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

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/12fBLJ8WenlxABQDzYEAJ6NtOhIXnZBPiGQLuhFUCyq/edit

- Requirements and design – Spring 2015
  - https://docs.google.com/document/d/1CMY1np3GyvPD8LcKvXljXcRO04V2zu3n_Jcq19jqNOw/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/1cJ3mR8lqfZKRmSfAIoSpbqdp1OPI-U6YcUFIRn17rhQ/edit
The stack

- Human data community
- Rare disease community
- Marine community
- Plant community
- etc

ELIXIR (ELIXIR AAI)

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

Federations (eduGAIN)  
Google, ORCID…

Network (GEANT)
Design of ELIXIR AAI
ELIXIR AAI design

Relying services
- EGA
- eLearning
- wiki
- Cloud
- Intranet
- Data archive

ELIXIR AAI
- Credential translation
- Step-up AuthN
- ELIXIR Proxy IdP
- ELIXIR Directory
- eduGAIN IdPs
- Common IdPs

External authentication (e-infrastructures)
- Dataset authorisation management (REMS)
- Group/role mgmt (PERUN)
- Bona fide management
- Attribute self-management

Bona fide management
Dataset authorisation management (REMS)
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

EGA  wiki  ...  ...  ...
Cloud  Intranet  Data archive

Relying services

ELIXIR AAI

ELIXIR identity

tommi@elixir-europe.org (ELIXIR ID)

tommi.offinland@google (Google ID)

nyronen@csc.fi (eduGAIN)

0000-0002-3634-3756 (ORCID)

External authentication (e-infrastructures)
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

EGA | eLearning | wiki | ... | ... | Cloud | Intranet | Data archive | Relying services
ELIXIR Proxy IdP
eduGAIN IdPs | Common IdPs

Credential translation
Step-up AuthN
ELIXIR AAI design

Relying services

EGA

eLearning

wiki

Cloud

Intranet

ELIXIR Proxy IdP

eduGAIN IdPs

Common IdPs

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

(e-infrastructures)
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

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)
**ELIXIR AAI design**

**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**
- Dataset authorisation management
- Group/role management
- Bona fide management
- Attribute self-management

**External authentication (e-infrastructures)**
- eduGAIN IdPs
- Common IdPs
- Relying services
  - EGA
  - eLearning
  - wiki
  - Cloud
  - Intranet
  - Data archive
ELIXIR AAI design

Relying services

EGA  eLearning  wiki  ...  ...
Cloud  Intranet  Data archive

Credential translation

Dataset authorisation management (REMS)
- Sensitive human data
- Data access application needed

eduGAIN IdPs  Common IdPs

ELIXIR AAI

Dataset authorisation management
Group/role management
Bona fide management
Attribute self-management

External authentication (e-infrastructures)
The REMS concept

1. Apply for access
2. Commit to licence terms
3. Circulate to approver
4. Approve
5. Access

Dataset 1
Dataset 2
Metadata on dataset 1&2

Sp

IdP

Principal investigator
Applicant

Research group
Members of the application

DAC 1 Approver
DAC 2 Approver
ELIXIR AAI Dimensions of authentication and authorisation
The planned assurance levels for authentication

- **Basic (in place now)**
  - Self-registrated accounts, password authN
    - Google, LinkedIn, ORCID authentication

- **Raised**
  - Organisation-registed accounts, password authN
    - Log in with Home Organisation IdPs (eduGAIN)
    - Requires Home Organisation complying to a minimal assurance level

- **Strong**
  - Face-to-face proof of identity, two factor authN
    - Step-up authentication
    - Possibly rely on external sources (e.g. eID)
The three authorisation layers

<table>
<thead>
<tr>
<th>User qualification</th>
<th>Service type</th>
<th>Example service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endorsed ELIXIR user</td>
<td>Controlled</td>
<td>Sensitive human data</td>
</tr>
<tr>
<td>Bona fide ELIXIR user</td>
<td>Restricted</td>
<td>Beacon</td>
</tr>
<tr>
<td>Any ELIXIR user</td>
<td>Public</td>
<td>e.g. training, ...</td>
</tr>
</tbody>
</table>
The three authorisation layers

User qualification

<table>
<thead>
<tr>
<th>Endorsed ELIXIR user</th>
<th>Controlled</th>
<th>Sensitive human data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bona fide ELIXIR user</td>
<td>Restricted</td>
<td>Beacon</td>
</tr>
<tr>
<td>Any ELIXIR user</td>
<td>Public</td>
<td>e.g. training, …</td>
</tr>
</tbody>
</table>

Service type

Example service

Any user:
- Who has committed to ELIXIR Acceptable U
The three authorisation layers

<table>
<thead>
<tr>
<th>User qualification</th>
<th>Service type</th>
<th>Example service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endorsed ELIXIR user</td>
<td>Controlled</td>
<td>Sensitive human data</td>
</tr>
<tr>
<td>Bona fide ELIXIR user</td>
<td>Restricted</td>
<td>Beacon</td>
</tr>
<tr>
<td>Any ELIXIR user</td>
<td>Public</td>
<td>e.g. training, ...</td>
</tr>
</tbody>
</table>

"Bona Fide" researcher
- Must commit to a Code of Conduct
- Must be a researcher
  - have publications
  - be vouched for by a person who has publications
  - home organisation confirms
The three authorisation layers

<table>
<thead>
<tr>
<th>User qualification</th>
<th>Service type</th>
<th>Example service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endorsed ELIXIR user</td>
<td>Controlled</td>
<td>Sensitive human data</td>
</tr>
<tr>
<td>Bona fide ELIXIR user</td>
<td>Restricted</td>
<td>Beacon</td>
</tr>
<tr>
<td>Any ELIXIR user</td>
<td>Public</td>
<td>e.g. training, …</td>
</tr>
</tbody>
</table>

**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.**

- Intranet (Drupal)
- Mailing lists (Mailman)
- ELIXIR Proxy IdP
- ELIXIR Directory
- Google
- EduGAIN IdPs
- Group management (Perun)
- Group memberships (batch job)
- Credential translation

**Authentication**

**ELIXIR AAI**

- Dataset authorisation management
- Group management (Perun)
- Home site management
- Attribute self-management
- External authentication (e-infrastructures)
Use case:
ELIXIR Beacon
Beacon idea – public access

Do you have samples with A in position 1234567 in chromosome 2?

• 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 Archiving

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

Services and Coordination

Managing Access
- Data Owner
- Data Access Agreement
- Data Access Committee
- Data Request
- Authorization Management Tools (EGA and CSC REMS)

Supporting sample logistics

Data Generation
- ESGI
- Sequencing centers

Secure Compute Clouds

EGA at CRG

High speed encrypted data transfer
- GridFTP/Globus/A spera

Bringing users to data

Data Users

Data Archiving

EGA at CRG

Secure data access remote API (GA4GH)
On-demand secure replication of:
- Datasets
- Dataset metadata to the EGA mirror sites ("Local EGA")

Managing Access:
- Data Owner
- Data Access Agreement
- Data Access Committee
- Data Request
- Authorization Management Tools (EGA and CSC REMS)

Supporting sample logistics:
- Federated Authentication
- Authorization
- Dataset registry
- Data transfer hub
- Policy and Legal Framework

High speed encrypted data transfer:
- GridFTP/Globus/A spera

Bringing users to data

Data Archiving

Secure Compute Clouds

Data Users
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

Managing Access
Data Owner
Data Access Agreement
Data Access Committee
Data Request
Authorization Management Tools (EGA and CSC REMS)
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

Supporting sample logistics

Managing Access
Data Owner
Data Access Agreement
Data Access Committee
Data Request
Authorization Management Tools (EGA and CSC REMS)
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
Sensitive human data

Architecture being developed in EXCELERATE WP9.

ELIXIR AAI

External authentication (e-infrastructures)

Credential translation

Step-up AuthN

ELIXIR Proxy IdP

eduGAIN IdPs

Google

Dataset access

Local EGA

Cloud

EGA

Dataset transfer

ELIXIR Directory

Dataset authorisation management

Group management (Perun)

Bona fide management

Attribute self-management

External authentication

Dataset permissions

Dataset transfer