EuroHPC Summit

2023 Göteborg

Federated HPC, cloud and data infrastructures

Introduction

Dirk Pleiter (KTH), 2023-03-21

Goals of this Session

- Create an understanding of different aspects of federation and related R&I topics
- Look at concrete use cases
- Initiate a discussion on how to realise a federation of HPC-based infrastructures in Europe

Speakers

Use cases

- Sandra Diaz (FZJ)
- Xavier Espinal (CERN)

Technical topics

- Nicolas Liampotis (GRNET)
- Javier Bartolome (BSC)
- Anders Sjöström (LUND)
- Utz-Uwe Haus (HPE)
- Enzo Capone (GEANT)

Federated HPC, cloud and data infrastructures

Leveraging Fenix for Brain Research Workflows

Sandra Diaz (FZJ), 2023-03-21

- EBRAINS -- European infrastructure for Brain Research created by the EU-funded Human Brain Project (HBP)
- Comprises a set of tools and services addressing a variety of requirements from the neuroscience community
  - User interfaces
  - Platform middleware
  - Scientific tools

Human Brain Project | Co-funded by the European Union | EBRAINS

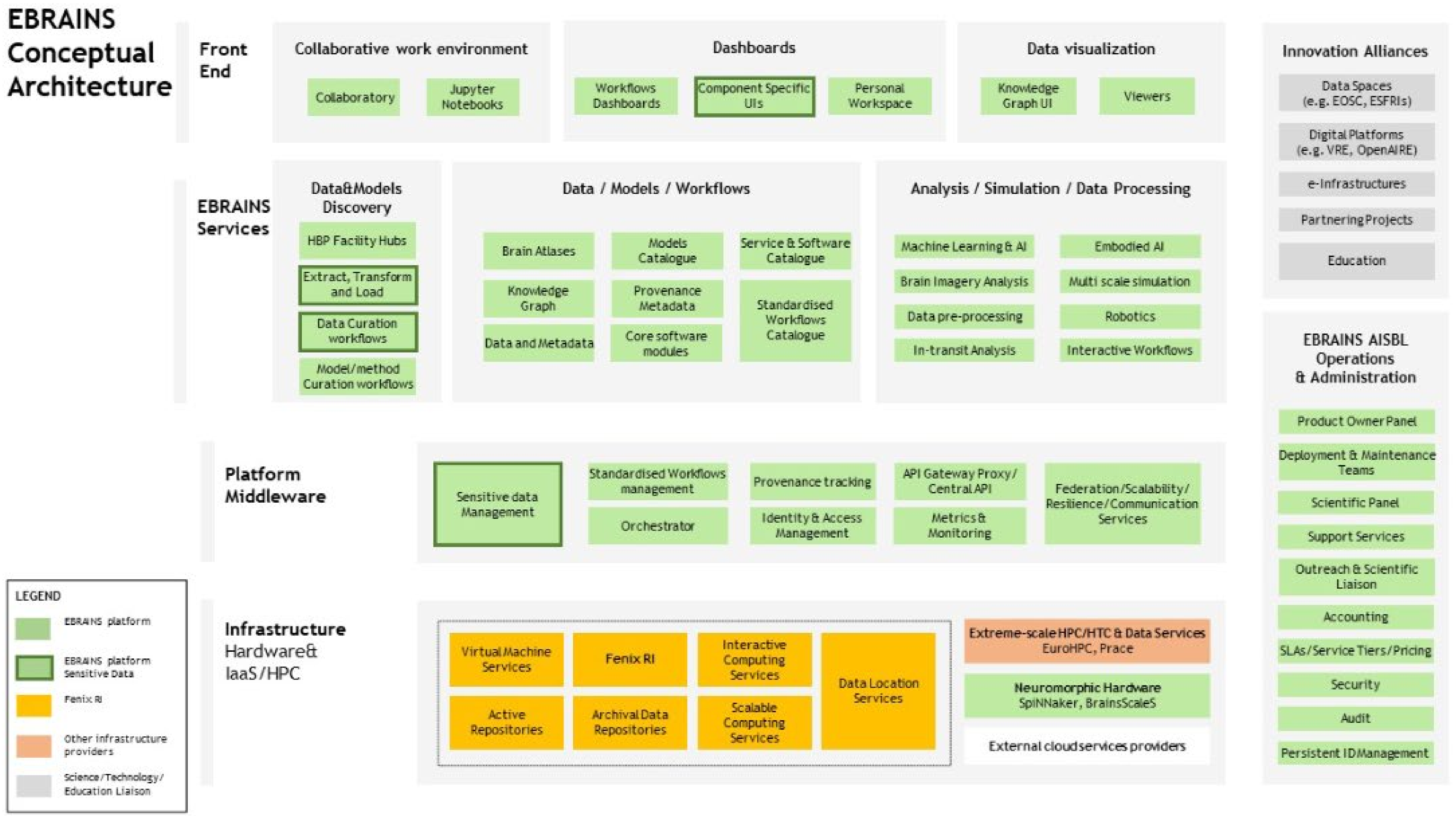
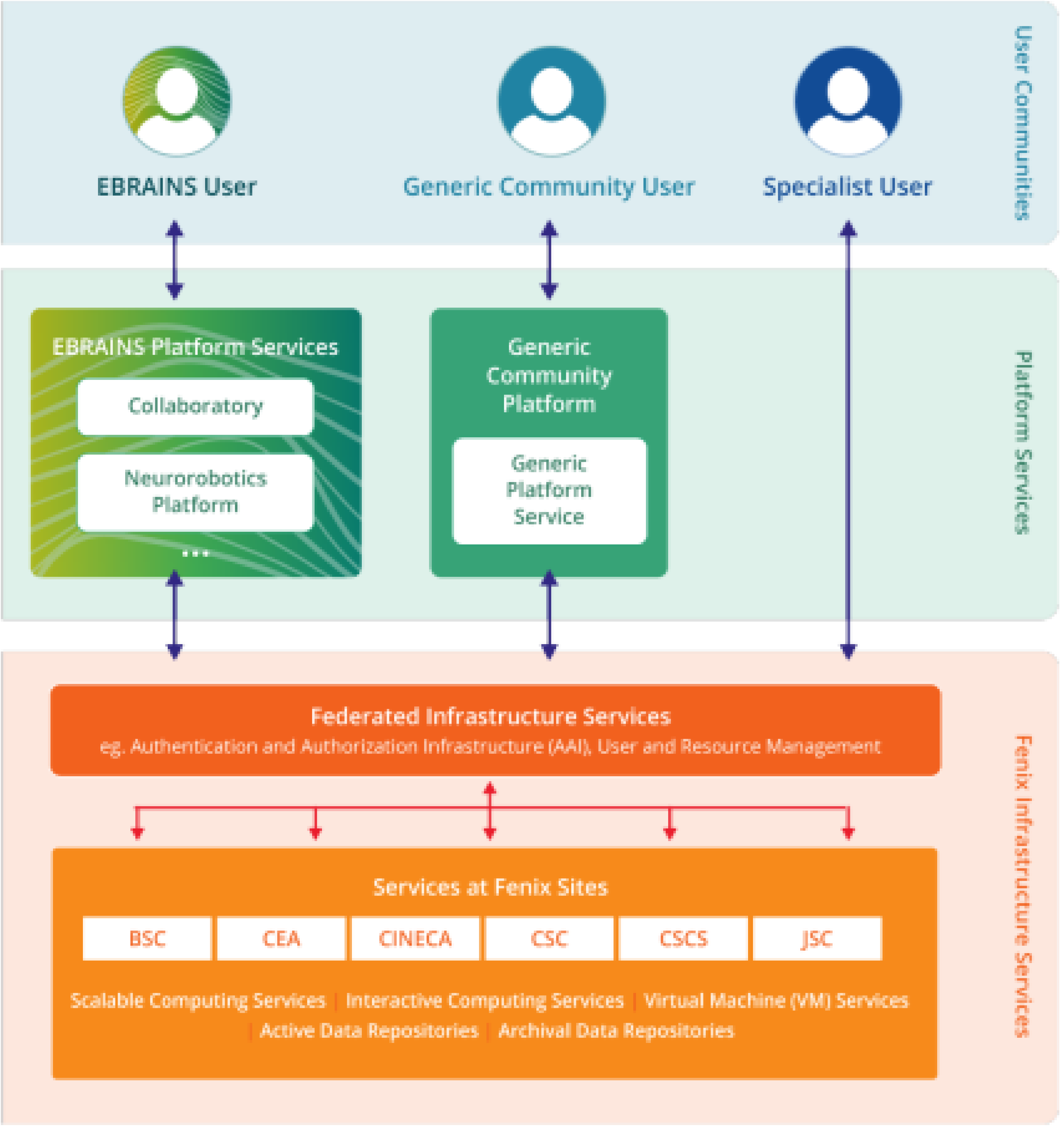


Figure 3: EBRAINS Conceptual Architecture

D5.4 (D51) SGA3 M21 SUBMITTED 211220.docx | PU = Public | 20-Dec-2021 | Page 32 / 86

https://sos-ch-dk-2.exo.io/public-website-production-2022/filer_public/02/2d/022d2b2e-a8ec-4caa-8db3-d8e4e925b59a/d54_d51_sga3_m21_accepted_220520.pdf https://fenix-ri.eu/



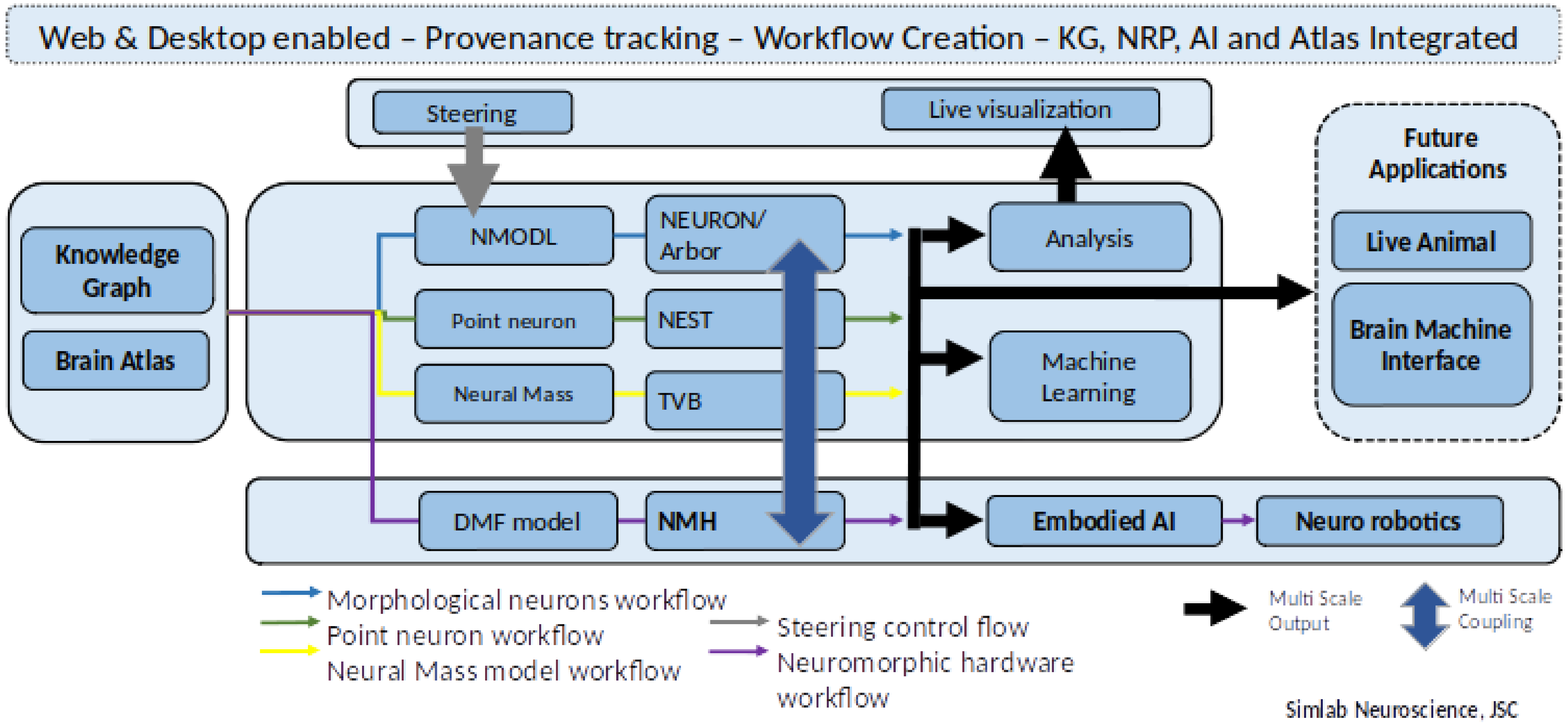
- EBRAINS users can access the Federated Infrastructure Services offered by FENIX to execute complex scientific compute and data workflows

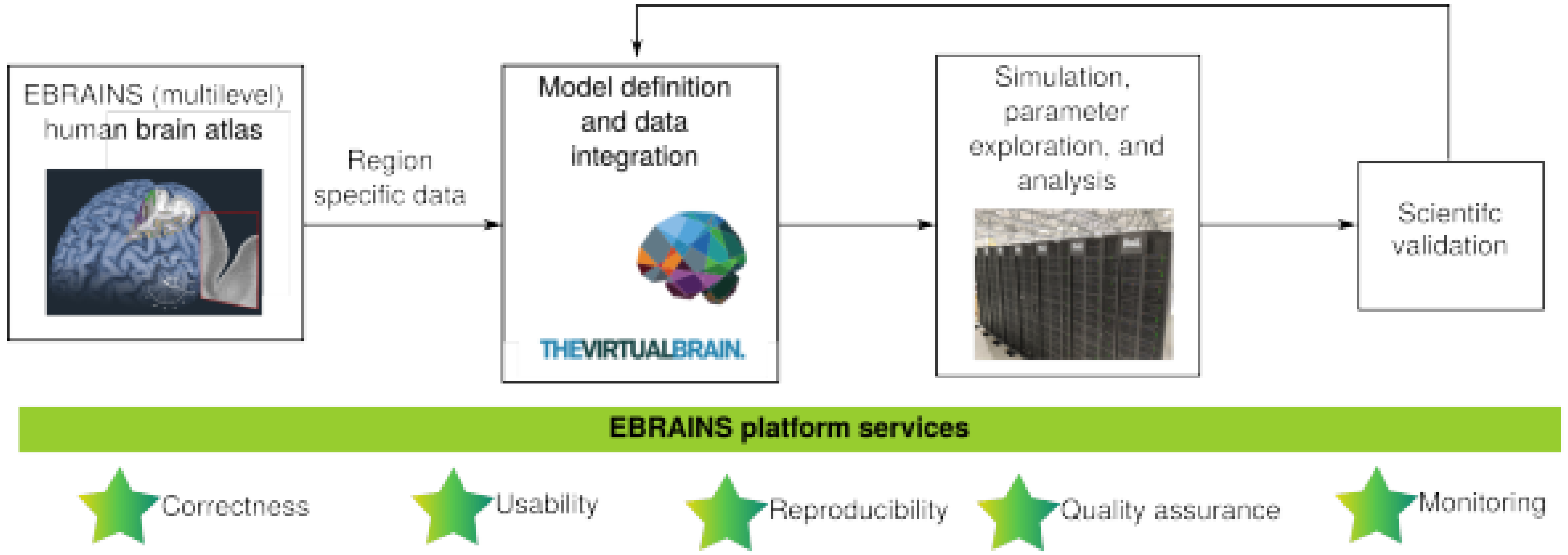
EBRAINS usage of Federated infrastructure

- Store and access large amounts of data
  - Long term storage and sharing using object storage
  - File storage close to computational and visualization resources
- Cloud infrastructure
  - Used to deploy specific scientific tools as a service as well as platform services like collaborative work environments, information catalogues, image services, etc.
- High performance computing
  - Opens a new avenue to simulate, optimize, visualize and integrate brain models at all scales
  - Larger and more complex models and workflows are emerging – new science
  - Scientific software developed to leverage hardware accelerators
  - Interactive computing services

Brain modelling workflow

- Explore, visualize, query and import data at different spatial and temporal scales -- Knowledge graph and brain atlas (Large data)
- Generate models at different scales and workflows (Cloud)
- (Co-)simulate the models using different dedicated simulators (HPC and NMH)
- Analyse, optimize and connect to applications in robotics, BCI and experimental neuroscience (HPC or dedicated modules)
- Orchestration, monitoring and steering (Cloud and HPC)





Showcase 1 and 2 – Human Brain Project

FENIX RESEARCH INFRASTRUCTURE

EBRAINS usage of Federated infrastructure - advantages and future

- Uniform access
  - EBRAINS users can deploy workflows on any of the FENIX sites using UNICORE
  - FENIX AAI
- Building a new scientific community of HPC users
  - Sharing and working on the same workflows on different systems is possible
  - Enhances the integration of expertise and cooperation between research groups
  - Uncomplicated access to compute and data infrastructure
  - Support and examples help get the community into the new infrastructure usage patterns
- Future
  - More homogeneous deployment infrastructure between sites (containers)
  - Homogeneous accounting and project setup (FURMS)
  - API for easy software testing and deployment on all sites
  - GDPR compliance for sensitive data processing
  - Build a platform for education in all related fields to neuroscience

EuroHPC Summit

2023 Göteborg

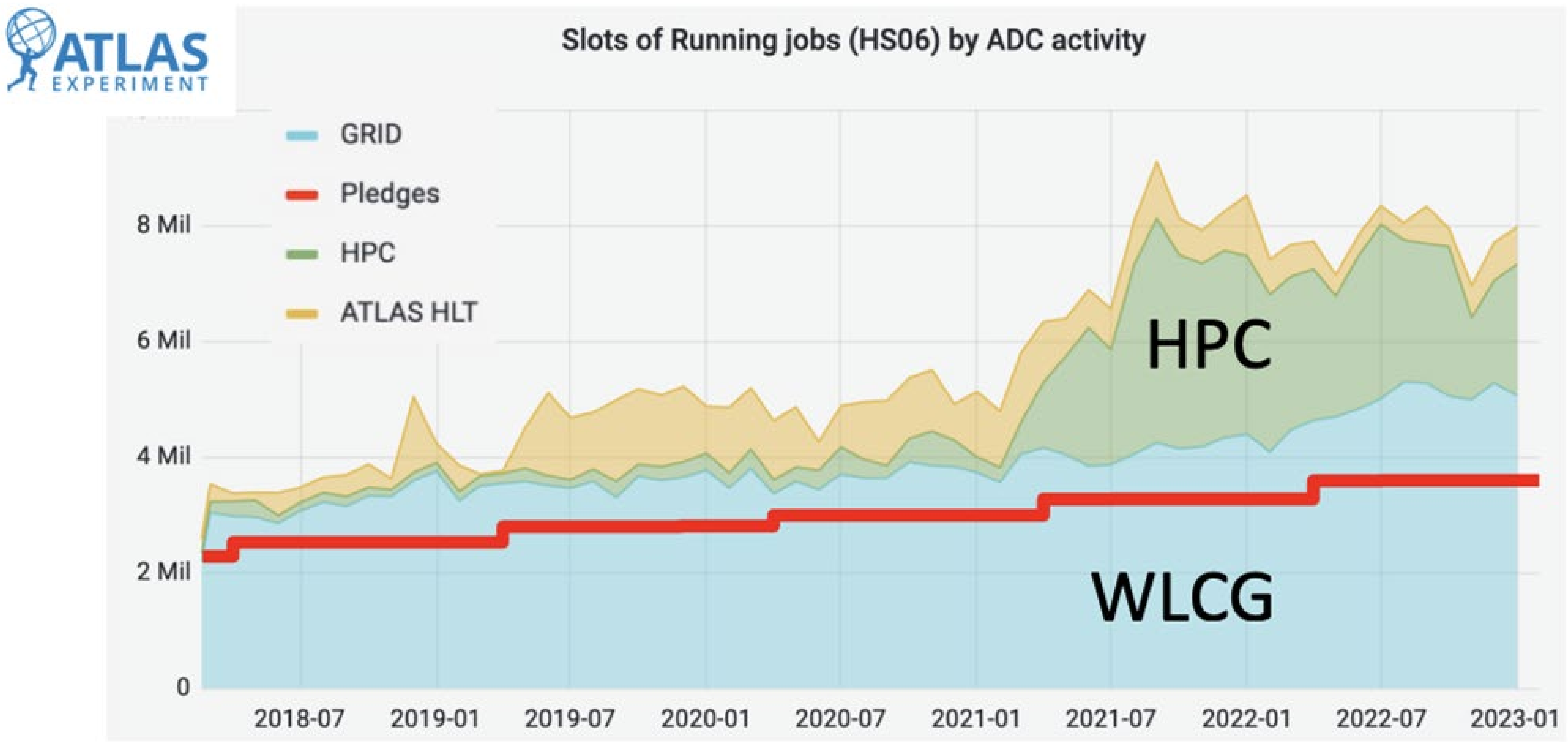
Federated HPC, cloud and data infrastructures

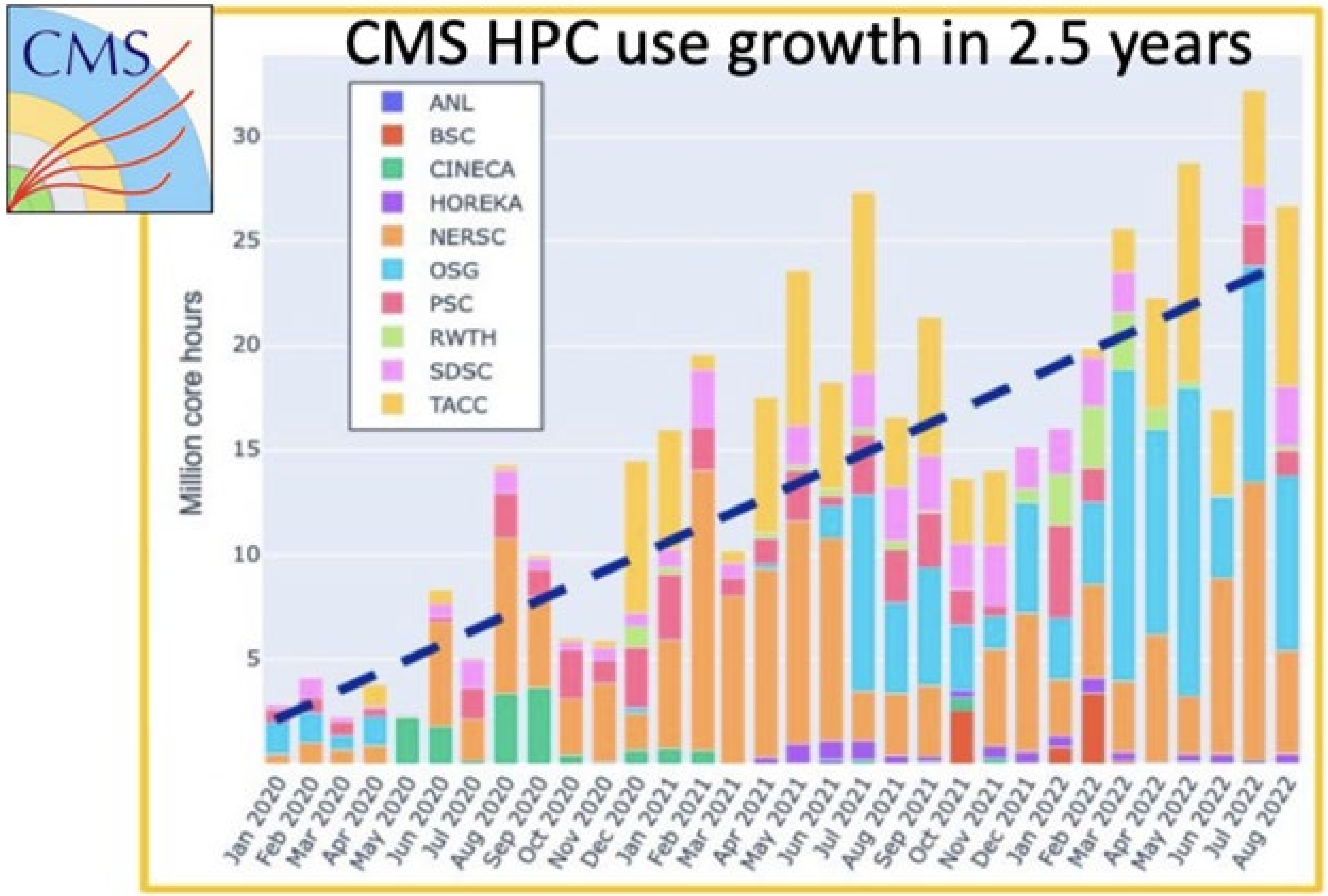
HPC-based Data Processing in Particle Physics and Astronomy

Xavier Espinal (CERN and ESCAPE), 2023-03-21

Introduction (1/2)

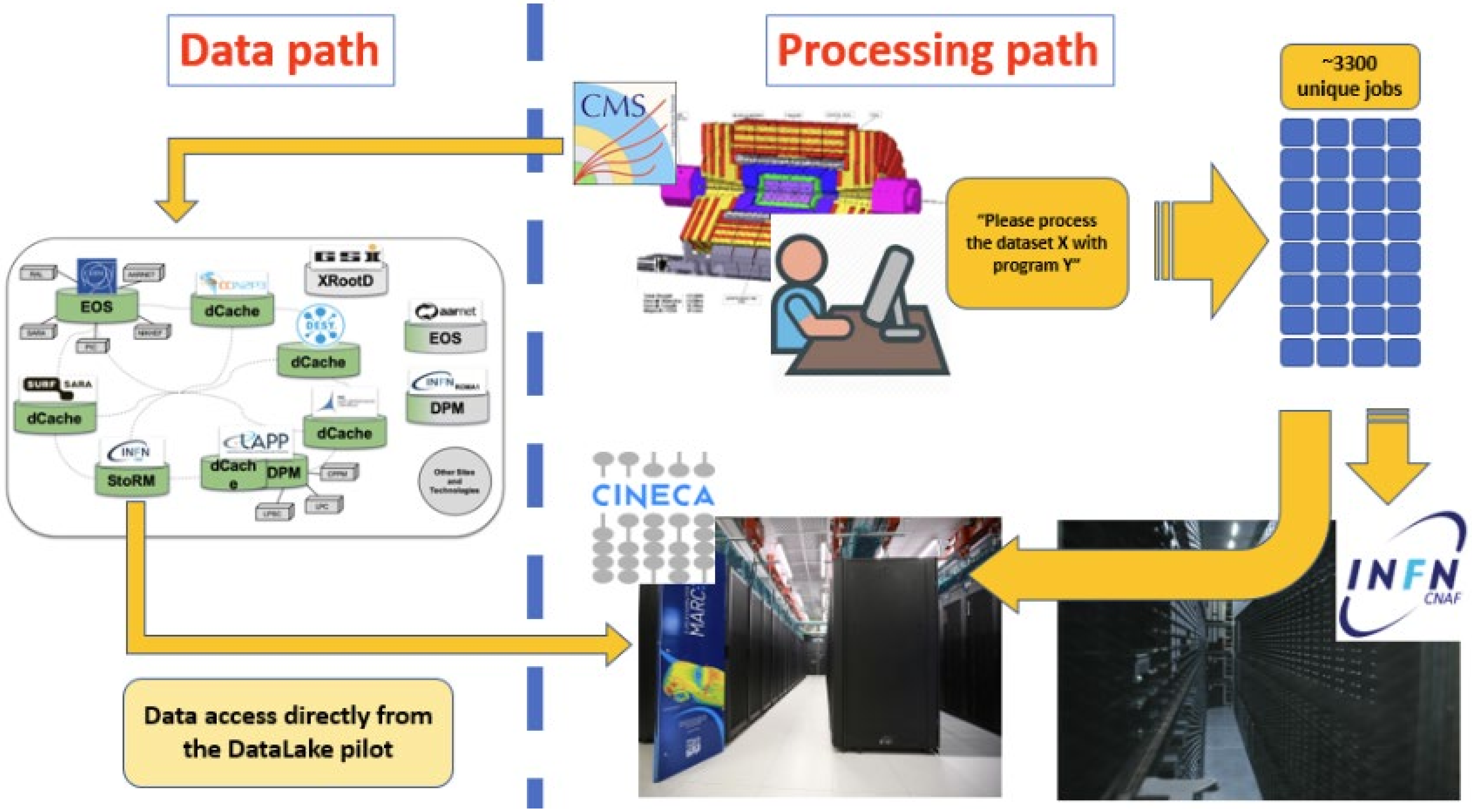
- Astronomy and HEP see potential large benefits in exploiting HPCs
- Substantial technical investment during the last years which increased its usage
- The use of HPC facilities increased considerably in the last years

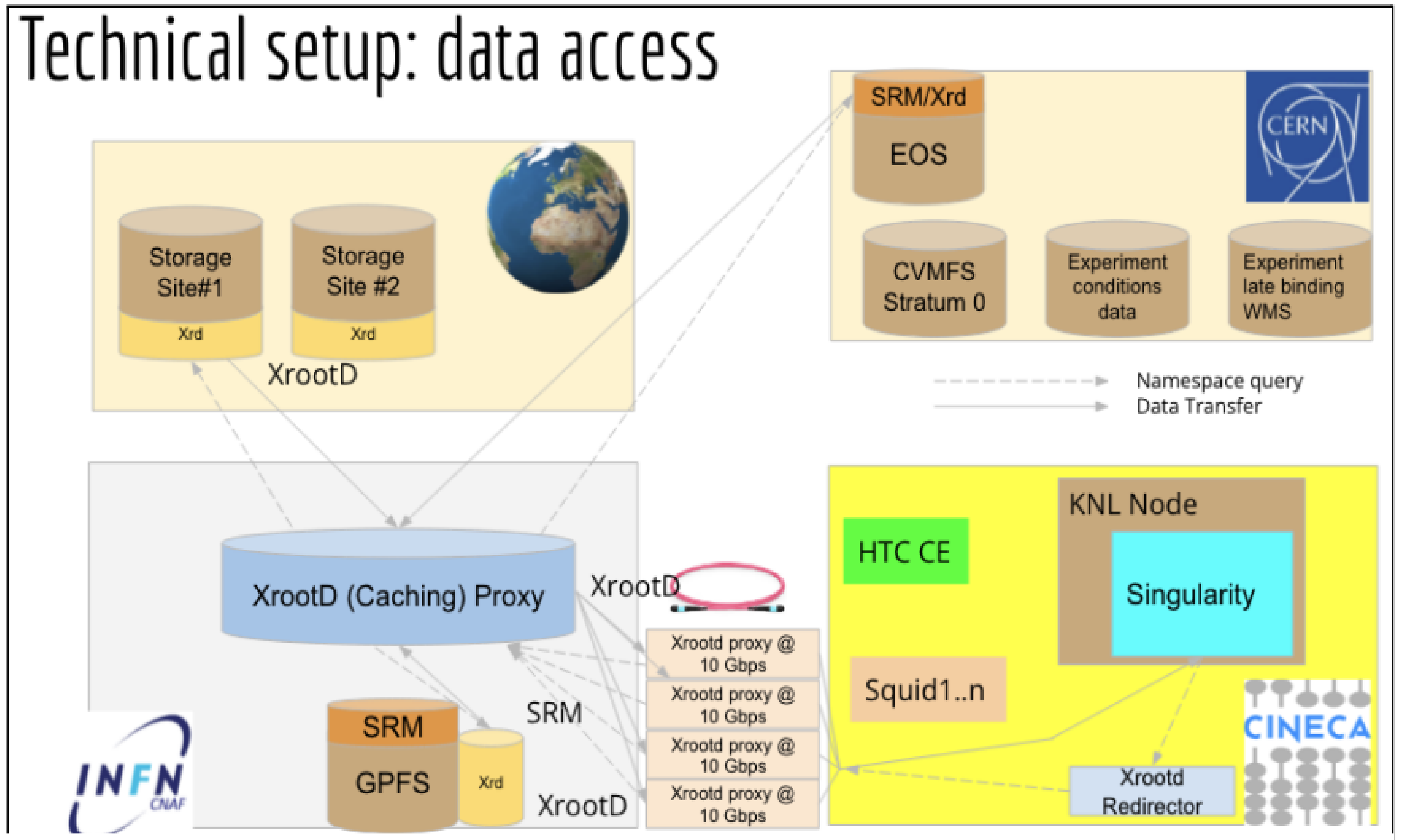




Introduction (2/2)

- Integration of HPC centers as extensions of sites providing storage and cpu to the experiments is the so far a successful approach[1]
- Standing collaborations and joint work eg. WLCG, ESCAPE, FENIX, InterTwin, EuroHPC. Instrumental in gaining experience together





[1]Example: Marconi A2 with XCache was used at the time of ESCAPE as CNAF (WLCG Tier-1) extension. Tier1 manage the WLCG storage. The transparent extension make the experiment operations much easier as you might see. From the infrastructure perspective this is fully in line with the DataLake

Use Cases (1/2)

- CPU usage, ‘standard candles’ workloads
  - Physics process simulation: 80-95% CPU/wall (CPU intensive)
  - Interaction with detectors/reconstruction/reprocessing: 50-80% CPU/wall (CPU and IO intensive)
  - Derivation and analysis: 30-80% CPU/wall (IO intensive)
  - Training/inference of machine learning models (GPU intensive)
- Architectures
  - Software largely developed and build for x86 CPUs, eg. processing events in parallel in HEP - jobs originally code single-threaded)
  - Successful efforts porting/re-writing multi-threaded versions on modern multi-core nodes and successful efforts to port to non-x86 (GPUs, ARM, and Power)
- Software distribution
  - Large software stacks with quick release cycles, many versions/releases in uses simultaneously
  - Heavily relying on CVMFS[2] for software distribution. Mounted as a read-only file system and http-sync-ed at node level (daemon), typical O(10 GBs)
- Identity management:
  - Common trust is fundamental.
  - Integration with Data Management and Workload Management systems, provide user access, token-based? IdP federations?

Use Cases (2/2)

- Workload Management
  - Jobs getting better to run on multi-node/multi-core. Change of paradigm, from independent-few-core nodes (classical sites) vs interconnected-multi-nodes (HPC)
  - Integration with experiments workload management systems is required:
    - Compute-Edge service? interfacing experiment job-distribution system and HPC batch systems
  - For HPCs push mode (fully defined jobs with data preplaced) favoured over pull mode (pilot/fetch workload)
- Data Management and Data Access
  - Applications require input data.
    - Producing output as a result of the computing task. Typical IO rates per core can vary depending on the workflow, from O(100KB/s) O(few MB/s)
  - Data access possibilities:
    - Remote streaming (possibly via a latency-hiding layer, cache/buffer)
    - Managed cache or “edge” service
      - Downloads to local cache from remote storage (Data Lake)
      - Possible to integration with experiment Data Management frameworks (managed cache)
    - Static (dedicated) storage
  - Other-data access requirements
    - Access to auxiliary data (e.g. calibrations) potentially in remote locations (antennas, telescopes) input/output

Next steps

- Current challenges and constraints
  - Effort is spent integrating the different machines as single entities, requiring specific integration strategies and developments.
  - Access and usage policies, available services, system architectures and machine-lifetime.
  - Resource allocation and resources availability: burst vs. the preferred continuous usage

- Goal (dream?)
  - Towards a General Purpose HPC by design? Common model “architecture”?
    - HPC machines are integrated and used as an alternative and standard backend, together with local-batch system, computing grids or hyperscalers.
    - Allow to flexibly and elastically expand the resources available to the experiments and the scientists performing analysis

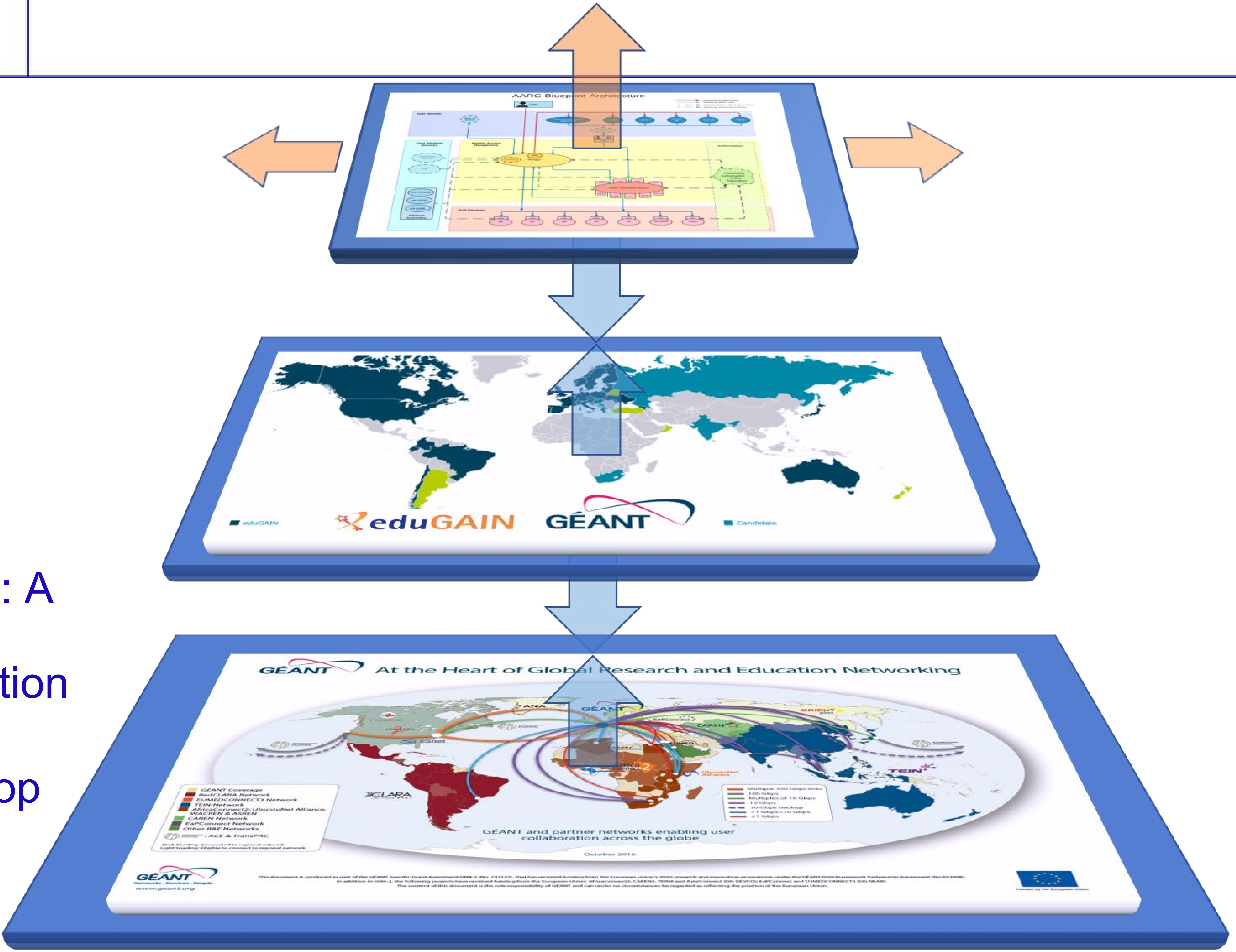
Federated HPC, cloud and data infrastructures

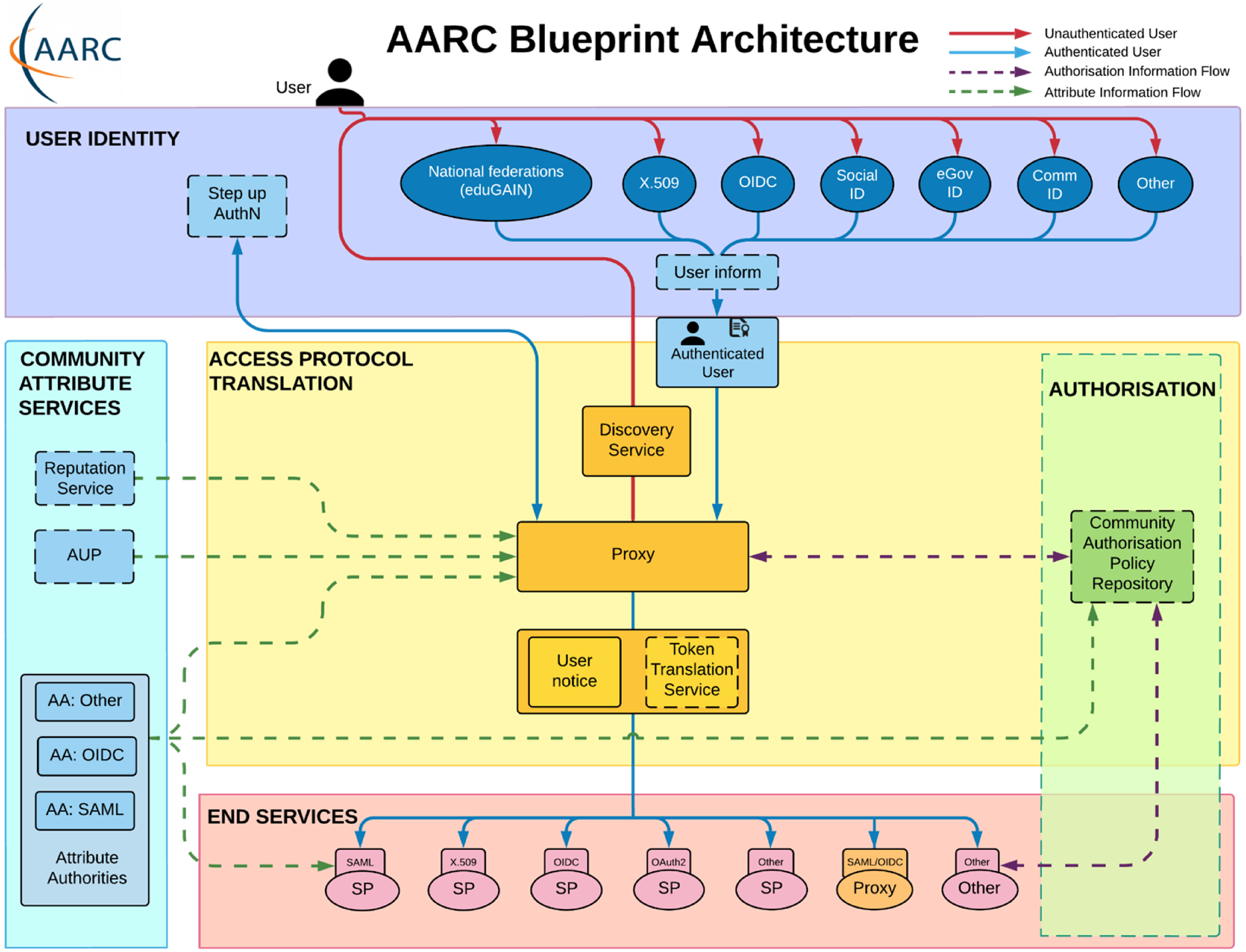
Authentication and authorisation infrastructure

Nicolas Liampotis (GRNET), 2023-03-21

AARC Blueprint Architecture

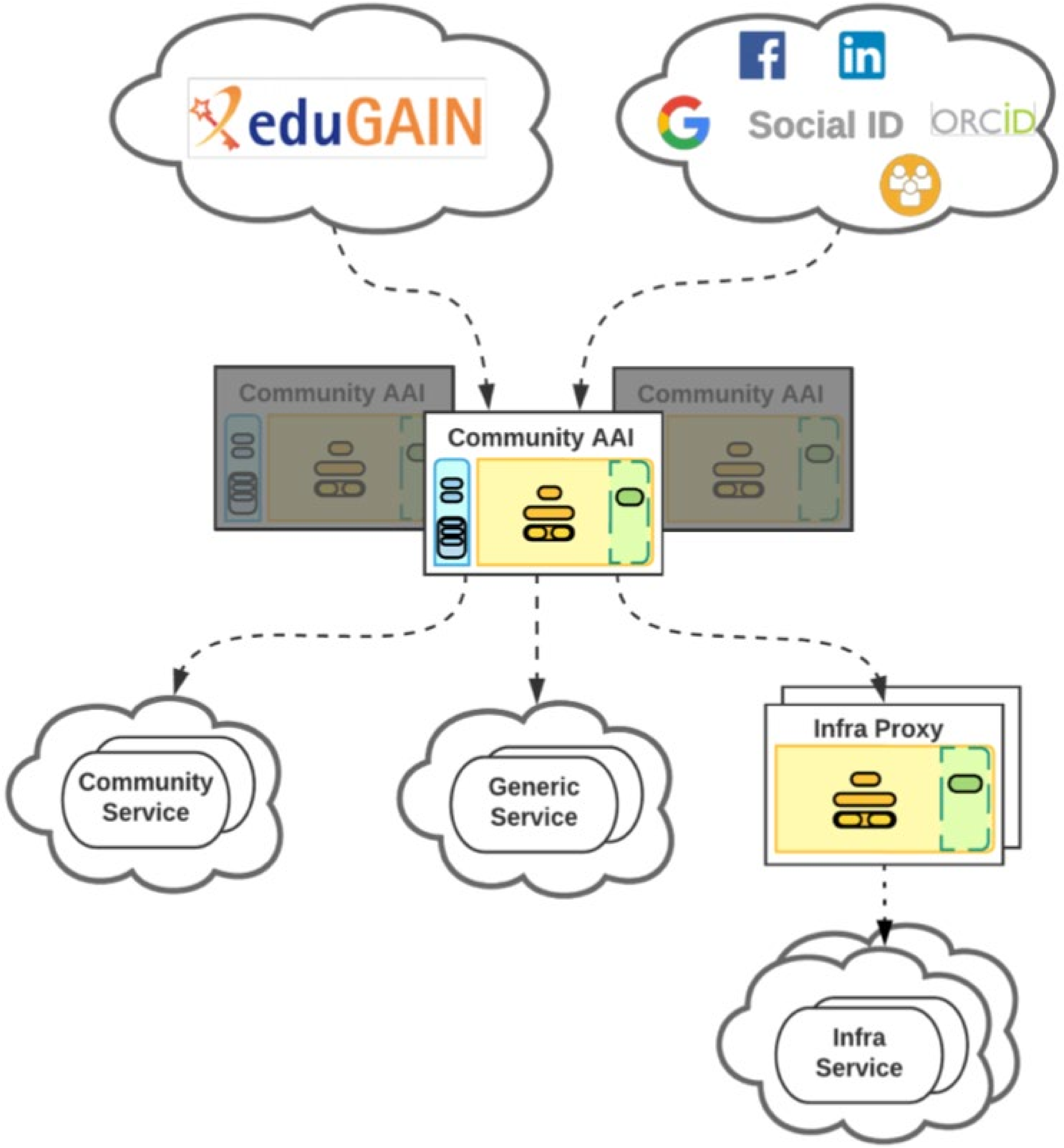
- eduGAIN and the Identity Federations
  - A solid foundation for federated access in Research and Education
- AARC Blueprint Architecture: A reference architecture for authentication and authorisation
  - A set of architectural and policy building blocks on top of eduGAIN





https://aarc-community.org/architecture

- **User Identities** – Services for the identification and authentication of users
- **Community Attribute Services** – Components related to managing and providing information (attributes) about users
- **Access Protocol Translation** – Single integration point between the Identity Providers from the User Identity Layer and the Service Providers in the End Services Layer
- **Authorisation** – Components for controlling access to services and resources
- **End-services** – The services and resources users want to use



Community AAI

Streamlines researchers' access to services, both those provided by their own infrastructure as well as the services provided by infrastructures that are shared with other communities

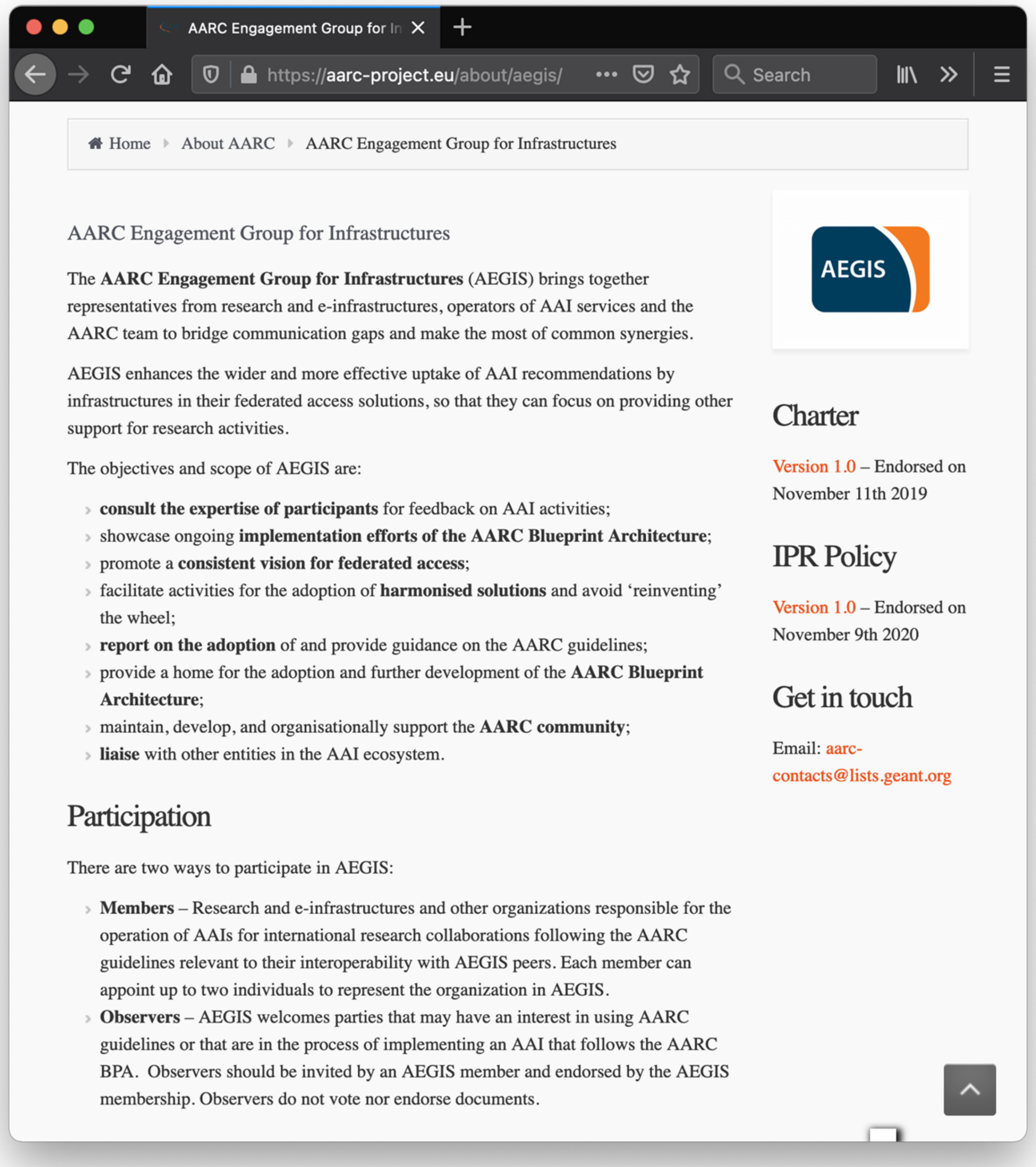
Infrastructure Proxy

Enables Infrastructures with a large number of resources to provide them through a single integration point, where the Infrastructure can maintain centrally all the relevant policies and business logic for making available these resources to multiple communities

AARC Interoperability Guidelines Approved by AEGIS

Created by Christos Kanellopoulos, last modified by Nicolas Liampotis on Jan 14, 2022

#	Document	AARC Identifier	Date first presented	Date approved	Status
1	Guidelines on expressing group membership and role information	AARC-G002	2017-11-13	2017-11-15	Current
2	Exchange of specific assurance information between Infrastructure	AARC-G021	2018-03-12	2018-03-12	Current
3	Guidelines for evaluating the combined assurance of linked identities	AARC-G031	2018-05-14	2018-07-09	Current
4	Specification for expressing resource capabilities	AARC-G027	2018-12-10	2018-12-10	Current
5	Implementing scalable and consistent authorisation across multi-SP environments	AARC-I047	2019-03-11	2019-03-11	Current
6	A specification for IdP hinting	AARC-G049	2019-03-11	2019-04-08	Superseded by AARC-G061
7	Guidelines for expressing affiliation information	AARC-G025	2019-03-11	2019-10-14	Current
8	AARC Blueprint Architecture 2019	AARC-G045	2019-11-11	2020-02-10	Current
9	Inferring and constructing voPersonExternalAffiliation	AARC-G057	2020-07-13	2021-02-08	Current
10	A specification for IdP hinting	AARC-G061	2020-05-11	2021-02-08	Current
11	Guidelines for expressing community user identifiers	AARC-G026	2019-09-09	2021-06-14	Current
12	Specification for hinting an IdP which discovery service to use	AARC-G062	2021-09-13	2021-10-11	Current


AARC Engagement Group for In
https://aarc-project.eu/about/aegis/
Search

Home › About AARC › AARC Engagement Group for Infrastructures

AARC Engagement Group for Infrastructures

The **AARC Engagement Group for Infrastructures** (AEGIS) brings together representatives from research and e-infrastructures, operators of AAI services and the AARC team to bridge communication gaps and make the most of common synergies.

AEGIS enhances the wider and more effective uptake of AAI recommendations by infrastructures in their federated access solutions, so that they can focus on providing other support for research activities.

The objectives and scope of AEGIS are:

- **consult the expertise of participants** for feedback on AAI activities;
- showcase ongoing **implementation efforts of the AARC Blueprint Architecture**;
- promote a **consistent vision for federated access**;
- facilitate activities for the adoption of **harmonised solutions** and avoid 'reinventing' the wheel;
- **report on the adoption** of and provide guidance on the AARC guidelines;
- provide a home for the adoption and further development of the **AARC Blueprint Architecture**;
- maintain, develop, and organisationally support the **AARC community**;
- **liaise** with other entities in the AAI ecosystem.

Participation

There are two ways to participate in AEGIS:

- **Members** – Research and e-infrastructures and other organizations responsible for the operation of AAIs for international research collaborations following the AARC guidelines relevant to their interoperability with AEGIS peers. Each member can appoint up to two individuals to represent the organization in AEGIS.
- **Observers** – AEGIS welcomes parties that may have an interest in using AARC guidelines or that are in the process of implementing an AAI that follows the AARC BPA. Observers should be invited by an AEGIS member and endorsed by the AEGIS membership. Observers do not vote nor endorse documents.

AEGIS

Charter

Version 1.0 – Endorsed on November 11th 2019

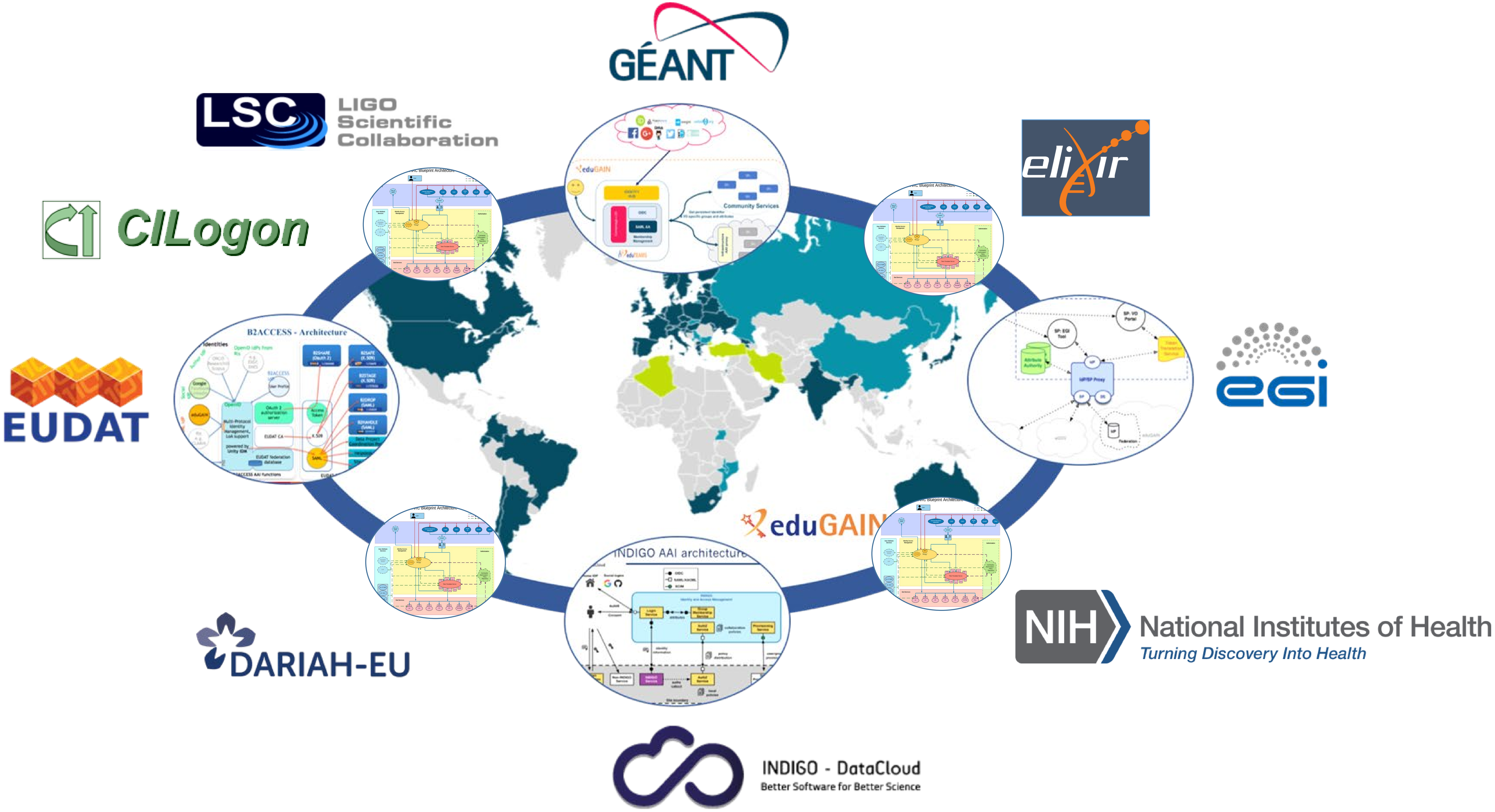
IPR Policy

Version 1.0 – Endorsed on November 9th 2020

Get in touch

Email: aarc-contacts@lists.geant.org

https://wiki.geant.org/display/AARC/AARC+Interoperability+Guidelines+Approved+by+AEGIS

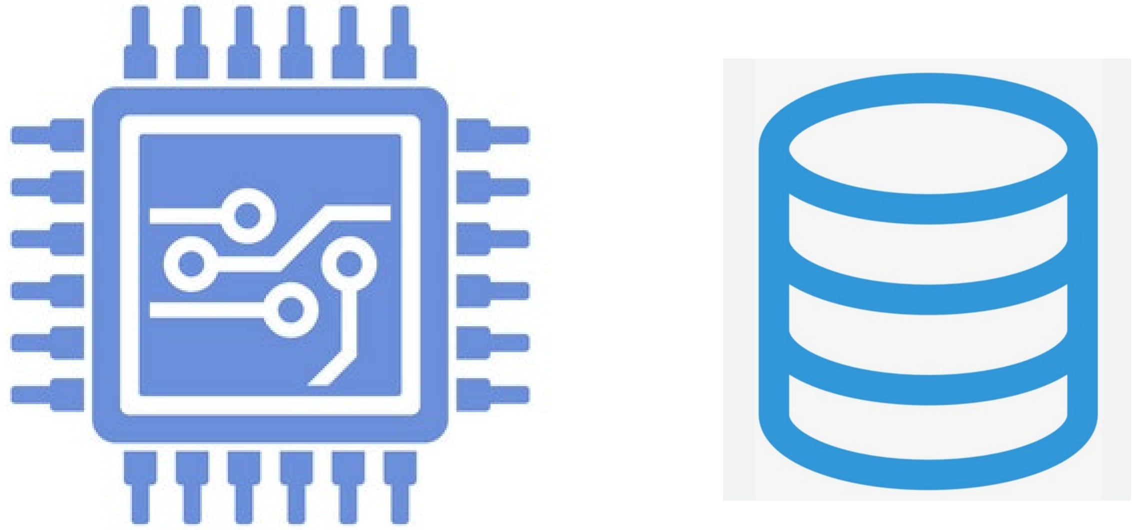


AARC Blueprint Architecture Implementations

Federated HPC, cloud and data infrastructures

Integration of HPC- and Cloud-based Compute and Data Services

Javier Bartolome (BSC), 2023-03-21



HPC Cluster

POSIX Parallel Filesystem

Protocol Access	SSH or similar shell Access
Access Authentication	Password, ssh keys, …
Management of resources	Batch Scheduling system
Compute Resources	Node-hours, core-hours
Storage Access	POSIX, direct …
Workload	simulation - result based

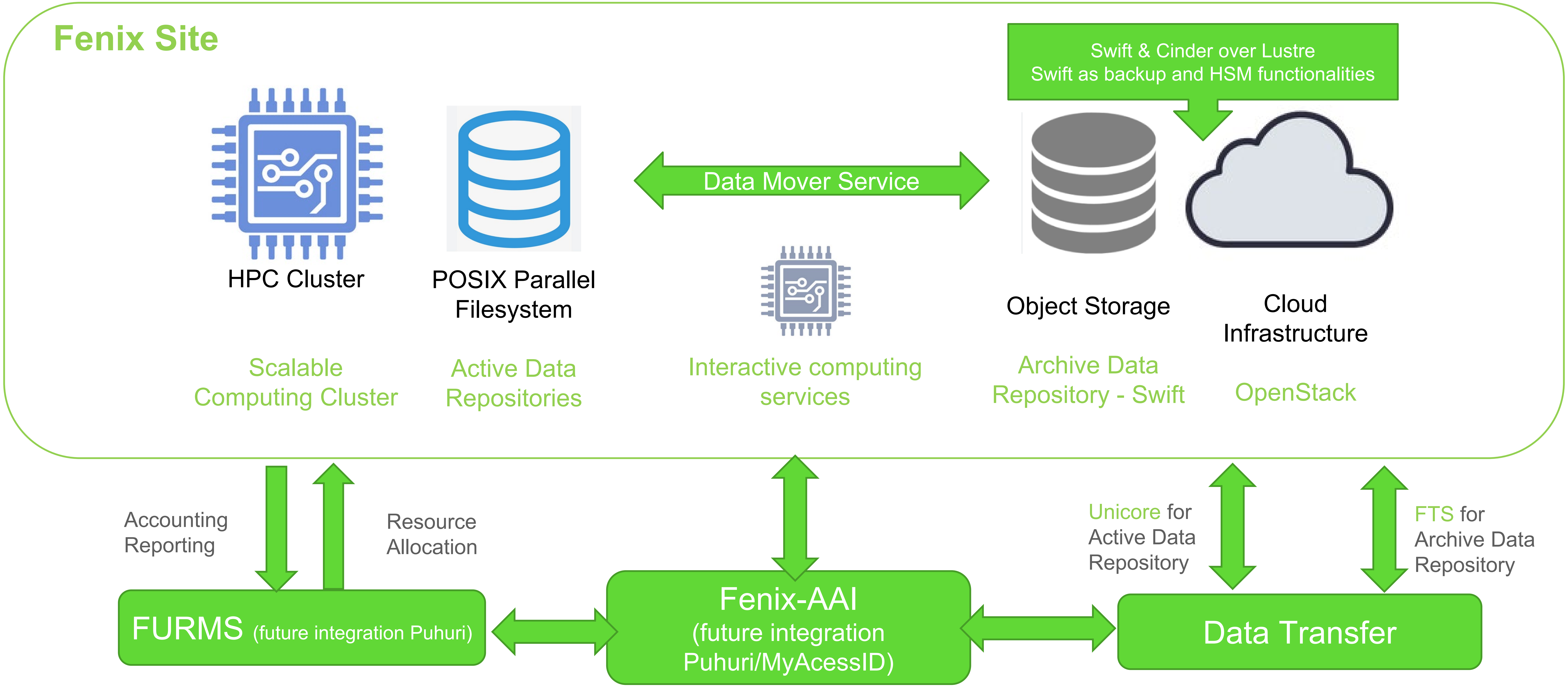
Object Storage

Cloud Infrastructure

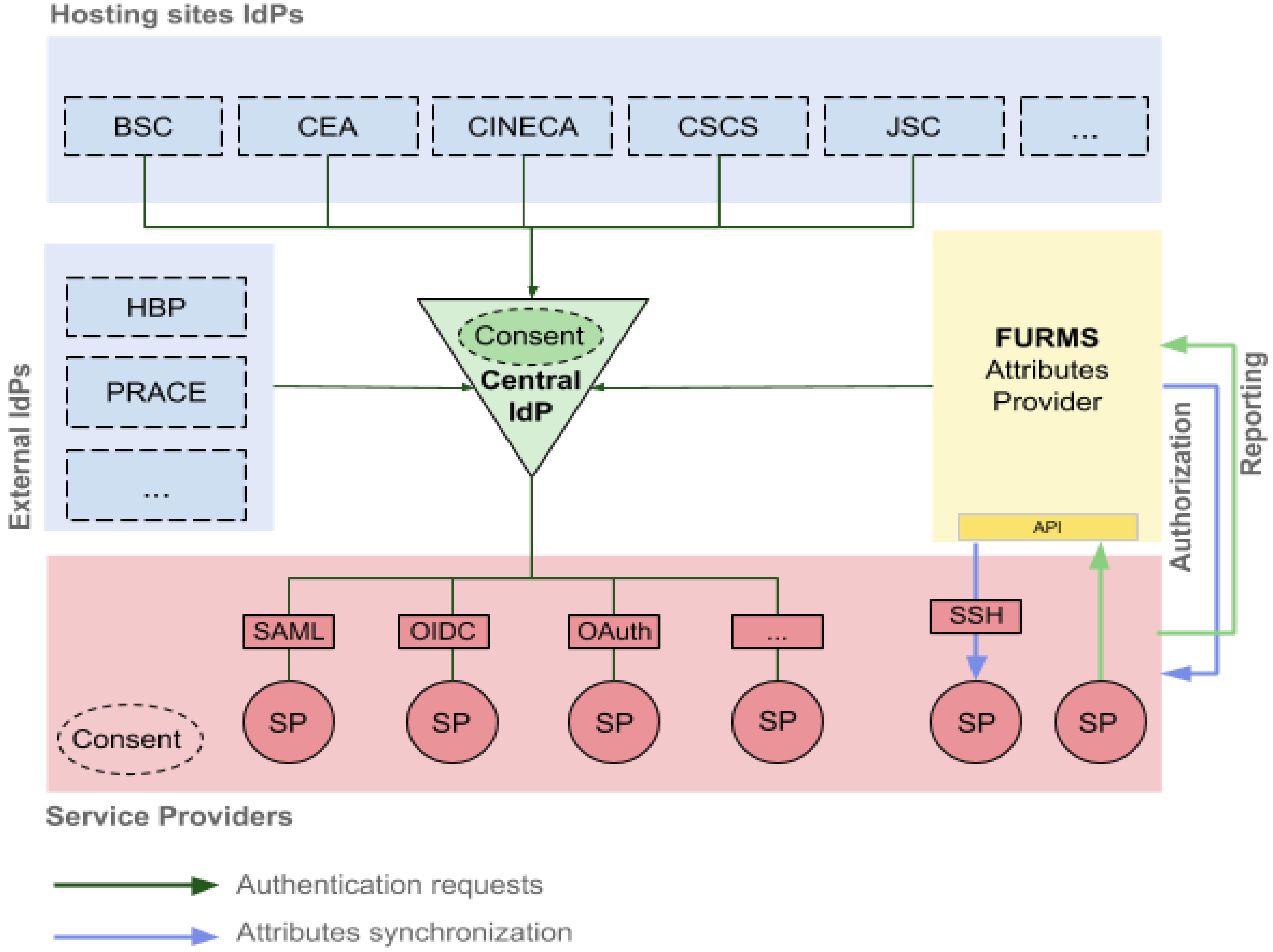
Protocol Access	HTTP
Access Authentication	OIDC, SAML, HTTP session token
Management of resources	Cloud Management software
Compute Resources	RAM, vCores
Storage Access	S3, Swift (based on HTTP)
Workload	Service-oriented

Fenix Infrastructure

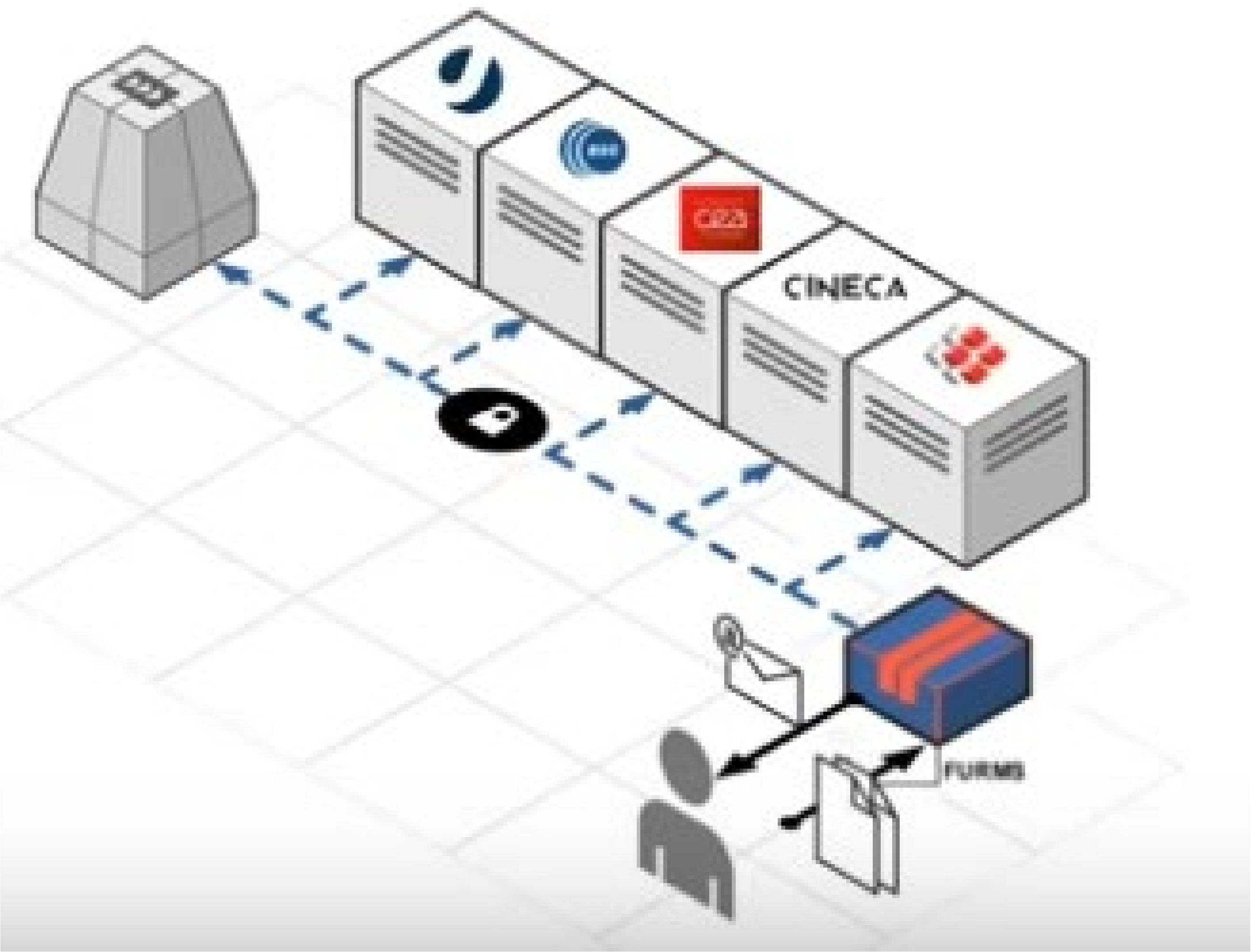
- Long term effort of European supercomputing centres on harmonizing and federating HPC, Cloud and storage services
  - Become: Infrastructure service providers (ISP) committed to a jointly agreed set of e-infrastructure services
- Based on MoU, currently 6 European supercomputing centres BSC, CEA, CINECA, CSCS, CSC, JSC


Fenix Site
Swift & Cinder over Lustre
Swift as backup and HSM functionalities
Data Mover Service
HPC Cluster
POSIX Parallel Filesystem
Object Storage
Cloud Infrastructure
Scalable Computing Cluster
Active Data Repositories
Interactive computing services
Archive Data Repository - Swift
OpenStack
Accounting Reporting
Resource Allocation
Unicore for Active Data Repository
FTS for Archive Data Repository
FURMS (future integration Puhuri)
Fenix-AAI
(future integration Puhuri/MyAcessID)
Data Transfer

Fenix-AAI


Hosting sites IdPs
BSC
CEA
CINECA
CSCS
JSC
...
External IdPs
HBP
PRACE
...
Consent
Central IdP
FURMS
Attributes Provider
API
Authorization
Reporting
SAML
OIDC
OAuth
...
SSH
SP
SP
SP
SP
SP
SP
Consent
Service Providers
Authentication requests
Attributes synchronization

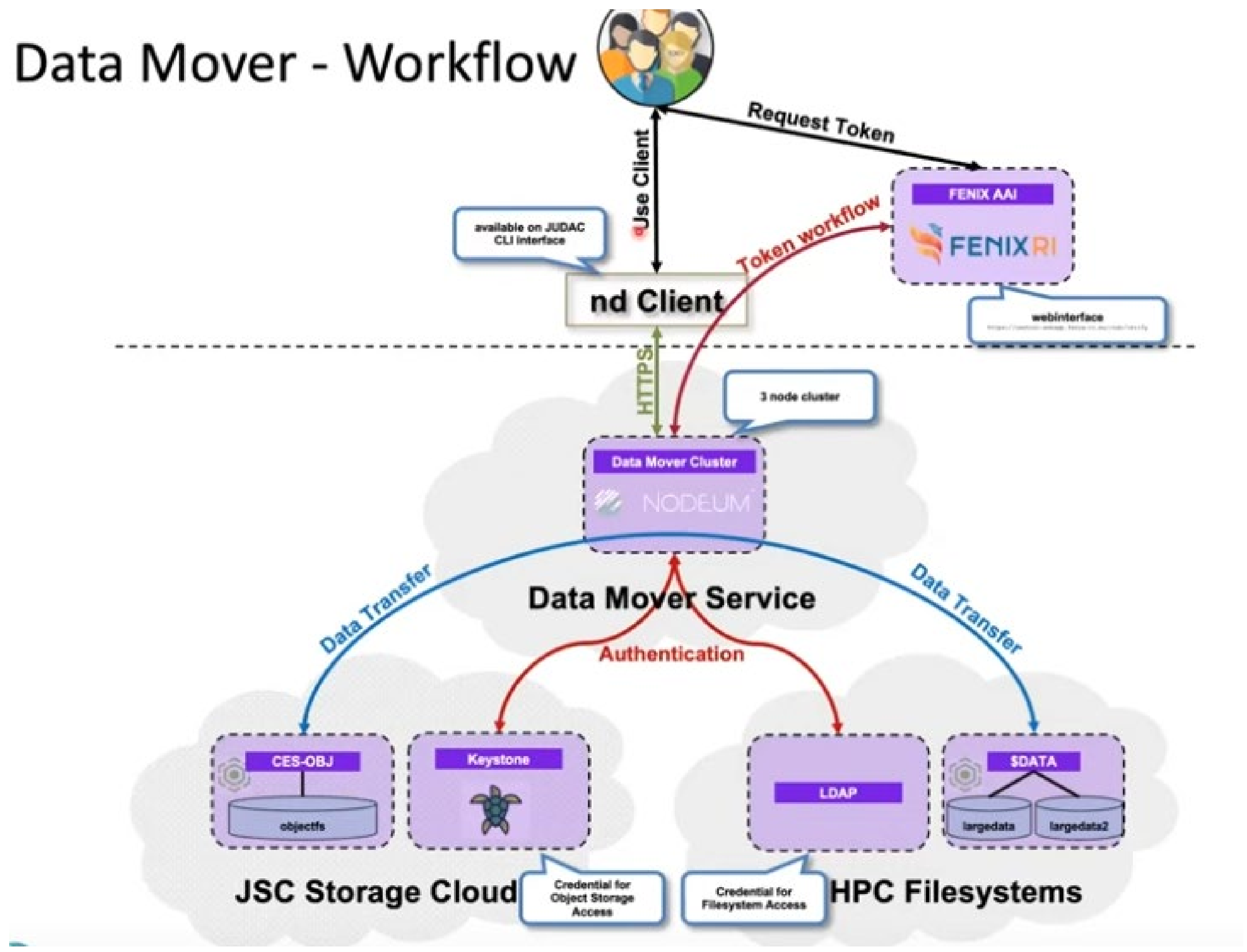
FURMS

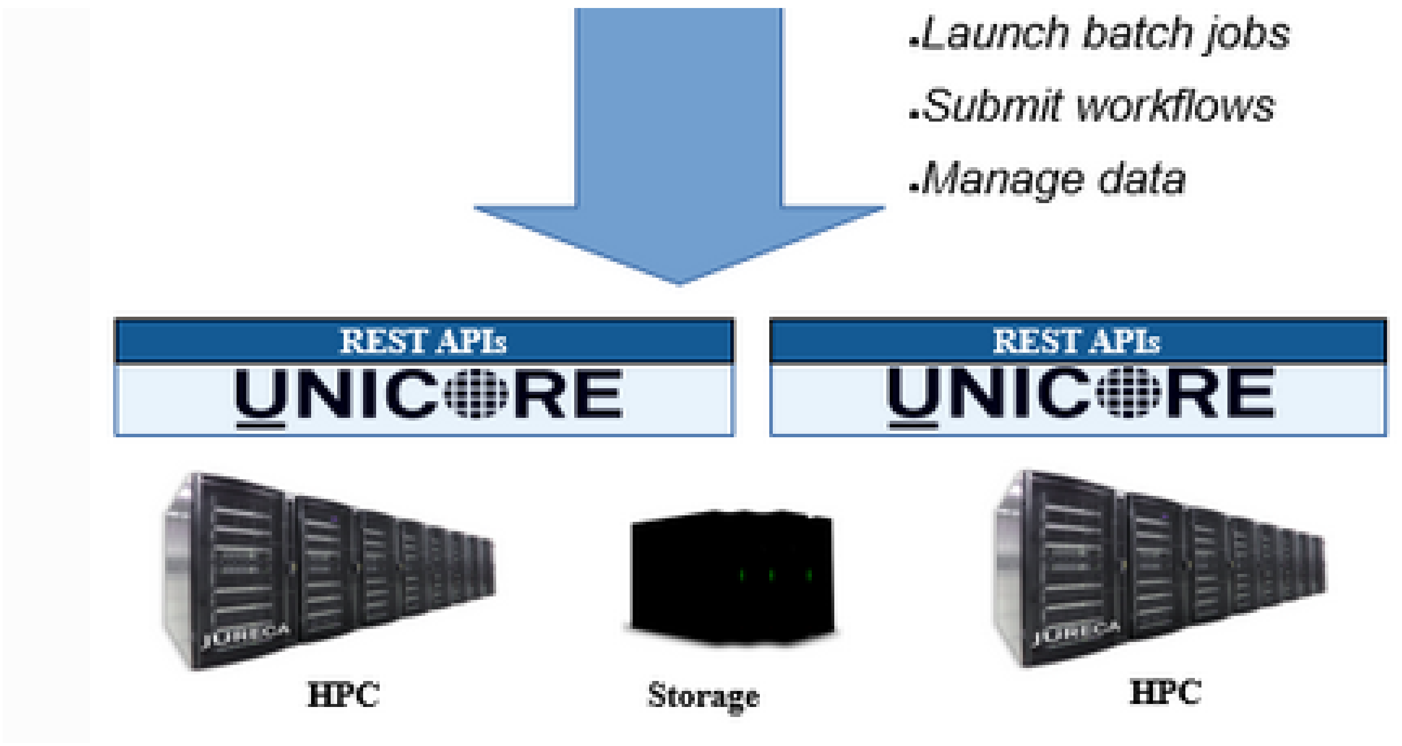


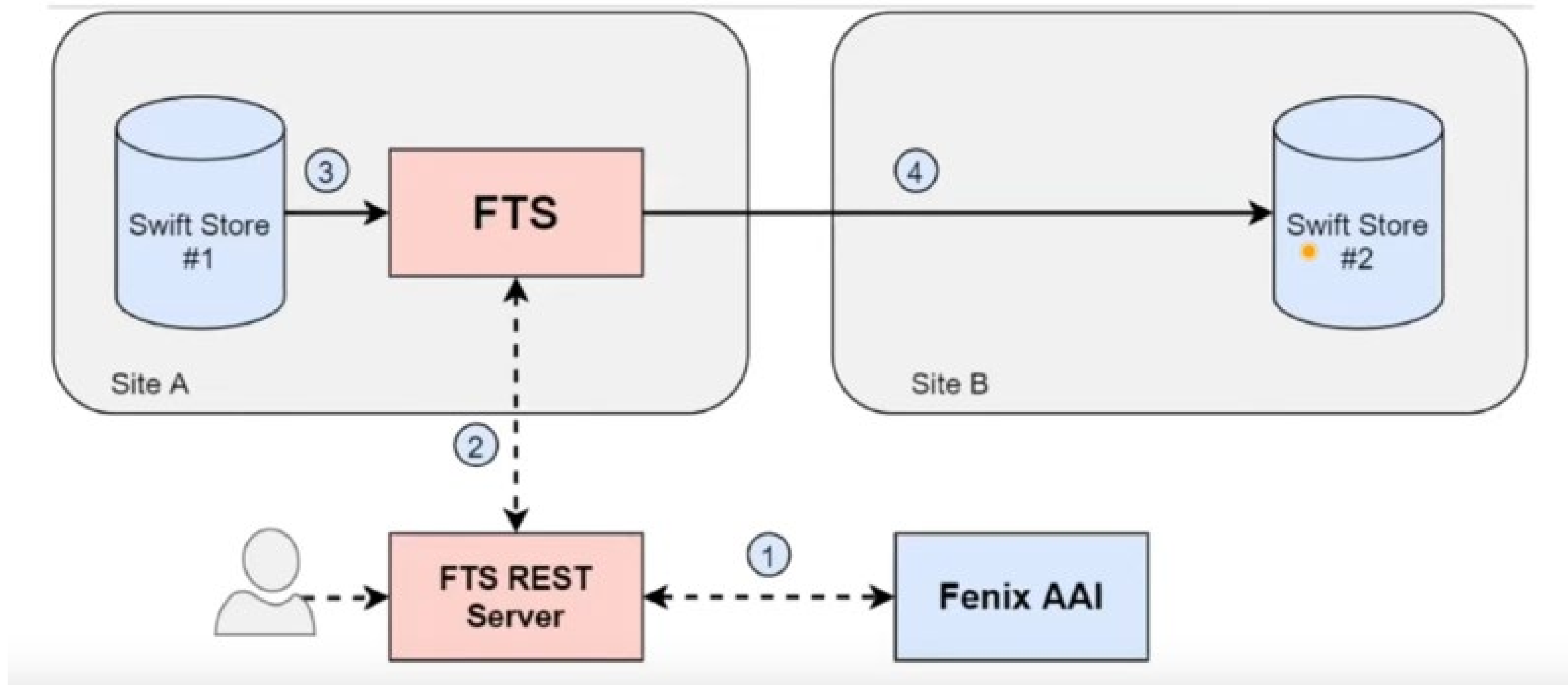
- Central portal, common WebUI and REST API
    - Community: (Virtual) Organization entitled to use resources
    - Project: Communities creates projects, and assign resources to them
    - User: Users are associated to projects and communities
- FURMS local-agent deployed in each center to interact with local infrastructure

Data Mover/Transfer

- Data Mover : transfer between POSIX & Object locally
- Unicore : transfer from/to Active storage repositories
- FTS : transfer from/to Archive storage repositories







EuroHPC Summit

2023 Göteborg

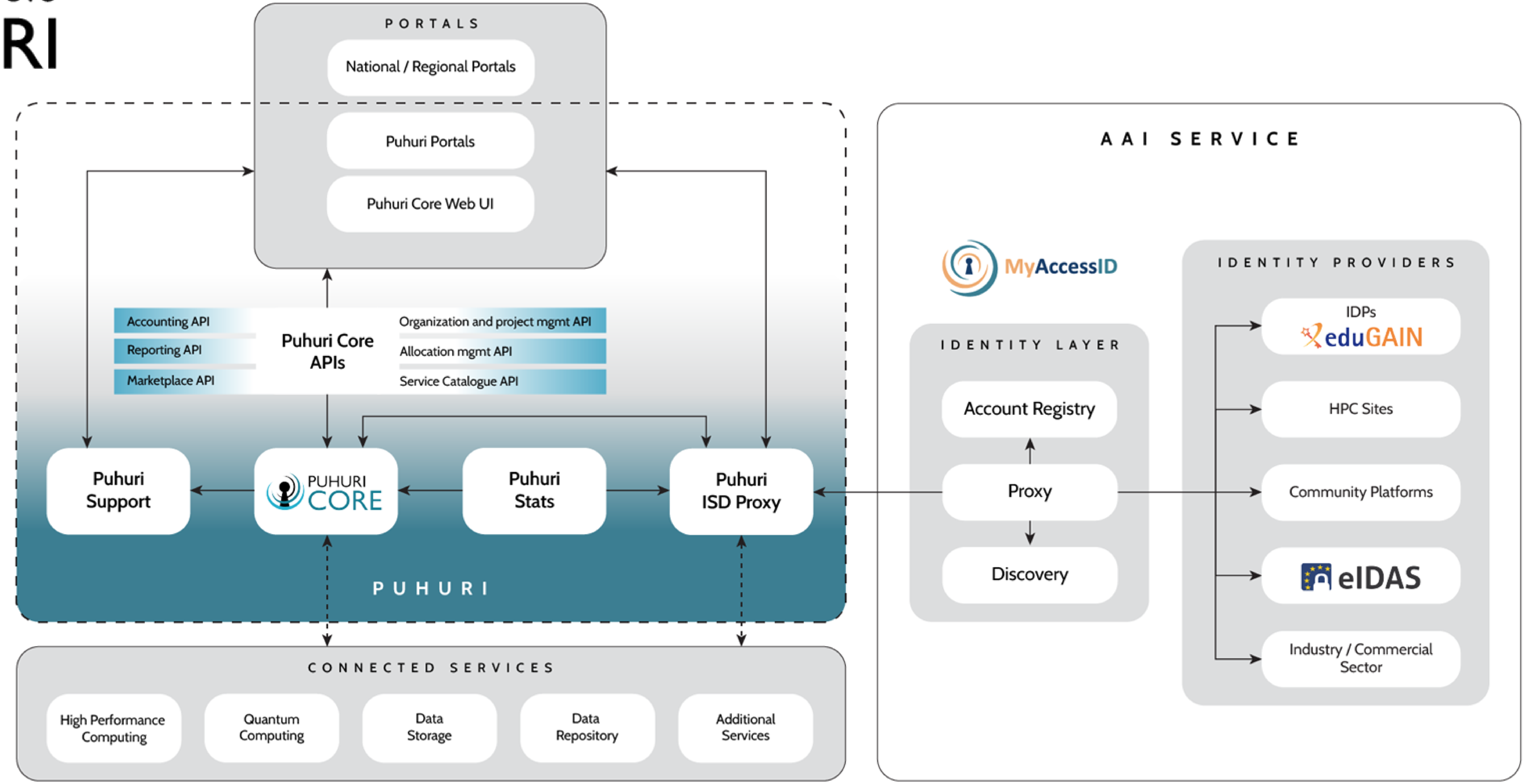
Federated HPC, cloud and data infrastructures

Resource Management, Allocation and Accounting

Anders Sjöström (LUND), 2023-03-21

Puhuri services:

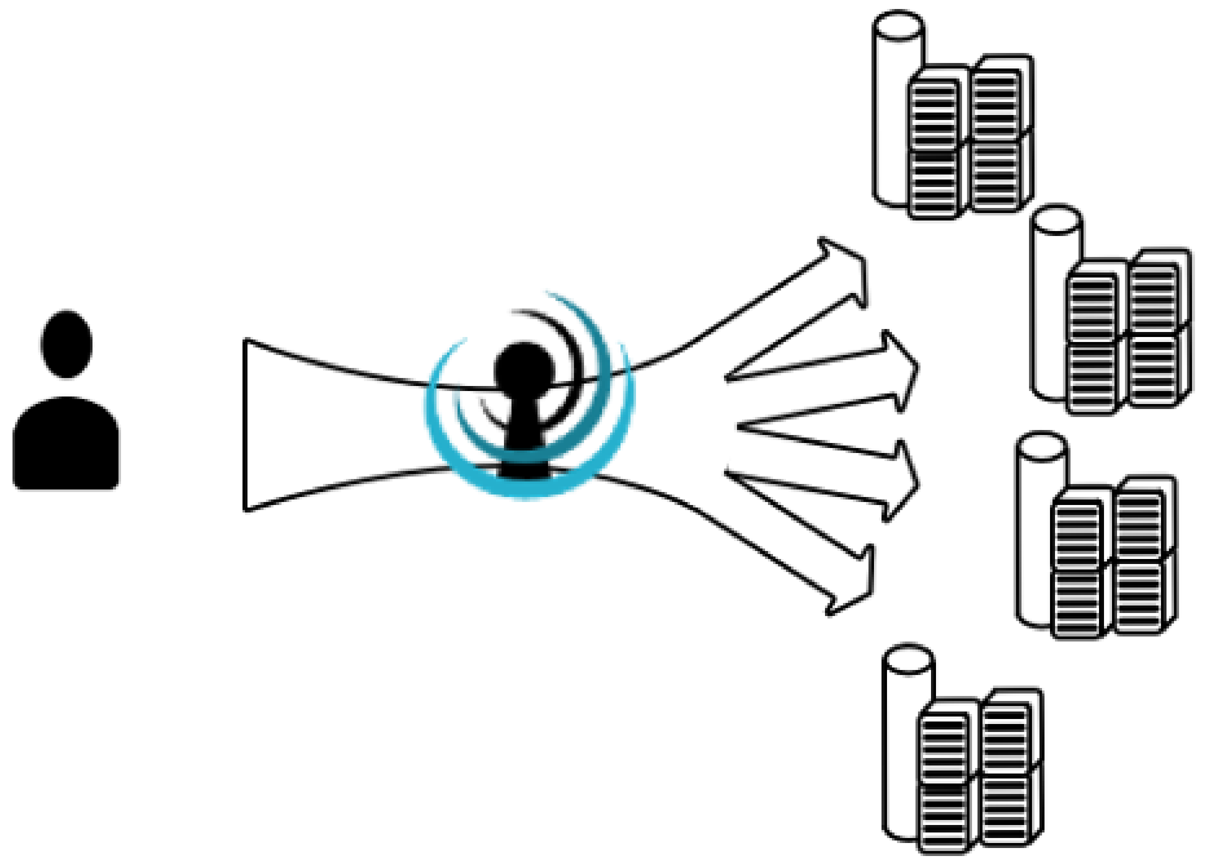
the aggregation of
Puhuri Core
Puhuri AAI
Puhuri Portal
and corresponding APIs



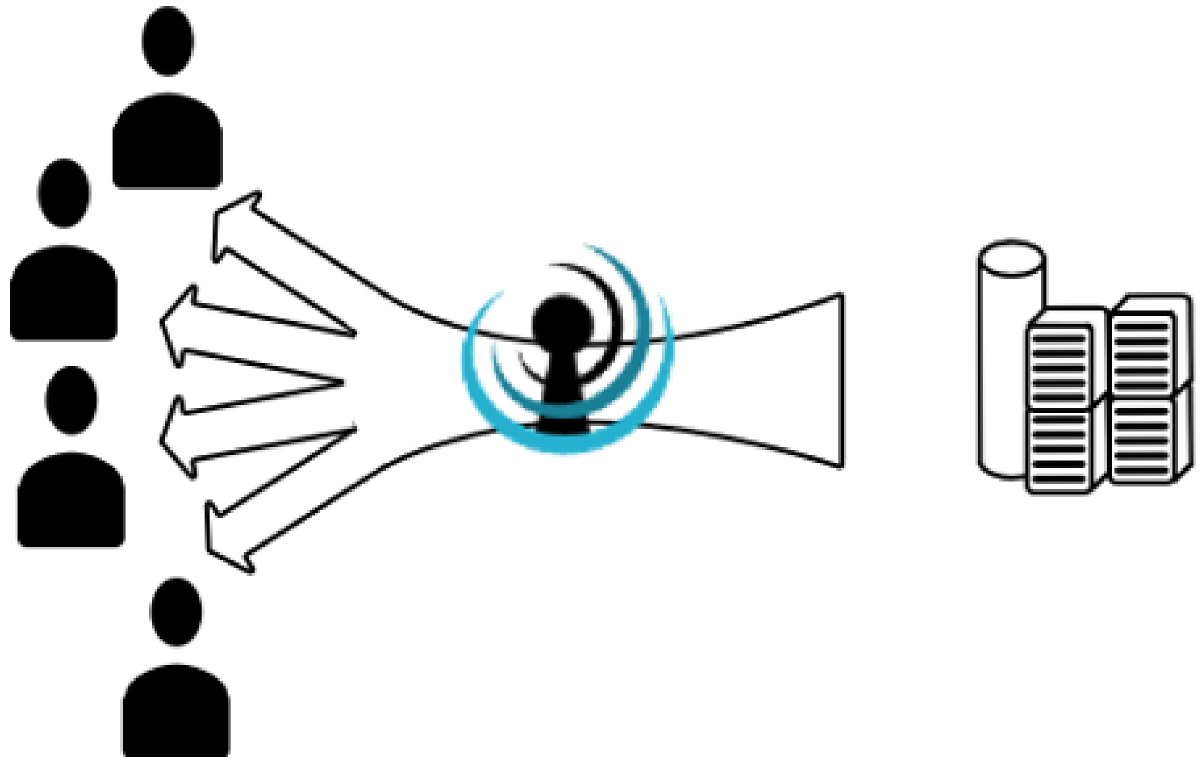
Resource management - who are the actors? - what are their interests?

- Resource Providers
  - Maximise resource utilisation, get reporting
  - Limit costs and ensure sustainability of the resource management
- Res. Allocation Service Providers, e.g. Puhuri, FENIX
- Resource Allocators
  - Allocate resources in a controller manner, reporting
- Users
  - Apply for, and access resources, manage groups, view accounting
- Identity providers
  - Enable authentication of their users, increase usage and uptake

Bringing resources to users



Bringing users to resources

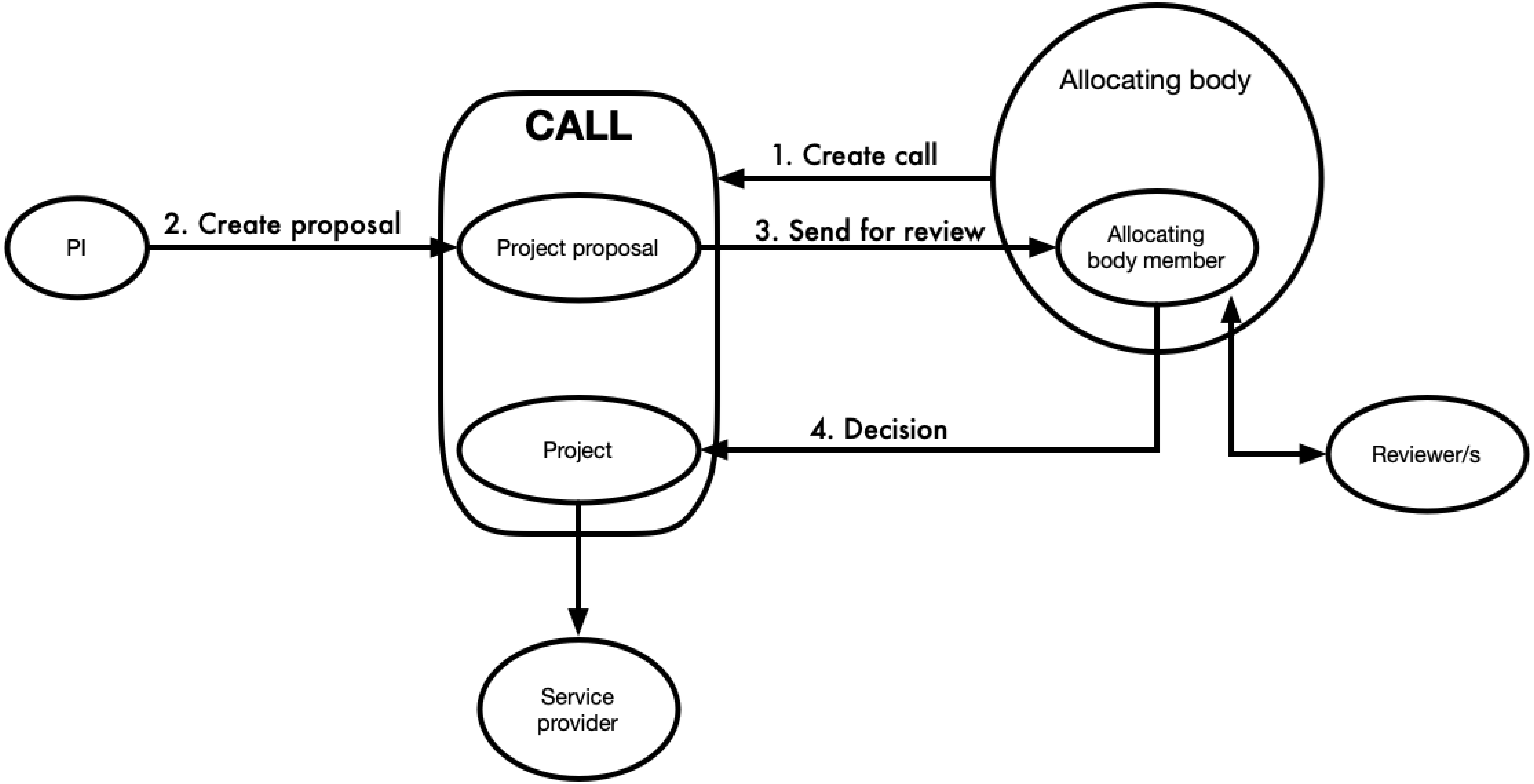


Flexibility of resource allocation

- With multiple resources and multiple service providers the actors requirements must be met
  - CPU-h, GPU-h, TB, TB-h, quantum computing allocation units, etc.
  - Data management, quota management

- Reporting
  - Puhuri aggregates reporting from resources, presenting the data to PI:s, resource allocators, service providers

Allocation, who does what where and when?



Industry vs. academic users

- Federated authentication is feasible for many academic users via eduGAIN
- Industry users must be able to register and authenticate as well
- Verifying user's identity can be a challenge for both groups
    - There is a need for an automatic user identity vetting solution
- Academic users usually get resources for free when as industry users may need to pay per usage
- Might have different expectations (user experience, SLA, security incl. MFA..)

A radical idea

- What if we have a completely different view on allocation? i.e. market driven, token-based, user focused marketplace

EuroHPC Summit

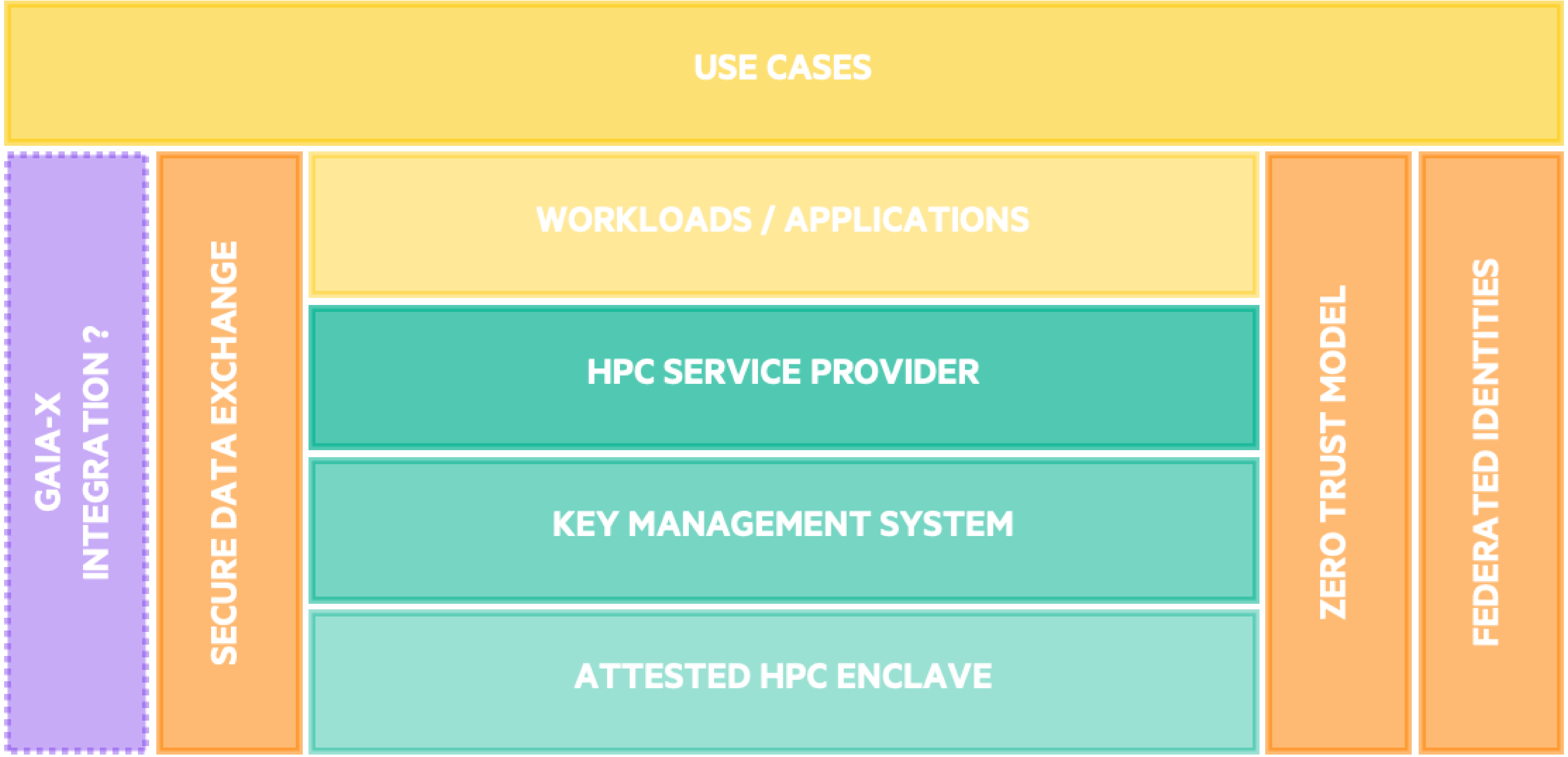
2023 Göteborg

Federated HPC, cloud and data infrastructures

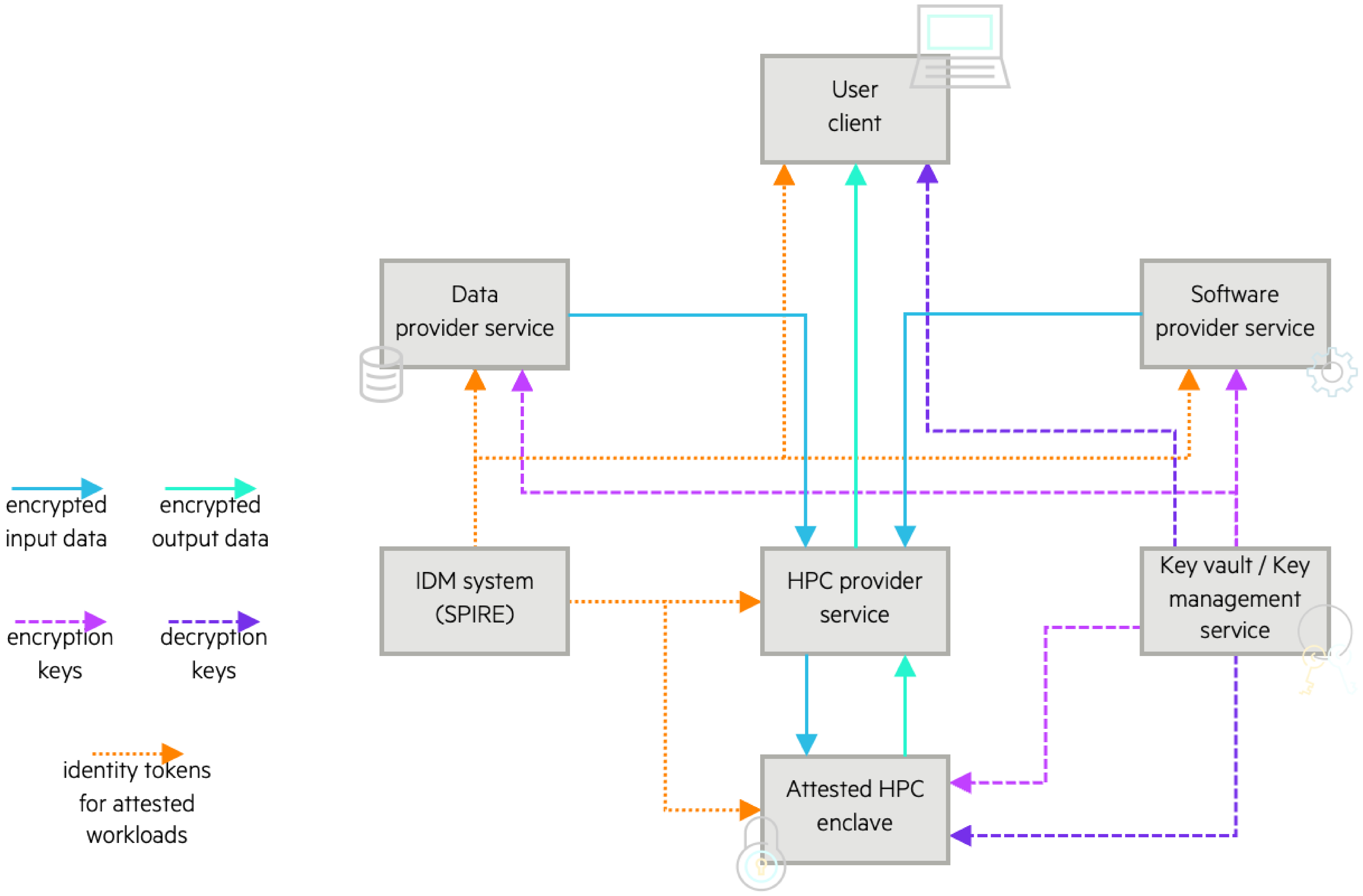
Trust, Security and Data Compliance

Utz-Uwe Haus (HPE), 2023-03-21

Vision of a secure federated HPC workflow

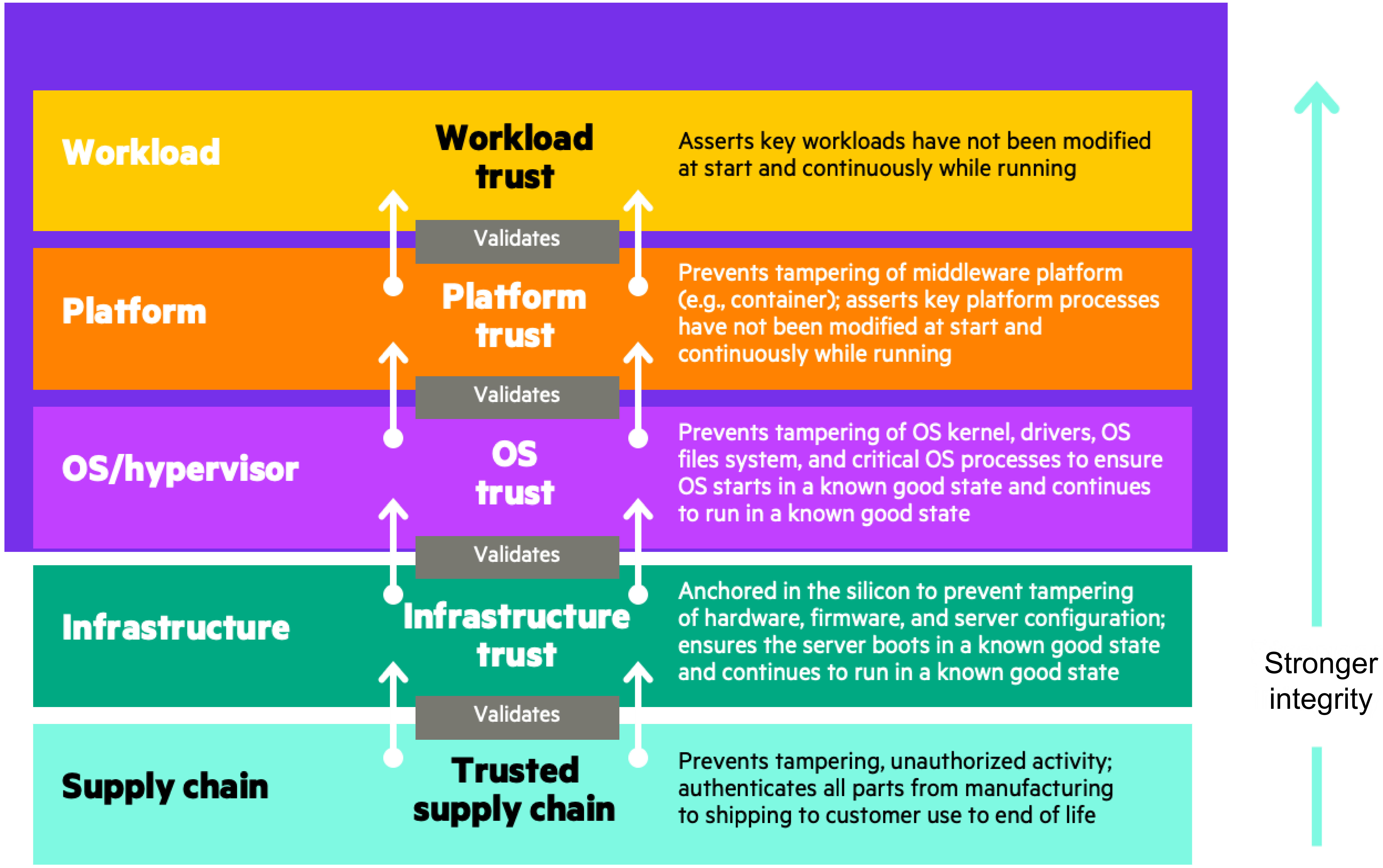


Possible Architecture



- **Connectivity** – connect data sources at edge, in the core, in the cloud
- **Identity life cycle** – SPIFFE IDs for all micro services
- **Attestation** – secure attestation of workloads and processing nodes likewise
- **Authentication** – mutual authentication between all services
- **Control** – access control to service endpoints, keys and data
- **Openness** – integrating open standards

Trust


Workload
Workload trust
Asserts key workloads have not been modified at start and continuously while running
Validates
Platform
Platform trust
Prevents tampering of middleware platform (e.g., container); asserts key platform processes have not been modified at start and continuously while running
Validates
OS/hypervisor
OS trust
Prevents tampering of OS kernel, drivers, OS files system, and critical OS processes to ensure OS starts in a known good state and continues to run in a known good state
Validates
Infrastructure
Infrastructure trust
Anchored in the silicon to prevent tampering of hardware, firmware, and server configuration; ensures the server boots in a known good state and continues to run in a known good state
Validates
Supply chain
Trusted supply chain
Prevents tampering, unauthorized activity; authenticates all parts from manufacturing to shipping to customer use to end of life
Stronger integrity

Security

- Many well-known levels to cover
- For containers specifically: consider NIST SP 800-190
  - use container-specific (restricted) OS
  - Only group containers with the same purpose, sensitivity, and threat posture on a single host OS kernel
  - Adopt container-specific vulnerability management tools and processes for images to prevent compromises
  - Consider using hardware-based countermeasures to provide a basis for trusted computing
  - Use container-aware runtime defense tools.

Compliance: an ugly duckling in HPC

GDPR – personally identifiable information; covering information concerning anyone physically in Europe at the time of data collection

HIPAA – protected health information; protecting US citizen anywhere

SOC 2 – defines how companies should manage, process, and store customer data

ISO 27001:2022 – standard to manage information security

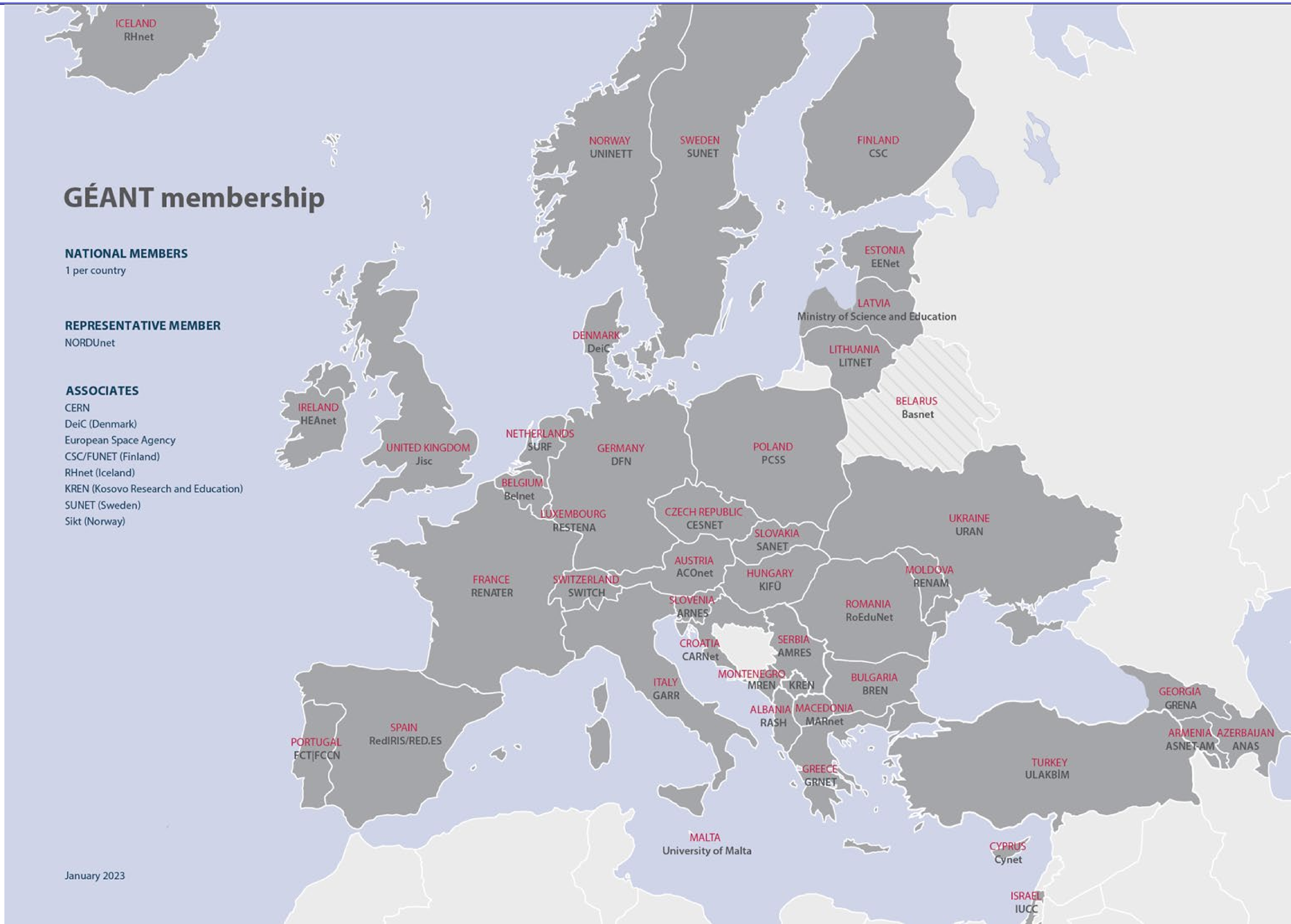
plus country/domain/contract specific regulations

Federated HPC, cloud and data infrastructures

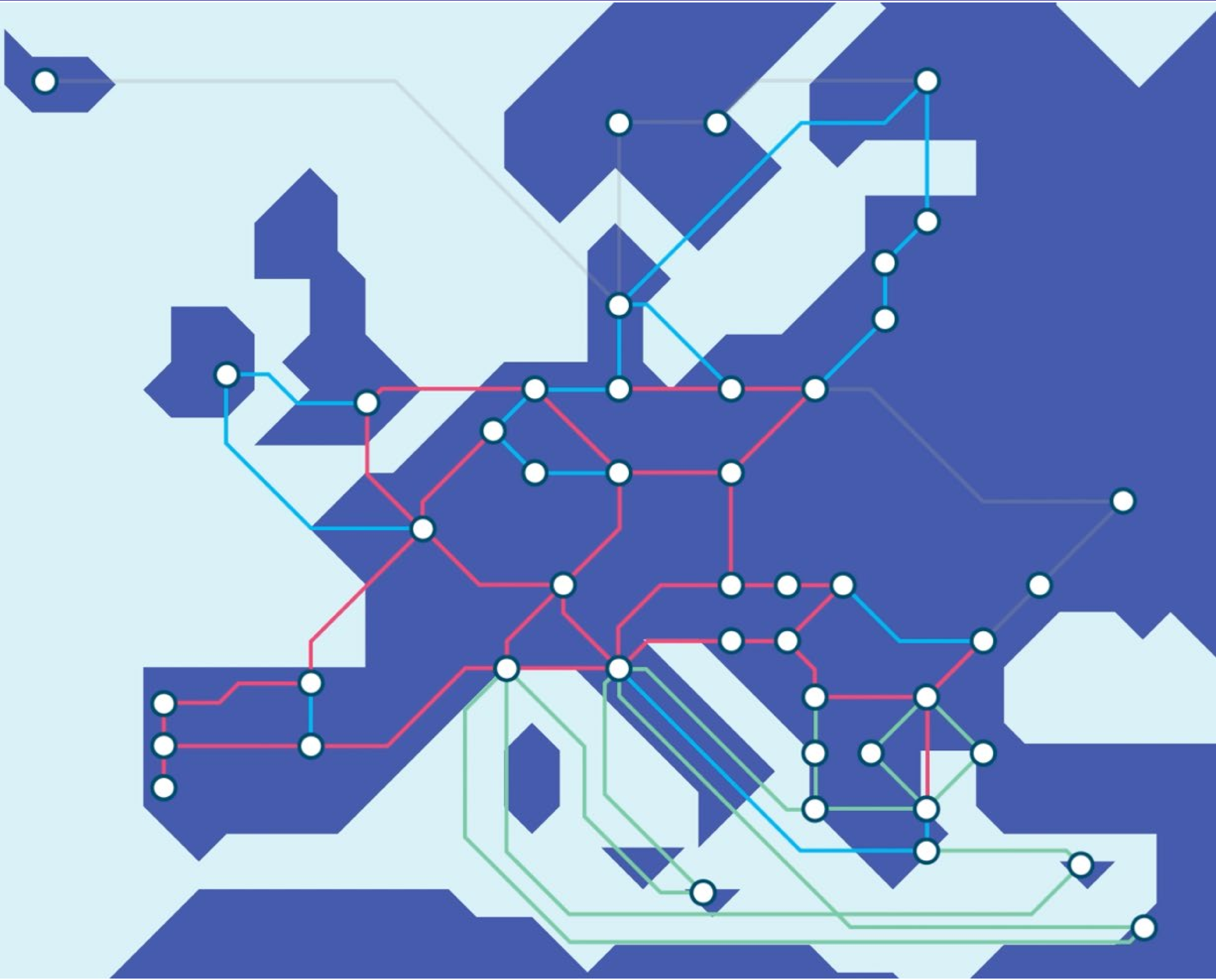
Connectivity

Enzo Capone (GEANT), 2023-03-21

The GÉANT Community



The new GÉANT network



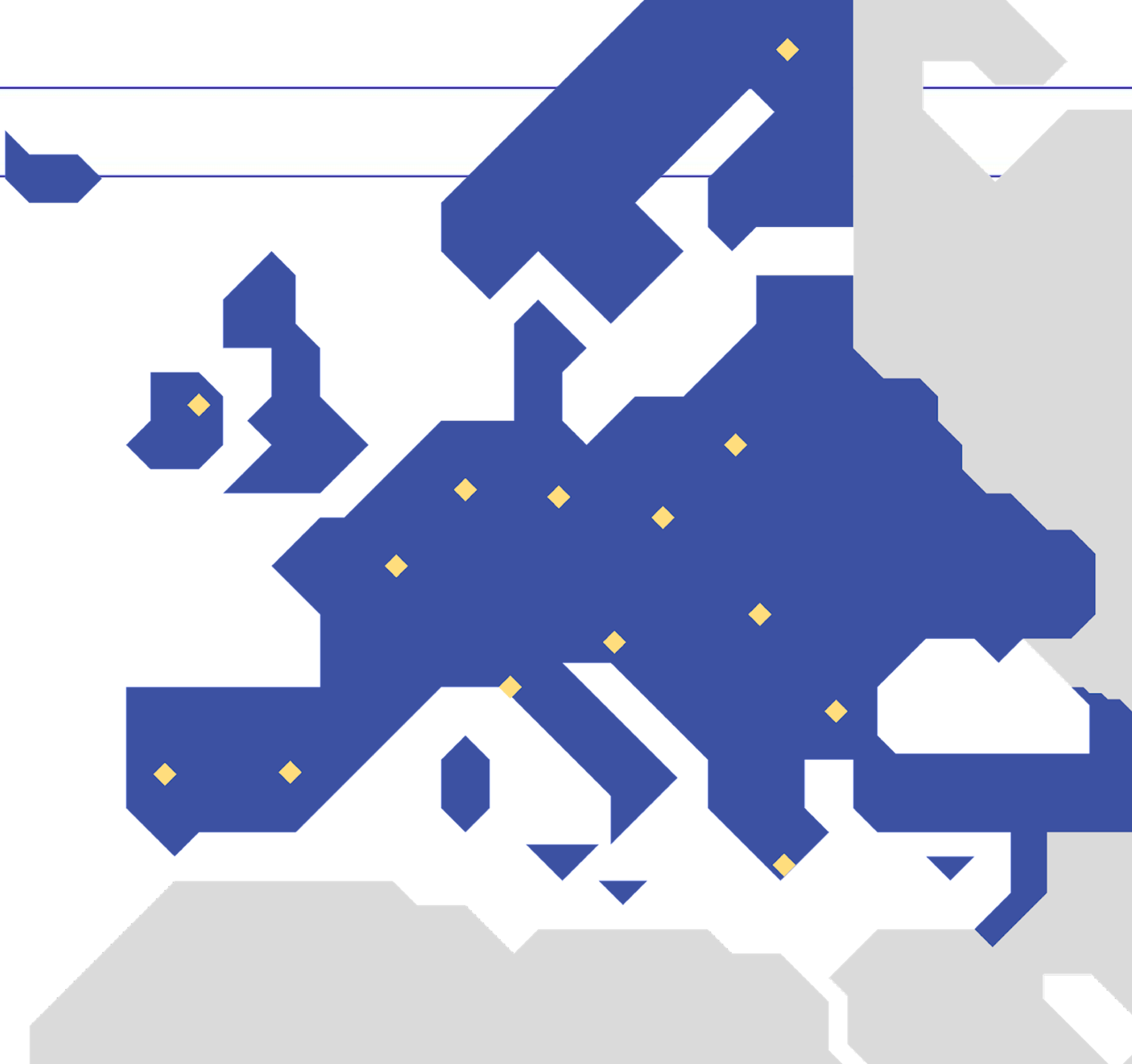
Funded by the EU

EuroHPC JU sites

Exascale
Julich (JUPITER)
TBC

preExascale
CINECA (Leonardo)
BSC (MareNostrum5)
CSC (LUMI)

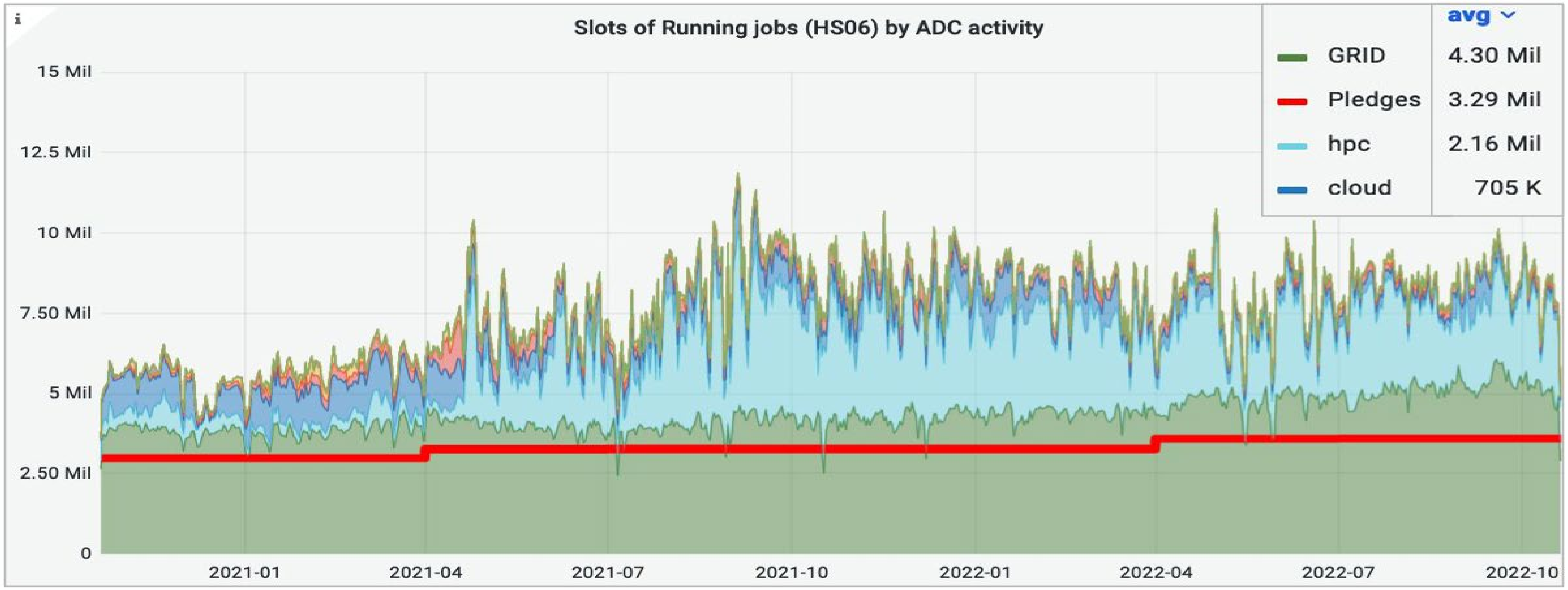
Petascale
Sofiatech (Discoverer)
MACC (Deucalion)
IT4I (Karolina)
LuxProvide (MeluXina)
IZUM (Vega)
GRNET (DAEDALUS)
KIFU (Levente)
NUI-ICHEC (CASPIr)
CYFRONET (EHPCPL)

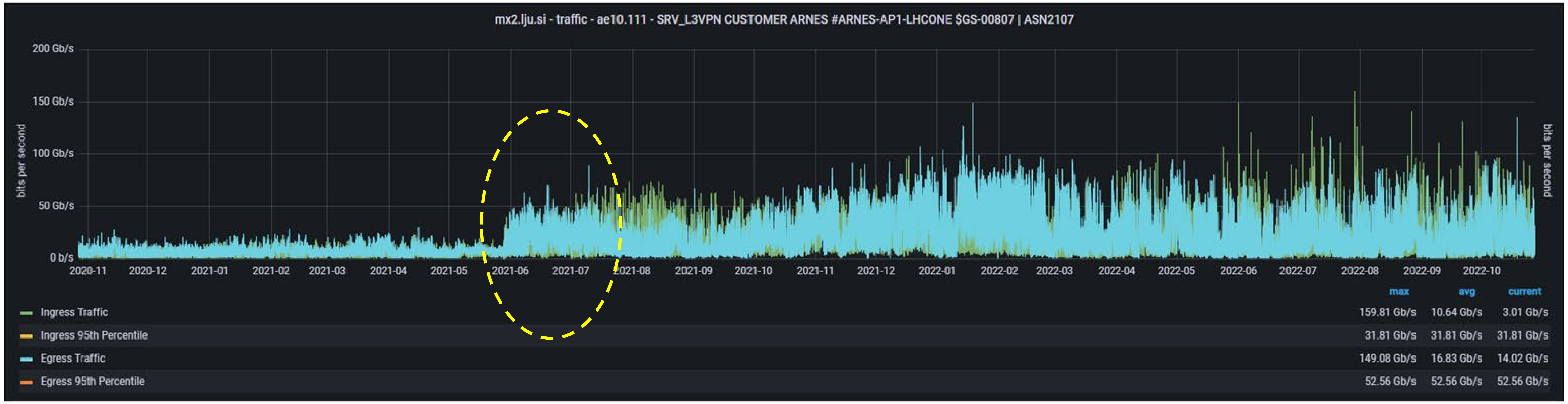


EuroHPC sites	NREN (Country)	NREN access to GÉANT (Gbps)	Site access to NREN (Gbps)
Exascale Julich (JUPITER)	DFN (DE)	2x300G	2x100G
Exascale site	*RENATER (TBC)*	*4 x 100G*	*TBC*
preExascale CINECA (Leonardo)	GARR (IT)	2x200G	*N* x 100G
preExascale BSC (MareNostrum5)	RedIRIS (ES)	2x100G	2x100G
preExascale CSC (LUMI)	FUNET/NORDUnet (FI)	100G+60G	*N* x 100G
Petascale Sofiatech (Discoverer)	BREN (BG)	30G+10G (2x100G Q3-23)	50G
Petascale MACC (Deucalion)	FCCN (PT)	2x100G	2x100G
Petascale IT4I (Karolina)	CESNET (CZ)	2x100G	2x100G
Petascale LuxProvide (MeluXina)	RESTENA (LU)	2x100G	2x100G
Petascale IZUM (Vega)	ARNES (SI)	2x100G	5x100G
Petascale GRNET (DAEDALUS)	GRNET (GR)	2x100G (2x200G planned)	2x100G
Petascale KIFU (Levente)	KIFU (HU)	2x100G	3x100G
Petascale NUI-ICHEC (CASPIr)	HEANET (IE)	2x100G	TBC
Petascale CYFRONET (EHPCPL)	PSNC (PL)	2x100G	N x 400G (2023 onwards)

ATLAS experiment started using **Vega** in **Slovenia**.

1 single HPC site now provides more than 50% resources and completes half the number of ATLAS jobs





Federated HPC, cloud and data infrastructures

Discussion and Q&A

Dirk Pleiter, 2023-03-21