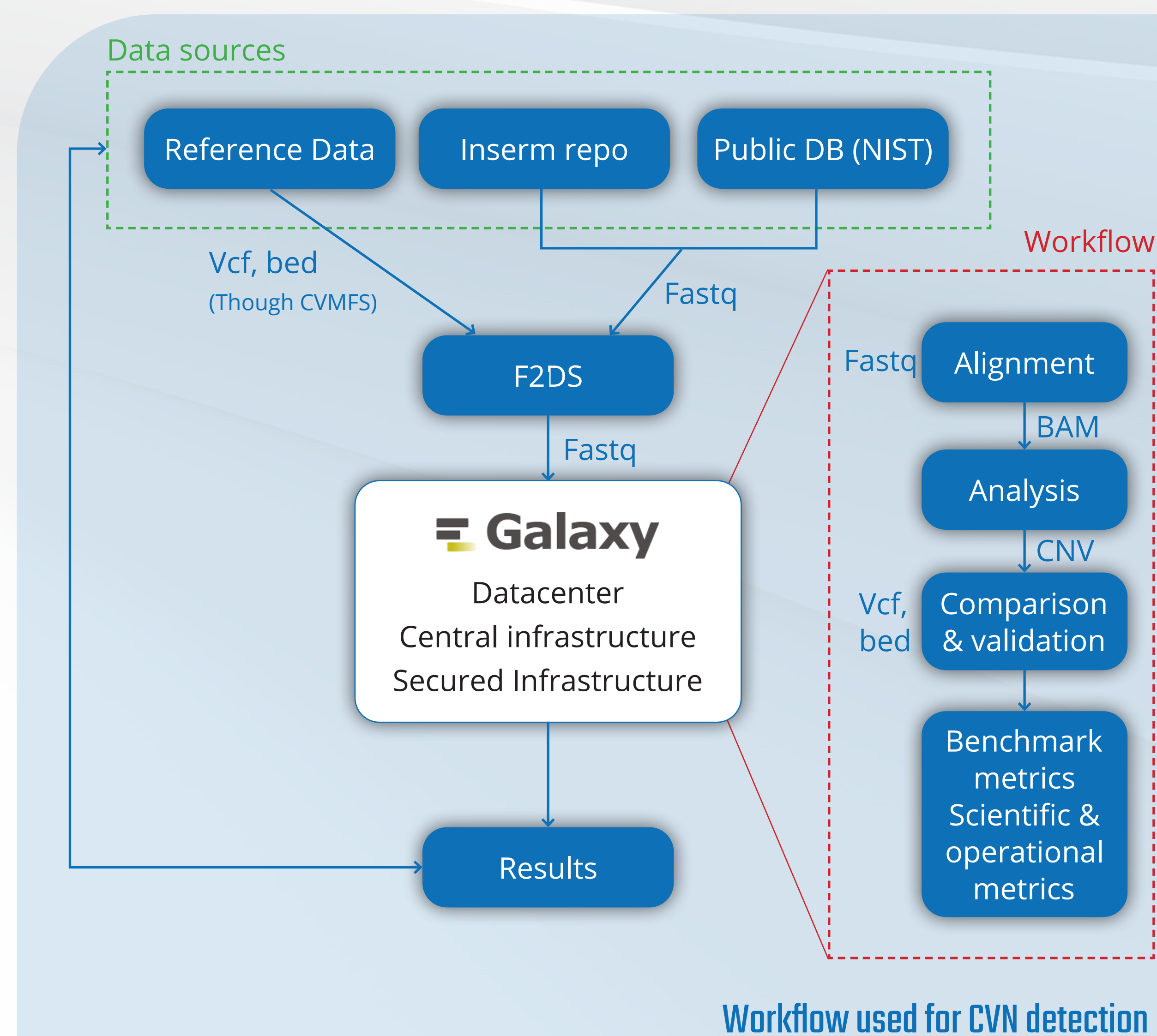
Laniakea- Galaxy as a service:

https://laniakea-elixir-it.github.io
» Open solution
» Provide Galaxy instance over heterogeneous cloud infrastructures

Results:

» Easily deployable Galaxy instances close to the data
» Using a simpler and lighter workflow due to the large volume of data from NIST (CNV)
» Connect source repositories to F2DS
» Connect Galaxy to F2DS: difficulty with the authentication
» Deploying Galaxy in a secured environment: Strong restrictions

For more details :

» Official website of EOSC-Pillar: https://www.eosc-pillar.eu
» UC6.6 description: https://www.eosc-pillar.eu/use-cases/exploring-reference-data-through-existing-computing-services-bioinformatics-community
» UC6.6 Galaxy demo: https://www.youtube.com/watch?v=SK9C_MWz6Yk

Participating Institutions: