



# ELIXIR and the e-Infrastructure Commons

Steven Newhouse, EMBL-EBI

*European Life Sciences Infrastructure for Biological Information*

*[www.elixir-europe.org](http://www.elixir-europe.org)*



## What is ELIXIR?

**Sustainable** European infrastructure for  
**biological research data**

**Facilitate research**

**Safeguard data and**  
**build sustainable data services**

Deliver **services** through **ELIXIR Nodes**  
building on **national strengths** and **priorities**

**ELIXIR Hub** drives coordination

# ELIXIR Members

Connects **national centers** and **EMBL-EBI**

Participated by major bioinformatics service providers (~**130**)  
and supported by **19 EU member states**



- **15 Members**
  - *Belgium , Czech Republic, Estonia, EMBL-EBI, Denmark, Finland, France, Israel, Netherlands, Norway, Portugal, Spain, Switzerland, Sweden, UK*
- **4 Observers**
  - *Greece, Ireland, Italy, Slovenia*

# ELIXIR Infrastructure within Excelerate

- Scientific Use Case Driven
  - Marine meta-genomics infrastructure
  - Integrating Genomic and Phenotypic Data for Crop and Forest Plants
  - Infrastructure for Rare Disease research
  - Framework for secure archiving, dissemination and analysis of human access-controlled data
- Platform Oriented
  - Training
  - Interoperability
  - Tools
  - Data
  - Compute

# ELIXIR Compute Platform

- What are our constraints?
  - Accelerate Project: Coordination and incremental delivery to the Scientific Use Cases and beyond
  - Financial: Use & extend existing services as no budget for large-scale middleware development
  - Resources: Investment in national Elixir nodes, European e-Infrastructures, & elsewhere
- First steps
  - Identify common technical aspects across the scientific use cases
  - Map technical use cases to available software/services/solutions
  - Instantiate and integrate technical use cases for scientific use

# Defining Basic Technical Use Cases

- AAI:
  - Federated ID, Other IDs, Elixir ID
  - Credential Translation, Group/Attribute Management
  - Endorsed Personal Data Attributes
- Software Environments
  - VM Library, Container Library, Module Library
- Compute
  - Cloud IaaS, HTC Cluster, PRACE
- Storage
  - Network File Storage, Cloud Storage
- Data
  - File Transfer, Data Set Replication, PID & Meta-data Registry
- Infrastructure
  - Federated Cloud, Federated HTC, IS Directory, IS Registry, Operations & Accounting

# Initial Prioritisation and Grouping Analysis

- AAI
  - 0: Federated ID, Other ID
  - 1: Elixir ID
  - 2: Credential Translation, Group/Attribute Mgmt, Endorsed Attributes
- Cloud Compute
  - 1: Cloud IaaS, HTC Cluster, Cloud Storage, Federated Cloud IaaS & HTC
  - 2: IS Registry, IS Directory, VM Library
  - 3: Operational Integration, Resource Accounting
- Data Transfer
  - 1: Network File Storage, File Transfer
  - 2: Data Set Replication, PID & Meta-data Registry

# Leverage existing e-Infrastructures

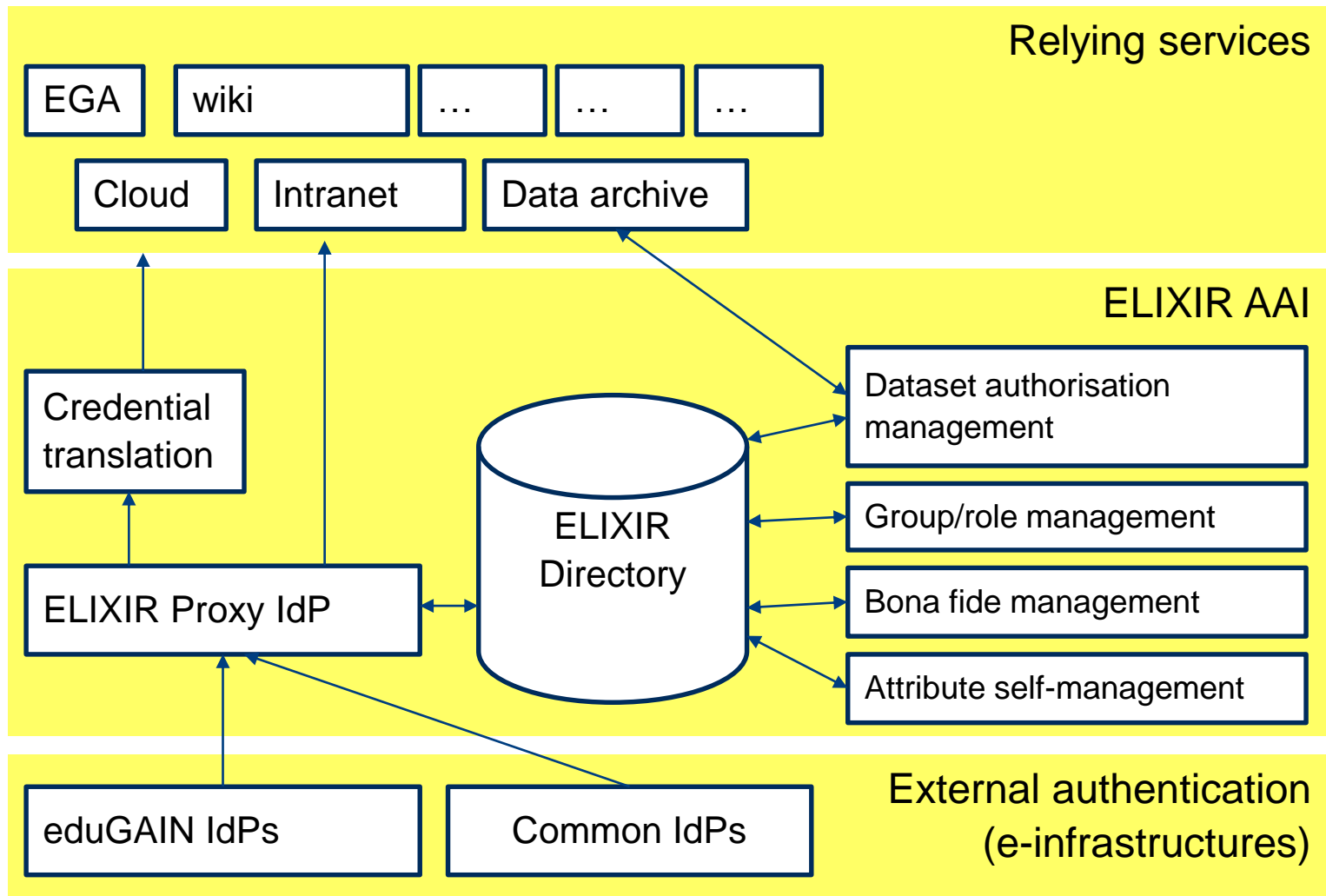
- Networks – Geant
- Data – EUDAT
- Compute infrastructure – EGI, PRACE
- Cloud infrastructure – EGI, Helix Nebula



# Leverage existing e-Infrastructures

- Networks – Geant
  - Federated ID not universal across all user communities
  - Integration with other IDs not currently supported
  - Credential Translation (e.g. to X.509) not provided as a service
- Data – EUDAT
- Compute infrastructure – EGI, PRACE
- Cloud infrastructure – EGI, Helix Nebula

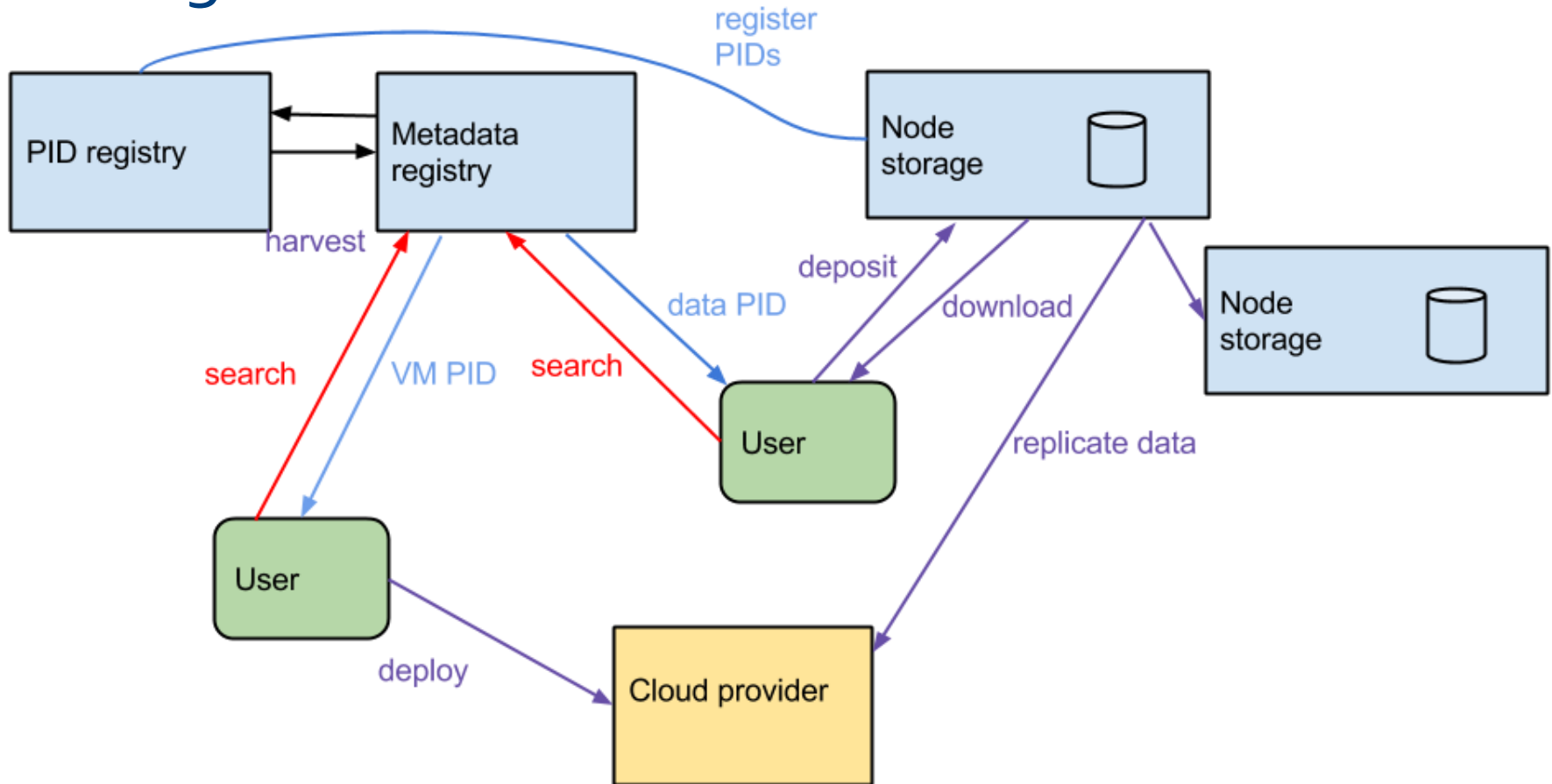
# AAI Architecture



# Leverage existing e-Infrastructures

- Networks – Geant
- Data – EUDAT
  - Engagement model complicated by 'joining the CDI'
  - Services tuned to long-tail users with no community infrastructure
  - EMBL-EBI working within EUDAT to explore adoption within ELIXIR
- Compute infrastructure – EGI, PRACE
- Cloud infrastructure – EGI, Helix Nebula

# Storage Architecture



ELIXIR already has:

- Model for community submissions
- Generating unique identifiers
- Meta-data catalogues and discovery services

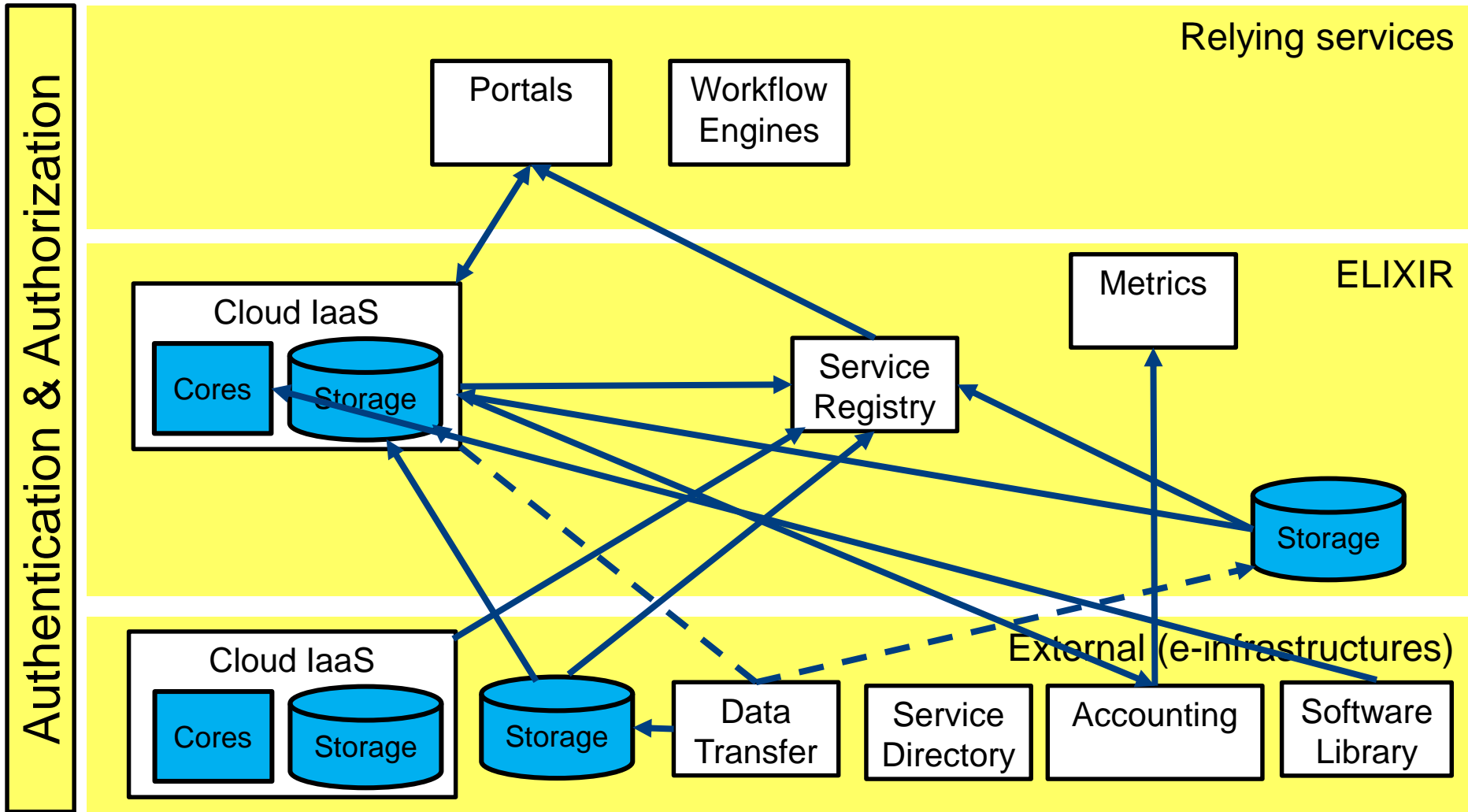
Exploring with EUDAT:

- Reference Data Set Replication
- Targeting cluster & cloud sites

# Leverage existing e-Infrastructures

- Networks – Geant
- Data – EUDAT
- Compute infrastructure – EGI, PRACE
  - Excelerate scientific use cases not 'PRACE' class
  - Many analysis pipelines built around clusters with fast storage
- Cloud infrastructure – EGI, Helix Nebula
  - Cloud Infrastructures need at both small and large scale

# Cloud & Compute Architecture



# The e-Infrastructure Commons

'... as an integrated living ecosystem of resources and services that is open, user friendly and accessible to European researchers and scientists, and continuously adapts to the changing requirements of research and science.'

Currently, from a user perspective...

- Services are not integrated
- Services are not user friendly
- Services are not accessible
- Great variability of quality, capacity & policy across countries
- Rate of change to improve this situation appears slow

# But... plenty of useful components

- Authentication & Authorisation
  - Adoption of Federated Identity is not ubiquitous – but it works
  - Different e-infrastructures have their own working systems
  - Systems from one e-Infrastructure not supported by others
  - There are investments that may improve this (e.g. AARC)
- Compute & Data Infrastructures
  - Lack of integration between current projects (e.g. EGI & EUDAT)
  - Some useful services within EGI for ELIXIR compute platforms
  - EGI, HNI & Cloud solutions provide different resourcing solutions

The burden of integration falls incorrectly on the user



# European Open Science Cloud (EOSC)

- Services for Open Science in Europe
  - Sustainable, Elastic, Reliable, Available, ...
- Long-term preservation & availability of research data and tools
  - Both for specific science domains and generally
- Vendor neutral, collaborative, secure and trusted environment
  - Delivery from both the public and private sector
- Supporting all research projects
  - National, international, public-private partnerships and private enterprises.

# Doing research with the EOSC

- A single, central open user-driven governance
  - Representation across commercial and public sectors
- Authentication & Authorisation (common across all areas)
  - Transparent identity mapping
- Data Infrastructure (locating, storing, moving)
  - FAIR: Findable, Accessible, Interoperable and Reusable
- Compute Infrastructure (primarily cloud)
  - High bandwidth links within the data and compute infrastructure
- Research Reproducibility (papers, software & data)
  - Embed the research life-cycle into the infrastructure

# Summary

- The e-Infrastructure commons as it exists is not usable
  - We have a walled commons which you cannot cross
- Research communities need usable services for open science
  - This investment is currently coming from the community
- The European Open Science Cloud 'may' address this
  - Put users in the driving seat with the ability to influence funding