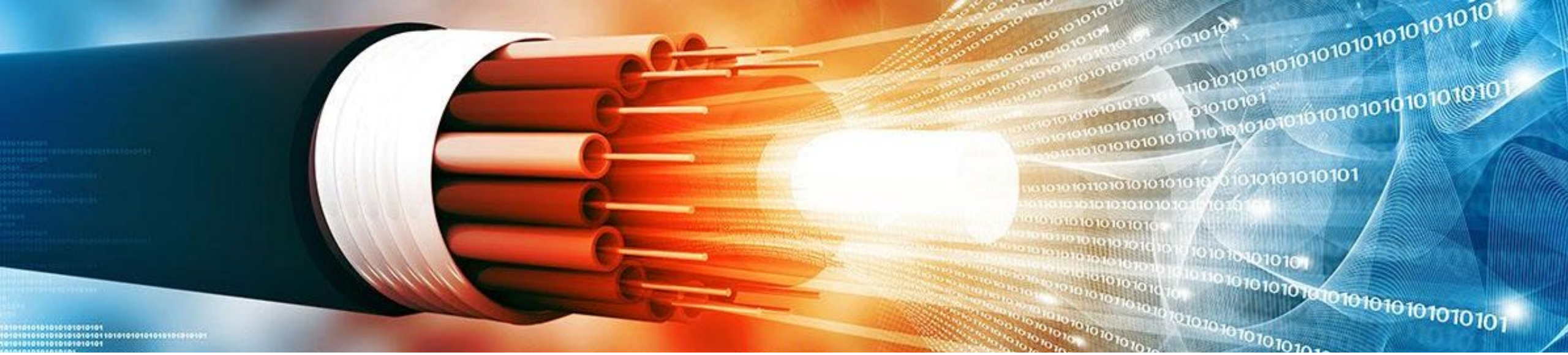


EOSC NORDIC

Cross-border Computing

Abdulrahman Azab, Bergman Tommi
Helmut Neukirchen

The EOSC-Nordic project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857652.

Cross-border computing

- Problem, e.g.: How to give users from one institution access to an HPC centre from another institution?
  - When you use remote resources: software needs to be available (reproducibility).

- Preferably user-friendly:
  - Web-based scientific portals,
  - Some scientific portals support submitting jobs to HPC clusters,
    - but: not every portal users has a user account for that cluster.

Options to give portal user access to HPC cluster

A. Each portal user has a matching account on that cluster.
   - Tedious: need to apply for each portal user for an account on that cluster.

B. One single "robot' user account on that cluster to submit jobs of *all* portal users.
   - Quota issues: one community can consume the whole HPC resource quota.
   - Security: HPC administrators do not like an account shared between all users.

C. Separate "Robot' user account for *each* scientific community of the portal.
   - Quota shared within community that works anyway together on same problem.
   - Security concerns still apply, but the group of users gets narrowed.

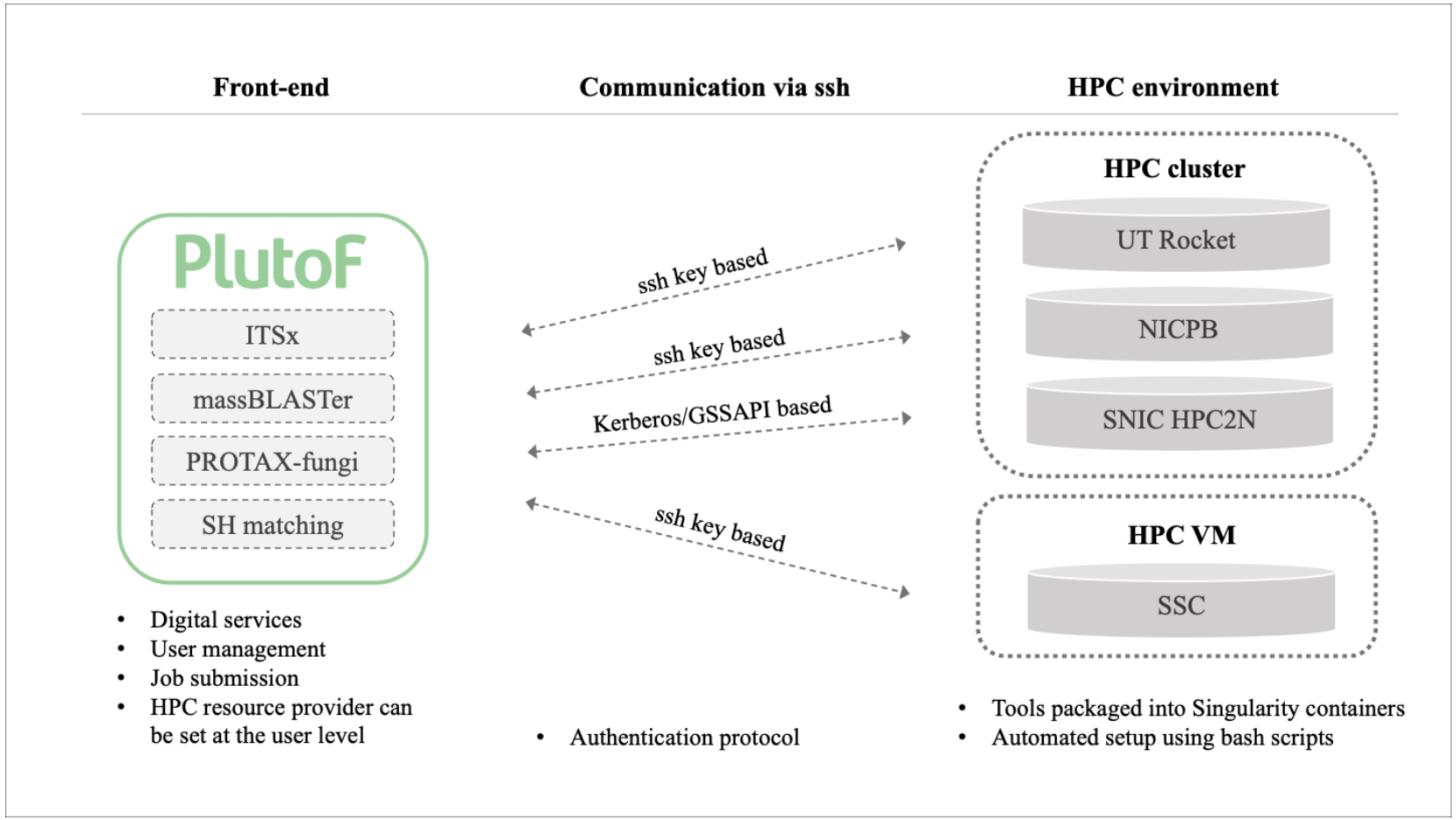
- Both for options B and C, security can be improved:
  - Users never get the credentials to access HPC cluster directly —only via portal.
  - Portal will log for each HPC cluster access who was the responsible portal user.

Pilot: Enhancing PlutoF portal

- PlutoF: a scientific web portal for bioinformatics.
  - Did support already to submit jobs to HPC clusters.

- Added for cross-border computing & reproducibility:
  - Package software to be executed using containers, automated setup (from GitHub).
  - Selecting more than one HPC cluster.
  - Authentication beyond existing ssh support: Kerberos.
  - Logging of which portal users submitted what HPC job via community robot user.
  - Support for clouds (instead of only HPC): create VMs using the containers.
- Tested on Swedish SNIC cloud & HPC and Estonian ETAIS HPC:
  - SNIC HPC does not allow robot user, so used only a 1:1 portal/HPC users mapping.

Pilot: Enhancing PlutoF portal



- Services packaged into Singularity containers.
- Support for sending analysis jobs to different HPC clusters and HPC VMs was added.
- Worked out recommended procedures on how users can apply for HPC resources and how to set up access to EOSC HPC resources.

Enhancing Galaxy Climate portal

- Galaxy is a generic scientific web portal:
  - Galaxy Climate Europe running on European Galaxy instance.
  - Did support already workflows, packaging, remote jobs (cloud & HPC).
- Added for cross-border computing & reproducibility:
  - Packaged software based on EOSC-Life best practises (Conda, containers).
    - Scientific collaboration agreement between EOSC-Life and EOSC-Nordic.
  - Automated setup (fetch from GitHub).
  - Exposed remote storage resources (S3 API) to run jobs independently from location.
  - Configure Galaxy to use cross-border cloud resources.
- Tested on Finnish CSC cloud (cPouta) & Czech CESNET cloud.
  - No HPC, because the targeted HPC clusters did not allow robot accounts.

Summary and Issues

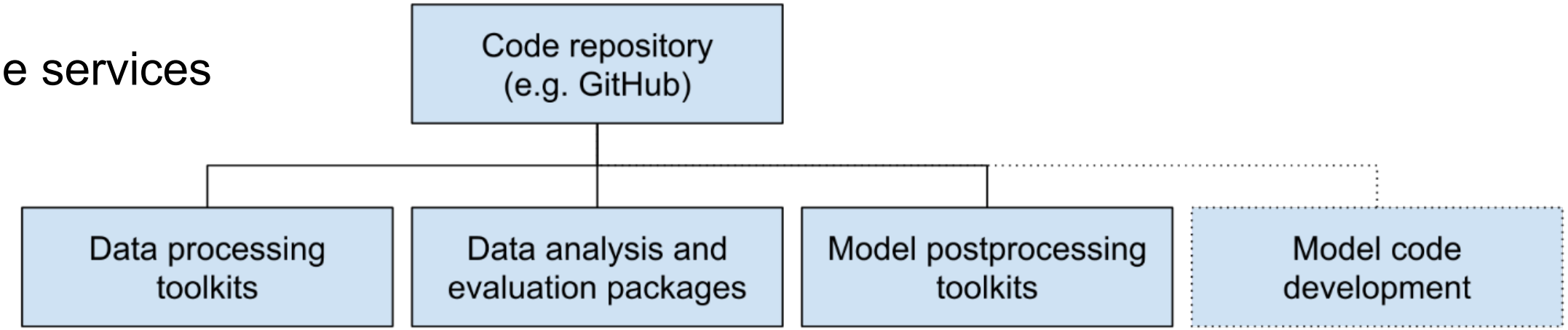
- Summary:
  - All software is FOSS+ documented procedures (on GitHub),
  - Improved reproducibility:
    - Rigorous use of container and automated setup from GitHub.
  - Robot account to access HPC cluster on a per-community-basis:
    - Logging which portal user is using the HPC cluster to address security-concerns.

- Unsolved issues:
  - Convince HPC centres within EOSC to allow community-specific robot users.

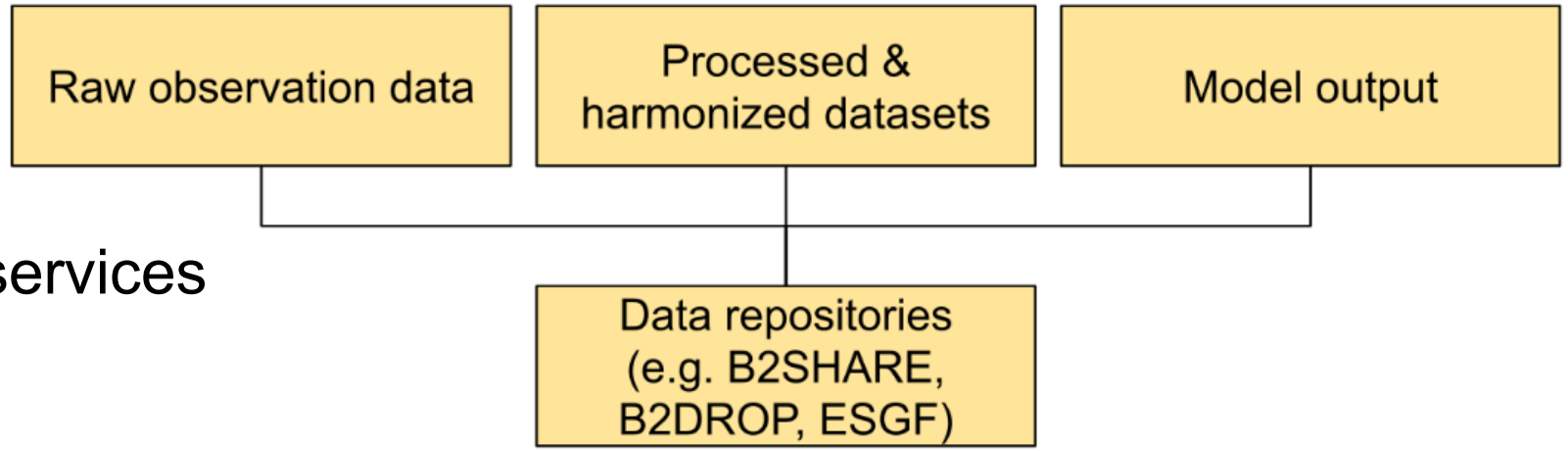
Abarenkov, Fouilloux, Neukirchen, Azab: Reproducible Cross-border High Performance Computing for Scientific Portals, 2nd Workshop on Reproducible Workflows, Data Management, and Security. eScience, Salt Lake City. 11-14 October 2022. To appear. DOI: 10.48550/arXiv.2209.00596 (based on Deliverable D5.2)

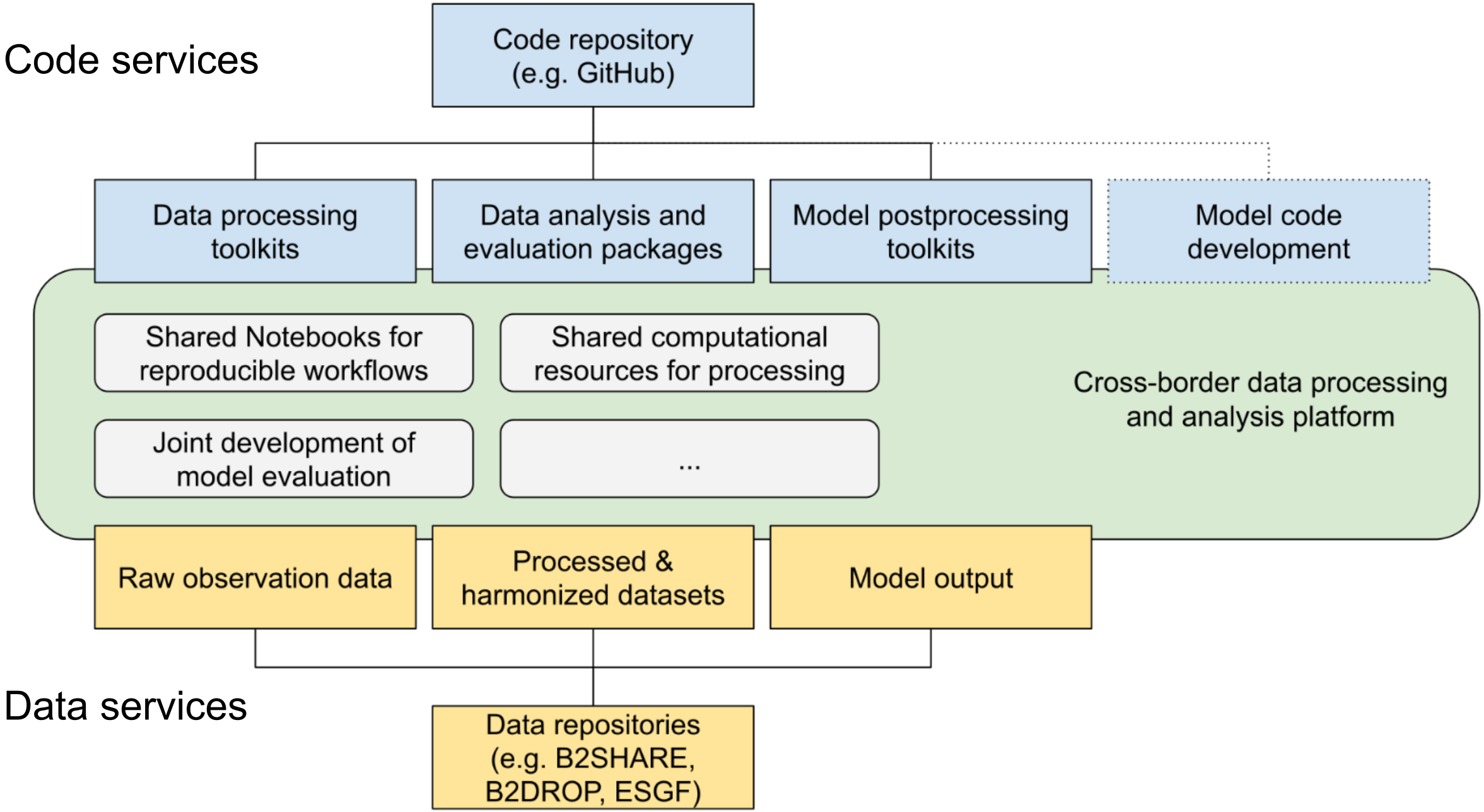
Cloud-based computation and analysis

Code services



Data services




Code services
Code repository (e.g. GitHub)
Data processing toolkits
Data analysis and evaluation packages
Model postprocessing toolkits
Model code development
Shared Notebooks for reproducible workflows
Shared computational resources for processing
Joint development of model evaluation
...
Cross-border data processing and analysis platform
Raw observation data
Processed & harmonized datasets
Model output
Data services
Data repositories (e.g. B2SHARE, B2DROP, ESGF)

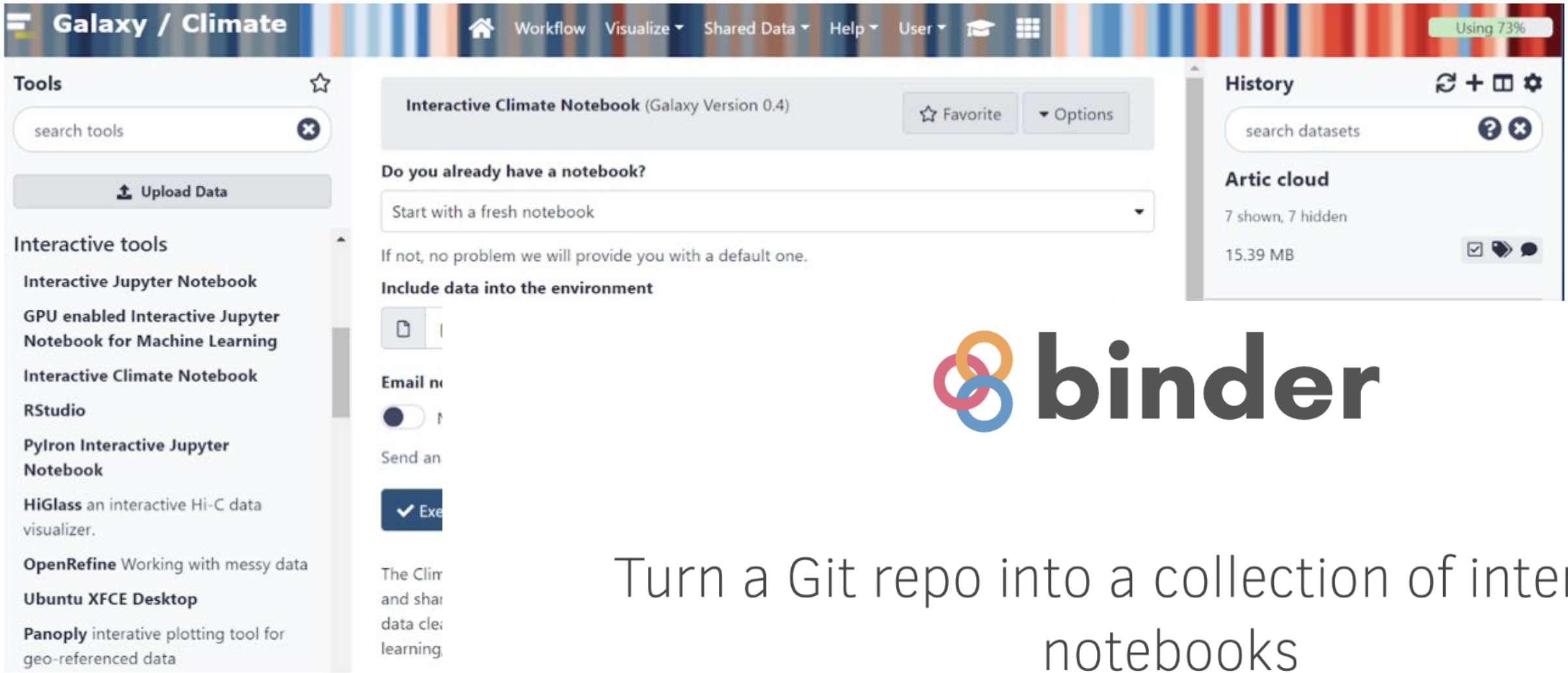


Fig. 2: Galaxy Interfac

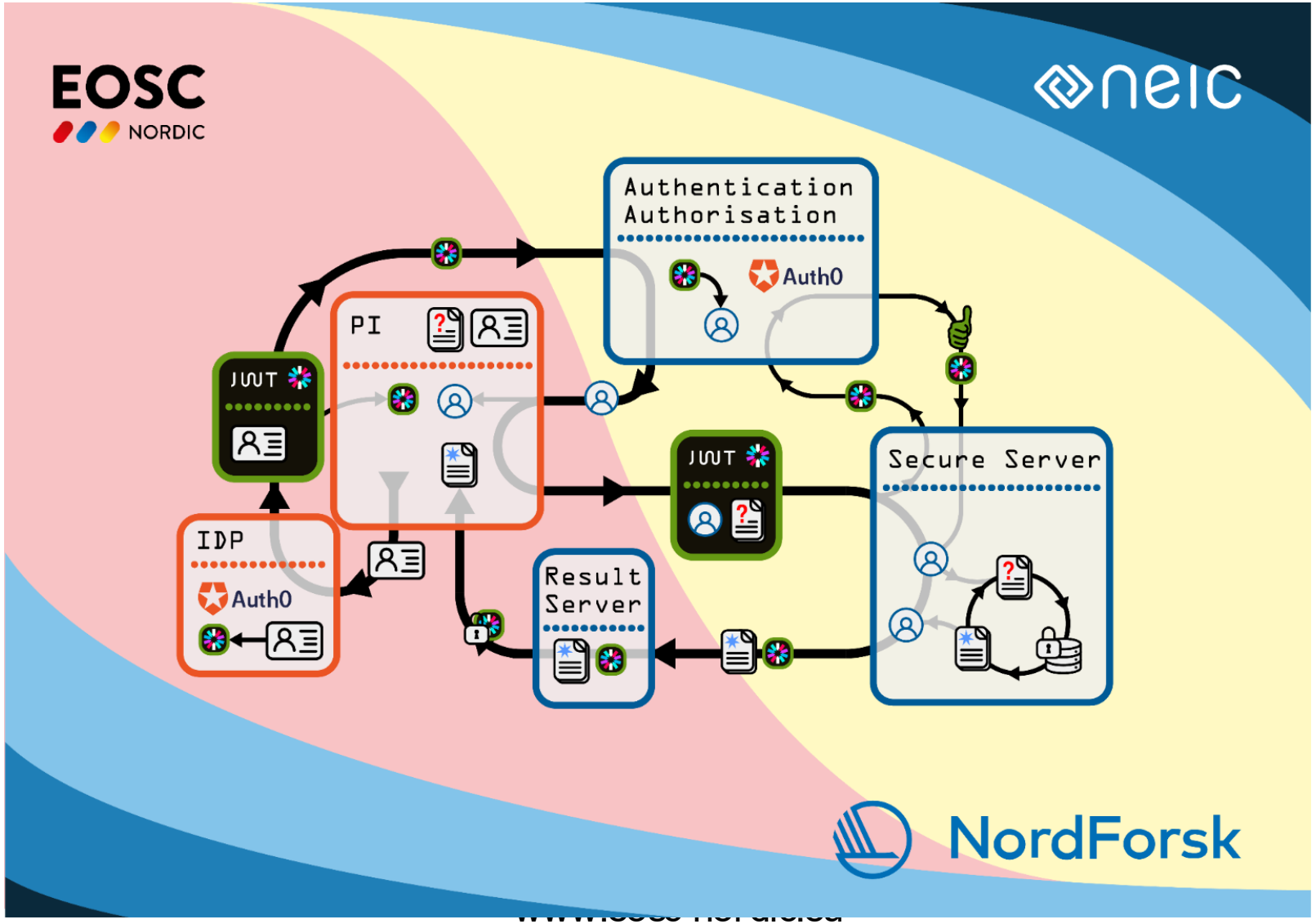
binder

Turn a Git repo into a collection of interact notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks executable environment, making your code immediately reproducible by anyo anywhere.

B2DROP, ESGF)

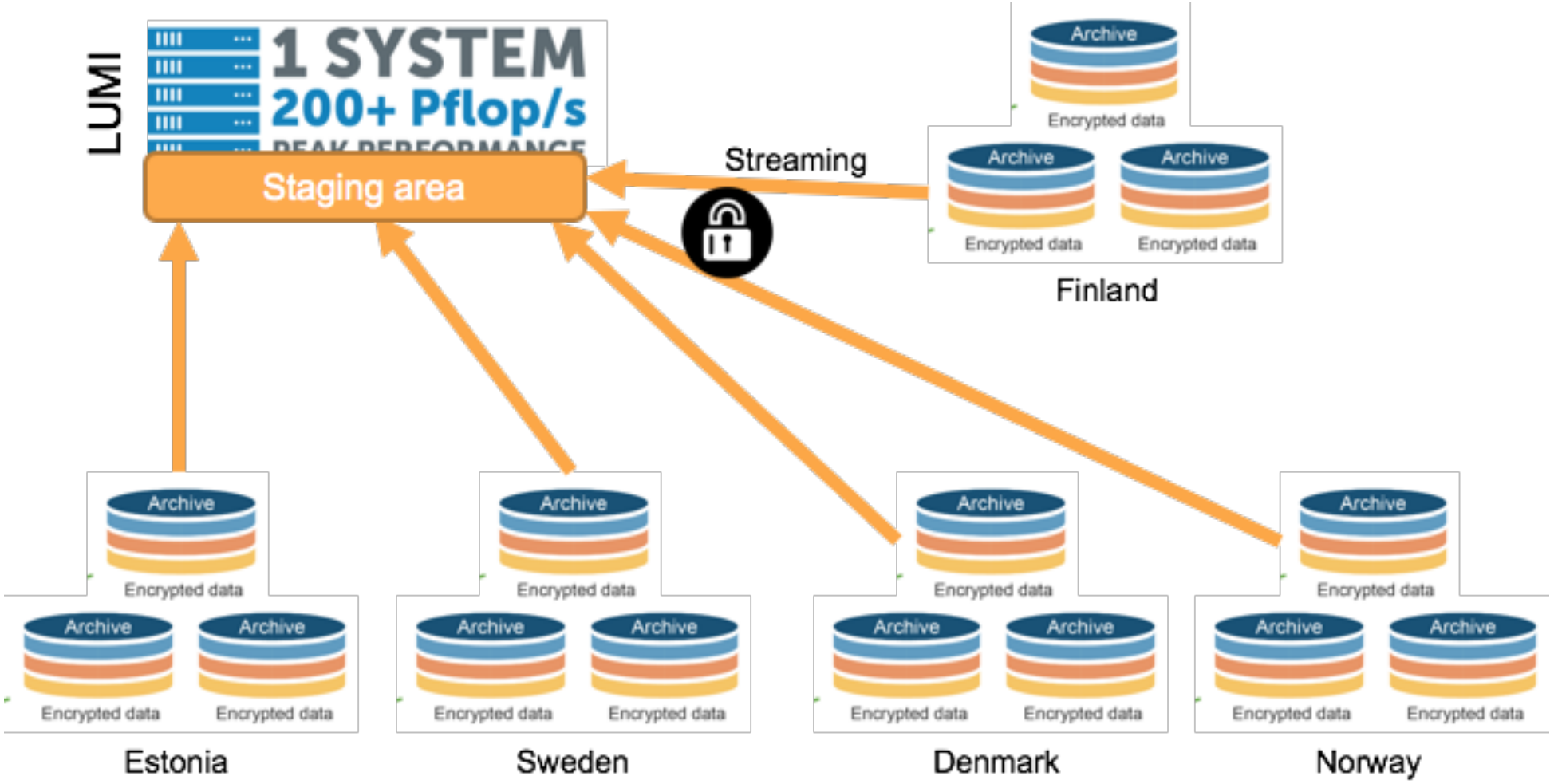
Federated access to sensitive data



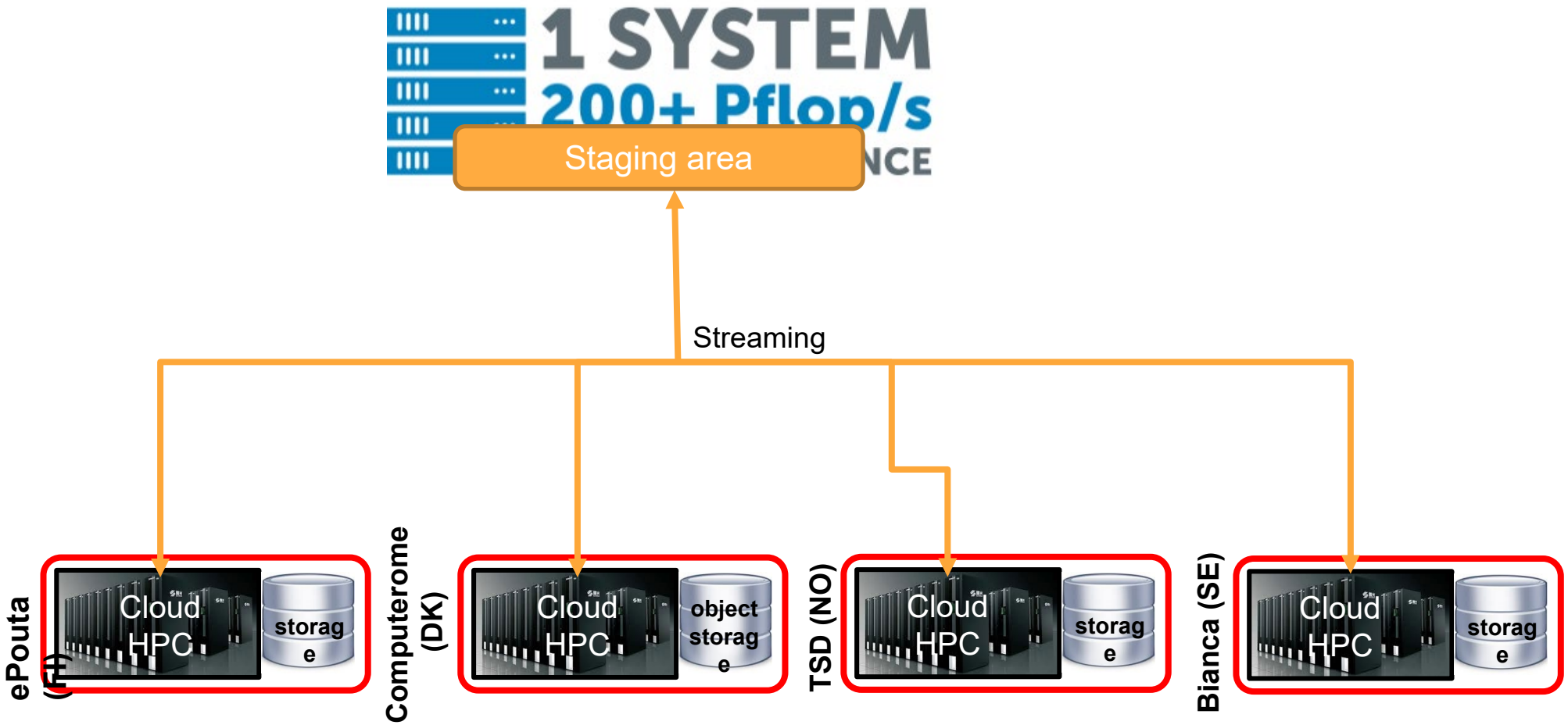
Unresolved issues

- Level of assurance
- Identity verification
- Secure data transfer

LUMI for sensitive data - vision



LUMI for sensitive data - vision



Ongoing work

- Solution for Access control (AAI)
- Solution for workload isolation
- Solution for tool porting mobility

People:

- Tewodros Deneke (CSC)
- Anne fouilloux (UiO)
- Kessy Abarenkov (UT)
- Jean Iaquinta (UiO)
- Helmut Neukirchen (UIce)
- Ernir Erlingsson (UIce)
- Andrey Kutuzov (UiO)
- Sabry Razick (UiO)
- Lorand Szentannai (Sigma2)
- Bergman Tommi (FMI)
- Abdulrahman Azab (NeIC/UiO)

Q&A

EOSC-Nordic WP5

wp5@eosc-nordic.eu

EOSC NORDIC

www.eosc-nordic.eu

https://twitter.com/EOSC_Nordic

https://www.linkedin.com/groups/13756550/

https://www.linkedin.com/groups/13756550/