Access Control to Research Data in the Frame of FAIR Principles and Open Access

Online Workshop The Future of Managing Osteological Data in Biological Anthropology

March 26, 2021

Peter Gietz
DAASI International

Agenda

- Ecosystem research infrastructures
- Ecosystem Open Science and why Open Source matters
- FAIR and Open Access
- Open Access and authentication
- Federated Identity Management / authentication and authorisation infrastructure

DAASI International

Ecosystem Research Infrastructures, VRE and VRI, a European Perspective

late 20th century -> early 21th century

Print -> Digital -> Remote

(Remote = any time, any place, any device)

- This development also holds true for research infrastructures
- In many fields of reseach not only virtual research environments (VRE) were established but also virtual research infrastructures (VRI), that provide general services for such VREs, such as collaboration tools and research data platforms
- Generic VRIs could be used for different fields of study, e.g.
  - **EGI** for advanced computing resources (https://www.egi.eu/)
  - **EUDAT** a collobarative data infrastructure (https://eudat.eu/)
  - **OpenAIRE** a driver of open science (https://www.openaire.eu)
  - These VRIs are now part of **EOSC** (https://eosc-portal.eu/) "EOSC - a tool for enabling Open Science in Europe"

see https://www.egi.eu/wp-content/uploads/2020/09/2020-09-17-EOSC-SRIA-Cluster-and-e-infra-statement-1.pdf

DAASI International

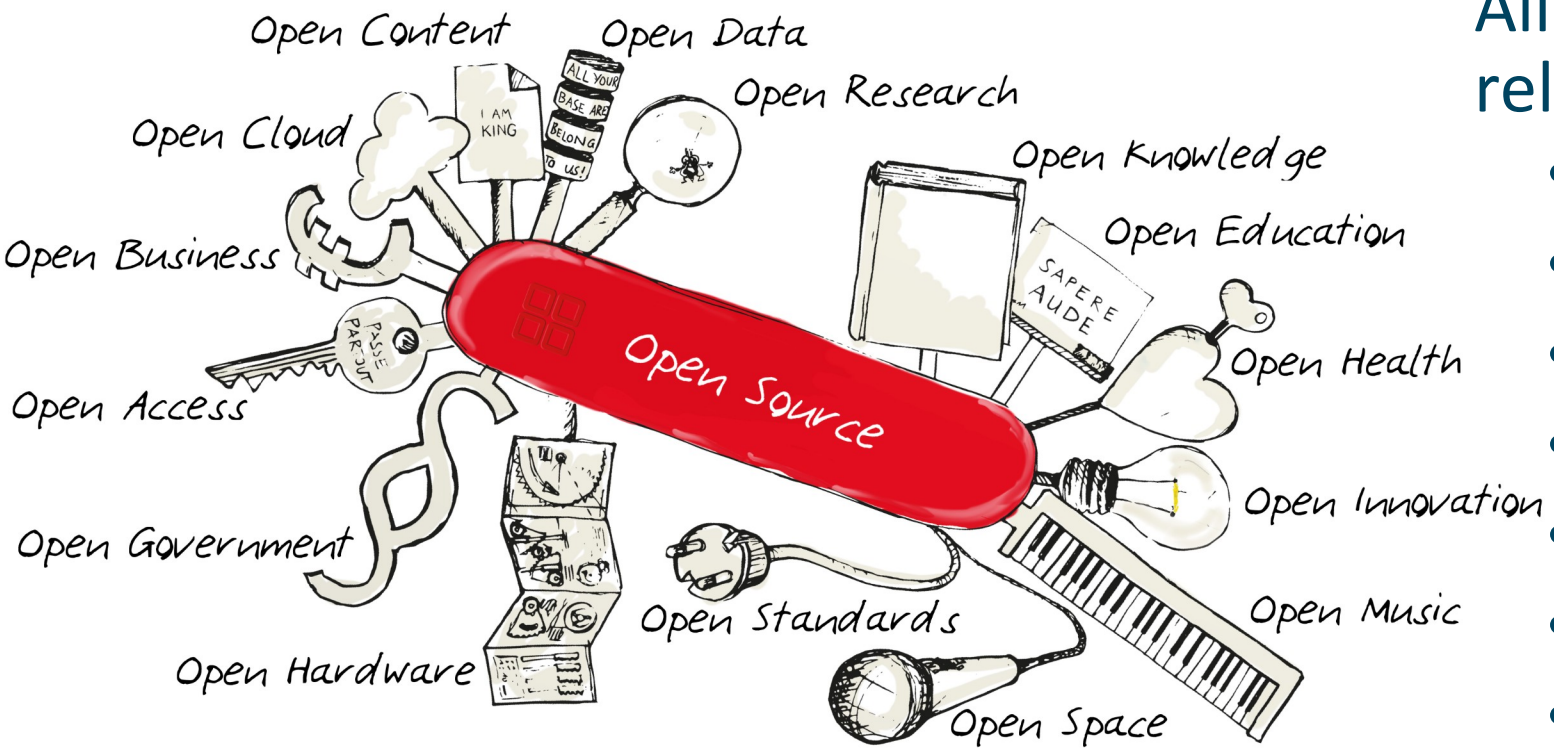
Ecosystem Research Infrastructures: EOSC and Domain Specific VRIs

- **EOSC** wants to "enable a trusted, virtual, federated environment in Europe to store, share and reuse digital outputs from research (including publications, data, metadata and software) across borders and scientific disciplines" (see https://www.eoscsecretariat.eu/sites/default/files/open_consultation_booklet_sria-eosc_20-july-2020.p)
- Besides such general infrastructures different domains have their own specialised VRIs, e.g.:
  - **ELIXIR:** "unites Europe's leading life science organisations in managing and safeguarding the increasing volume of data being generated by publicly funded research" (https://elixir-europe.org/)
  - **DARIAH:** "The Pan-European infrastructure for arts and humanities scholars" (https://www.dariah.eu/)

Ecosystem Research Infrastructures: Services

- **General** services:
  - Computing resources, batch run infrastructure (scheduler), storage resources, research data annotation, data replication, long term preservation, persistent identifiers, cloud storage, data repositories, data management planning, data discovery, identity and authorisation etc.
- **Domain specific** services:
  - Domain specific metadata standards, collaboration platforms, research software, repositories, authorisation, etc.
- Both VRI types also stand for “Democratization of Data” (Franklin and Sachin on Wednesday)

Ecosystem Open Science (= Open Research) and Open Source



Open Source model application domains

All parts of the "Open Movement" rely on Open Source for:

- Transparency
- Flexibility
- Independence
- Intersubjectivity
- Sustainability
- Affordability
- etc.

FAIR Principles and Access

- FAIR
  - **Findable** (persistent ID, rich metadata, indices)
  - **Accessible** (retrieval interface, standardised communication protocol, authentication and authorisation, long term preservation of metadata)
  - **Interoperable** (standardised language for knowledge representation, controlled vocabularies, relations between (meta)data
  - **Reusable** (accurate and relevant attributes, clear data usage license, detailed provenance, domain-relevant metadata standards)

Open Access and Authentication and Authorisation Infrastructure (AAI)

- Open Access is defined “as a comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community” (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities)
  - Basically any information funded by the state and therefore by means of tax payer money should be openly accessible for the tax payer (in a global sense). This also leads to conflicts of interest of libraries with scholarly publishers and to new business models for the latter
- If it should be open for all, why would you need AAI then?

Open Access and Authentication and Authorisation (A&A)

- Even with open access platforms, access control makes sense:
  - Proof of authorship: if anybody could anonymously upload content to an Open Access platform, the authorship cannot be proven. Thus, at least the content upload needs to be secured with (A&A)
  - De-anonymise scholarly discussions: as soon as an Open Access platform allows for content to be commented on, again, it is important for everyone to know who takes part in the discussion, just as it was in the times when reviews were made in printed journals.
  - Secure pre-publication content: many OpenAccess platforms allow for restricting the access of yet to be published content, to allow to share the content with some colleagues ahead of the official publication.
  - Part of the business model of commercial Open Access platforms is that the author pays, which again needs A&A
- If we need it, let‘s do it right:
  - Use standards
  - Be interoperable with other research groups
  - Reuse existing infrastructures

Authentication and Authorisation Infrastructure (AAI)

- **AAI** is mostly used in the sense of an infrastructure that allows for Federated Identity Management (FIdM)
- **IdM**:
  - Systems for the management of computer records that map identities of persons. Such records contain unique names (e.g. login name), credentials (e.g. password) and any other necessary information about the person, such as name, email, group memberships, etc.
  - IdM takes place within an organisation
- **FidM:**
  - Systems that allow for using identities beyond cross organisational borders
  - Such infrastructures allow the user to authenticate (prove their identity to a technical system) and services to decide about access based on authorisation information

AAI Federation

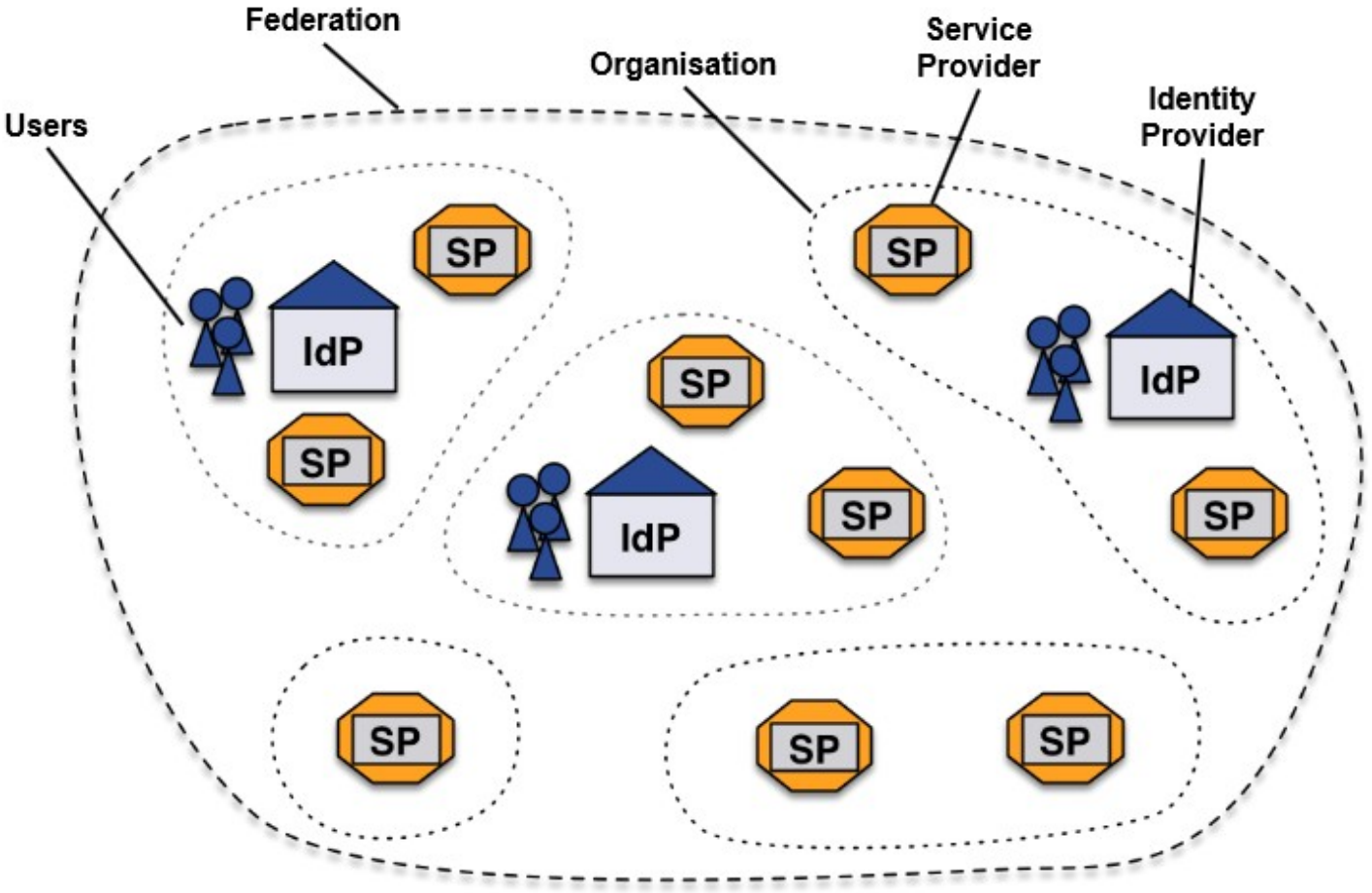
- Two or more organisations can form a federation to share resources
  - Trust within the federation is established by contracts
- The organisation can have one or two roles:
  - **Identity Provider** (**IdP**)
    - having an identity management, so that their users can authenticate
    - state in contract that the information about the identities is correct
  - **Service Provider** (**SP**)
    - providing any kind of resource to authorised users within the federation
    - state in contract that they will use the data provided by the IdP only for the agreed upon use case

AAI Federation

- A federation also needs **central management** to
  - Manage **contracts** (1 to n instead of n to n)
  - Maintain the **membership list**
  - Manage technical data about the computer systems involved (URLs, **server certificates**)
  - Manage a central **Discovery Service** (DS)

- Summary: *A federation is a group of organisations running IdPs and SPs that agree on a common set of rules and standards*
  - It's a label - to talk about such a collection of organisations
  - An organisation may belong to more than one federation at a time

AAI Federation



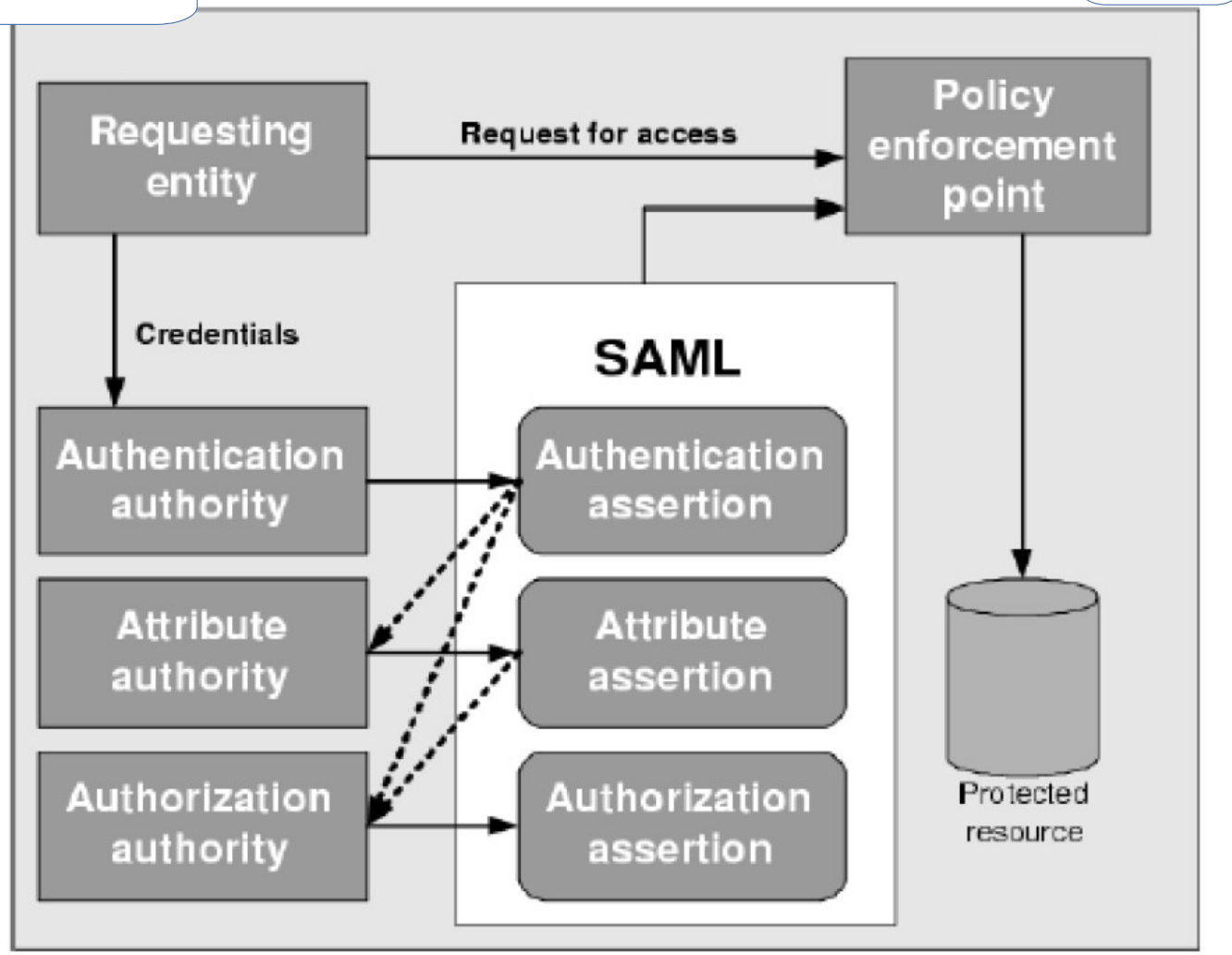
DAASI International

SAML: an XML Based Standard Allowing for Federations and SSO

User

SP

IdP



(c) RUBENKING, NEIL J.: Securing web services

SAML: Security Assertion Markup Language

SP: Service Provider

IdP: Identity Provider

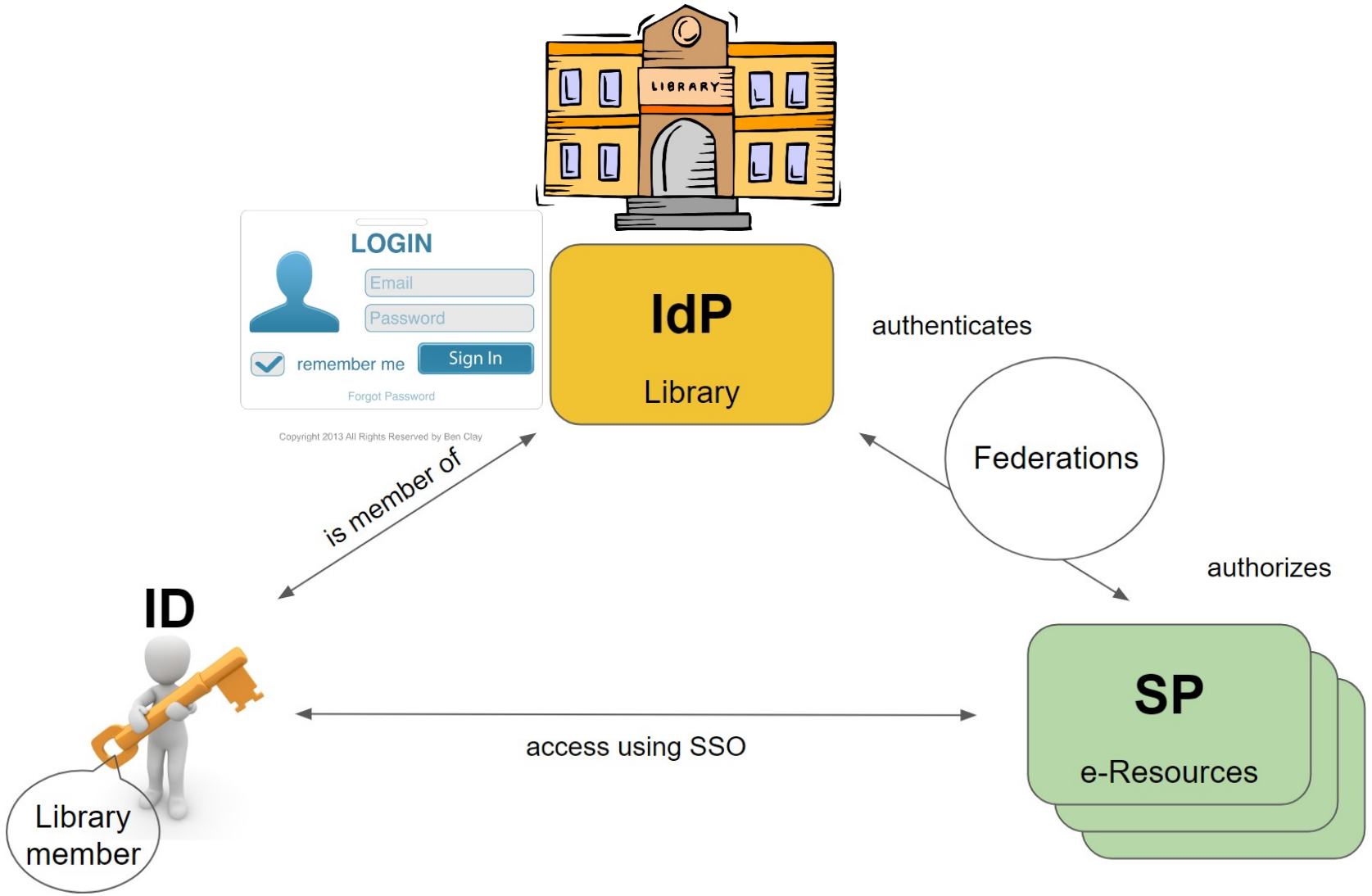
Attribute Assertions potentially contain sensitive personal information (name, email, etc.)

DAASI International

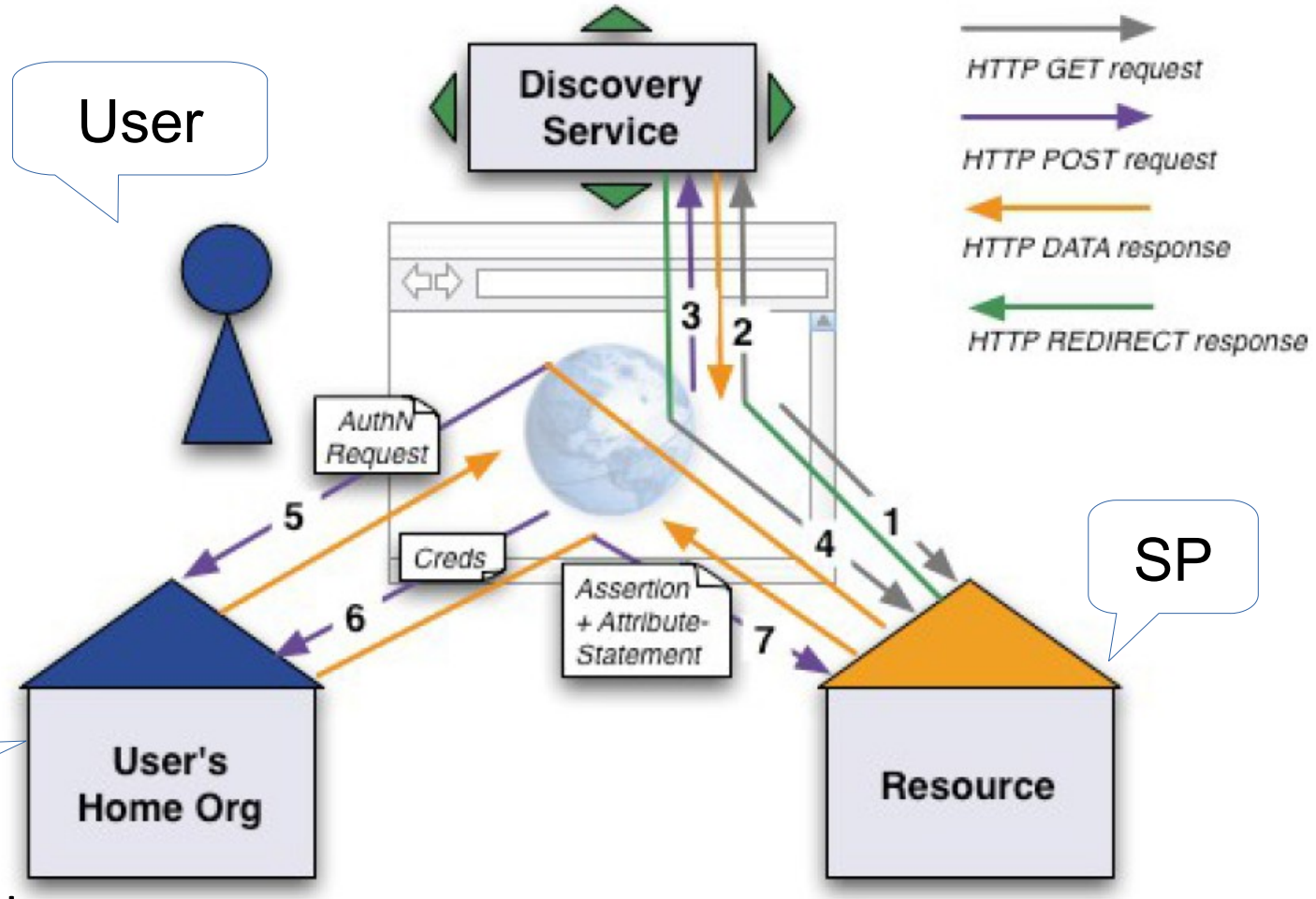
SAML Statements, Examples and Personal Identifiable Information (PII)

- Authentication assertion:
  - e.g.:
    - The user you just sent to me could prove their identity (PII is only the session ID)
  - The method of proof was a correct password in combination with a hardware token (no PII)
- Attribute assertion:
  - e.g.: The user you just sent to me
    - belongs to the staff of our institution (no PII)
    - Is student at our institution (no PII)
    - Is just a user of our library (no PII)
    - Is member of the research group XXX (no PII)
    - Has the name Joe Doe (**PII**)
    - Has the email address john.doe@university-x.edu **(PII**)
  - ***Caution:*** *as soon as you sent one piece of PII, the non PII also get PII*
- Authorisation assertion:
  - e.g.:
    - The user you just sent to me is entitled to use your service (no PII)
  - Such assertions are very seldomly used or even implemented in software

Federated Access

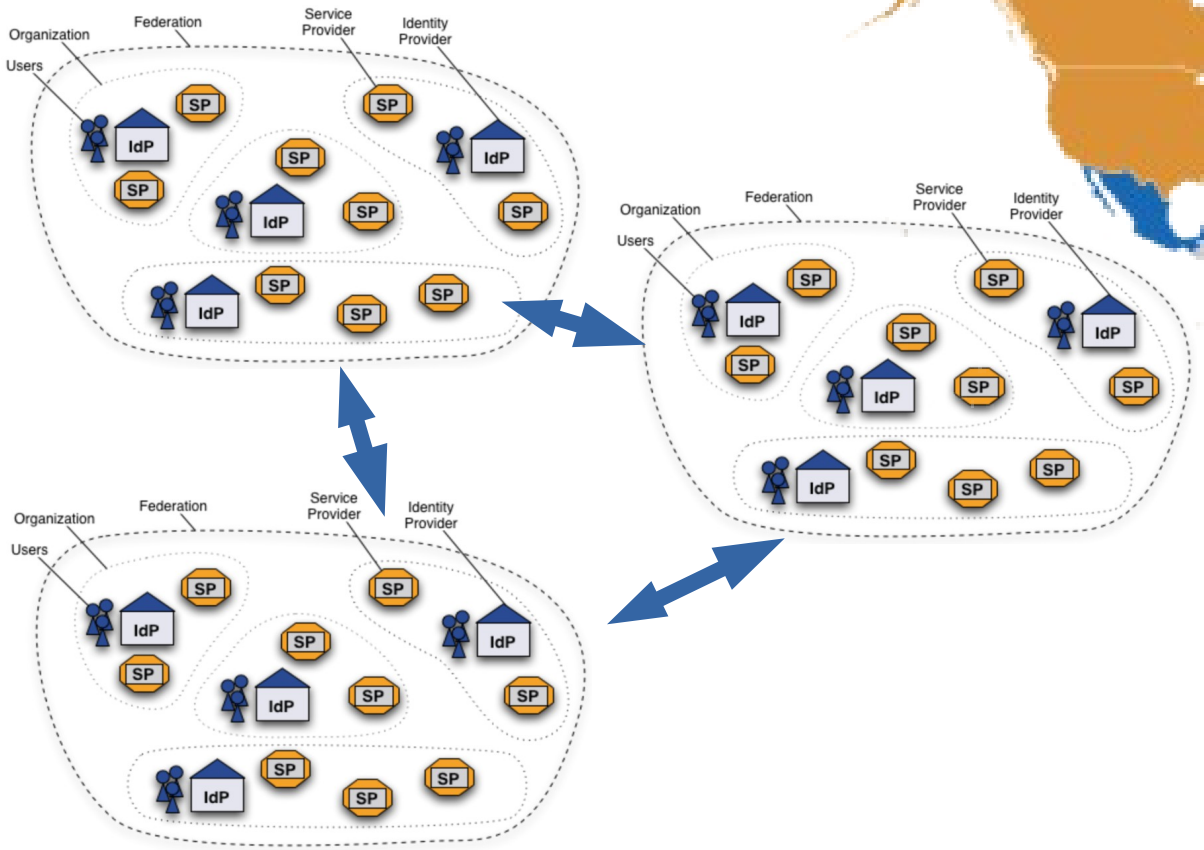


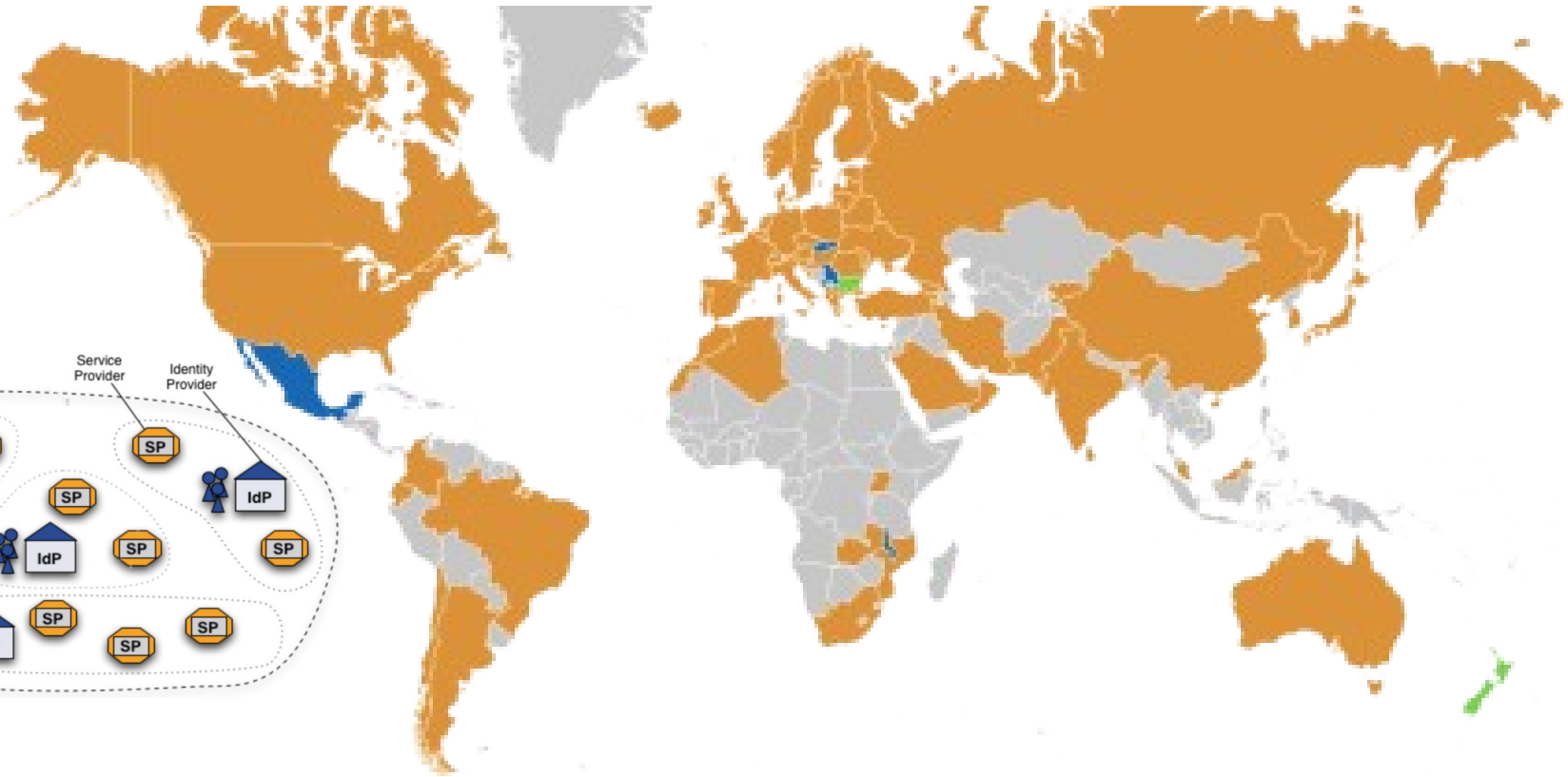
Shibboleth: an Open Source Implementation of SAML



SSO Enabler

SAML based Inter-Federation eduGAIN





Advantages of FIdM

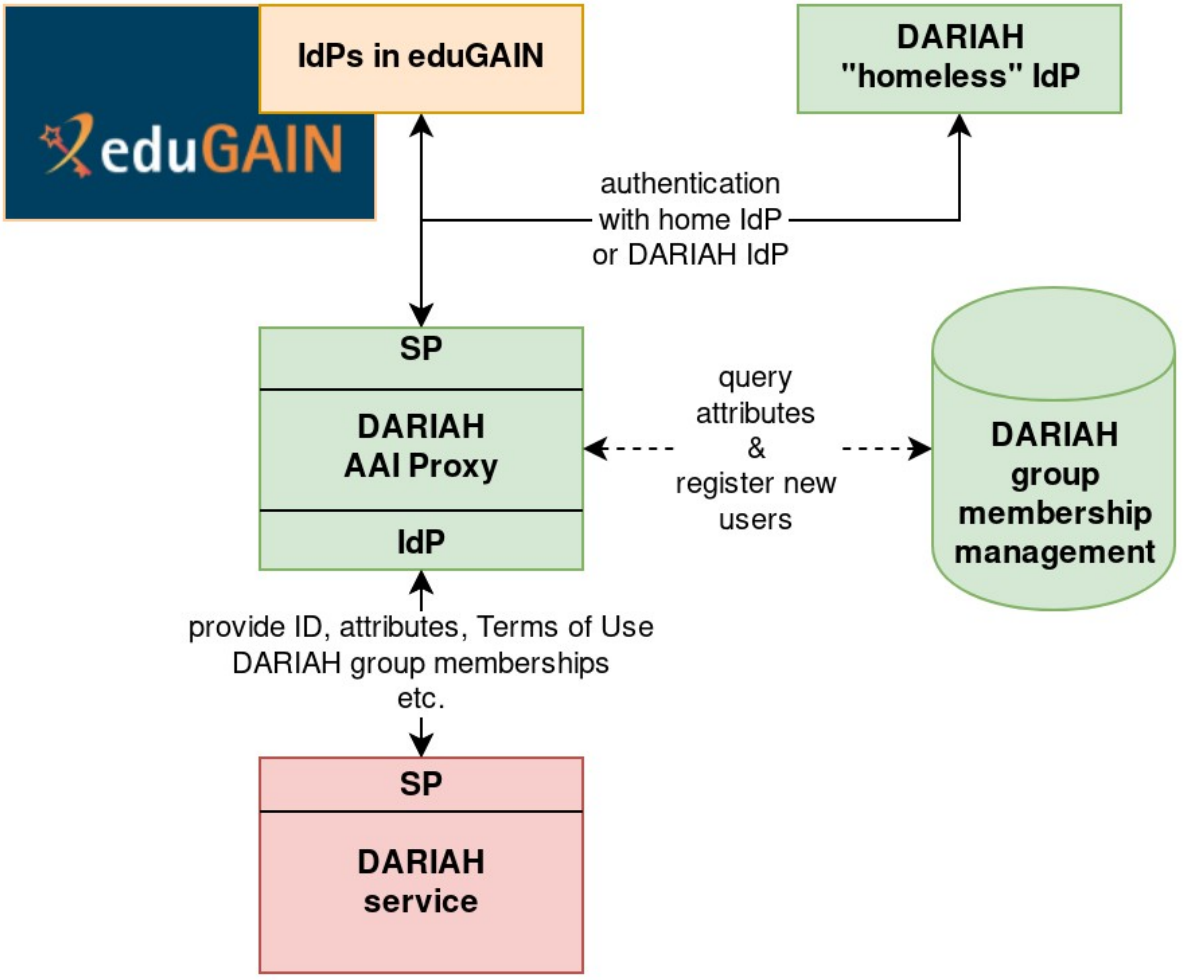
- Users
  - Can use their own campus login credentials
  - Generally find the resulting single sign-on experience to be easier than logging in numerous times
  - Like that the authentication process is consistent regardless of the service accessed
- Identity providers
  - Keep control of personal data and can decide what data to send to which service
  - Experience a simplified process of integrating new services
- Service Providers
  - Do not need to operate their own user management
  - Still in control of who is permitted to use their service
- Existing reusable infrastructure:
  - There are a lot of higher-ed national federations operated by the NRNs
  - There is a world-wide interfederation called eduGAIN

Thanks!

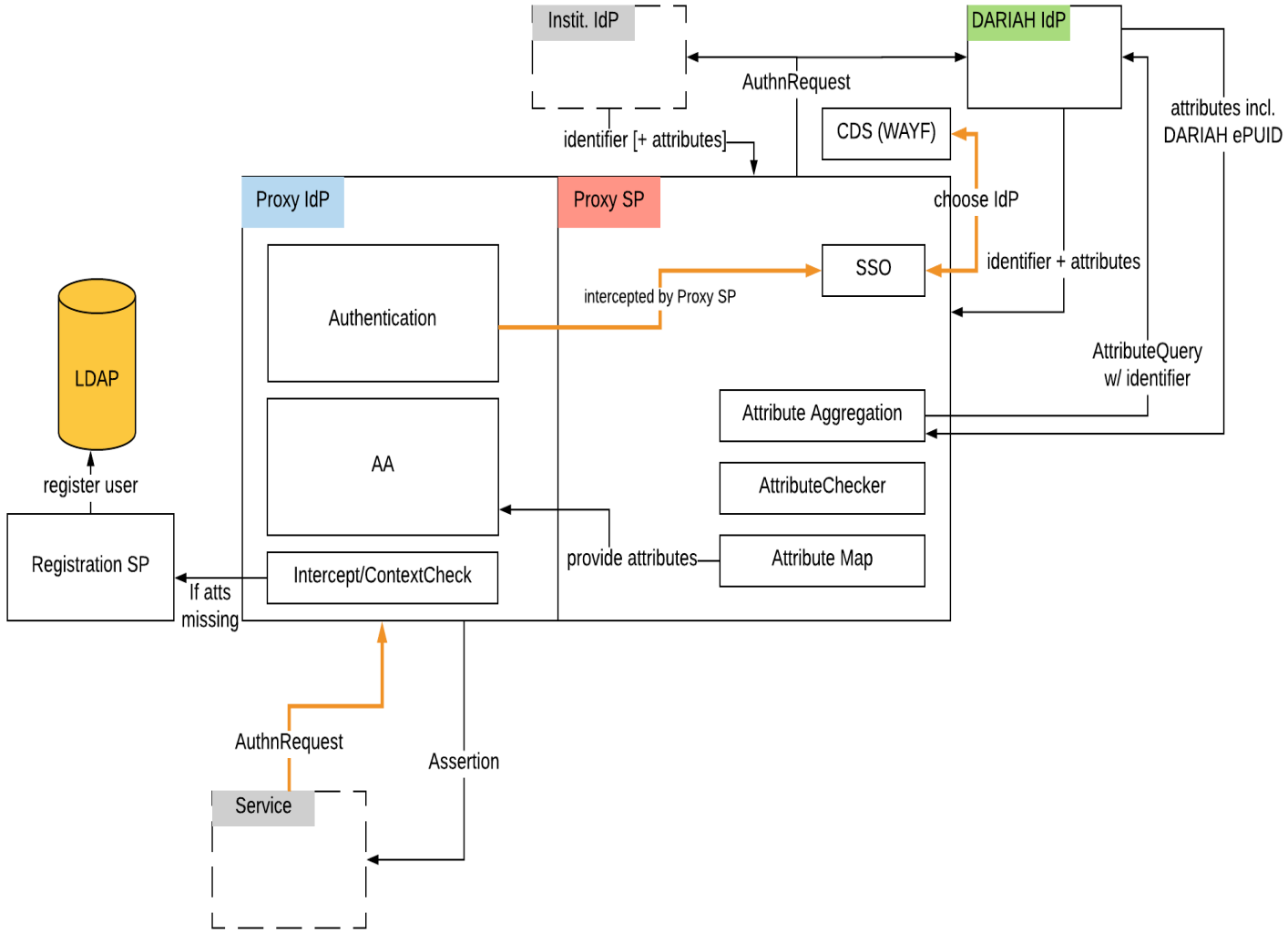
- Questions?
- Comments?
- Additions?

Peter Gietz
Peter.gietz@daasi.de

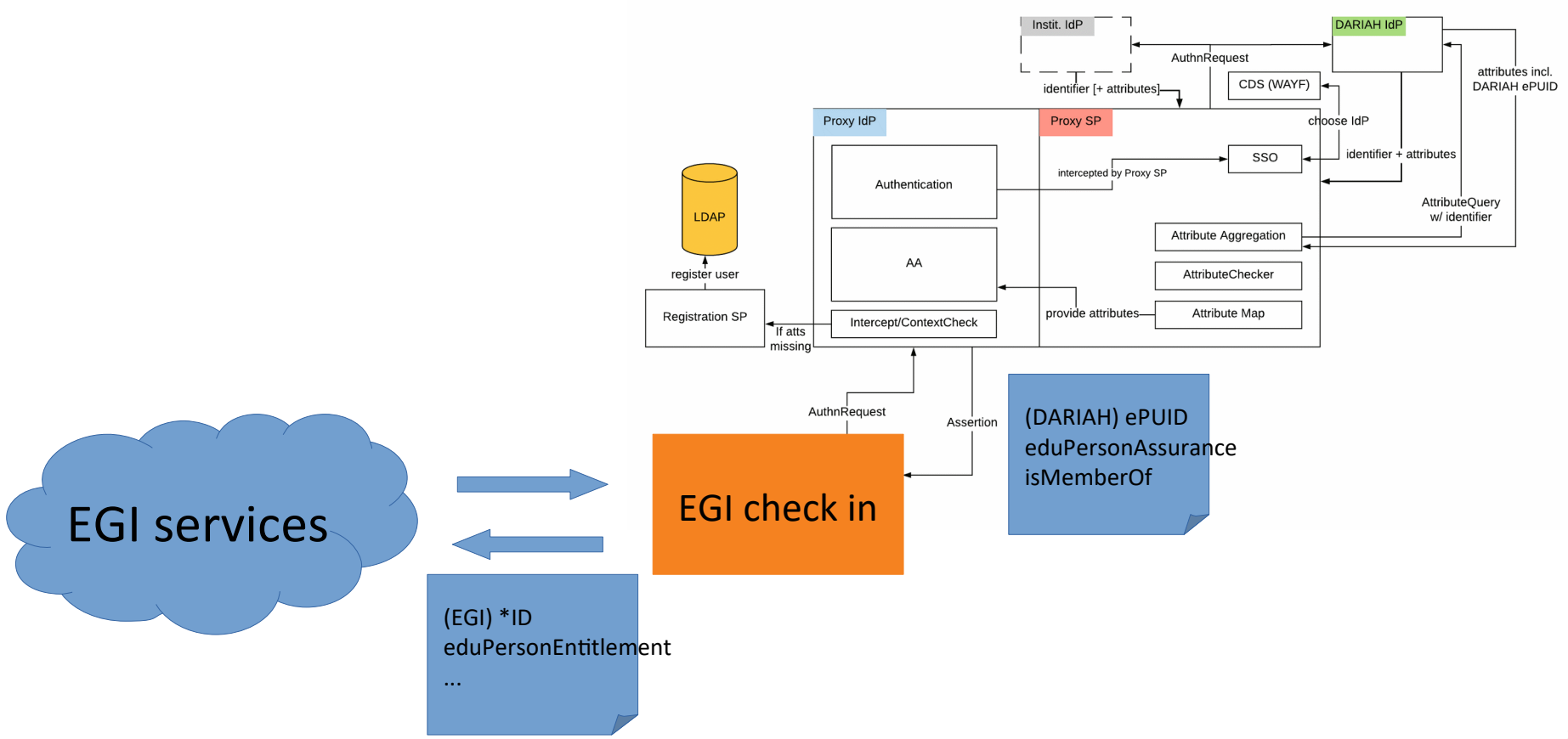
DARIAH IAM with Proxy Architecture according to AARC Blueprint Architecture



DARIAH IAM with Proxy Architecture according to AARC Blueprint Architecture more technical



Integration der DARIAH IAM mit EGI-Platform



DARIAH IAM with Proxy Architecture according to AARC Blueprint Architecture