Karlsruhe Institute of Technology

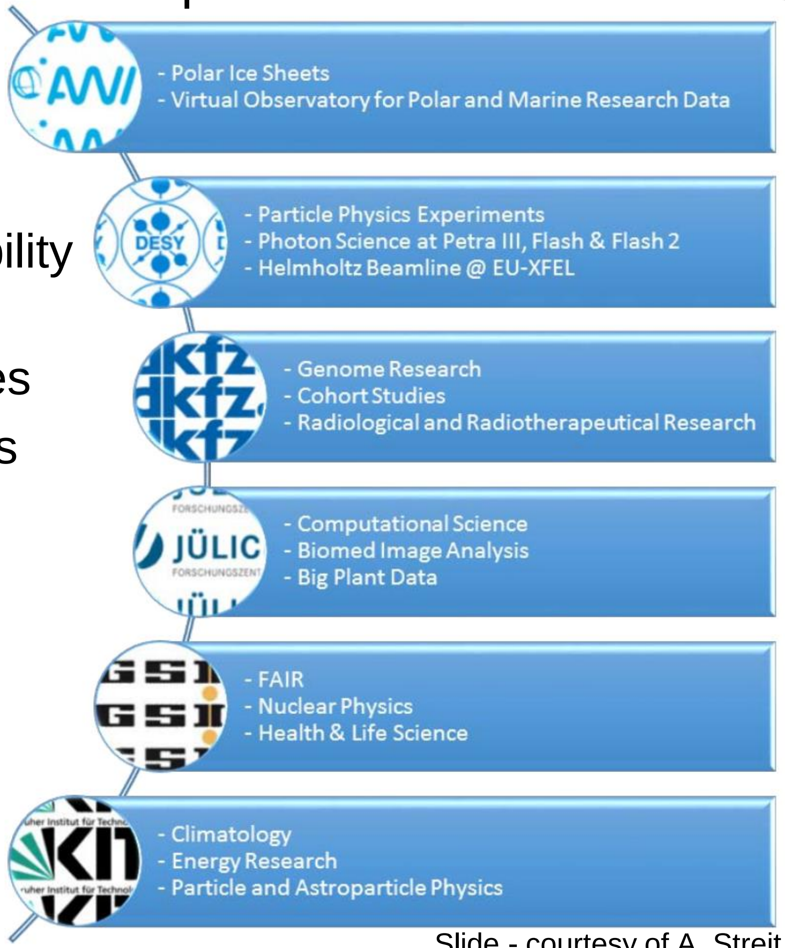
Policies in the European context
Overview

U. Stevanovic, M. Hardt (uros.stevanovic@kit.edu)

Steinbuch Centre for Computing - SCC

Helmholtz Data Federation (HDF) coordinated by KIT

- Helmholtz Association is developing a federated research data infrastructure in Germany for the benefit of and open to the full German science system
  - Finding, accessing, linking, analyzing scientific multi-disciplinary data
  - Long-term preservation, curation, availability and usability of research data
  - Data ownership remains with communities
  - Secure federation of existing data centers with uplinks to EU/Intl.
- National building block for the EOSC
- Comprising three elements
  1) Innovative software technologies
  2) Excellent user support and joint R&D
  3) Modern storage and analysis hardware



- Polar Ice Sheets
- Virtual Observatory for Polar and Marine Research Data

- Particle Physics Experiments
- Photon Science at Petra III, Flash & Flash 2
- Helmholtz Beamline @ EU-XFEL

- Genome Research
- Cohort Studies
- Radiological and Radiotherapeutical Research

- Computational Science
- Biomed Image Analysis
- Big Plant Data

- FAIR
- Nuclear Physics
- Health & Life Science

- Climatology
- Energy Research
- Particle and Astroparticle Physics

Slide - courtesy of A. Streit

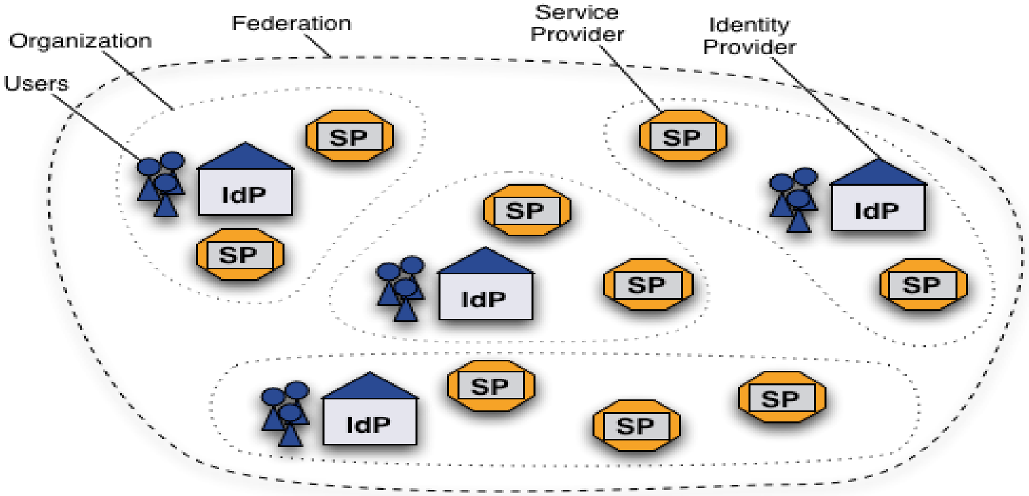
Federated Identity

- “Federated identity is an arrangement that can be made among multiple organisations that lets subscribers use the same identification data to obtain access to the secured resources of all organisations in the group. “ (FIM4R)
- Why federated identity (FI)?
  - Collaboration
  - Multiple benefits:
    - Convenience for the users
    - Cost effective
    - Security, privacy
    - Simplified user management
    - Minimisation of personally identifiable information (PII)
- Requirements for FI
  - Organizational
  - Technical
  - Trust! → Policies

Federated Identity Management Worldwide

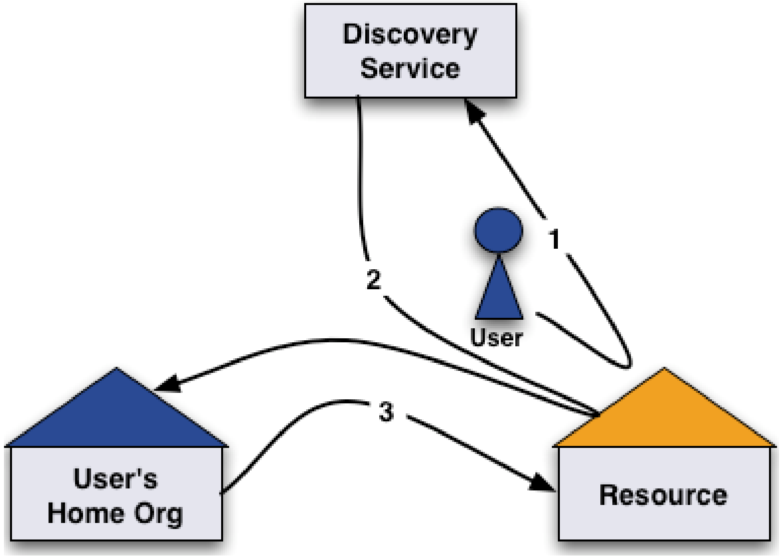
What is a Federation?

- Federated Identity Mangement (FIM) is the concept of groups of Service Providers (SPs) and Identity Providers (IdPs) agreeing to interoperate under a set of policies.
- Federations are typically established nationally and use the SAML2 protocol for information exchange
- Each entity within the federation is described by metadata



Credit to Alessandra Scicchitano – GEANT for this slide

FI usage

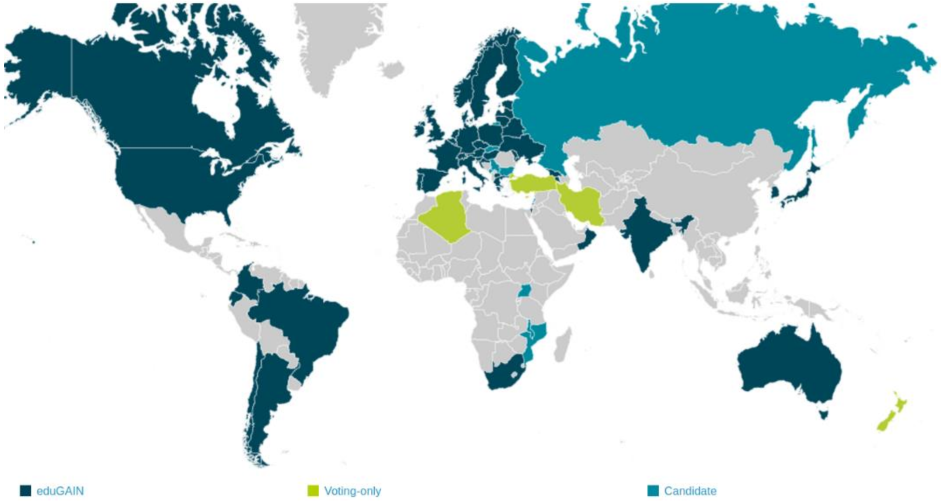


Courtesy of SWITCHaai

- Home organisation manages user credentials, serves as Identity Provider (IdP)
- End user wants to access the service, through Service Provider (SP)
  - User goes to the service, gets redirected to the IdP
  - User authenticates at IdP
  - IdP releases the attributes/credentials to SP
  - SP, upon evaluating the credentials, grants the access to the service (or not)

FI – current status

- Research infrastructures:
  - HBP
  - ELIXIR
  - DARIAH
  - LIGO
  - WLCG
- e-Infrastructures:
  - EGI
  - PRACE
- National Federations:
  - Germany → DFN
- Interfederation
  - eduGAIN
  - 45 Members, 2507 IdPs, 1697 SPs



Helmoltz Data Federation

- Helmholtz Association – 18 members
- HDF (currently) has 6 members:
  - KIT – Karlsruhe
  - FZJ – Jülich
  - DESY - Hamburg
  - AWI - Bremerhaven
  - GSI - Darmstadt
  - DKFZ – Heidelberg
- Practically all HDF members are running IdPs

Policies in FI – current overview

- Security policies
- Operational policies
- Acceptable use policies
- Accounting policies
- Recent policies and efforts, from AARC/Community organisations → harmonizing policies across administrative domains, enable wider adoption of FI
  - SIRTFI - Security Incident Response Trust Framework for Federated Identity
  - SCI – Trust Framework for Security Collaboration among Infrastructures
  - Research and Scholarship (R&S) Entity Category
  - SNCTFI – Scalable Negotiator for a Community Trust Framework in Federated Infrastructures
  - Accounting use policies + GDPR effects

SIRTFI

Operational Security

- Require that a security incident response capability exists with sufficient authority to mitigate, contain the spread of, and remediate the effects of an incident.

Incident Response

- Assure confidentiality of information exchanged
- Identify trusted contacts
- Guarantee a response during collaboration

Traceability

- Improve the usefulness of logs
- Ensure logs are kept in accordance with policy

Participant Responsibilities

- Confirm that end users are aware of an appropriate AUP

SCI – main points

- Operational Security
- Incident Response
- Traceability
- Participant Responsibilities
  - Individual users
  - Collections of users
  - Resource providers, service operators
- Legal issues and Management procedures
- Protection and processing of Personal Data/Personally Identifiable Information

Research and scholarship Entity Category

- Support the release of attributes to SPs
- Attribute bundle
  - shared user identifier
  - person name
  - email address
  - affiliation – optional
- Shared user identifier (persistent, non-reassigned, non-targeted identifier):
  - eduPersonPrincipalName (if non-reassigned)
  - eduPersonPrincipalName + eduPersonTargetedID
- Person name (either/both):
  - displayName
  - givenName + sn
- Affiliation:
  - eduPersonScopedAffiliation

GDPR – overview and effects

- General Data Protection Regulation (GDPR)
  - Adopted 14th April 2016, goes into force 25th May 2018
  - Legally binding for all Member States, without the need for parliament ratification
  - Data Protection Directive 94/46/EC – still valid

GDPR

- Apply to the processing of personal data by controllers and processors in the EU, regardless where it takes place
- Penalties – up to 4% of annual global turnover or 20M€ (whichever is greater)
- Consent – conditions are strengthened (clear and plain language, explicitly related to the processing, easy to withdraw)
- Breach notification
- Privacy by design
- Right to be forgotten
- Data Protection Officers
- Right to access

Grounds for processing

- Six distinct grounds for processing (Article 6), with two most relevant for research use cases
  - Article 6.1(a) - User consent: “the data subject has given consent to the processing of his or her personal data for on or more specific purposes”
  - Article 6.1(f) – Legitimate interest: “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.”
- GDPR has strict requirements for user consent
  - Freely and clearly given, revocation at any time, given for a limited purpose
  - However, if a researcher needs access to services to perform their job, is the consent really “freely given, specific, informed and unambiguous indication of the data subject’s wishes” (from Article 4(11))
  - Data protection by design and by default (Article 25), data minimisation
- Relying only on the user consent is not good enough

Release of personal data to third parties

- Article 6.1(f) “in effect requires a balancing of the legitimate interests of the controller, or any third parties to whom the data are disclosed, against the interests or fundamental rights of the data subject” (Opinion 06/2014, Article 29 Data Protection Working Party)
- Any sharing of personal data within the infrastructure must be considered “release” of personal data to a third party (this makes almost all entities data controllers)
  - Log files, accounting records, community membership information
- Release of personal data is permitted, under certain safeguards:
  - Informing the user
  - Performing a balancing test, i.e. the stronger the legitimate interest and the less harm the processing does to the interest of the data subject, the greater the likelihood the activity will be lawful
  - Examples: attribute release has a positive impact on a user (accessing the service); security incidents are legitimate interests of a service provider
  - Legitimate interest should NOT be treated as a last resort, nor be automatically applied

Release of personal data outside the EU

- Infrastructure operation almost invariably involves transfer of personal data outside EU
- Recognized countries (Andorra, Isle of Man, Argentina, etc.) – not many, US is not recognized
- Conditions for transfer:
  - User needs to be informed of the safeguards
  - Custom exchange model needs an explicit approval from a data protection authority
  - “Binding Corporate Rules” (BCR)
  - “Standard Data Protection Clauses”
  - “Approved Codes of Conduct”
- Explicit user consent – an option, requirements may not make it suitable
- “Safe Harbor” – invalidated by ECJ in 2015
- “EU-US Privacy Shield” – not applicable to research environment

GDPR

- Exchange of personal data within EU – framework is already present
- Model contracts, BCR, code of conduct – release of data to non EU countries
- No clear winner, probably all should be used with modifications/caveats
- NOT BCR, but BCR-like approach
- One possible solution – Code of Conduct, Article 46.2(e) “an approved code of conduct … with binding and enforceable commitments of the controller … in the third country to apply the appropriate safeguards …” without requiring any specific authorisation from a supervisory authority
  - Discussion still ongoing
  - REFEDS effort for attribute release outside EU

Levels of assurance with federated identity

- Quality and levels of assurance with FI
- MNA3.1 – “Recommendations on Minimal Assurance Level Relevant for Low-risk Research Use Cases”
  - Non-shareable accounts
  - Persistent user identifiers
  - Documented identity vetting procedure
  - Password authentication
  - Prompt changes in user’s affiliation
  - Self-assessment
- IGTF
- Kantara
- Industry standards
  - ISO
  - NIST

Survey participants:

- CLARIN
- DARIAH
- ELIXIR
- LIGO
- photon/neutron facilities
- WLCG
- EGI
- PRACE

IGTF Levels of assurance

- ASPEN, BIRCH, CEDAR
  - Documented id vetting process
  - Identity vetting and validation should be based on an in-person appearance before a trusted agent of the authority with presentation of a reliable photo-ID and/or valid official documents; or
  - be validated using notary-public attestations and/or official government data sources and supported by remote live video conversation; or be performed according to Kantara LoA 2 or better.
  - Identity vetting can also be based on a current and ongoing relationship with the Applicant
- DOGWOOD
  - Uniqueness requirement

REFEDS Assurance Framework

- Under development
- Assurance profile Cappuccino – low risk research use cases
- Assurance profile Espresso – verified identity + 2FA

Value	Cappuccino	Espresso
$PREFIX$/ID/unique	X	X
$PREFIX$/ID/no-eppn-reassign		
$PREFIX$/ID/no-eppn-reassign		
$PREFIX$/IAP/local-enterprise	X	X
$PREFIX$/IAP/assumed	X	X
$PREFIX$/IAP/verified		X
$PREFIX$/AAP/good-entropy	X	
https://refeds.org/profile/mfa		X
$PREFIX$/ATP/ePA-1m	X	X

HDF – Policy requirements

- AUP
- Security policies:
  - Service operations, etc
- Security Traceability and Logging Policy
- Security Incident Response Policy
- Policy on the Processing of Personal Data
- Policy on Acceptable Authentication Assurance

Possible merging of policies in:

- Community Membership Management Policy
- Community Operations Security Policy

HDF – Policy requirements

- AUP
- Security policies:
  - Service operations, etc
- Security Traceability and Logging Policy
- Security Incident Response Policy
- Policy on the Processing of Personal Data
- Policy on Acceptable Authentication Assurance

Possible merging of policies in:

- Community Membership Management Policy
- Community Operations Security Policy

→ Templates for all these policies exists
→ Most (all) of them are community “vetted”

HDF – Policy requirements

- Data protection – processing of personal data
- LoA
- Security
- Community membership management
- Community operations management

Data protection – processing of personal data

- Reason for processing
- GDPR influence
- Privacy by design
- Documented policies
- DPO
- Release of attributes
  - Minimum attribute release R&S + SIRTFI
- Documented processes
- Types of data
- Accounting, monitoring, logging -

LoA

- Minimal requirements (from survey):
  - Non-shareable accounts
  - Persistent user identifiers
  - Documented identity vetting procedure
  - Password authentication
  - Prompt changes in user’s affiliation
  - Self-assessment
- Do we need more? Need to identify minimum LoA for our use case
- DFN AA- Advanced

Security

- Incident response
- Risk assessment
- Cooperation
- CSIRT teams
- People, teams, etc?

Policies for community membership management and security

- Membership management
  - Individual Users
  - Community Manager and other roles
  - Community
  - Protection and processing of Personal Data
  - Audit and Traceability Requirements
  - Registry and Registration Data
- Security Operations
  - Providing AUP
  - Security contacts
  - Compliance with the resource providers security policies, etc.
  - Security incidents mitigation policies
  - Liabilities, etc.

HDF – Policy requirements

- Data protection
- LoA
- Security
- Community membership management
- Community operations management

HDF – Policy requirements

- Data protection
- LoA
- Security
- Community membership management
- Community operations management

→ Templates for all these policies already exists

Questions?

References

- https://refeds.org/category/research-and-scholarship
- https://wiki.refeds.org/display/ASS/Assurance+Home
- https://wiki.refeds.org/display/GROUPS/SIRTFI
- https://aarc-project.eu/policies/
- https://aarc-project.eu/
- https://gdpr-info.eu/