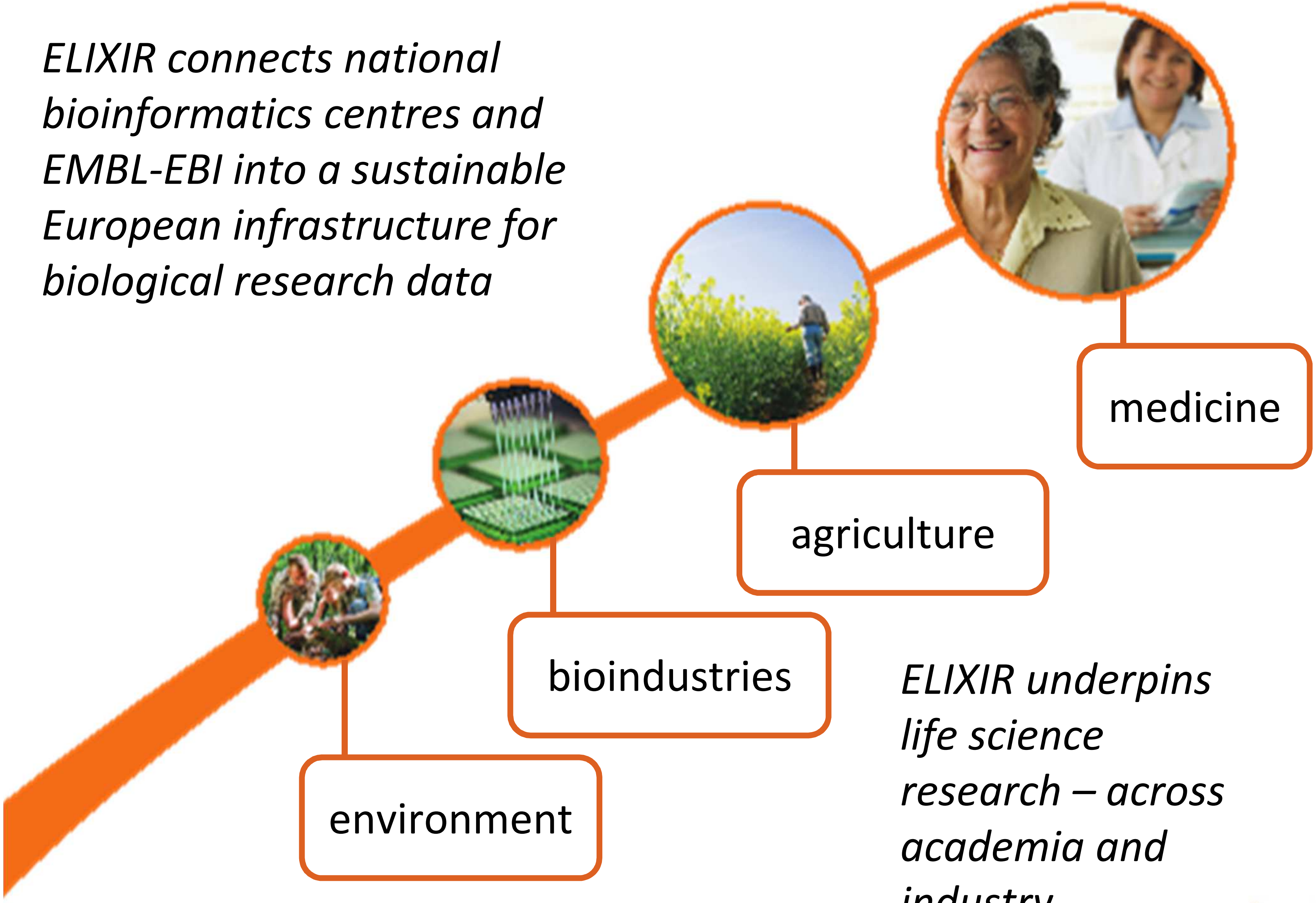
ELIXIR AAI

CORBEL/AARC AAI workshop
31 May - 1 Jun 2016

Mikael Linden,
ELIXIR AAI task

www.elixir-europe.org

ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data



environment

bioindustries

agriculture

medicine

ELIXIR underpins life science research – across academia and industry

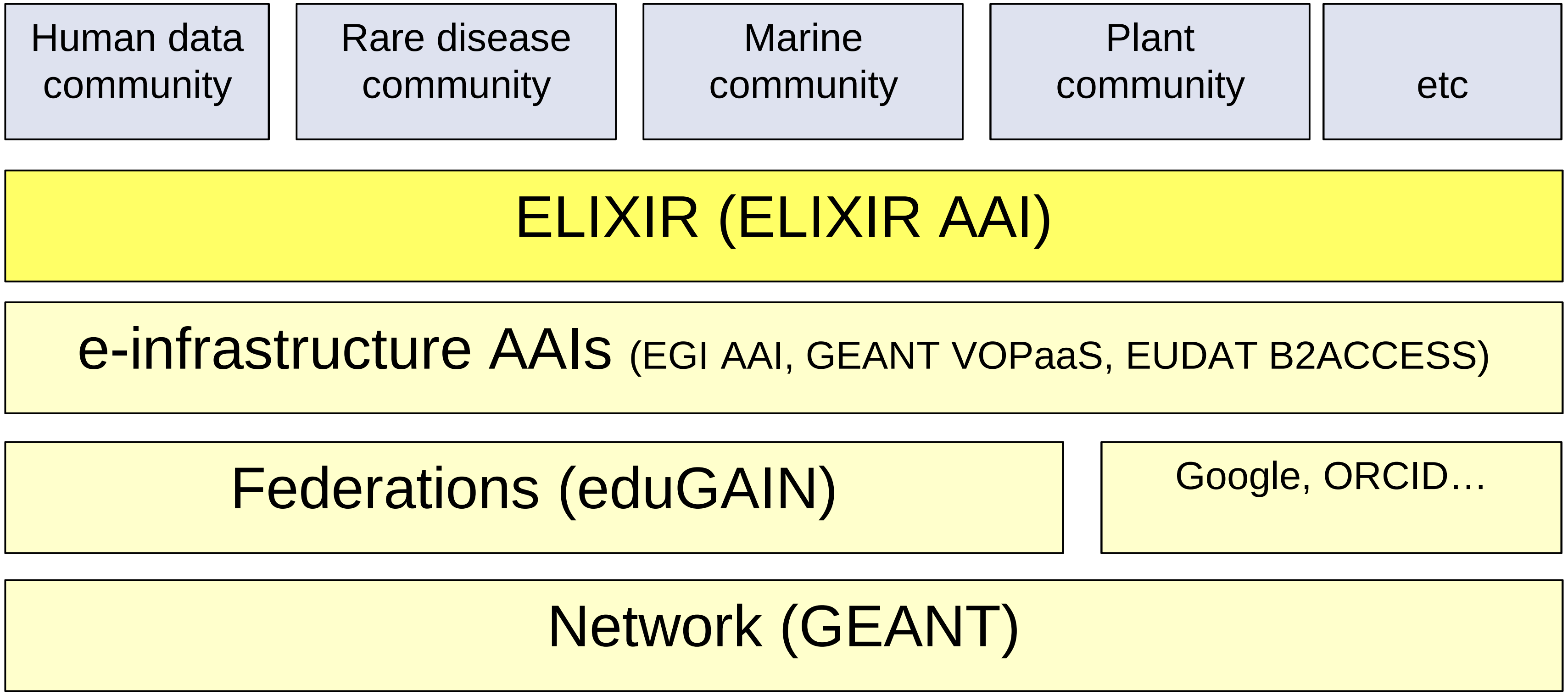
ELIXIR AAI history – where we are now

- Use case gathering -- Autumn 2014
  - https://docs.google.com/document/d/12fBLl8WenlxABQDzYEAJ6NtOhIXnZBPiGQLuhFUCyqo/edit
- Requirements and design – Spring 2015
  - https://docs.google.com/document/d/1CMY1np3GyvPD8LcKvXljXcRO04V2zu3n_Jcg19jgNOw/edit
- Deployment starts – Autumn 2015 – EXCELERATE WP4.3.1
  - Part of ELIXIR Compute platform
- First release -- August 2016
  - Until that ELIXIR AAI in pilot status
  - Key components up and running already

High level stuff:
ELIXIR AAI strategy (proposal)

- Covers
  - ELIXIR AAI under the responsibility of the hub
  - Relations to e-infrastructures (collaborate, make use of)
  - Relations to other BMS research infrastructure (common AAI)
  - ELIXIR AAI policies for end users, relying parties and AAI operators
- To be presented to ELIXIR Heads of Nodes in June
- Proposal (submitted to Heads of Nodes for agreement):
https://docs.google.com/document/d/1cJ3mR8lqfZKRMvSFalSmPbqd1OPU-L6YcUFIRnh1rhQ/edit

The stack



Human data community

Rare disease community

Marine community

Plant community

etc

ELIXIR (ELIXIR AAI)

e-infrastructure AAIs (EGI AAI, GEANT VOPaaS, EUDAT B2ACCESS)

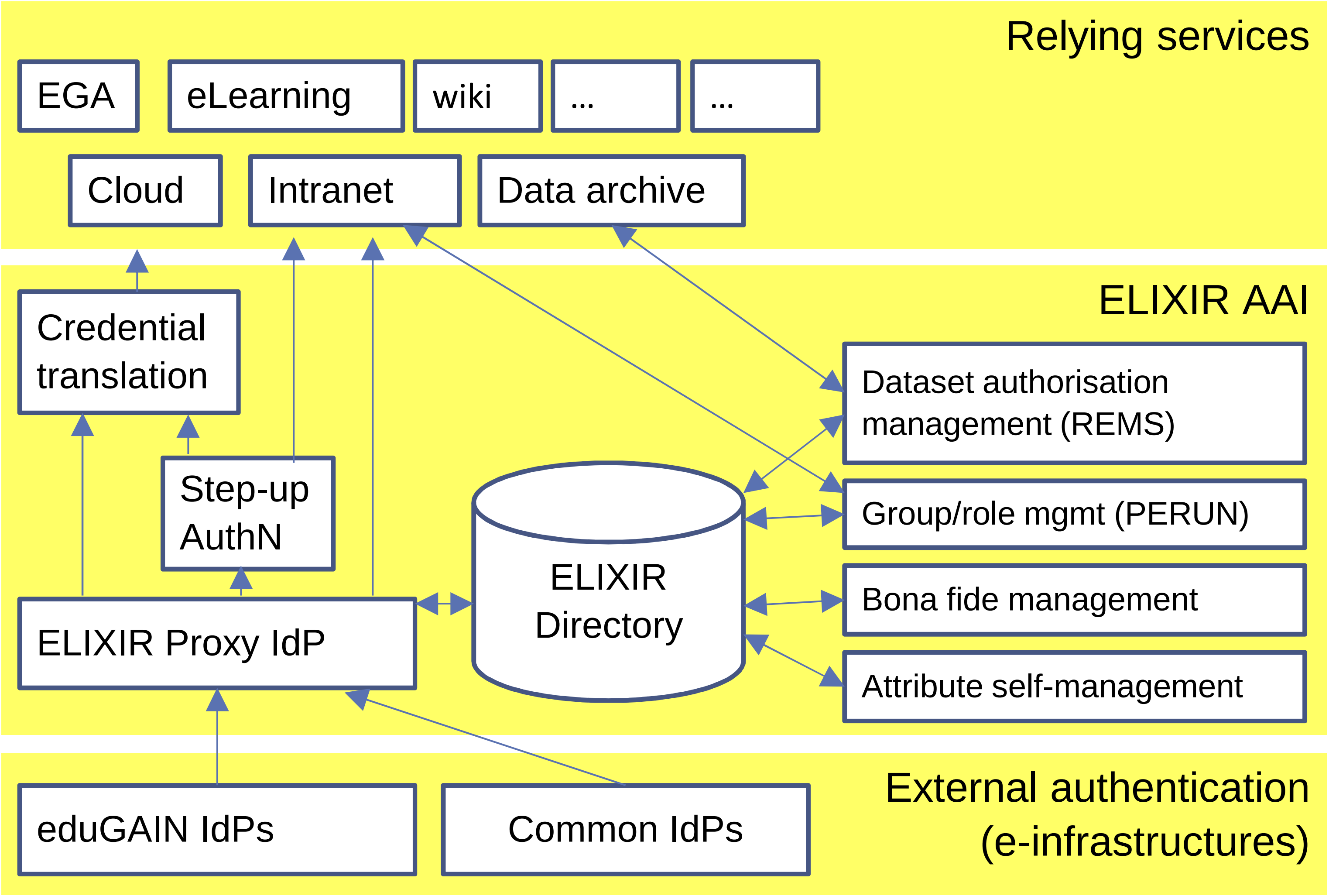
Federations (eduGAIN)

Google, ORCID…

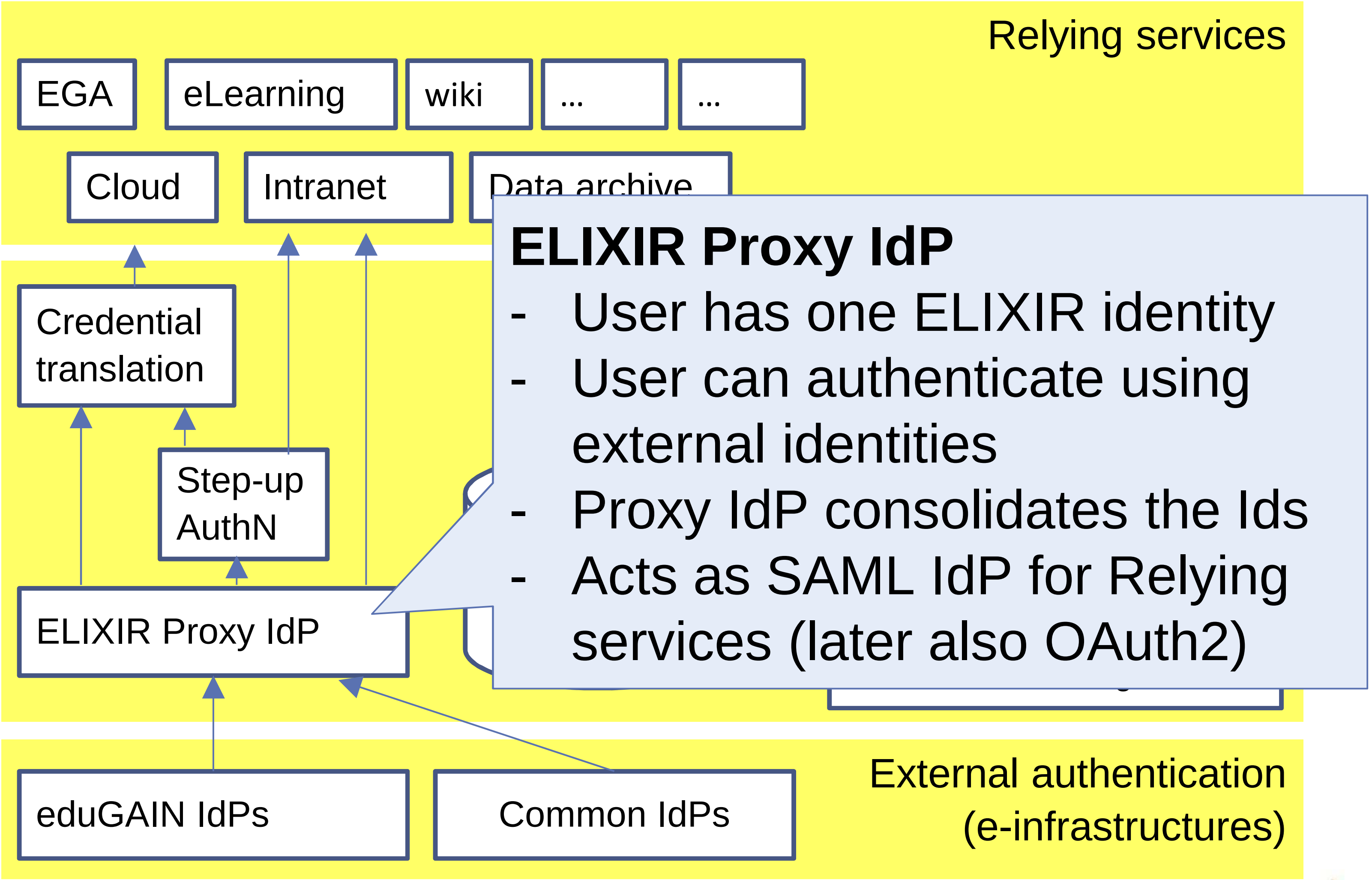
Network (GEANT)

Design of ELIXIR AAI

ELIXIR AAI design



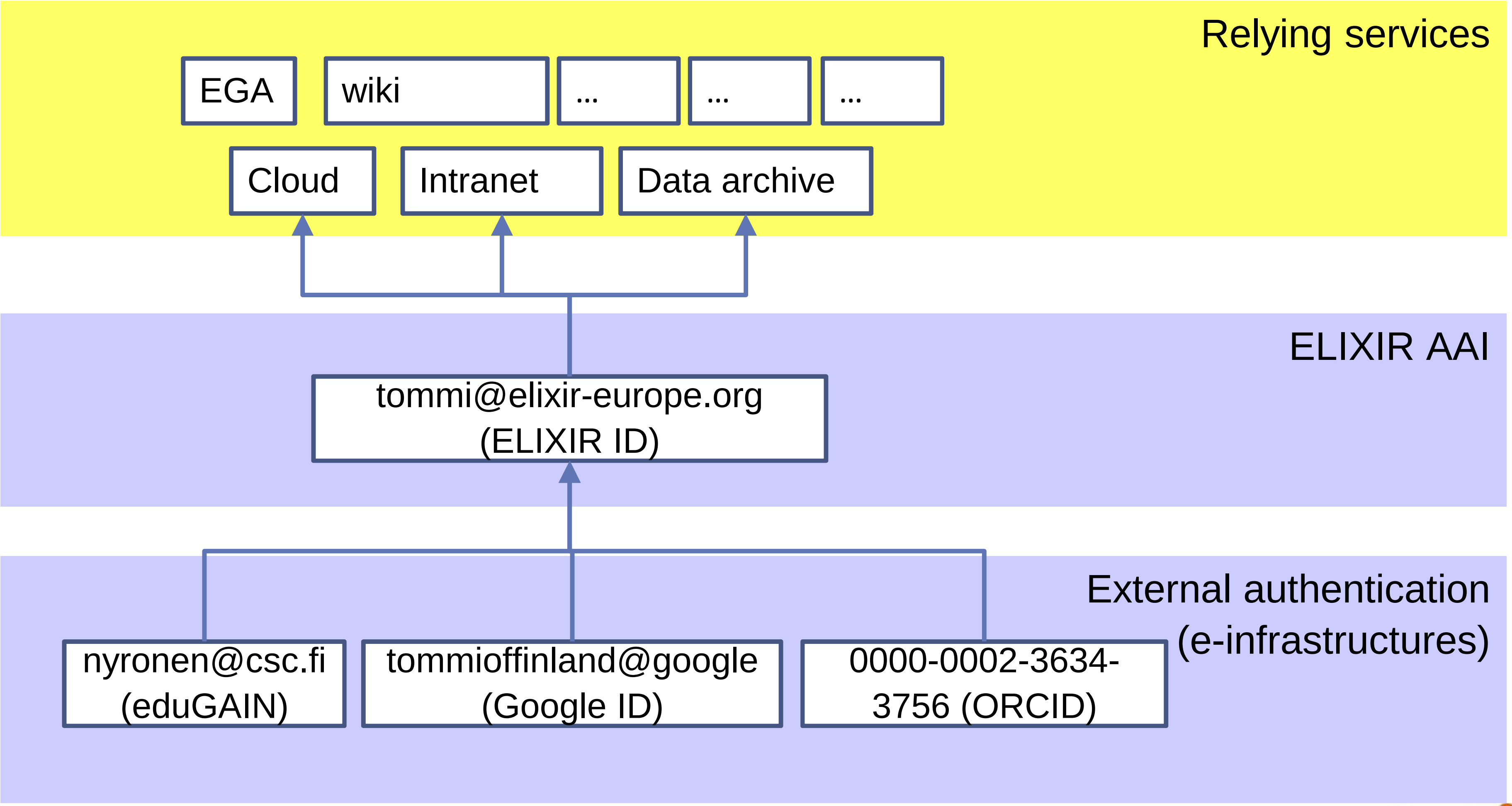
ELIXIR AAI design



ELIXIR Proxy IdP

- User has one ELIXIR identity
- User can authenticate using external identities
- Proxy IdP consolidates the Ids
- Acts as SAML IdP for Relying services (later also OAuth2)

ELIXIR identity



Relying services

EGA

wiki

…

…

…

Cloud

Intranet

Data archive

ELIXIR AAI

tommi@elixir-europe.org (ELIXIR ID)

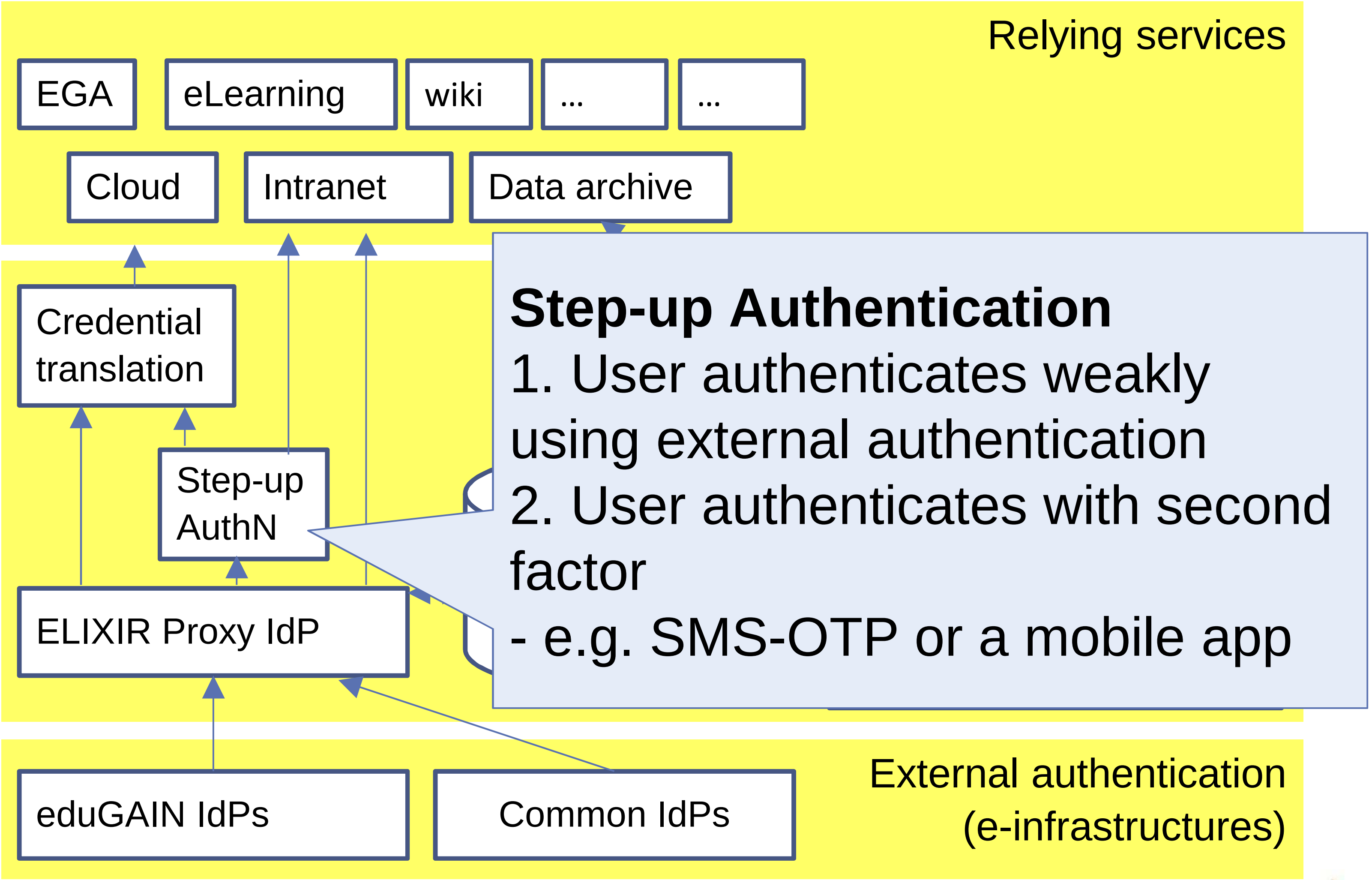
External authentication (e-infrastructures)

nyronen@csc.fi (eduGAIN)

tommioffinland@google (Google ID)

0000-0002-3634-3756 (ORCID)

ELIXIR AAI design



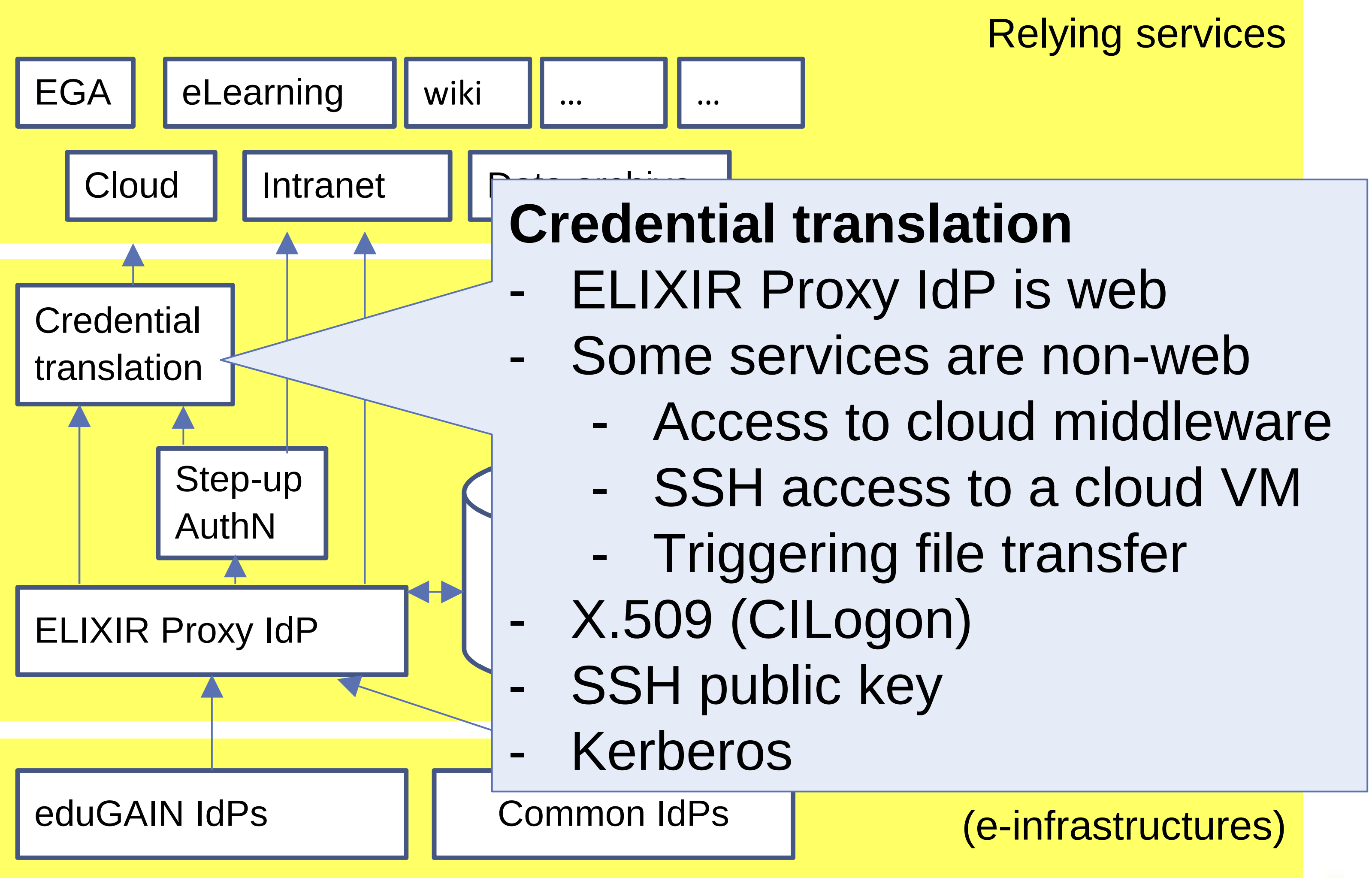
Step-up Authentication

1. User authenticates weakly using external authentication

2. User authenticates with second factor

- e.g. SMS-OTP or a mobile app

ELIXIR AAI design



Relying services

EGA

eLearning

wiki

…

…

Cloud

Intranet

Data archive

Credential translation

Step-up AuthN

ELIXIR Proxy IdP

eduGAIN IdPs

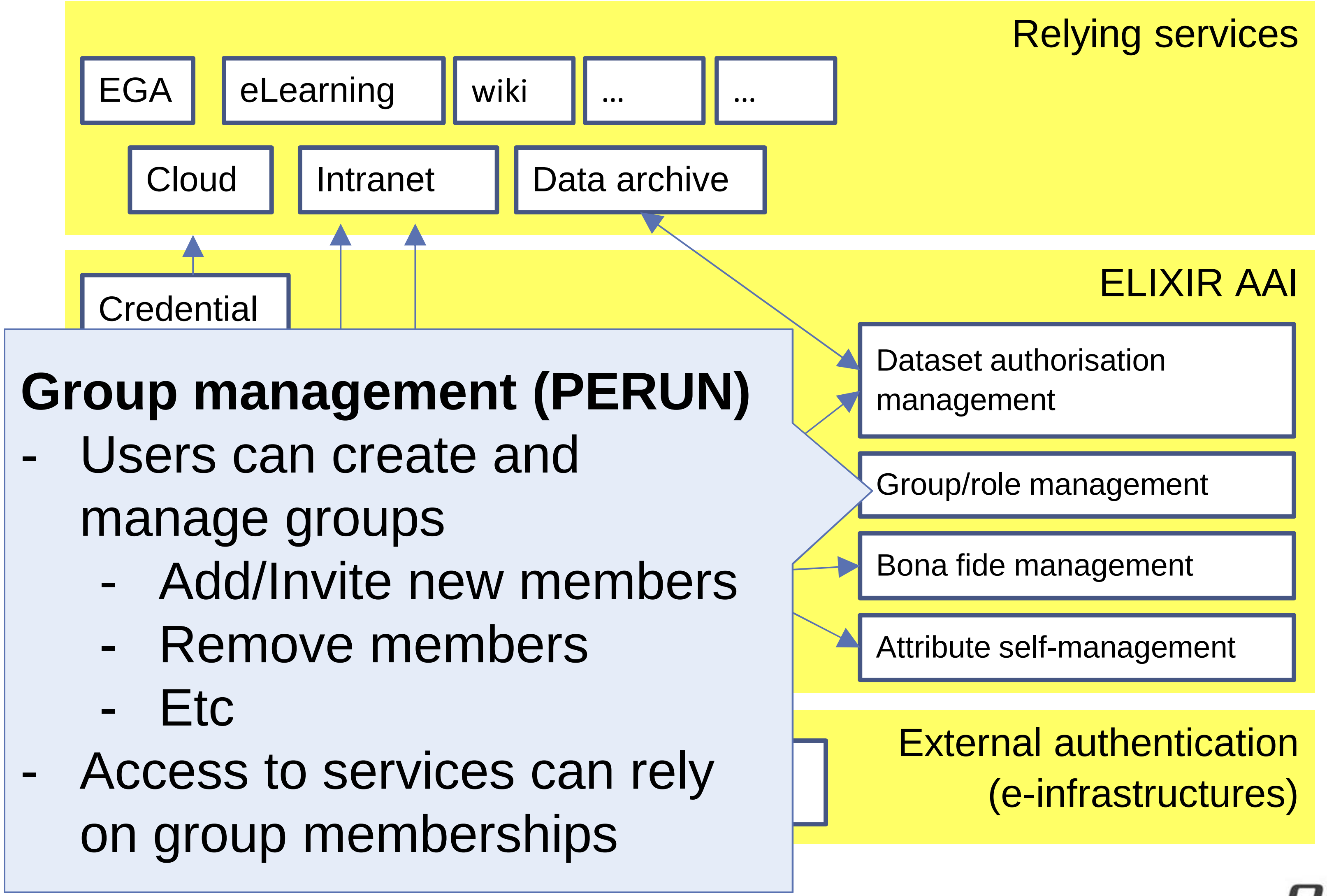
Common IdPs

(e-infrastructures)

Credential translation

- ELIXIR Proxy IdP is web
- Some services are non-web
  - Access to cloud middleware
  - SSH access to a cloud VM
  - Triggering file transfer
- X.509 (CILogon)
- SSH public key
- Kerberos

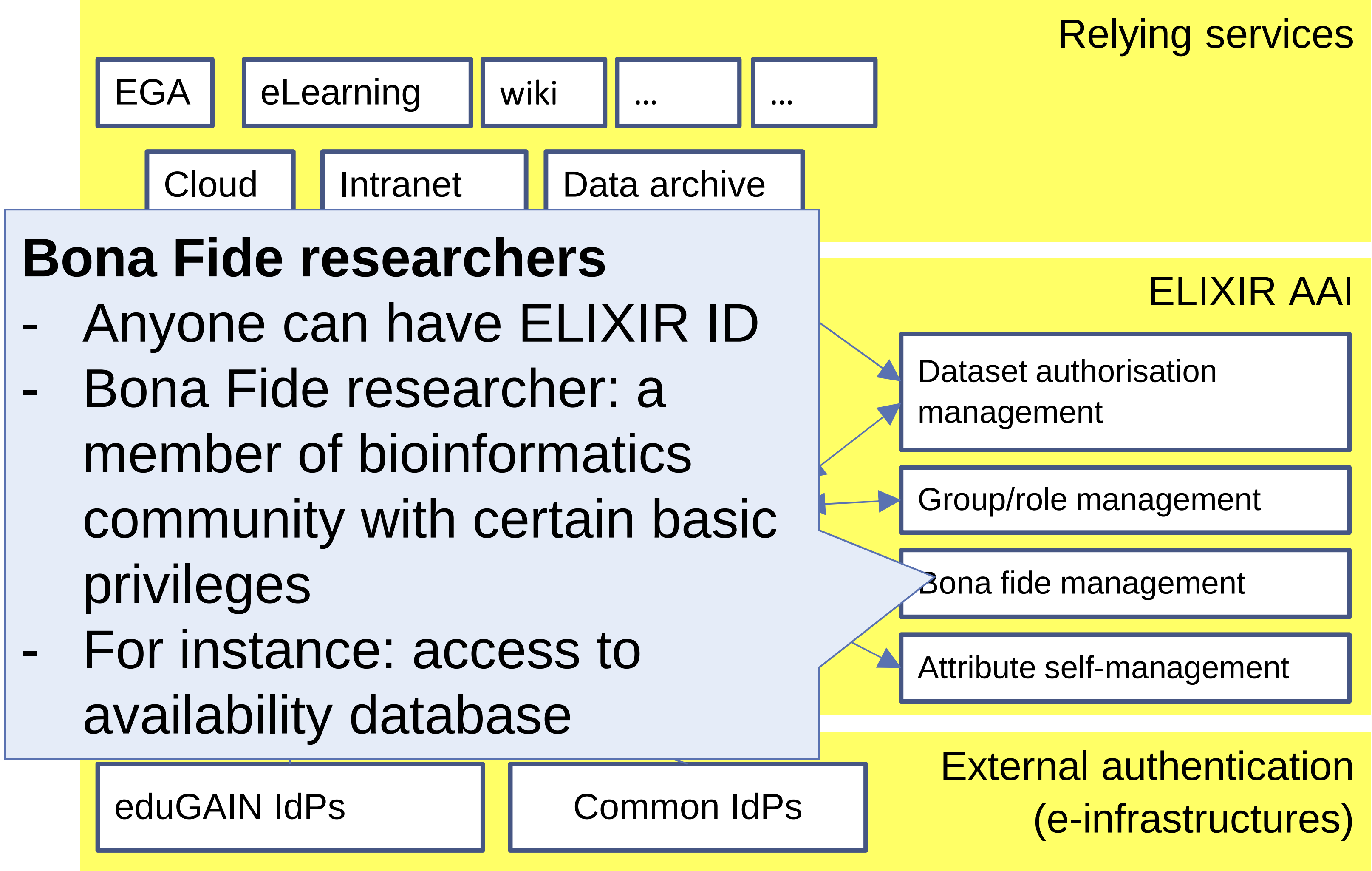
ELIXIR AAI design



Group management (PERUN)

- Users can create and manage groups
  - Add/Invite new members
  - Remove members
  - Etc
- Access to services can rely on group memberships

ELIXIR AAI design



Relying services

EGA

eLearning

wiki

...

...

Cloud

Intranet

Data archive

ELIXIR AAI

Dataset authorisation management

Group/role management

Bona fide management

Attribute self-management

External authentication (e-infrastructures)

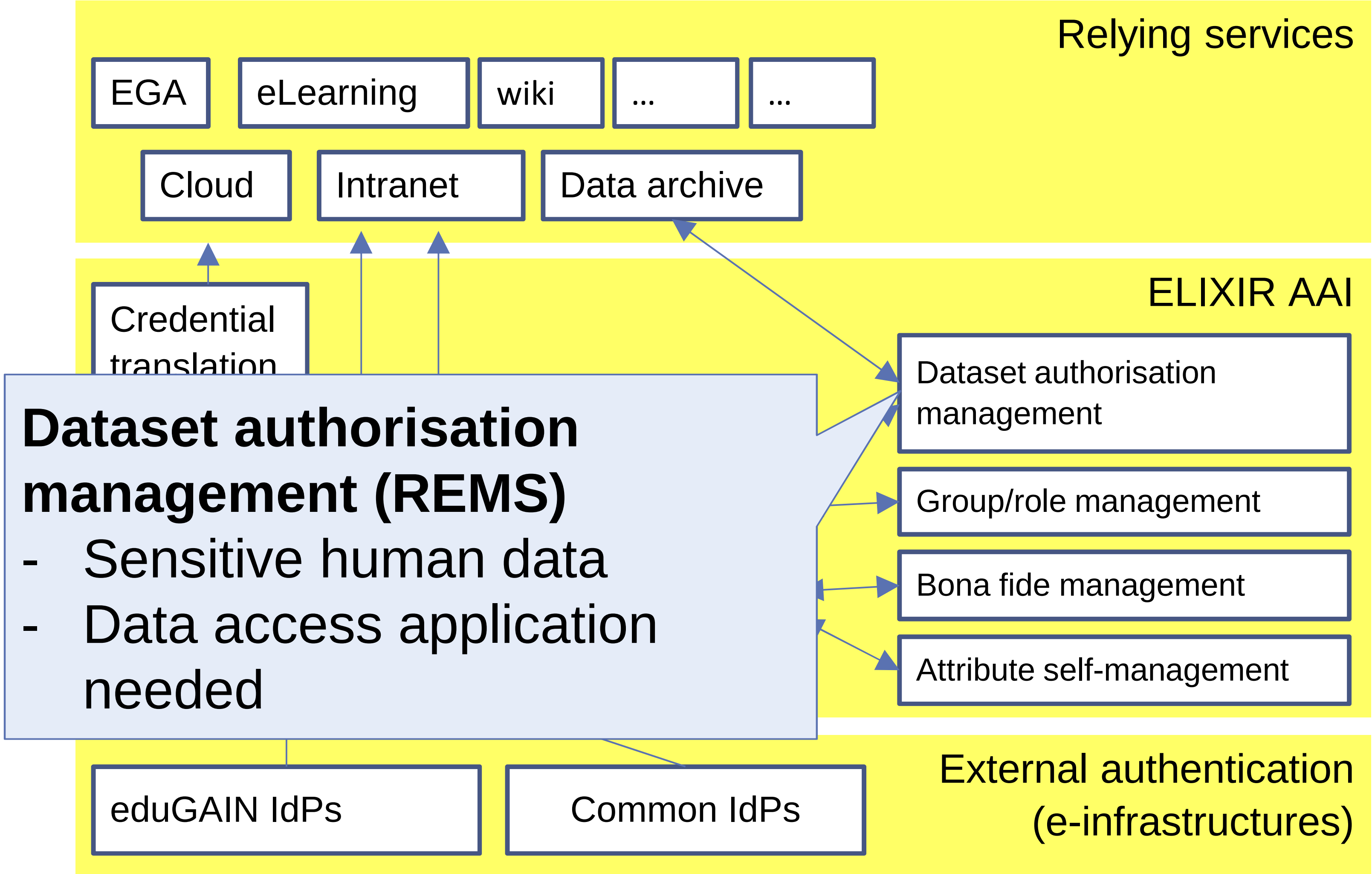
eduGAIN IdPs

Common IdPs

Bona Fide researchers

- Anyone can have ELIXIR ID
- Bona Fide researcher: a member of bioinformatics community with certain basic privileges
- For instance: access to availability database

ELIXIR AAI design



Relying services

EGA

eLearning

wiki

...

...

Cloud

Intranet

Data archive

ELIXIR AAI

Credential translation

Dataset authorisation management

Group/role management

Bona fide management

Attribute self-management

eduGAIN IdPs

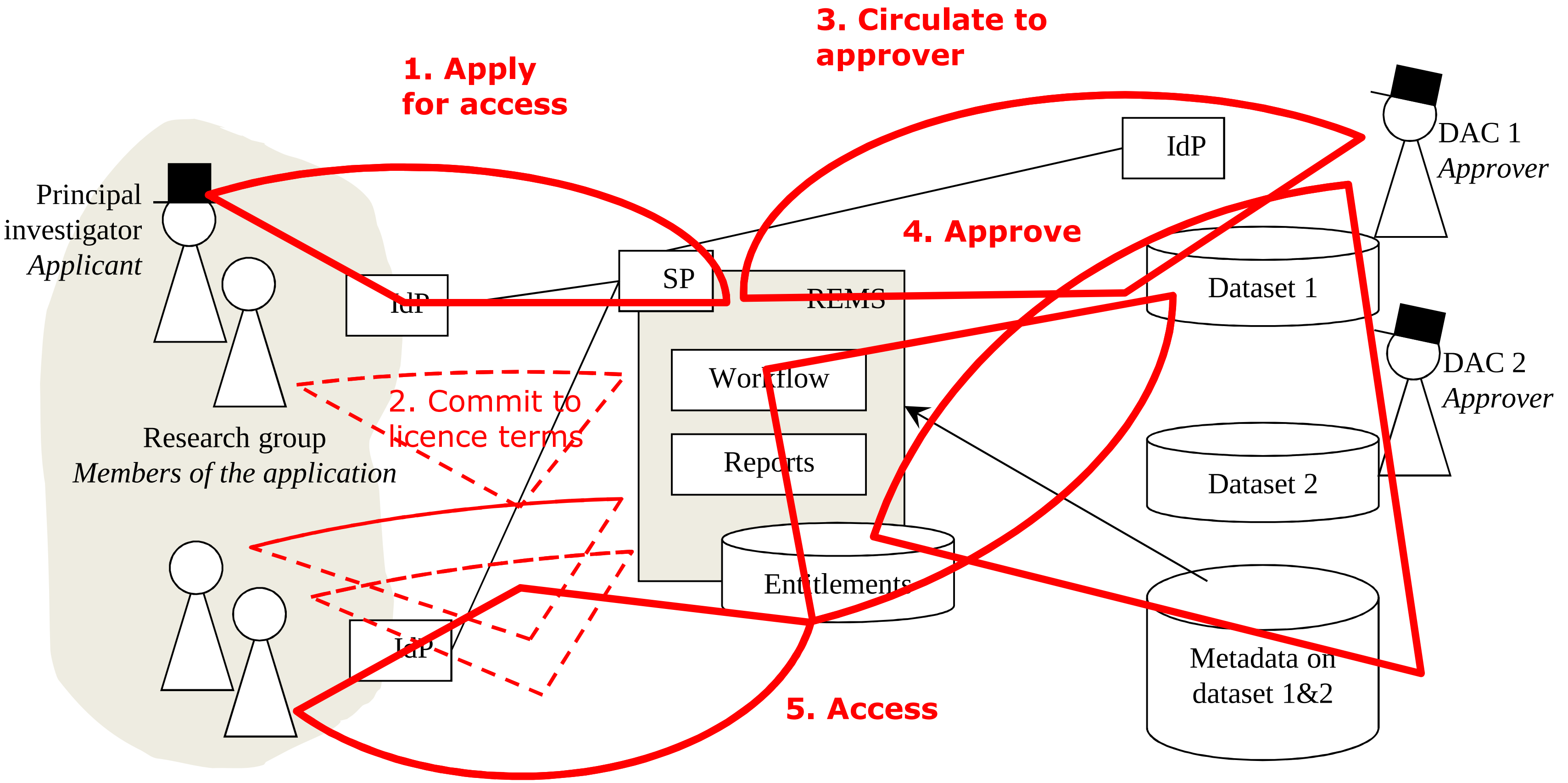
Common IdPs

External authentication (e-infrastructures)

Dataset authorisation management (REMS)

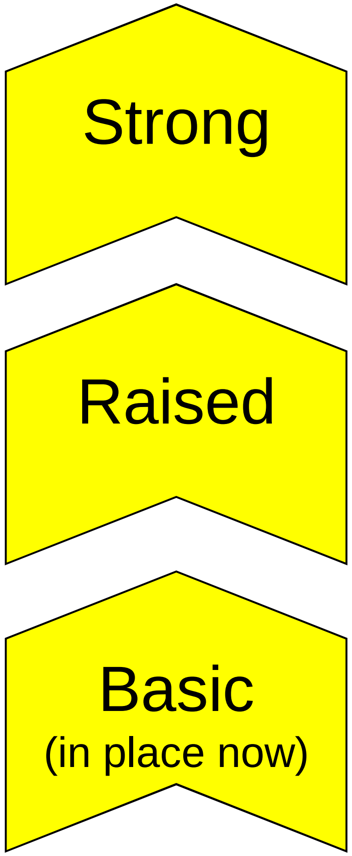
- Sensitive human data
- Data access application needed

The REMS concept



ELIXIR AAI Dimensions of authentication and authorisation

The planned assurance levels for authentication



Strong

Face-to-face proof of identity, two factor authN

- Step-up authentication
- Possibly rely on external sources (e.g. eID)

Raised

Organisation-registed accounts, password authN

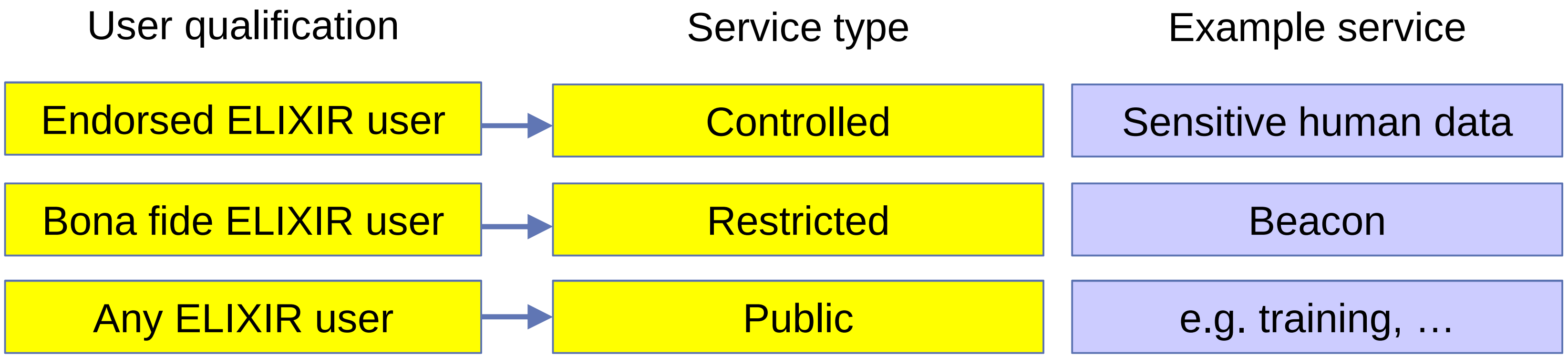
- Log in with Home Organisation IdPs (eduGAIN)
- Requires Home Organisation complying to a minimal assurance level

Basic
(in place now)

Self-registrated accounts, password authN

- Google, LinkedIn, ORCID authentication

The three authorisation layers



User qualification	Service type	Example service
Endorsed ELIXIR user	Controlled	Sensitive human data
Bona fide ELIXIR user	Restricted	Beacon
Any ELIXIR user	Public	e.g. training, …

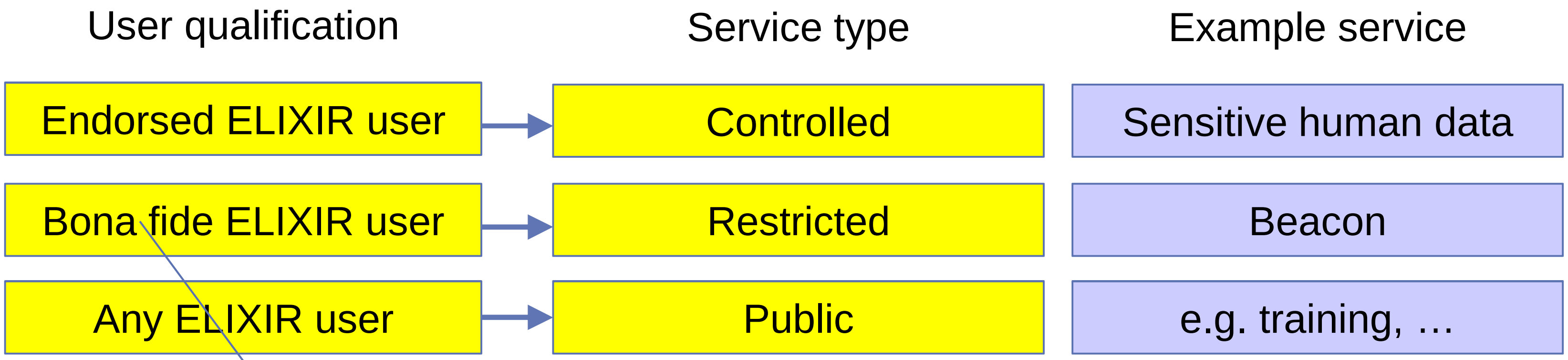
The three authorisation layers

User qualification	Service type	Example service
Endorsed ELIXIR user	Controlled	Sensitive human data
Bona fide ELIXIR user	Restricted	Beacon
Any ELIXIR user	Public	e.g. training, …

Any user:

- Who has committed to ELIXIR Acceptable U

The three authorisation layers



User qualification	Service type	Example service
Endorsed ELIXIR user	Controlled	Sensitive human data
Bona fide ELIXIR user	Restricted	Beacon
Any ELIXIR user	Public	e.g. training, …

”Bona Fide” researcher

- Must commit to a Code of Conduct
- Must be a researcher
    - have publications
    - be vouched for by a person who has publi
    - home organisation confirms

The three authorisation layers

User qualification	Service type	Example service
Endorsed ELIXIR user	Controlled	Sensitive human data
Bona fide ELIXIR user	Restricted	Beacon
Any ELIXIR user	Public	e.g. training, …

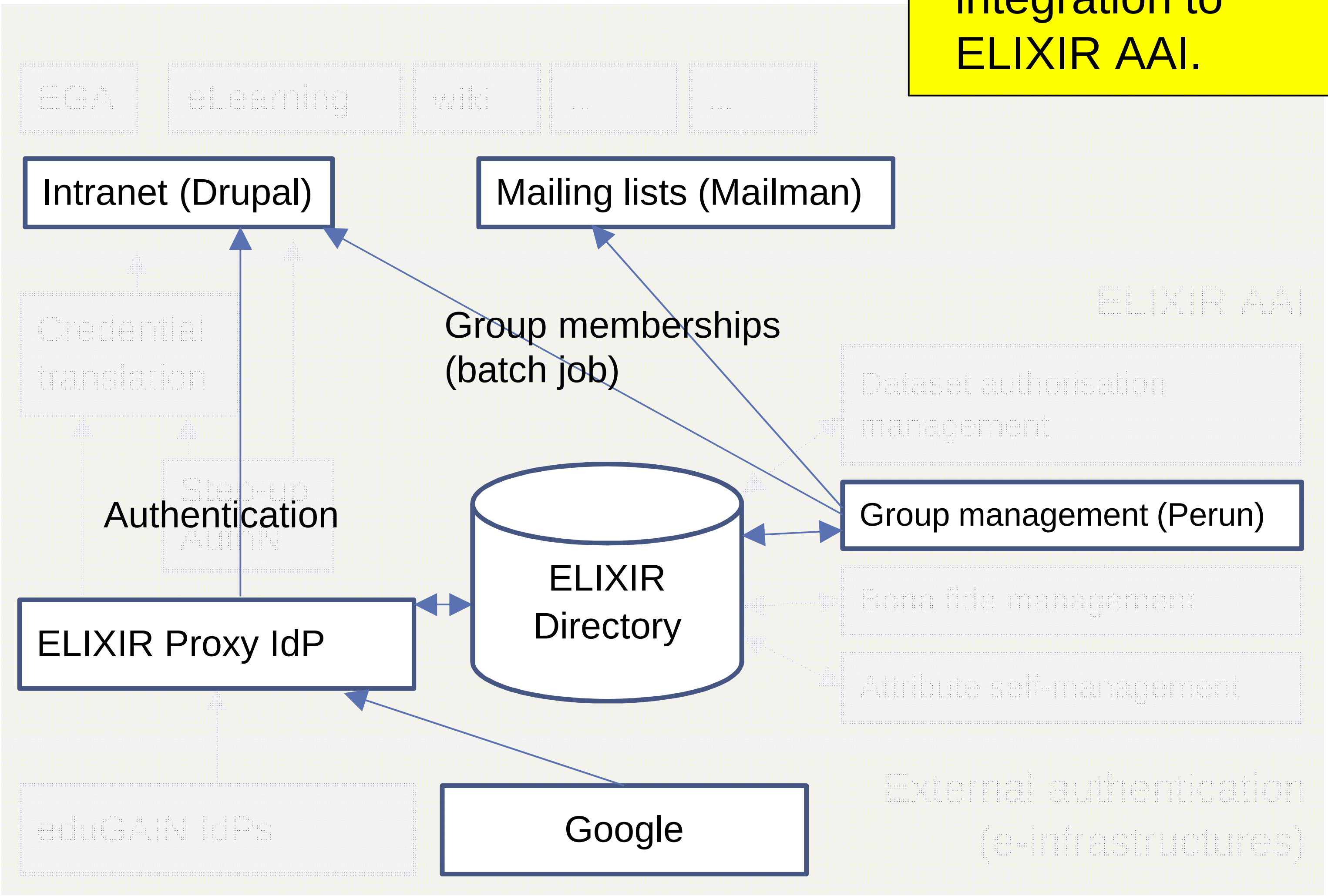
Endorsed user

- The user needs to apply for access
    - attach a research plan
- Each application is screened individually (e.g

Use case:
Collaborative service

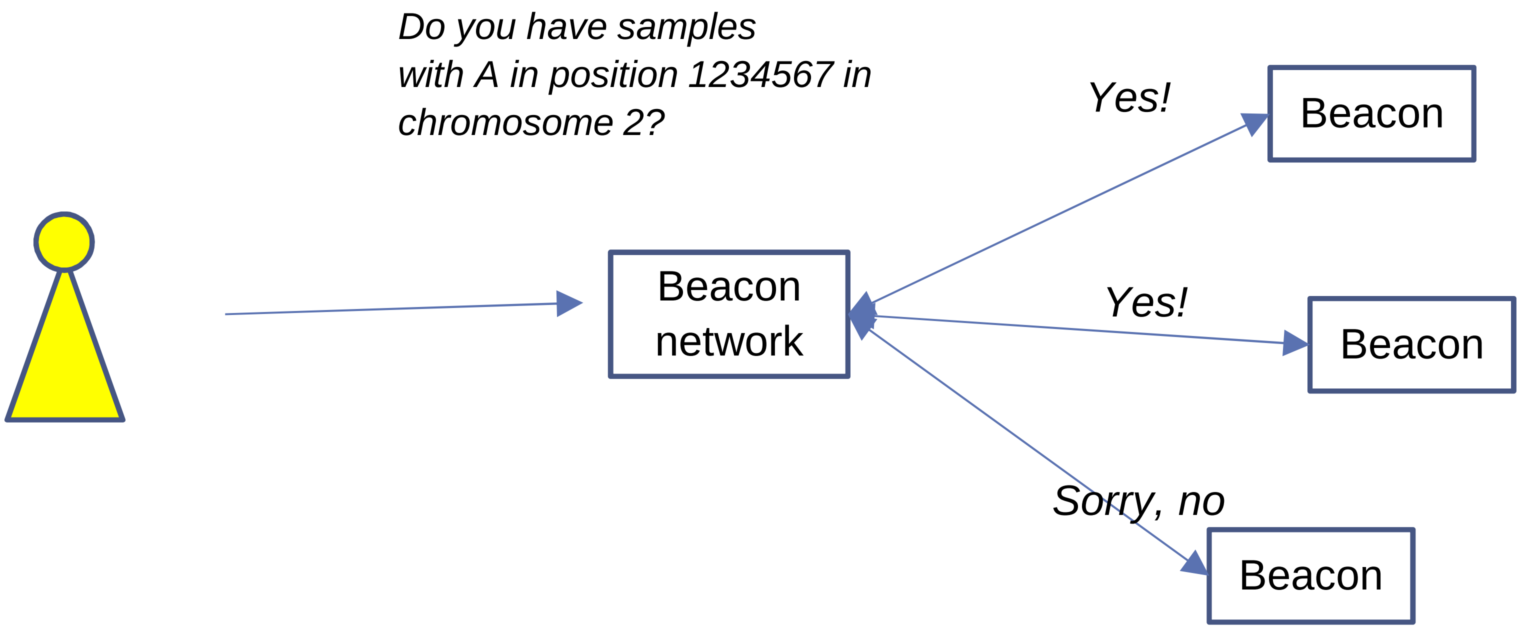
ELIXIR Intranet and mailing lists

Current operational integration to ELIXIR AAI.



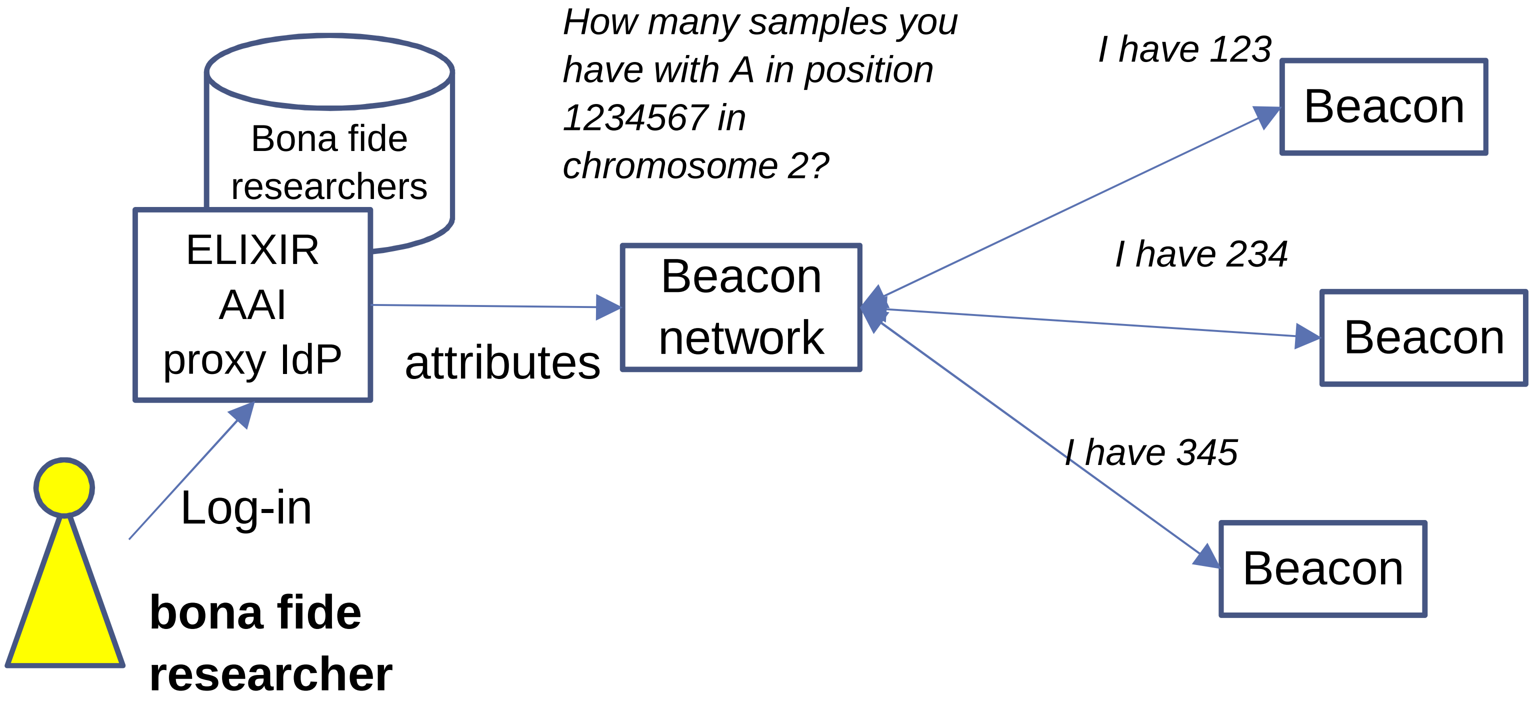
Use case:
ELIXIR Beacon

Beacon idea – public access


Do you have samples with A in position 1234567 in chromosome 2?
Beacon network
Yes!
Beacon
Yes!
Beacon
Sorry, no
Beacon

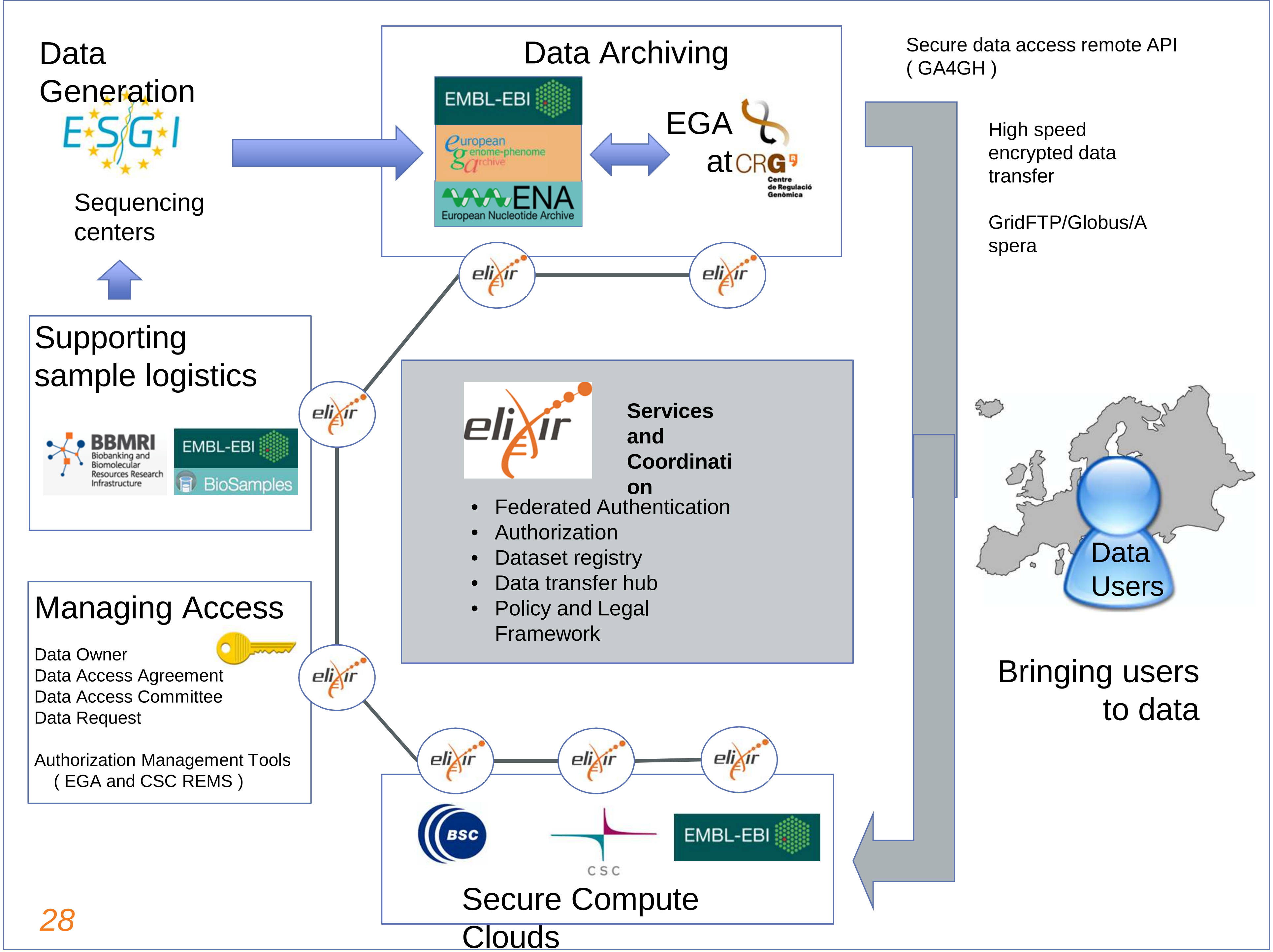
- Simple REST API for the query
- https://genomicsandhealth.org/work-products-demonstration-projects/beacons

Beacon – restricted access



- ELIXIR AAI would keep record on bona fide researchers
- ELIXIR AAI would use OAuth2/OpenID Connect to deliver the bona fide attribute to the Beacon network & beacons
- Each beacon would enforce access control

Use case:
Sensitive human data



Data Generation

Sequencing centers

Supporting sample logistics

Managing Access

Data Owner
Data Access Agreement
Data Access Committee
Data Request

Authorization Management Tools
(EGA and CSC REMS)

Data Archiving

EGA at CRG

Services and Coordination

- Federated Authentication
- Authorization
- Dataset registry
- Data transfer hub
- Policy and Legal Framework

Secure Compute Clouds

Secure data access remote API (GA4GH)

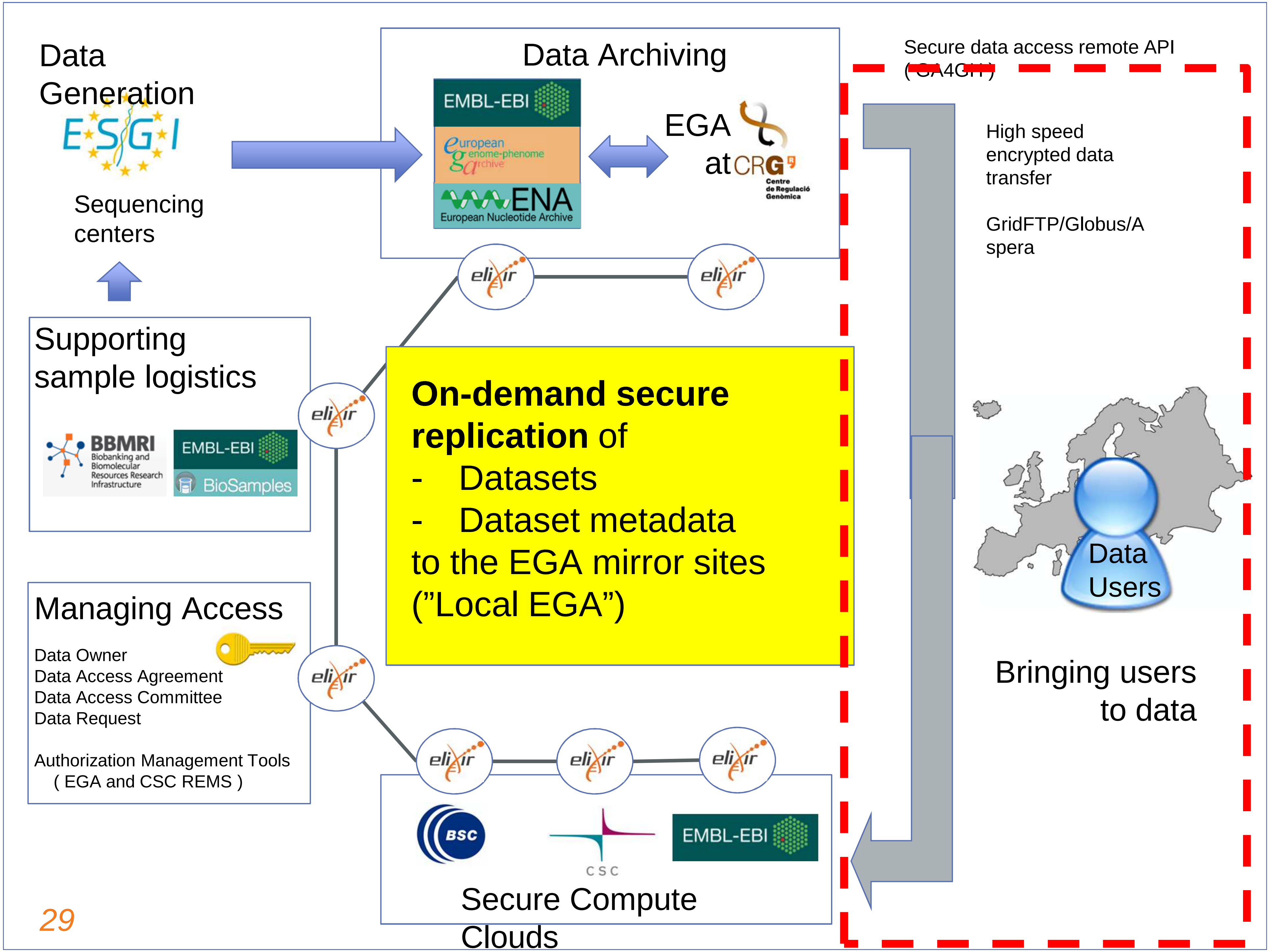
High speed encrypted data transfer

GridFTP/Globus/Aspera

Data Users

Bringing users to data

ESGI
EMBL-EBI
european genome-phenome archive
ENA European Nucleotide Archive
Centre de Regulació Genòmica
BBMRI Biobanking and Biomolecular Resources Research Infrastructure
BioSamples
elixir
BSC
CSC



Data Generation

Sequencing centers

Supporting sample logistics

Managing Access

Data Owner
Data Access Agreement
Data Access Committee
Data Request

Authorization Management Tools
(EGA and CSC REMS)

Data Archiving

EGA at CRG

On-demand secure replication of
- Datasets
- Dataset metadata

to the EGA mirror sites ("Local EGA")

Secure Compute Clouds

Secure data access remote API (GA4GH)

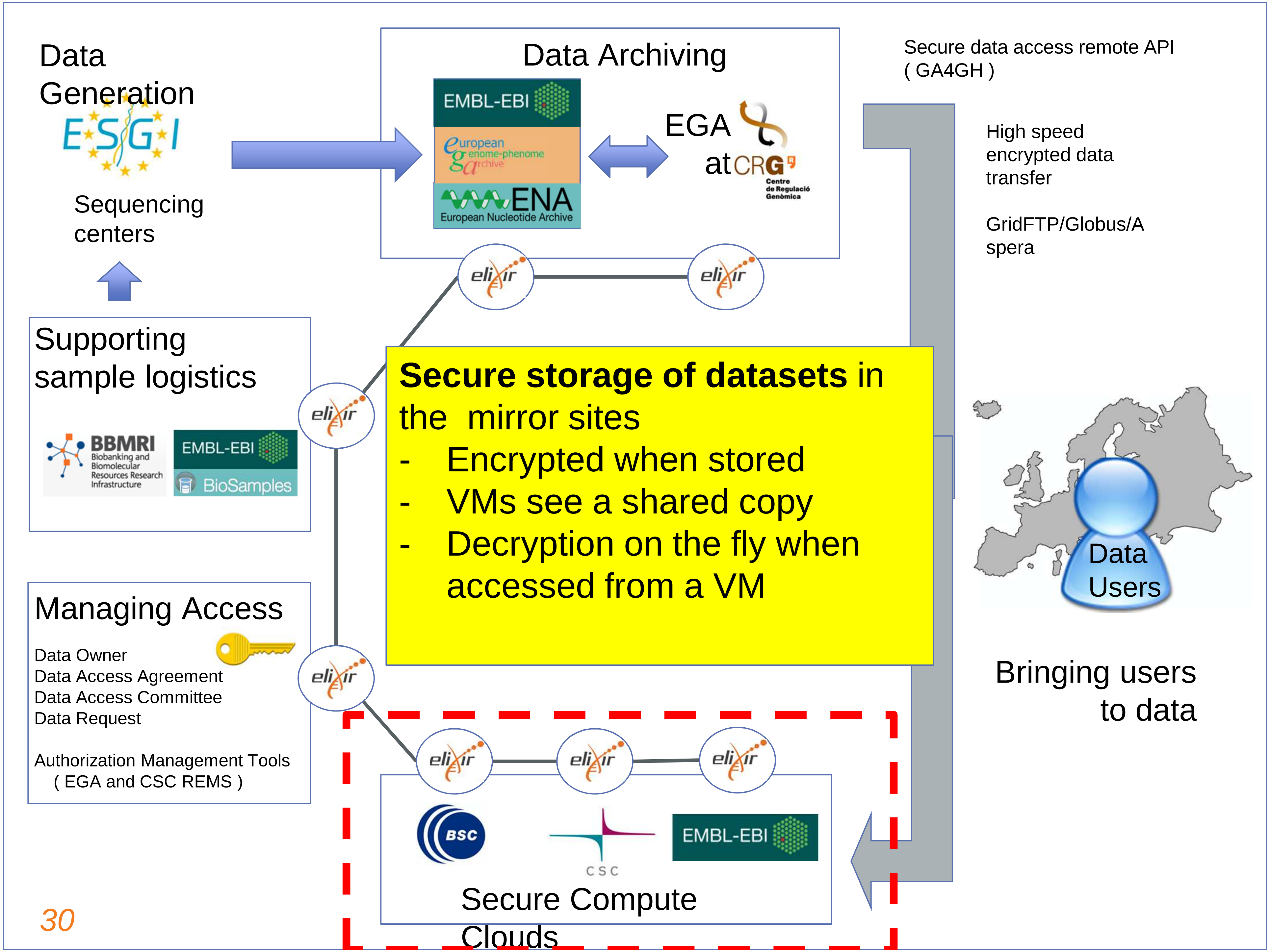
High speed encrypted data transfer

GridFTP/Globus/A
spera

Data Users

Bringing users to data

ESGI
EMBL-EBI
european genome-phenome archive
ENA
European Nucleotide Archive
Centre de Regulació Genòmica
BBMRI
Biobanking and Biomolecular Resources Research Infrastructure
EMBL-EBI
BioSamples
elixir
BSC
CSC
EMBL-EBI



Data Generation

Sequencing centers

Data Archiving

EGA at CRG

Secure data access remote API (GA4GH)

High speed encrypted data transfer

GridFTP/Globus/Aspera

Supporting sample logistics

Secure storage of datasets in the mirror sites

- Encrypted when stored
- VMs see a shared copy
- Decryption on the fly when accessed from a VM

Data Users

Bringing users to data

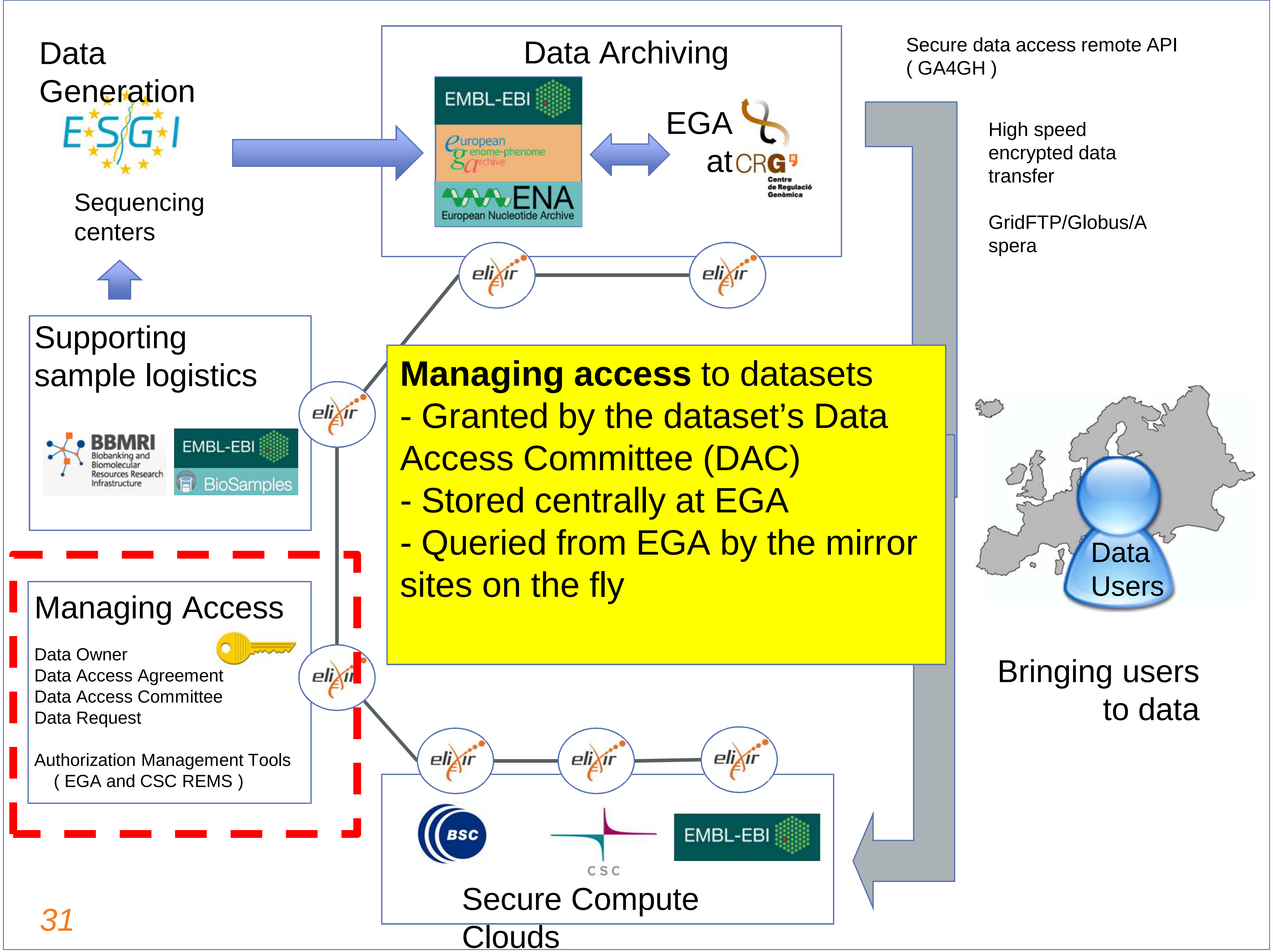
Managing Access

Data Owner
Data Access Agreement
Data Access Committee
Data Request

Authorization Management Tools
(EGA and CSC REMS)

Secure Compute Clouds

ESGI
EMBL-EBI
european genome-phenome archive
ENA European Nucleotide Archive
Centre de Regulació Genòmica
elixir
BBMRI Biobanking and Biomolecular Resources Research Infrastructure
EMBL-EBI
BioSamples
BSC
CSC
EMBL-EBI



Data Generation

Sequencing centers

Supporting sample logistics

Managing Access

Data Owner
Data Access Agreement
Data Access Committee
Data Request

Authorization Management Tools
(EGA and CSC REMS)

Data Archiving

Secure data access remote API (GA4GH)

High speed encrypted data transfer

GridFTP/Globus/Aspera

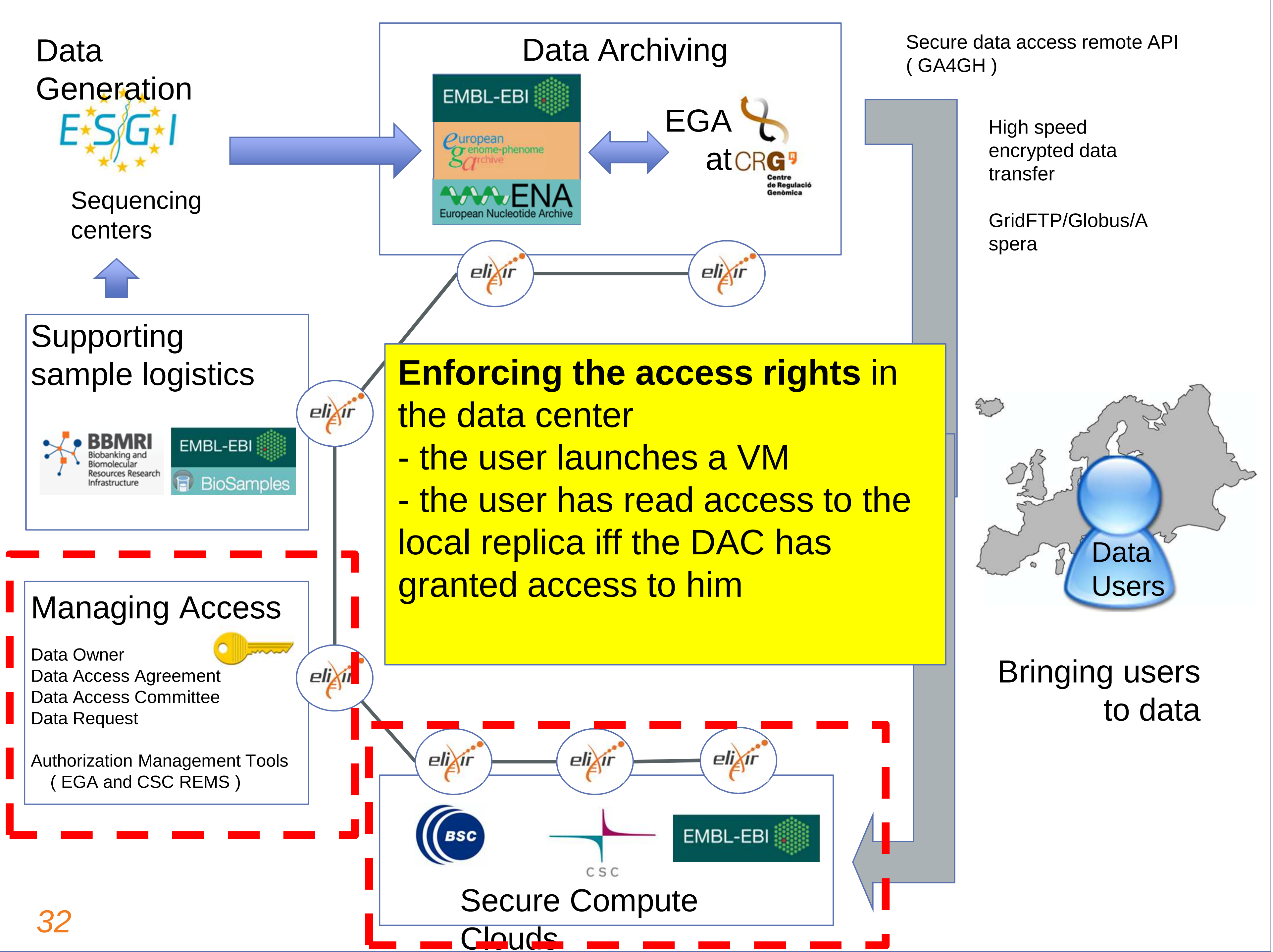
Managing access to datasets
- Granted by the dataset's Data Access Committee (DAC)
- Stored centrally at EGA
- Queried from EGA by the mirror sites on the fly

Data Users

Bringing users to data

Secure Compute Clouds

ESGI
EMBL-EBI
european genome-phenome archive
ENA
European Nucleotide Archive
EGA at CRG
Centre de Regulació Genòmica
elixir
BBMRI
Biobanking and Biomolecular Resources Research Infrastructure
EMBL-EBI
BioSamples
BSC
CSC
EMBL-EBI



Data Generation

Sequencing centers

Data Archiving

EGA at CRG

Secure data access remote API (GA4GH)

High speed encrypted data transfer

GridFTP/Globus/Aspera

Supporting sample logistics

Enforcing the access rights in the data center

- the user launches a VM
- the user has read access to the local replica iff the DAC has granted access to him

Managing Access

Data Owner
Data Access Agreement
Data Access Committee
Data Request

Authorization Management Tools
(EGA and CSC REMS)

Data Users

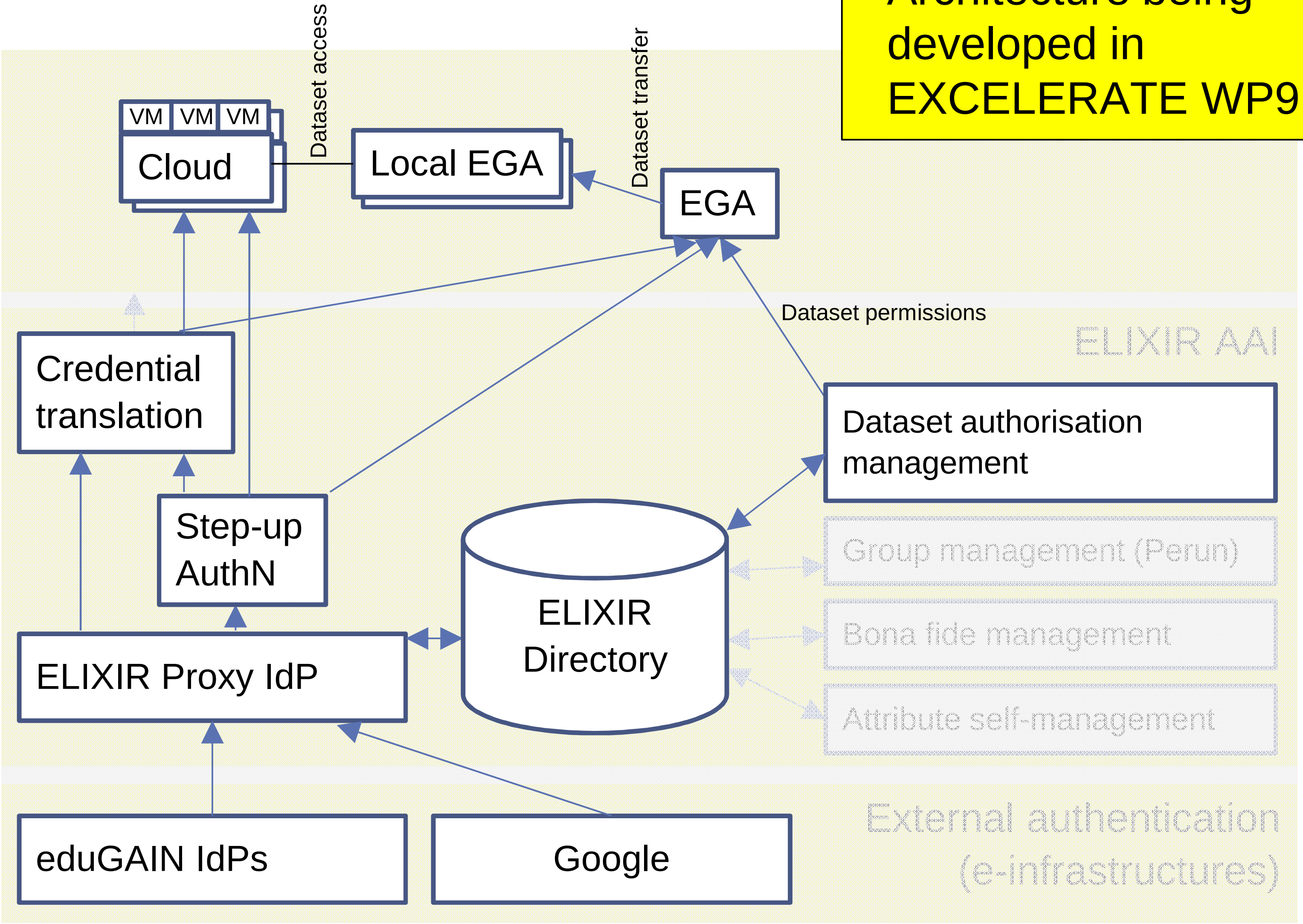
Bringing users to data

Secure Compute Clouds

EMBL-EBI
european genome-phenome archive
ENA
European Nucleotide Archive
Centre de Regulació Genòmica
ESGI
BBMRI
Biobanking and Biomolecular Resources Research Infrastructure
EMBL-EBI
BioSamples
elixir
BSC
CSC
EMBL-EBI

Sensitive human data

Architecture being developed in EXCELERATE WP9.



VM
VM
VM
Cloud
Dataset access
Local EGA
Dataset transfer
EGA
Dataset permissions
ELIXIR AAI
Credential translation
Dataset authorisation management
Step-up AuthN
ELIXIR Directory
Group management (Perun)
Bona fide management
Attribute self-management
ELIXIR Proxy IdP
External authentication (e-infrastructures)
eduGAIN IdPs
Google