GEANT Data Protection Code of Conduct (DP CoC)

20 March 2013

FIM for Research Collaboration

Mikael.Linden@csc.fi

The Issue

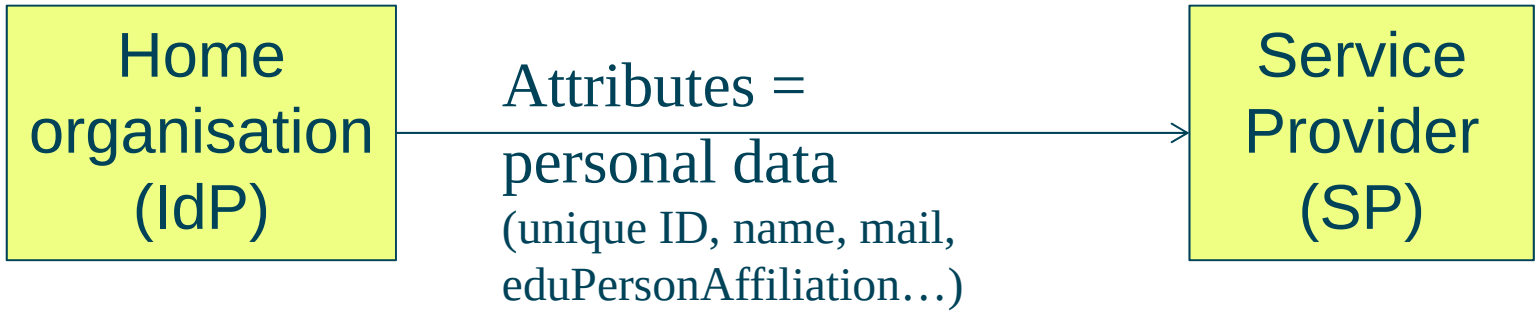
Federated Identity Management for Research Collaborations
Date of this version: 23rd April 2012

"Flexible and scalable IdP attribute release policy. *Different communities and indeed SPs within a community are likely to require a different set of attributes from the IdPs. The IdP policy related to the release of user attributes and the negotiation mechanism needs to be able to provide this flexibility. Bilateral negotiations between all SPs and all IdPs is not a scalable solution."*

"Attributes must be able to cross national borders. *Data protection considerations must allow this to happen."*

"Privacy and data protection to be addressed with community-wide individual identities. *There are many use-cases identified which will require the release of personal data to identify individual users."*

The data protection risk



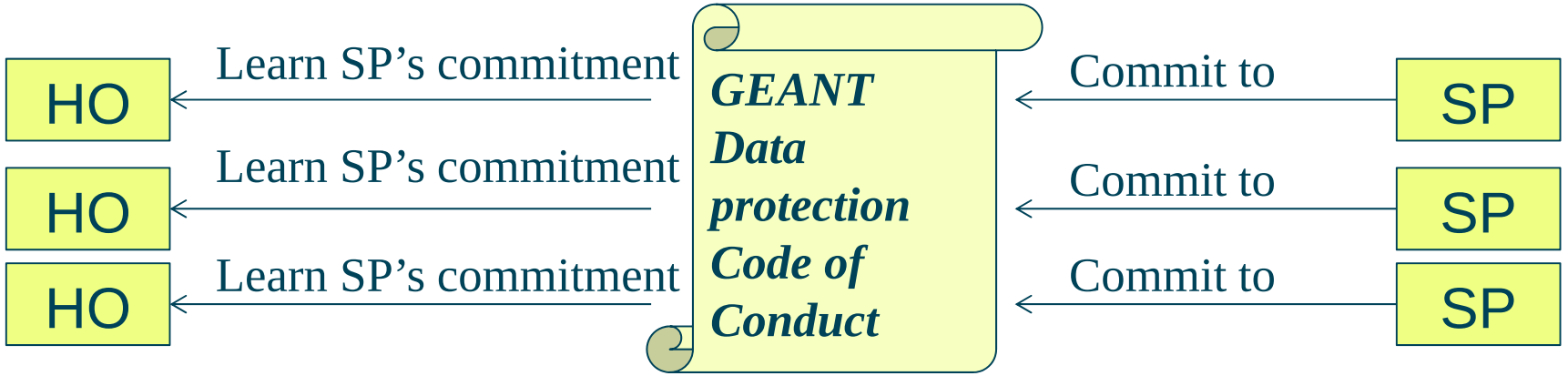
- A Home organisation takes a risk when it releases attributes to an SP
  - Home organisation may become partly liable if the SP is hacked and personal data is spilled to the Internet

  => Home Organisations hesitate to release attributes

Goal of the Data Protection Code of Conduct

- **Ease the release of attributes**
  - Try to reduce Home Organisations' **hesitation to release attributes**
- Try to be sufficiently **compliant** with the EU data protection laws
  - Balance risks with benefits of easy collaboration
- **Scalable**
  - Thousands of HOs and SPs

Data Protection Code of Conduct approach



- Voluntary to SPs (but SPs have an interest to sign to receive attributes)
- Voluntary to Home Orgs to rely on (but may increase Home Org's scientific output and reduce the IdP admin's work)
- It's simple!

Code of Conduct: Service Providers commits to

Directly from the DP law

- Data minimisation
- Data rentention
- Information security
- Informing the end user
- Data release out of EU/EEA

Further specifies the DP law

- Purpose: ”enabling access”
  - Deviating purpose: on user consent only
- Release to 3rd party: on user consent or 3rd party committed to CoC
- Report security breaches

Read the full Code of Conduct text in:
https://refeds.terena.org/index.php/Code_of_Conduct_for_Service_Providers

SAML 2.0 metadata profile supporting the Code of Conduct

Standard SAML 2.0 metadata elements used to convey SP's relevant properties to a Home Organisation:

- Indication of SP's commitment to the CoC (Entity Category element)
- List of attributes the SP needs (md:requestedAttributes)
- Link to the SP's Privacy policy document (mdui:privacyStatementURL)
- SP's name (mdui:displayName)
- SP's description (mdui:Description)

Informing the end user on release of his/her personal data

Service name Weblicht

Description WebLicht is a service for language research. It provides an execution environment for automatic annotation of text corpora.

Privacy policy

Your following information will be released

Unique ID	mlinden@csc.fi
Name	Mikael Linden

OK

The Code of Conduct timeline

- 1st public call for comments 6-8/2012
- 2nd public call for comments 11-12/2012
- Pilot with the CLARIN community started in June 2012

- Production Q2/2013
- Submission to Article 29 Working Party Q2/2013
  - the EU body contributing to the uniform application of the Data protection directive

Code of Conduct pilot with CLARIN

- Identity Federations: DFN-AAI (Germany). Haka (Finland), SWAMID (Sweden)
- Home Organisations: DFN (connected to DFN-AAI), Institut für Deutche Sprache (DFN-AAI), CSC (Haka), Uppsala university (SWAMID)
- Service Providers
  - LAT – Language Archive Tools (CSC, connected to Haka)
  - IDS – CLARIN services (IDS, DFN-AAI)
  - IDS – repository (IDS, DFN-AAI)
  - CLARIN Catalog (MPI for Psycholingusitics, DFN-AAI)
  - MPI second SP (MPI for Psycholingusitics, DFN-AAI)
  - Weblicht – annotation tool (Tübingen university, DFN-AAI)
  - Filesender (CSC, Haka)

Preliminary findings in the pilot

- Pilot Service Providers happy to commit to the Code of Conduct
- Pilot Home Organisations happy to release attributes to committed SPs
- More documentation, templates and training needed
  - How to write a Privacy Policy document
  - What attributes are necessary for a service
- Some sanity checks seem necessary by a third party (federation operator)
  - Privacy policy document and SAML 2.0 metadata are consistent
  - Service name and description understandable and useful for common users
    - *“Lux17 Service Provider”*
    - *“Max Planck Institute for Psycholinguistics second Service Provider”*
- Draft final report: https://refeds.terena.org/index.php/CocPilotReport

Questions?

Backup slides

EU Data protection directive Definitions

- **Personal data**: ” any information relating to an identified or identifiable natural person”
  - Lawyer: assume any attribute (ePTID and even eduPersonAffiliation) counts as personal data
- **Processing of personal data**: ”any operation or set of operations on personal data, such as collection, …, dissemination,… etc”
  - Both IdP and SP processes personal data
- **Data Controller**: organisation which alone or jointly with others determines the purposes and means of the processing of personal data
  - IdP and SP (usually) are data controllers
  - Federation (and interfederation) may be joint data controller

EU Data protection directive
Obligations to data controllers (1/3)

Security of processing

- The controller must protect personal data properly
- Level of security depends e.g. on the sensitivity of attributes

=> *Federation policies, use of TLS and endpoint authentication, federation operator's practices…*

Purpose of processing

- Must be defined beforehand
- You must stick to that purpose

=> *Purpose of processing in IdPs: ~to support research and education*

=> *SPs' purpose of processing must not conflict with this*

EU Data protection directive
Obligations to data controllers (2/3)

Relevance of personal data

- Personal data processed must be adequate, relevant and not excessive
- *SPs must request and IdPs must release only relevant attributes*
- *=> md:RequestedAttribute*

Controller must inform the end user

- when attributes are released for the first time
- *SP's name and identity (=>mdui:Displayname, mdui:Logo)*
- *SP's purpose (=>mdui:Description)*
- *Categories of attributes processed (=> uApprove or similar)*
- *Any other information (mdui:PrivacyStatementURL)*
- Layered notice!

EU Data protection directive
Making data processing legitimate

a. User consents, or
b. Processing is necessary for performance of a contract to which the user is a subject, or
c. The controller has a legal obligation to process personal data, or
d. Necessary for vital interests of the user, or
e. Necessary for a task carried out in public interest, or
f. Necessary for the legitimate interests of the data controller

- Lawyer: Use (f): the SP has legitimate interests to provide service to the user
  - When the user expresses his willingness to use the service (e.g. by clicking ”log in” link)

Summary: EU data protection directive in very short

- Process personal data **securely**
- Use personal data only for a pre-defined **purpose**
- **Inform** the user
- Data **minimisation** (Minimal disclosure)
- Service Provider's **legitimate interests** as the legal grounds
- If attributes released **out of EU/EEA**, some more paperwork needed

- We seem to be coverging on these interpretations
- The proposed General Data Protection Regulation does not change the big picture, but there are some updates