Notes from the

DARIAH/DASISH Workshop on a Federation for eHumanities and eSocial Science

Cologne 18.10.2013

Note takers: Peter Gietz and Lukas Hämmerle

Organizational matters

Participants

Following Communities were represented in the workshop:

- CENDARI
- CESDA
- CLARIN
- DARIAH
- DASISH
- DFN-AAI
- EHRI
- EUDAT
- GÉANT/EduGAIN
- LSDMA

Acronyms used in the following

- A: Answer
- C: Comment
- Q: Question

Problem space and aims of the workshop (Peter Gietz, DARIAH)

- Common interests on AAI in the communities CESDA, CLARIN, and DARIAH
  - Easily accessible computing and storage facilities
  - Sustainable data repositories with PIDs and long term preservation (from bit stream up to data curation)
  - Common metadata and ontology needs
  - Licence and data privacy issues
  - Service infrastructure needs
  - User that want intuitive and secure access to all this
  - Thus also authentication and authorization
- The communities should combine forces for political impact on
  - scholars in our fields
  - funders
  - user managers at the single campuses

  - NRENs and TERENA
- A lot already has happened:
  - CLARIN/DARIAH Call for action on federated identity
  - DASISH Manifesto: The New Global Data Generation - Simplifying and Guaranteeing Access and Sharing in e-Science Scenarios
  - FIM4Research Workshops
  - Plans to start a follow up activity within RDA
  - TERENAQ VAMP workshops
  - EduGAIN and GÉANT 3 plus activity with a DARIAH pilot
- Major aim of the workshop is to discuss different ways forward to reach a European federation

No discussion took place after this short presentation

Data protection aspects of federated Identity management in the light of proposed EU privacy regulation (Ville Oksanen, CLARIN)

- Current European data privacy legislation was defined way before Facebook and Google search features
- The current initiative for a modernisation of privacy legislation will define a General Data Protection Regulation that will supersede the Data Protection Directive
- There has been yet unseen interest of US government and US companies like Facebook and Google to water down the current proposal
  - in all almost 5.000 amendmend proposals
- Basic principles of the current draft:
  - Data protection by design
  - Data may be provided to 3rd parties for “scientific research”
  - Consent is either subject perceived behavior or result of an active choice
  - There is a right to be forgotten, that also applies to data given to 3rd parties
  - Accountability
  - Notification
  - Portability
- There is also an international scope
  - Safe harbor rules for the US was not really successful

Discussion:
Q: Does this mean that universities can provide data to research infrastructures without consent?
A: Yes

Q: Even for student discounts to companies like Microsoft?
A: Possibly, but it might be a good idea to ask user in advance

Q: What exactly is meant by “scientific research”, only research on the personal data or also research infrastructures that need personal data

A: The term is not yet properly defined

Q: When will the Regulation be finalised?
A: planned to be finalised and adopted in 2014, actually before the next European elections in 7 month

C: There might be an impact on the trade agreements with the US

Q: How clear will the legislation be?
A: It will become a soft law. Thus European research infrastructures are advised to write a document on how to interpret the legislation.
But beware for not respecting the law there will be a fine of 5% of the annual income of an organisation

Q: As to the right to be forgotten an IdP has to remember and know to which SPs which personal data had been sent, and ti implement a mechanism to ask for revoking those data
A.: Yes

C. Such is also included in the Code of Conduct (to anonymise or delete)

Attribute Release - Technical and Legal Issues - Contractual Matters (Wolfgang Pempe, DFN-AAI)

- It is often difficult for University IdPs to add new or correct existing attribute filter rules because
  - IdP admin is not sufficiently familiar with the IdP
  - IdP admin fears changing the (running) system for 10k+ users
  - IdP in some setups has to be restarted for changing the filter rules
- Different legal layers (federal, state, inter-organisational) for data protection -> situation is different for almost each IdP
- Result is a general reluctance to release personal data
- Federation operator can immprove situation by:
  - telling IdP admins to make filter a reloadable resource
  - providing examples for filter policies
  - advertising the use of opaque identifiers (eduPersonUniqueId vs ePPN)
  - support and promote Code of Conduct in federation
- SP operator can improve situation by:
  - Supporting code of conduct
  - agree to a small set of attributes
  - Declare required attributes in metadata
- The more popular and bigger a service becomes the more likely IdP admins are willing to release personal data
- Even though CoC summarizes "only" existing EU law/regulations it still is helpful because it's brief and easy to understand and makes IdPs see that the SP cares about privacy
- DFN-AAI contract framework

    - IdPs are DFN member organisations and thus sign the framing contract for connectivity and as an add-on to that an IdP contract with assurance of identity vetting and process for keeping the data current
    - SPs are often not DFN member and thus sign an independent contract with assurance of data privacy processes

Discussion:

Q: Doesn't the IdP agreement include acceptance of providing attributes as recommended by the DFN-AAI paper?
A: There is no obligation to release the recommended attributes

Q: does the Code of Conduct already show any impact on the willingness of IdPs to release attributes?
A: Since CoC is in place only for a short time and since there has not yet been advertisement of the feature, it is to early to answer this question

Q: CoC distinguishes between what is retrieved, what is stored and what is self provided by the user. How can you put this into the privacy statement of the local SP policy
A: You can describe this in the privacy statement

C: That statement should be as easy understandable and as concise as possible, clarification of legal issues are not important here. There are short guidelines around.

Introduction to EduGain (Brook Schofield, GÉANT)

- eduGain: educational Global Authentication Infrastructure
- second attempt started 2 years ago from scratch
- Current status
    - 19 Federations have joined, 6 are in the process of joining
    - Total of 143 IdPs and 73 SPs
- Current Potential:
    - 43 federations
    - 2390 IdPs, 4654 SPc
- Number of Federations is increasing
    - New federations also include Non-European
- Code of Conduct has been designed for Europe and compatible countries like Canada, Switzerland, New Zealand, Chili, Israel

Discussion:
Q: What is the exact scope of eduGain? What about additional services like identity vetting, Guest IdPs, VO platforms?
A: Current scope is: support by Lukas team, Metadata management, Code of Conduct. eduGain will not provide a guest IdP, but will give recommendations e.g. to DARIAH. Surfnet stepup-IdP is operated for The Netherlands only.

Q: Why are there more Brazillian IdPs in eduGain than European?

A.: In Brazil it is very common to give away personal data. Thus there is not Opt-In-Modell for IdPs, but all IdPs are automatically added to eduGain

Requirements from EHRI(Kepa Rodriguez, EHRI)

- To add and update holocaust archives authentication is needed
- Scholarly researches: 50% established researchers (incl. undergraduate and research students), 30% attached to university, rest from others from museums, free lances, other research centers etc.
- Sept. 2014 fundings to run own infrastructure runs out
- AAI Requirements are: Sustainability, suitability for all users, easiness to join, interoperation with other infrastructures like DARIAH
- Current setup: complex non-automated accreditation process, OpenID based authentication, local group based authorization

Discussion:
Q. Can EHRI-AAI be part of DFN-AAI?
A: may be, not sure

Q: Is the documentation of the accreditation process available?
A: Will be published soon, after it has been finalized

Digital Services Infrastructure for Social Sciences and Humanities (Daan Broeder, DASISH)

- Implementation of common solutions for a cluster of ESFRI infrastructures in the field of "Social Sciences and Humanities
- Consortium: 18 partners from 10 EU countries + Norway
- 5 ESFRI infrastructures:
  - center network oriented: CESSDA, CLARIN, DARIAH
  - single center: ESS, SHARE
- DASISH Budget: 700PMs
- It is a project
  - and thus sustainability is not here
  - but in the Research infrastructures with sustainable ERIC as organisational unit
  - DASIH only investigates, recommends, brokers
- Sustainable survices can be provided:
  - by commercial services (e.g. PID)
  - ESFRI projects
  - generic EU funded infrastructure services like EUDAT
- Need to create common infrastructures and not strengthen community specific ones
- Traditions vary considerably
- WP5 = Data Access, includes AAI
- SSH trust federation would use eduGAIN if it worked:

    - Opt-in policy per IdP is the biggest problem
- SSH FIM Status
    - CESSDA: Existing collaboration with two Identity federations.
    - CLARIN: Multi IDF, multi-SP federation with CLARIN homeless IDP
    - DARIAH: Robust homeless IdP
- DASISH AAI Strategy:
    - Create SSH trust federation for SSO based on SAML2 using eduGAIN/NRENs/etc *if* this is a workable solution
    - Alternative is creating a proper SSH federation
- Whatever model is choosen:
    - Insufficient attribute release problem will remain
    - Home for the homeless still needed (community managed or even commercial solution?)
    - Level of assurance for accessing sensitive data
- CLARIN SPF agreement could be basis for SSH federation
- Careful communication with Home Orgs required which federation they should join: CLARIN SPF/eduGAIN/SSH Federation?
- CLARIN now is now better off joining all federations than eduGAIN because local federations release attributes easier to local SPs than to eduGAIN.

Discussion
C: instead of creating own federation introduction of entity category could achieve almost the same when it comes to attribute release

C: local pressure is the important factor to get attributes released
A: especially when high profile users. They need good information how to approach their local IdP

Q: What is the schedule for deciding the AAI strategy?
A: End of next year. First feasability of extending Clarin SP federation will be checked but eduGain solution will be needed soon

C: eduGain (re)started 2,5 years ago, and now actual marketing is needed. We have a chicken egg problem: SPs look for IdPs and vice versa

C: we should distinguish dedicated vs. general marketing but do both

C: special entity categories for CLARIN?

C: DFN can do entity categories already

DARIAH AAI (Peter Gietz, DARIAH)

- DARIAH operates a central LDAP (used by JIRA and Confluence), Admin portal and self-service password reset
- LDAP is replicated to RZ Garching where LDAP server is used by an IdP
- Shibboleth SPs aggregate attributes from Campus IdP and central LDAP IdP.

- Shibboleth Attribute Checker ensures that all required attributes are available, if not users are sent to central registration SP that asks users to provide missing data
- LDAP Directory contains branch from homeless users, federated users and groups.
- The groups are managed by the Administration portal
- Solution would also work with persistentID only because all persistentIDs can be mapped in DARIAH ldap via a redirection to the central registration page.
- Flat group-based authorization
- 3-level hierarchy (top - country - organisation) to delegate group and homeless management permissions
- Groups are managed not by competences but by country and organisation
- Software is there, what is missing are processes
- Infrastructure is currently only in DFN-AAI, goal is to add services to eduGAIN, either via existing federations or via own federation

Discussion: No notes were taken.

CLARIN AAI (Dieter Van Uytvanck, CLARIN)

- AAI Services:
  - Service Provider Federation
  - CLARIN Identity Provider
  - Easy-to-use discovery service
- SP Federation
  - is only a club represented by one organisational entity to prevent n to n contract signing
  - 11 SPs, connected to 6 IdFS: SURFconext, DFN-AAI, HAKA and KALMAR. -> connection to 200+ IdPs
- CLARIN ERIC becomes legal entity that can take over the contract signing
- Plan is to connect to more national federations
- Joining multiple federations means a lot of paperwork
- exploring eduGAIN
  - Pros:
    - you only need to join your national Identity Federation
    - no need for SPF (less administration)
    - access to non-EU user base
  - Con:
    - is not large enough yet for CLARIN: too few IdPs not yet ready
    - reason is the Opt-In
- Open issues:
  - Missing attributes (no release of attributes by some IdPs),
  - Trust delegation (hopefully SAML/OAuth2 could solve this issue)
- Next steps:
  - Code of Conduct to get more attributes,
  - extend SPF (join more federations),

    - mutual recognition of homeless IdPs

Discussion:
Q: has the process of convincing people to sign become easier?
A: Yes It is better documented now and a requirement of more people

Q: what is the effort to run the SPF?
A: estimate of about 0.5 FTE for communication, legal expertise for contracts, changing agreements

C: Metadata management and ingest is different at each Id federation.

Q: could the current SP federation club be deployed in parallel to all SPs also joining national ID federations?

C: The SP federation is only about trust: level of trust and whom to trust

Options for Joining eduGAIN (Lukas Hämmerle, GÉANT GN3plus)

- WP of GÉANT3plus is SA5 Application Services For global collaboration, with subtask “enabling users” that promotes eduGAIN
    - Help communities integrate their services into eduGAIN
    - three pilot projects with 3 communities: DARIAH, CRISP, Elixir
- DARIAH:
    - likely to operate many services (SPs) in different countries
    - Hand-full of services and Homeless IdP are already SAML-enabled
    - DARIAH LDAP and Admin Portal for permission and authorization management
    - Main questions are:
        - How to best integrate services/SPs into eduGAIN?
        - How to ensure SPs receive required attributes?
- eduGAIN steering group is like a parliament: every federations sends one member
- some federations offer a homeless IdP
- Options to join eduGAIN
    - Option A: Join via existing federation#Let each service join eduGAIN via an existing federation (e.g. the national federation that already exists)
    - Option B: Create your own federation#Organize all services in an own federation and join with the whole federation
    - Option C: Use a hub/proxy#Place all services behind a hub/proxy and add that proxy via an existing federation to eduGAIN
- Recommendations for DARIAH:
    - Choose option A or B
        - It's basically a question of commitment: How much long-term financial and personal resources are available to operate an own federation?
    - Implement the GÉANT Data Protection Code of Conduct
    - Be patient

Discussion

After Lukas presentation a longer general discussion took place which can only be summarized here.

A fourth option was discussed: to not interfederate at all, but to just use one big data base, just as the Umbrella project seems to do. This would make some things easier, but would also have to be managed – and DARIAH distributed management interface could be used here. But from the user perspective it would not be ideal, because of yet another registration process and yet another password to remember. Thus this should only be the last resort.

The majority seemed to agree that Option A and B could be followed in parallel: first connect to the national federations and in the same time work on a social science and humanities federation.

It will be interesting to see how CESSDA will decide.

Also the Feide OPenIdP was discussed, a SAML 2.0 Identity Provider for users that do not have an account in Feide. All you need in order to create a user account is to click the button below, and verify your e-mail address

All agreed that both days of the workshop were very helpful and that there should be follow up activities again with participation of eduGAIN representatives. Such activities could well be organized within the DASISH umbrella.