Towards a Pan-European Research Storage Infrastructure

Bo Bai, DeiC

GN5-2 WP4

Data and storage requirements from a research perspective

Data needs protection

Researchers need seamless collaboration

Institutions need trusted data management

Research infrastructures need seamless data exchange

Everybody needs competitive pricing

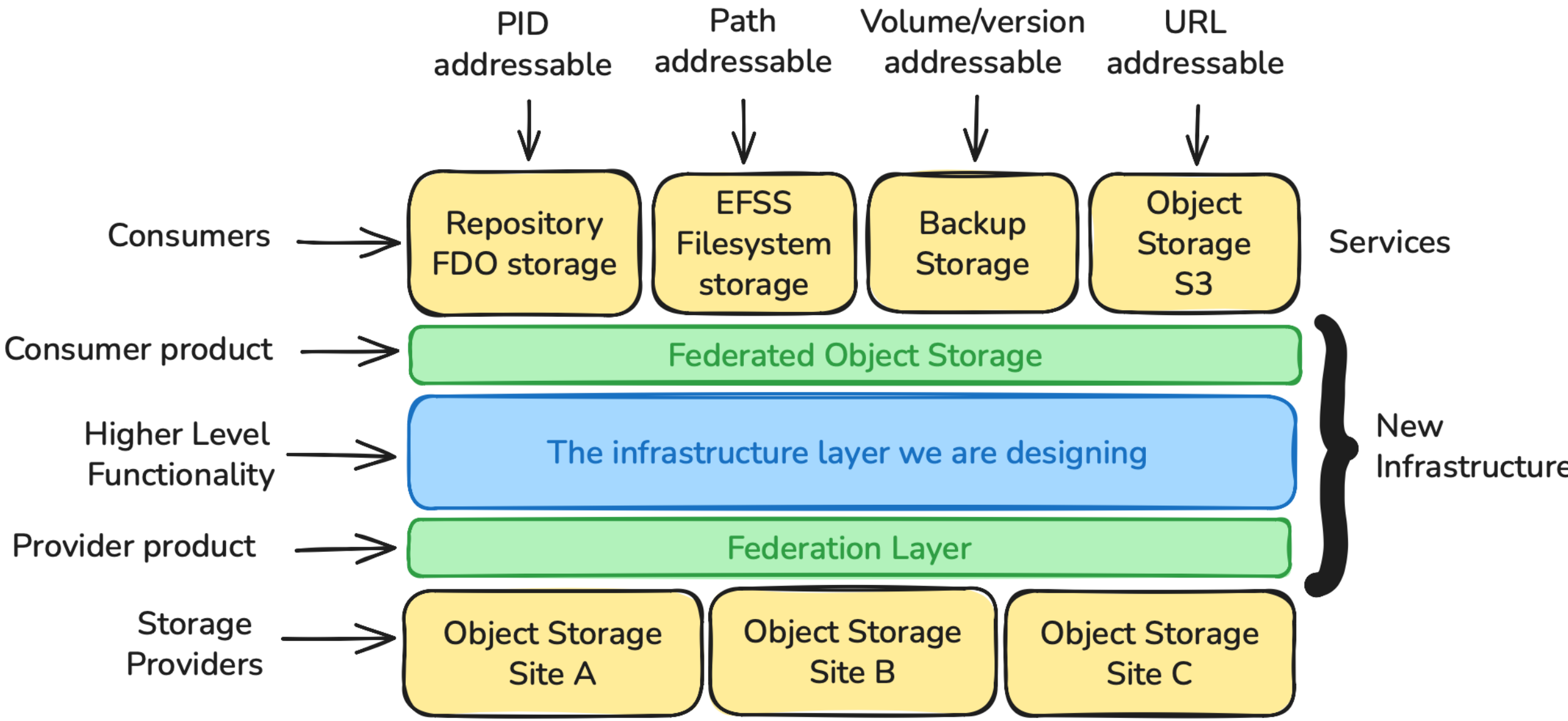
Problem statement – for the storage of research data

- Storage needs for research data are growing
- Datasets are increasing in volume, complexity, and retention requirements
- Compliance has become a necessity
- Rising user expectations
- Sovereignty has regained its importance

Stated Goals for a Pan-European Research Storage Infrastructure

- A sovereign, coherent infrastructure for European research
- Easy to use as a foundation for other services
- Long-term sustainability
- Trusted by data controllers
- A feasible roadmap

High Level Architecture



Infrastructure Design Principles

1

Build Infrastructure

Infrastructure is a building block for someone else's service

2

Use Existing Components

Utilise existing standards and components when possible

3

Cut to the Core

Do not internalise what can be externalised

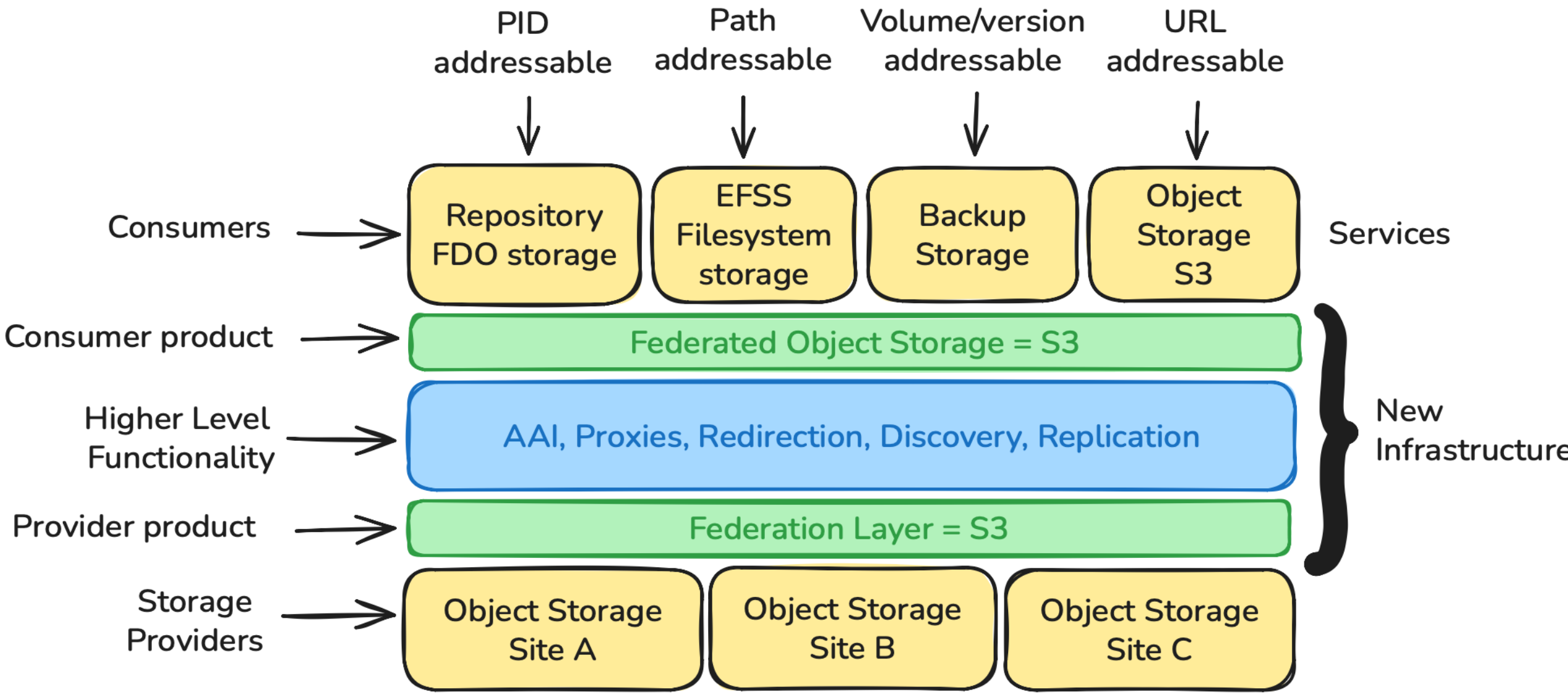
"Solutions should come from the communities that own the problems”

Pieces of the puzzle

- Object storage protocol: S3
  - Ceph Community, SNIA, OIF
  - CS3 Community, Ceph Science Group, EOSC
- Transport layer
  - HTTP, Servers, Proxies, Redirects
  - W3C
- Identity federations: EOSC AAI, eduGAIN, AARC
  - FIM4R, REFEDS, Tiime
- Object Storage Federation
  - SIG-CISS
- Umbrella federation: EOSC
  - EOSC Federation

High Level Architecture - starting point

8



Why do this through GÉANT?

GÉANT is already a **key European federation actor** for research and education

Building a Pan-European Research Storage Infrastructure is not primarily a problem of storage technology. It is a **federation, trust, identity, procurement, governance, interconnection, and operations** problem

GÉANT's sovereign network infrastructure is a **key enabling component** for creating a Pan-European Storage Infrastructure

Roadmap overview



Foster convergence among providers



Establish cross-site functionality



Introduce higher level functions



Establish a federated service

11

Year 1-3 Blueprints for Object Storage

Blueprints for a Ceph Storage Cluster

Emphasis on:

Convergence

Compatibility

Sustainability

Compliance

Year 3-5 Higher Level Functions

Higher level functionality

Emphasis on:

Storage Abstractions

Data Management

Data Release

Cross-site Functionality

Years 5+ Federation

Federated service

Emphasis on:

Governance

Certification

Sovereignty

Productisation

Providers first!

- Shared blueprints, tooling, and operational practices reduce duplicated work
- Providers keep there business model

- A defined interoperability specification
- A test set against the interoperability specification

- Certifications accepted by data controlers
- Bilatteral peering of storage
- Improve recillience

Increase Integration and Functionaity

- Cooperation
- Building a service ecosystem on top of the object storage infrastructure
- Data management functions
- Data locality awareness
- Cross-border data-replication
- Common Tiers

Federation

- Policy
- Organisation
- Governance

- Lean on / align with, other federations

- Recognition and focus

15



Poznańskie Centrum Superkomputerowo-Sieciowe
Poznan Supercomputing and Networking Center

Ron Trompert - SURF
Kalle Happonen - CSC
Jan Meijer - SiKT
Maciej Brzezniak - PCSS
Bo Nygaard Bai - DeiC
Morten Kjeldgaard - DeiC











Towards a Pan-European Research Storage Infrastructure

Please join the discussion:
Join SIG-CISS
Follow: https://codeberg.org/SAGE_object_storage/

GÉANT Clouds: **Sovereign Object Storage**, OCRE 2024 Framework, and Cloud Sovereignty

www.geant.org

Co-funded by
the European Union