DATA PROTECTION CODE OF CONDUCT FOR IDENTITY FEDERATIONS

Mikael Linden
CSC – IT Center for Science, P.O. BOX 405, FI-02101 ESPOO, Finland
e-mail: mikael.linden@csc.fi
Steven Carmody
Brown University, CIS, Box 1885, Providence, R.I. USA 02912
e-mail: Steven_Carmody@brown.edu

Paper type
Research paper

Abstract
The Data protection Code of Conduct designs an approach to meet the requirements set by the EU Data Protection Directive for parties that are using federated identity management. It defines behavioral rules for Service Providers which want to receive user attributes from the Identity Provider servers managed by the Home Organisations. It is expected that Home Organisations will be more willing to release attributes to the Service Providers who declare that they conform to the Code of Conduct. The Code of Conduct is a unilateral declaration by a Service Provider, rather than a bilateral contract between the parties.

The Code of Conduct has been developed together by the REFEDS attribute release workgroup and the GÉANT eduGAIN project. A successful pilot was carried out in the CLARIN community during the autumn 2012 and spring 2013. In the pilot, the Code of Conduct was found to be a promising approach to encourage the Home Organisations to release attributes to the Service Providers in the EU and European Economic Area in the SAML 2.0 Web single sign-on scenario.

This paper presents the legal and community requirements for an attribute release, and the Code of Conduct and how it addresses the requirements. The key findings of the pilot are presented as well.

Keywords
Federated identity management, SAML 2.0, data protection, data protection directive, CLARIN

1. Introduction

Federated identity management has been a subject of interest for the research and education networking community since the beginning of the last decade. Already 34 academic identity federations have emerged (REFEDS, 2013), allowing end users to log in to remote services using the usernames and passwords their Home Organisations have issued to them. Currently, federated access is enabled to more than 2500 services and 1500 Identity Providers are serving the end users affiliated to their Home Organisations (TERENA, 2012).

When federated access takes place, the Home Organisation not only authenticates the end user but also releases some of his/her personal data (called attributes) to the Service Provider. For authorisation in a simple service, information of the end user being a student or a faculty member may be enough to enforce the access control, but there are also Service Providers which want to provide a personalised service and therefore need to identify the user or at least recognise that he or she is the same person who visited the service last week. All those services are processing end user's personal data and, in the European Union, are subject to the Data protection directive (EU, 1995).

The directive holds the data controllers, such as the Home Organisations managing their users' personal data, responsible for protecting the personal data appropriately. That namely includes having the adequate safeguards in place when the Home Organisation releases the user attributes to a Service Provider. If something unexpected happens to the attributes – for instance, the service is hacked and the personal data spills to the Internet – a court or regulator may hold the Home Organisation at least partly liable because it has decided to release personal data to such a poorly managed Service Provider.

In practice, this has made many Home Organisations hesitate to release attributes to Service Providers. To be on the safe side, many Home Organisations have decided not to release attributes to a Service Providers with whom they do not have a contractual relationship and which do not belong to the same identity federation. Services,

especially in an interfederation such as eduGAIN, are reporting problems in obtaining necessary user attributes (Broeder, 2012; CLARIN-D, 2012). That has made some Service Providers even question the benefits of federated identity management. It is easier to issue local usernames and passwords to the user than try to contact potentially hundreds of Home Organisations and convince them to release attributes.

The directive often provides multiple ways to approach each of the issues. For instance, there are several "legal grounds" a Service Provider can choose from when asking for the release of attributes by a Home Organization. But, the different choices impose differing requirements on the two parties. For each attribute, the two parties must know the associated legal grounds, and then each party must be convinced that the other party has met its responsibilities before it can proceed.

E-research and the services researchers and research communities share with each other can be seen as a major user of federated identity management and interfederations (Broeder, 2012). Research is highly international and researchers establish dynamic collaborations. In order to persuade the research services to make use of federated identity management, the attribute retrieval from the Home Organisation should be as simple as possible. The researchers do not have the time and patience to wait until the university counsellor goes through complicated legal agreements to establish contractual relationships between the Service Providers and Home Organisations.

On the other hand, many of the attributes shared between the Home Organisations and Service Providers are relatively innocuous and low risk, such as a person's name, unique identifier and e-mail address. For instance, every time someone sends an e-mail message, those attributes are shared by the SMTP servers processing the e-mails. Moreover, researchers have seldom a desire to hide their identity. Quite the opposite, having his/her name attached to his/her contribution is a merit for him/her as a researcher and is beneficial for his/her academic career. The perceived risk is much lower than the perceived gain.

Having these findings in mind, the REFED attribute release working group and the eduGAIN task of the GN3 project joined their forces to design a Data protection Code of Conduct (REFEDS, 2013b) which intends to ease the attribute release to services in the European Union and European Economic Area. The Code of Conduct was developed during the spring 2012. In order to get a general approval in the community, the Code of Conduct was exposed to a public commenting in the summer and autumn, and its applicability was tested in a pilot with CLARIN, the European e-research infrastructure for language research technology.

This paper presents the Data protection Code of Conduct. In the section 2, some previous work is first introduced. The section 3 introduces the non-technical and technical requirements on the Code of Conduct. The section 4 presents the Code of Conduct and how it relies on the SAML 2.0 metadata exchange between the Identity and Service Providers. The section 5 compares how the requirements presented in the section 3 were met. The section 6 presents the main findings of the pilot carried out together with the CLARIN community. The section 7 presents some future work items and the section 8 concludes the paper.

2. Previous Work

Internet Service Providers, including national research and education networks, have traditionally limited their role to the pure traffic exchange. Anything related to the actual traffic content, such as any copyright infringements, has been out of the scope for them and something the content provider alone is responsible for. That was also the data protection approach adopted by the early identity federations. The federations required that the Home Organisations and Service Providers respect the data protection laws, but how they did it was left to them to decide. Some federations, especially those focusing to serve access management to the licensed contents acquired by the university libraries, also discouraged the release of any personal data because it was not necessary for the type of Service Providers in their focus area.

The first results regarding the data protection issues in the federated identity management were presented in 2005 (Liberty, 2005; Linden, 2005). In the research and higher education community, the focus shifted to consent and how it should be used as the legal basis for the attribute release (Simonsen, 2009; Wegdam, 2011). Solutions such as Shibboleth uApprove and SimpleSAMLphp consent module were presented. Later, in year 2011, the European data protection authorities questioned consent as a valid legal basis for the attribute release (WP29, 2011), which led the community to seek for more comprehensive solutions to satisfy the regulatory requirements (Cormack, 2011).

In the eduGAIN task of the GN3 project, a specification called Data Protection Good Practice Profile (Linden, 2011) was developed to introduce a division of the data protection responsibilities between the Home Organisations and Service Providers. The profile was criticised by lawyers for not providing strong enough legal guarantees to actually reduce the Home Organisations' risks in the attribute release. Nevertheless, the profile was the first wave in the development which led to the introduction of the Data protection Code of Conduct.

Alternative approaches have also been presented to ease the attribute retrieval by the Service Providers. Some federations have started from an assumption that Home Organisations are more willing to release attributes to the Service Providers if they know more about their properties, especially if the service qualifies to a certain "category" of the service types, such as the "research and scholarship" services. The federation assigns a category tag to a Service Provider, and the Home Organisations are free to filter services based on the tags. This approach is adopted in the InCommon federation of the United States (InCommon, 2013) where the data protection laws do not have as significant role as they do in Europe. Also some European federations are adopting service categories (SWAMID, 2013; RENATER, 2013).

3. Requirements for the Data protection Code of Conduct

This section first provides a short introduction to the relevant provisions of the Data protection directive, supported by the legal opinion acquired by the eduGAIN project (van Eecke, 2011). A more complete introduction is provided in the supporting material of the Code of Conduct (REFEDS, 2013b). The section also introduces the specific requirements of the research and education sector.

3.1. Introduction to the Data Protection Directive

The Data protection directive and the national laws implementing it in the EU Member States are applied to any processing of personal data. **Personal data** is defined as any information relating to an identified or identifiable natural person. User attributes like name, e-mail address and eduPersonPrincipalName obviously qualify as personal data and it can be argued that also even pseudonymous identifiers (such as eduPersonTargetedID or SAML 2.0 persistent identifier) relate to an identified or identifiable individual. According to a conservative interpretation, to be on the safe side, a Service Provider should treat also the non-identifying attributes (like eduPersonAffiliation) as personal data when retrieved from an Identity Provider, because the identity of the end user can be revealed by merging the log files collected by the Identity and Service Provider (van Eecke, 2011).

The directive defines a **data controller** as a natural or legal person or any other body which *alone or jointly with others determines the purposes and means of the processing of personal data*. A data controller bears the legal responsibility of the personal data processing, even if it uses contractors (qualified in most cases as data processors). Home Organisations are obviously controllers of their end users' personal data. Some Service Providers (such as licenced content providers or Software as a Service providers) may in principle qualify as data processors, but often also the Service Providers have enough freedom in determining the purposes and means of the data processing to also qualify as data controllers (van Eecke, 2011). Thus, the release of attributes to a Service Provider is usually a release of personal data between two data controllers.

In a full-mesh federation, the attribute release takes place directly between the Home Organisation and Service Provider without the federation operator having access to the attributes. Nevertheless, there is a possibility that the federation qualifies as a **joint data controller** which *jointly with others determines the purposes and means* of data processing because the identity federations govern the policy that the federation members need to adhere to (van Eecke, 2011). This implies that an identity federation cannot simply wash its hands of the data protection issues in the federation.

The directive requires that personal data is *collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.* In the research and education sector, the **purpose of processing** is in general related to or restricted by supporting the research and education, and the Home Organisations cannot release attributes to a conflicting purpose. For instance, releasing end user's attributes to a gambling service is hardly supporting the research and education.

The directive also mandates the controller to **inform the end user** on processing of his/her personal data no later than the time when his/her data is first disclosed to a Service Provider. The end user must be told at least the name and contacts of the Service Provider, purposes of processing, categories of attributes concerned and how to access and rectify his/her data. This information must be given directly to the individual; it is not enough that the information is "available" somewhere (WP29, 2011, p.20). In the Internet, a standard practice to inform the end user is to provide him/her easy access to the service's Privacy Policy web page.

According to the directive, personal data *must be adequate, relevant and not excessive* in relation to the purpose of processing. In the federated identity management, this translates to the principle of **a minimal disclosure**; a Home Organisation must release only relevant attributes to the Service Provider.

Additionally, the directive defines various **legal grounds** for processing personal data, of which at least one must be satisfied. Personal data may be processed if it is necessary for a *legal obligation* or *for performance of a contract to which the data subject is a party* or *for performance of a task carried out in the public interest.*

Furthermore, data may be processed if it is necessary *for the purposes of the legitimate interests pursued by the third party to whom the data are disclosed,* as long as those interests are not conflicting with the end user's fundamental rights and freedoms. Finally, personal data may be processed if the end user has *unambiguously given his consent*. To qualify as valid legal grounds, the consent must be *freely given, specific and informed.*

Whether or not consent can be used as the legal grounds for the attribute release to a Service Provider has been a subject for lively discussions within the identity federations. The Article 29 working party (the EU body contributing to the uniform application of the Data protection directive) has used an employment relationship as an example where consent may not be valid legal grounds because it may not be freely given. An employee is in a situation of dependence on the employer and might fear that he could be treated differently if he does not consent (WP29, 2011, p.13). The same approach could be applied also to a student whose ability to receive a university degree may be endangered if he/she does not consent to the attribute release. This opinion of the Article 29 working party has severely limited the applicability of consent as the legal grounds for the attribute release.

Consequently, the legitimate interests of the data controller are now proposed as the legal grounds for the attribute release in an identity federation (van Eecke, 2011). When an end user has shown interest to log in to a Service Provider, it is a legitimate interest of the Service Provider to retrieve necessary user attributes from his/her Home Organisation. In a research and education federation, reliance on the legitimate interests legal grounds seems justified, taking into account the general privacy-protecting setup of the data flows, the low level risk posed by the personal data being exchanged and the innocuous (mainly scientific) types of services accessed by the end user (van Eecke, 2011).

However, several legal grounds for processing can be used simultaneously (WP29, 2011, p.8). In a federated identity setup, the attributes necessary for the identification and authorisation of the end user and for personalising the service for him can be released under the necessary for the legitimate interests legal grounds. But, in order to enjoy some higher service level, the end user can consent to the release of additional extra attributes. For instance, the eduPersonAffiliation attribute is released for authorisation and SAML 2.0 persistent identifier for personalisation based on the legitimate interests legal grounds without a user consent. Additionally, the user can consent to the release of his/her e-mail address in order to enjoy optional e-mail notifications sent by the service. Alternatively, the user may also type in his/her e-mail address to the service by him/herself, in which case the Home Organisation is not involved in that transaction.

Finally, for avoidance of doubt, it needs to be emphasised that using the user consent as the legal grounds does not exempt the Service Provider from the other requirements in the directive. Even if consent was a valid legal grounds the data controller must respect the other requirements, such as the data minimisation, relevance of the attributes and the obligation to inform the user.

3.2. Design Requirements Following from the Directive

During the design of the Data protection Code of Conduct, the requirements summarised above were compiled into concrete design requirements for technical deployment (REFEDS, 2013b). This section introduces the design requirements.

3.2.1. Security of Processing

To ensure the security of the processing

- the Home Organisation and Service Provider must take necessary measures to protect the personal data, in particular when it is transmitted over a network.
- a Home Organisation cannot release attributes to a Service Provider without implementing the appropriate organisational measures to ensure that the attribute release does not result in an unlawful processing.

3.2.2. Minimal Disclosure

All attributes exchanged between the Home Organisations and Service Providers are assumed to qualify as personal data (including role attributes, such as eduPersonAffiliation). To ensure a minimal disclosure of personal data to the Service Provider,

- A Service Provider must publish the list of the attributes that are adequate, relevant and not excessive to the service. A Home Organisation must have the confidence that all of the attributes requested by the Service Provider are relevant to the service. A Home Organisation must release only the relevant attributes to the Service Provider.

- A Service Provider must lower the risk for all parties by deciding to request lower risk kinds of data (for example, SAML 2.0 persistent identifier instead of eduPersonPrincipalName). For multi-valued attributes, the Service Provider must indicate the subset of values it needs.

3.2.3. Informing the End User

To cover the data controller's obligation to inform the end user of the personal data release,

- A Service Provider must publish a publicly readable Privacy Policy page which covers at least the identity and contacts of the Service Provider, the Service Provider's purpose of processing personal data, the categories of data to be released and the existence of the right of access to and rectify the data concerning him/her.
- The link to the Privacy Policy page must be provided to the user in the service's public facing landing page and must be mediated to the Identity Provider.

To make sure the end user gets informed before any personal data is released, it is recommended that the Home Organisation carries out the duty to inform him/her. In the front-channel binding of the SAML 2.0 web single sign-on, it can be done conveniently after the user is authenticated by the Identity Provider server and before his/her browser carrying the user's attributes is redirected back to the Service Provider.

To balance the privacy and usability, the Home Organisation may decide to apply a layered approach to fulfill the obligation to inform the end user:

- 1st layer: the Identity Provider server deploys an attribute release module which shows to the user the Service Provider's name, a short description of the Service Provider, a list of the attributes and their values to be released and a clickable link to the service's Privacy Policy page.
- 2nd layer: if the end user wants to know more about the Service Provider's privacy practices, he/she can click the clickable link that resolves to the Service Provider's Privacy Policy page.

3.2.4. Specifying the Legal Grounds

In order to ensure the proper legal grounds for the release of each attribute, the Service Provider must indicate the legal grounds separately for each attribute. Two alternatives are available

- Legitimate interests legal grounds for the attributes which are necessary for the service.
- User's consent for extra attributes which are not necessary but offer the user an optional higher service level.

It is the role of the Home Organisation to

- Inform the end user when attributes are being released due to the legitimate interests legal grounds (see above).
- Inform the end user and ask him/her to give his/her consent before optional extra attributes are released due to the consent legal grounds (see the next section).

3.2.5. Release of Optional Extra Attributes on User Consent

When optional extra attributes are released based on the user's consent

- The user must give his/her consent to the Home Organisation before optional extra attributes are released. The Service Provider must be confident that the Home Organisation has properly acquired user consent before accepting any attributes whose release is indicated to require the user's consent.
- The user's consent must be freely given (an end user must have an option to say no), specific (given to each Service Provider separately) and informed (an end user must understand to what he/she consents).
- If several attributes are released based on consent, the user must be able to give his/her consent individually to each attribute. Consent to "all or nothing" is not sufficient.
- A user can be asked to consent to the release of a named "group" of similar attributes (for instance, a user could be asked to consent to release "name", and this single consent would allow the release of commonName, surName, givenName and displayName). If an attribute has multiple values being released, consenting to the release of the attribute is sufficient to release all of the values.
- There must be a way for the user to withdraw his/her consent after having previously granted their

consent. If they then access the Service Provider in the future, he/she would be re-prompted. However, withdrawing the consent does not require a Service Provider to discard the previously received attributes. The user him/herself can use the Service Provider's Privacy Policy to contact the Service Provider in that matter.

- The Identity Provider may remember the user's consent decision and not prompt him/her again when he/she accesses the same Service Provider for the next time. For usability reasons, this may be desirable.
- For audit trail, the Home Organisation must ensure that reliable log entries are stored on users' decisions to consent to the attribute release.

3.3. Design Requirements from the Research and Education Community

This section introduces requirements that do not follow from the law but from the needs specific to the research and education community.

Scalability. The Code of Conduct should scale to the size where hundreds or thousands of Service Providers and Home Organisations exchange user attributes. Clearly bilateral negotiations and/or contracts between each Home Organisation and Service Provider are out of the question.

Balancing the risks with the easiness of collaboration. The Code of Conduct should find an appropriate balance between minimising the data protection risks and enabling easy collaborations for end users in different Home Organisations. Bilateral contracts would be the approach that guarantees best opportunities for the parties to assess and minimise their risks, but in many cases the researchers in universities are not willing to wait and put their work on hold until the lawyers have come to an agreement. However, the Code of Conduct should not exclude a Home Organisation entering bilateral negotiations with a Service Provider in a case where the guarantees provided by the Code of Conduct are not sufficient.

Tolerate and recover from misbehaving entities. In an ideal world, all the parties would meet their respective privacy-related requirements, and all parties can assume that all the other parties are also meeting their privacy-related responsibilities. There would be no security incidents; no end users would think that their privacy has been injured. In practice, that world does not exist. It is necessary to develop a design that provides all parties with sufficient assurance that other parties with which they are interoperating are meeting their privacy-related responsibilities, and is able to tolerate and recover from incidents and misbehaving entities.

Minimise the risk of a Home Organisation becoming liable for a Service Provider's misbehaviour. In general, it belongs to the mission of a Home Organisation to release attributes to a Service Provider when the attribute release supports its end user's research, teaching and study related activities. However, if it turns out that the Home Organisation may become liable for a data protection problem caused by a Service Provider, the Home Organisations become hesitant and refrain from their role as an Identity Provider. Instead, the users would need to register and learn local usernames and passwords, and the benefits of the federated identity management are lost. Thus, the Code of Conduct should articulate the responsibilities of a Service Provider in a way which avoids a situation where the Home Organisation becomes a (joint) data controller suffering from the liabilities caused by a Service Provider's misbehaviour, as long as the Home Organisation does its due diligence with respect to Service Provider practices.

Suggest good practices to the Home Organisations. The intention of the Code of Conduct is to reduce the Home Organisations' hesitation to release attributes to the Service Providers. This is achieved by introducing behavioural rules for Service Providers. The rules convince the Home Organisation that it is safe enough to release personal data to the Service Provider. This does not apply in a reversed order. The Code of Conduct is a Code of Conduct for Service Providers, not for Home Organisations. The Code of Conduct may introduce good practices which help a Home Organisation to reduce its data protection related risks, but if the Home Organisation decides to ignore them, it is first and foremost the Home Organisation, not the Service Provider, that faces the risk exposure.

Minimise the federation's role. The attribute exchange takes place between the Home Organisation and the Service Provider and the identity federation is not involved in the transaction. The Code of Conduct should minimise the federation's role.

A global approach needed. The research and education community is global and the collaboration crosses national and federation borders. The approach adopted should be applicable globally, or at least as widely as possible. Competing and conflicting approaches would increase the burden both for the Home Organisations and the Service Providers.

4. The Code of Conduct Approach

This section presents the key elements of the Data protection Code of Conduct, and how the Code of Conduct is used in practice. The full Code of Conduct text is available in (REFEDS, 2013b).

4.1. The Code of Conduct for Service Providers in EU/EEA

In general, the Code of Conduct does not add to the legal requirements that the Data protection directive and its national implementations already require. Rather, it defines a single approach where the directive may offer multiple options, each imposing different requirements on Home Organisations and Service Providers. A Service Provider's commitment to the Code of Conduct can be seen more as a statement that the Service Provider is aware of its legal obligations following from the national laws in European Union or European Economical Area. This section examines these further. The references in the parentheses refer to appropriate clauses in the Code of Conduct.

4.1.1. The Code of Conduct Reflecting the Data Protection Laws

The Code of Conduct emphasises that a Service Provider must always respect the local data protection laws (Clause a). If something is not covered or turns out to be in conflict with the law or its interpretation, the law has precedence.

Many of the clauses in the Code of Conduct simply repeat the related section in the Data protection directive such as the sections on data minimisation (Clause c), data retention (Clause e), security measures (Clause g) and information duty towards the end users (Clause h). In some sections the Code of Conduct further clarifies the data protection laws in the context of federated identity management. For instance, the data minimisation implies in practice that if several alternative attributes can be used, the Service Provider requests the least intrusive one, such as the SAML 2.0 persistent identifier instead of the eduPersonPrincipalName attribute.

4.1.2. The Code of Conduct Further Specifying the Data Protection Law

The Data protection laws leave still some room for the data controllers to decide the details of the data processing. To achieve the goals of the Code of Conduct, it was decided to further limit the Service Provider's options when it was believed to be necessary.

In the first phase of adoption, the Code of Conduct (Clause b) limits the legal grounds of data processing to the user attributes "that are necessary" for enabling access to the service (see discussion on the legal ground in section 3.1). The directive provides also other grounds. However, they impose different responsibilities on each of the parties, and both parties would need to know that the other party had met its responsibilities. For instance, if the user consent was used as the legal grounds for a Service Provider to obtain optional extra attributes from an Identity Provider, the Identity Provider would need to deploy a standard mechanism to signal to the Service Provider that it has properly acquired the consent. As discussed in section 3.1, the necessity legal grounds has a simpler set of requirements. In addition, a number of Service Providers were consulted and they felt that this would be sufficient to meet their needs. If the Service Providers need optional extra attributes, they can obtain them directly from the user.

The Service Provider must limit the attributes' purpose of processing to what is necessary for "enabling access to the service" (Clause b). This wording limits the functionality for which the Service Provider can use the attributes received from the Home Organisation. The wording was chosen to be a middle way which still leaves some freedom for the Service Provider. A stricter wording "to do access control in the service" could have excluded e.g. identification and personalisation of the end user and other important functionalities. A looser wording "to provide the service" would have left the door open for a Service Provider requesting more attributes than the Home Organisations feel comfortable to release.

The Code of Conduct (Clause d) also makes it explicit that the attributes can be used for purposes other than enabling access (e.g., for marketing) only if the user consents to the additional use. This was deemed appropriate because the user consenting to the new purpose is assumed to start a new and separate data processing where the Home Organisation is no longer involved and cannot become liable for violations occurring at the Service Provider.

The Code of Conduct allows an attribute release to third parties (Clause f) only under some special conditions. The Service Provider can use a contractor that provides the service on its behalf; in that case the ordinary data controller/processor relationship applies between them.

Like above, attribute release to a third party is possible if the user consents to the release. However, there are legitimate scenarios where the consent would not be freely given (see discussion in the section 3.1 above). For

instance, the European community for the language research technology has Service Providers where the user actually triggers a Web Service that passes his/her attributes to another data controller who together provide the service to the user. In this situation, the attribute release to a third party becomes possible if the third party commits to the Code of Conduct or a comparable framework. However, when using this possibility, the Service Provider must remember that the attributes shall not be used for deviating purposes.

When releasing attributes to a third party or when processing them for instance inside a multinational company or collaboration, the Service Provider must take into account that the Code of Conduct alone does not enable an attribute release out of the European Union, European Economic Area and countries (such as Switzerland and Argentina) or arrangements (such as the US safe harbour) that guarantee adequate protection for the user's personal data. In other words, if a US company is committed to the US safe harbour framework (a data protection framework negotiated between the European Commission and the US Department of Commerce) it can assert that it conforms to the Code of Conduct and can receive personal data from the European Home Organisations. However, apparently the US universities do not belong to the jurisdiction of the US Department of Commerce and cannot use the Code of Conduct. They need to supplement the Code of Conduct by another agreement that is based on the model contracts prepared by the European Commission (EC, 2004).

Finally, the Code of Conduct (Clause j) requires that the Service Provider reports any suspected security or privacy breaches to the Home Organisation. This requirement was inspired by the proposed General Data Protection Regulation (EC, 2012).

4.1.3. Other legal conditions

Finally, the Code of Conduct fixes some legal practicalities for the Service Provider, such as the jurisdiction (Clause m), to the extent possible. The jurisdiction is important for knowing which country's laws are applied and it also indicates the competent court and data protection supervisor. In a multinational collaboration or company, the jurisdiction is not necessarily obvious. The Service Provider indicates its jurisdiction in its Privacy Policy document.

The Service Provider (Clause k) warrants that it hold harmless the End User and the Home Organisation who has suffered damage as a result of the Service Provider's violation of the Code of Conduct. This section is important to reduce a Home Organisation's hesitation to release attributes.

The Code of Conduct defines also how a Service Provider can terminate its commitment to the Code of Conduct (clause o) and that any agreement (such as a data controller/processor agreement) between the Home Organisation and Service Provider takes precedence over the Code of Conduct.

The Code of Conduct (Clause n) requires that the individual who commits to the Code of Conduct on behalf of the Service Provider (by doing the configurations described in the next section) is eligible to execute such a commitment. This requirement is similar to the ordinary contractual text that the individual signing a contract on behalf of a legal person must be eligible to do that. As will be discussed in the next section, the Code of Conduct is not technically signed by anyone, but other means of commitment are used instead.

4.2. How the Code of Conduct Works Technically

The technical mechanisms used for implementing the Code of Conduct have been made simple for the Service Providers and Home Organisation. The Service Provider can commit to the Code of Conduct when it is registered to a federation; compliance is self-asserted by the Service Provider. A Home Organisation can then observe the Service Providers' commitments when they decide whether to release attributes (Figure 1). The Code of Conduct relies on the standard SAML 2.0 metadata shared between the Service Providers and Home Organisations. The necessary SAML 2.0 metadata elements are specified in a SAML 2.0 metadata profile for the Code of Conduct.

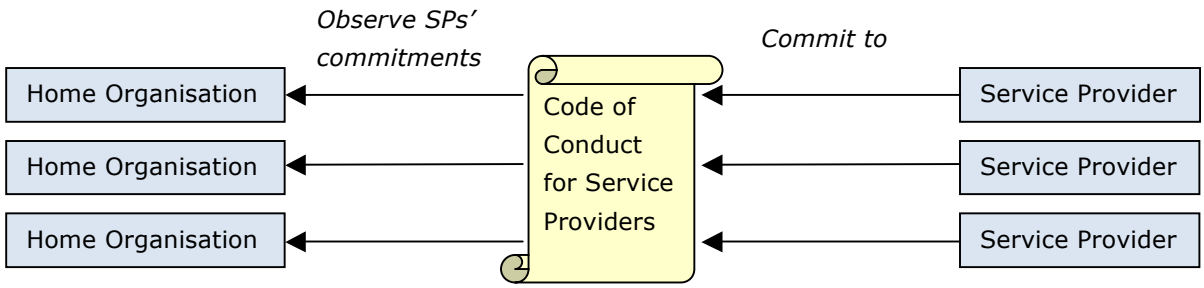

Figure 1. Service Providers commit to the Code of Conduct and Home Organisations observe the Service Providers' commitments.

The first step is taken by the Service Provider which wants to receive attributes from the Home Organisations. The Service Provider studies the Data protection Code of Conduct text and decides if it can assert that it is in compliance. The commitment to the Code of Conduct can be seen as a unilateral declaration (as opposed to a bilateral contract) given by the Service Provider to those Home Organisations from which it wants to receive attributes. If the Service Provider does not want to give the commitment towards a particular Home Organisation, it does not consume the SAML 2.0 metadata (i.e. does not recognise SAML assertions) from that Home Organisation. The approach scales because no bilateral contracts need to be signed.

To further simplify the Code of Conduct for Service Providers, it was decided that the Service Provider does not need to technically sign the Code of Conduct. Management and archival of the ink-signed copies of the Code of Conduct document was found difficult and replacing them by the digitally signed ones had its problems as well. Because (as discussed in the previous section) the Code of Conduct does not add strong requirements to the Service Provider but mostly reflects what the data protection laws already require, an indication weaker than a signature was found sufficient. Instead of a signature, the Service Provider needs to refer to the Code of Conduct in its Privacy Policy document, which is an unsigned statement made by the Service Provider on its privacy practices.

When a Service Provider has decided to commit to the Code of Conduct, it needs to

- Place a reference to the Data protection Code of Conduct into its Privacy Policy. Because the Privacy Policy document has the legal standing defined in the Data protection directive, a reference to the Code of Conduct there is believed to qualify as the evidence of the Service Provider's commitment to the Code of Conduct.
- Place a SAML 2.0 Entity Category attribute as defined by the Code of Conduct in its SAML 2.0 metadata. The presence of the Entity Category attribute enables automated processing in the Identity Provider server and supports scalable use of the Code of Conduct.

To enable the automated processing in the Identity Provider server, the Service Provider's SAML 2.0 metadata is the vehicle also for the other information related to the attribute release, including

- The list of the requested attributes. The Service Provider uses the `md:requestedAttribute` elements to indicate which attributes are *adequate, relevant and not excessive for enabling access to the service.* The Service Provider can further indicate which attributes are required (tagged as `isRequired="true"`) and which attributes can be released for an optional higher service level (tagged as `isRequired="false"`). See section 3.2.5 for details.
- The name (`mdui:displayName`) and a short description (`mdui:description`) that the Home Organisation can show to inform the user on the attribute release.
- A link to the Service Provider's Privacy Policy document, using the `mdui:PrivacyStatementURL` element. The Home Organisation can then use this link to provide additional information to the user.

It is then up to the Home Organisation to use local risk management practices to decide if they are willing to release attributes to the Service Provider that asserts conformance to the Code of Conduct. The SAML 2.0 metadata elements introduced above enable the Home Organisation to automate the attribute release to the extent where the attributes are automatically released to the Code of Conduct-committed Service Providers. The Home Organisations can also reduce their risks by limiting the attribute release only to the non-sensitive attributes even if the Service Provider requests more.

As described in section 3.2.3, the Code of Conduct recommends that Home Organisations reduce their data protection risks by installing an Identity Provider side module that informs the end user before the attribute release takes place for the first time. The attribute release module would show the end user the Service Provider's name (`mdui:displayName`), description (`mdui:description`) and the clickable Privacy Policy URL (`mdui:privacyStatementURL`) together with the list of the attributes the Service Provider has requested. If the Service Provider requests optional extra attributes for a higher service level, the module would take care of asking the user to consent to the attribute release, as described in section 3.2.5. However, the Code of Conduct proposes a two-phase adoption where the optional extra attributes are not requested and released in the first phase.

The Code of Conduct is not going to be mandatory for the Service Providers in the eduGAIN interfederation service. However, to encourage a wide adoption, the Code of Conduct documentation is published under a Creative Commons Attribution Share-alike 3.0 license. This allows federations to adopt the Code of Conduct also in their local federation policy, if they wish.

5. Comparison with the Requirements

Section 3 introduced legal and community requirements for the Data protection Code of Conduct. This section evaluates how the solution presented in the previous section maps to the requirements presented in the section 3.

5.1. Legal Requirements

The obligation to protect the personal data. The Code of Conduct for Service Providers obligates the Service Provider to protect personal data (Clause g). This covers, for instance, the secure configuration and patch management of the service and use of the Transport Layer Security (TLS) in the delivery over the network.

The obligation to implement adequate organisational measures to protect the attributes. The Code of Conduct itself is an organisational measure the Home Organisation uses to protect the users' personal data.

Minimal disclosure. The Code of Conduct obligates the Service Provider to develop a list of the attributes that are adequate, relevant and not excessive to the service. The SAML 2.0 metadata is used to relay the list to the Home Organisation.

Informing the end user. The Service Provider publishes a Privacy Policy document and uses the SAML 2.0 metadata to relay its link and the name and description of the service to the Home Organisation. In the front-channel binding of the SAML 2.0 Web Single sign-on, the Home Organisation can deploy an appropriate module to the Identity Provider server to provide this information to the end user before her/his attributes are released to the Service Provider for the first time.

Legal grounds. It is suggested that, after informing the end user, the Home Organisation releases the necessary attributes to the Service Provider. Future phases of the Code of Conduct may provide for the release of optional extra attributes if he/she consents to it. The distinction between the necessary and optional attributes is done using the `isRequired` XML attribute. In the beginning, to ease adoption, it is proposed that the release of the optional extra attributes is deferred to the phase two.

5.2. Community requirements

Scalability. The Code of Conduct does not require bilateral negotiations and contracts between the Service Provider and Home Organisation. Instead, a Service Provider can make a declaration on its commitment to the Code of Conduct and specify its attribute requirements to all Home Organisations in one step. The SAML 2.0 metadata is the vehicle to pass this information to the Home Organisations.

Balancing the risks with the easiness of collaboration. Home organisations can use their local risk management practices to assess and balance the potential privacy risk of the attribute release with the risk of hindering their researchers' use of important resources. The Home Organisation can then configure their Identity Provider server in a way that matches their risk appétit.

Tolerate and recover from misbehaving entities. If an end user or a Home Organisation has doubts that a Service Provider is not behaving as the Code of Conduct requires, there are several alternatives available. Anyone can contact the Service Provider and ask it to check if it has a compliance problem. Alternatively, he/she can also contact the Service Provider's home federation which, depending on the federation's local policy, may have provisions for enforcement in the local federation policy. It is also possible for anyone to lodge a complaint to the data protection authority of the country in which the Service Provider is established.

Minimise the risk of a Home Organisation becoming liable for a Service Provider's misbehaviour. It is expected that a Service Provider's commitment to the Code of Conduct is enough to demonstrate that the Home Organisation has complied with its data protection duties when releasing the attributes to the Service Provider. In clause k of the Code of Conduct the Service Provider also commits to hold harmless the end user or Home Organisation who has suffered damage as a result of the Service Provider's violation of the Code of Conduct.

Suggest good practices to the Home Organisation. Only the Service Providers commit to the Code of Conduct, not the Home Organisations. The Code of Conduct introduces good practices which help a Home Organisation to reduce its data protection related risks, but if the Home Organisation decides to ignore them, it is the Home Organisation, not the Service Provider that faces the risk exposure.

Minimise the federation's role. The Code of Conduct does not require existing federations to change their policies. Also the operational requirements for registering Service Providers' commitments to the Code of Conduct are modest. Instead, to ease adoption, the Code of Conduct introduces as an overlay to the existing federations. However, if a federation is willing to do it, there should be no obstacles for adopting the Code of Conduct into a federation's policy.

A global approach needed. The current Code of Conduct is limited to the Service Providers in the EU/EEA and countries with adequate data protection, such as Switzerland. Depending on the laws binding the Home Organisation, the Code of Conduct may convince also non-EU/EEA Home Organisations to release attributes to the Service Provider, but the current approach does not enable a Home Organisation in the EU/EEA to release attributes to a Service Provider outside EU/EEA and similar.

To summarise, the Data protection Code of Conduct mostly fulfils the requirements set in the section 3, except that the current approach does not allow the personal data to be released from the EU/EEA to third countries. Furthermore, the release of optional extra attributes on user consent is deferred to the phase two.

6. Pilot with the CLARIN community

To get some non-technical and technical experience on the applicability, deployability and use of the Code of Conduct, the eduGAIN project organised a pilot during the end of 2012 and the beginning of 2013. This section describes the pilot and its findings. For more information on the pilot, see the pilot's final report (REFEDS, 2013c).

6.1. Description

The Code of Conduct pilot was a joint effort of the eduGAIN and CLARIN projects and was carried out by four European identity federations. DFN-AAI, Haka and SWAMID were the full-mesh federations (Identity and Service Providers communicating directly) involved in the pilot and mostly relying on the Shibboleth implementation of the SAML 2.0 technology. The fourth federation was SURFFederatie, which is a hub and spoke federation (the Identity and Service Providers communicating through a centralised proxy server).

A handful of Service Providers (mostly from the CLARIN e-infrastructure for the language research technology) were invited to participate in the pilot. The Service Providers were asked to study the Data protection Code of Conduct and determine if they feel comfortable to commit to it (see section 4.1). After committing to the Code of Conduct, the Service Providers populated the necessary SAML 2.0 metadata elements and fulfilled the other requirements for the Service Providers (see section 4.2). In the end of the pilot, 7 Service Providers (2 from the Institute for the German Language, 2 from the Max Planck Institute for Psycholinguistics, 2 from CSC – the Finnish IT Center for Science and one from Tübingen University) had completed these steps.

In the pilot, some Home Organisations in the four federations were also contacted and asked to study the Code of Conduct. The Home Organisations were asked if they were willing to release non-sensitive attributes to a Service Provider who has committed to the Code of Conduct. The Home Organisations were also actually asked to configure their Identity Provider servers to release the required attributes to the Service Providers. In the end of the pilot, 3 Identity Providers (DFN-AAI, CSC and Uppsala University) were configured to release attributes to the Service Providers.

The role of the four federations was to coordinate the pilot towards the pilot participants in their federations and to mediate the SAML 2.0 metadata including the Code of Conduct elements from the Service Providers to the Identity Providers via the eduGAIN metadata service. In practice, the federations provided a mechanism (for instance, a web interface) to the Service Providers to register the Code of Conduct elements into the metadata. For Home Organisations, the federations provided a mechanism to pick up those Service Providers who have committed to the Code of Conduct. This was accomplished either by providing the Identity Provider administrators with instructions to develop scripting or filtering for the Code of Conduct Service Providers (DFN-AAI and SWAMID) or by implementing a web interface for the Identity Provider administrators where they can configure the Code of Conduct enabled Service Providers to appear in their SAML 2.0 metadata feed (Haka, SURFFederatie).

6.2. Findings

The Service Providers felt comfortable to commit to the Code of Conduct. They quickly discovered that most of the requirements flowed directly from the Data protection directive. However, as most of the persons operating the Service Providers are not experts on the data protection laws, some details of the Code of Conduct were new to them.

For instance, some Service Providers were not familiar with writing and publishing a Privacy Policy document and they needed to develop one for the pilot. To assist the Service Providers, a Privacy Policy template was provided. It also turned out that there is room for conflicts between the Privacy Policy document and the Code of Conduct related SAML 2.0 metadata elements; for instance, in the Privacy Policy document, the list of required attributes did not match the list of the attributes requested according to the SAML 2.0 metadata of the Service Provider. The Service Providers also introduced names and descriptions (e.g. "Lux17 Service Provider") which were not very descriptive and useful for informing common end users.

Furthermore, it was not obvious for Service Providers what attributes they can require from the Home organisations. Guidelines were provided to assist the Service Providers to develop their attribute requirements. However, it also appears that if nobody challenges the Service Provider's attribute list, the Service Providers tend to overlook the data minimisation principle.

Although the general intention of the Code of Conduct is to minimise the role of the federation that has registered the Service Provider (see section 3.3), it seems necessary to obligate the Service Provider's home federation to make some sanity checks to the Service Provider's Privacy Policy document and the other SAML 2.0 metadata elements. However, care must be taken to avoid the federation becoming liable for the Service Provider's omissions.

After studying the Code of Conduct, the Home Organisations involved in the pilot felt comfortable to release attributes to the Service Providers which committed to it. This required consultation with the Home Organisation's information security manager, chief information officer or information manager responsible for the identity management issues. In general, it was emphasised that the Code of Conduct balances the risks of the attribute release leading to legal troubles with the risk of the employees not being able to do their job effectively, and that the alternative would require bilateral negotiations with all Service Providers which would mean a lot of extra work for the Identity Provider administration.

To further convince the Home Organisation that the risks are acceptable the following arguments were presented:

- The Home Organisation can reduce its risks by limiting the maximum set of attributes released to the Service Provider committed to the Code of Conduct. In particular, there is no need to release sensitive personal data.
- The Service Providers are established in EU/EEA or similar where the local laws already provide protection similar to the Code of Conduct.
- There is no batch release of personal data. Instead, in the SAML Web Single sign-on, the personal data is released only when the user needs to access the service, and the Identity Provider server can inform the user on the release before it takes place.

Finally, the Home Organisation was reminded that, in practice, the alternative for the Code of Conduct is that every employee of the organisation creates a new local account at the Service Provider or use his/her social media (e.g. Google, Facebook) account to access the service. Those accounts are outside the employer's control and will not be closed when the user departs.

The pilot demonstrated that the Code of Conduct works from both the technical and non-technical perspective. More material and training is needed to ensure the Service Providers and Home Organisations know how to use it properly. The number of the Home Organisations and Service Providers in the pilot was not extensive, but instead of enlarging the pilot it was preferred to conclude the pilot and start encouraging organisations and federations to deploy the Code of Conduct in the production use.

7. Future Work

The Article 29 working party is the EU body contributing to the uniform application of the Data protection directive. The directive encourages trade associations and other bodies representing data controllers to develop **community codes** that take into account the specific features of the various sectors. The codes may be submitted to the national data protection authorities or to the Article 29 working party.

Currently the Code of Conduct is only a good practice with no legal guarantee. To further legitimise the Data protection Code of Conduct, it is proposed that the Code of Conduct is submitted to the Article 29 working party. The working party's blessing could confirm that the approach is valid, which would further encourage the Home Organisations to rely on it. The downsides of an Article 29 working party submission are that the process takes relatively long time and it can be expected that the working party is going to require amendments to the Code of

Conduct.

Currently, the Code of Conduct is limited to the Service Providers in the European Union and the European Economic Area. However, research is global and the European researchers need to also log in to non-European services; this would require an **attribute release out of the EU/EEA**. The standard approach to releasing personal data to a non-EU/EEA organisation is to use the EC's contractual clauses (EC, 2004); this is a model agreement where the non-European data controller commits to the European data protection laws even though it is not established in the EU/EEA. A potential approach to extend the Code of Conduct to the non-EU/EEA attribute release is to merge the current Code of Conduct with the EC's contractual clauses.

The current Code of Conduct limits the legal grounds to "attributes that are necessary" for enabling access to the service. The directive lists **other possible legal grounds** (e.g. user consent). However, these other grounds impose requirements on both parties, and require that both parties are confident that the other party is meeting its requirements before they proceed. This requires more interaction and signalling than the simpler "necessary" for enabling access to the service. However, the Code of Conduct should be extended to include some of these other legal grounds.

The European Union is reviewing its data protection legislation. The European Commission published its proposal on a **General Data Protection Regulation** on 25th January, 2012 (EC, 2012). The regulation is proposed to repeal the Data protection directive which has been the basis for the Code of Conduct. Although the regulation in its proposed format does not nullify the Code of Conduct, it is likely to imply some changes to it. However, it is expected to take still several years until the new regulation is effective.

The Code of Conduct introduced recommendations on a **graphical user interface** for the Home Organisation's Identity Provider server. In the technical track, a future work item is to facilitate the implementation of those features to the common Identity Provider attribute release modules, such as Shibboleth uApprove and SimpleSAMLphp's consent module.

8. Conclusions

The Data protection Code of Conduct introduces an approach to reduce Home Organisations' hesitation to release attributes to Service Providers. The approach relies on the Service Provider's assertion that it is committed to certain behavioural rules derived from the data protection laws, and a supporting SAML 2.0 metadata profile and other documents. Initially the Code of Conduct expects that the Service Provider is established in the European Union, European Economic Area or in a country that guarantees similar level of data protection.

A successful pilot was organised to test and evaluate the Code of Conduct together with the European language research e-infrastructure. In the pilot, a handful of Service Providers managed by the language researchers committed to the Code of Conduct and Home Organisations from different countries and identity federations relied on the Service Providers' commitment in their decision to release attributes. The pilot encouraged to take the next step towards a wider adoption among the identity federations.

The identified future work items are submitting the Code of Conduct to the Article 29 data protection working party and taking into account the new General Data Protection Regulation, when available. There are also needs to extend the Code of Conduct to cover attribute release also outside the EU/EEA and to cover the release of optional extra attributes on user consent. It is further suggested that the Identity Provider side attribute release modules are developed to support the Code of Conduct.

Acknowledgements

Acknowledgements to Esther Zysset, Nicole Harris, Andrew Cormack, Ian Young, Thomas Lenggenhager, Peter Schober and all the others who have provided fruitful comments on the work. Acknowledgments to Valter Nordh, Josh Howlett, Brook Schofield and Licia Florio for supporting the work and giving it a high priority. Finally, acknowledgements to the eduGAIN policy subtask members, the CLARIN pilot team and the REFEDS attribute release working group members.

The work leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 2007-2013) under Grant Agreement No. 238875 (GÉANT).

References

Broeder, D. Jones, B. Kelsey, D. Kershaw, P. Lüders, S. Lyall, A. Nyrönen, T. Wartel, R. Weyer, H., 2012 Federated Identity Management for Research Collaboration. CERN-OPEN-2012-006.

<https://cdsweb.cern.ch/record/1442597>

CLARIN-D, 2012

CLARIN-D, DARIAH-DE. Call for action on federated identity. <http://www.clarin.eu/system/files/clarin_dariah_call-for-action-aai.pdf>

Cormack, A., 2011

Federated Access Management. REFEDS White Paper. Oct, 2011. <https://refeds.org/docs/Art29WP_v0_194.pdf>

EC, 2004

European Commission. 2004/915/EC. Commission Decision amending Decision 2001/497/EC as regards the introduction of an alternative set of standard contractual clauses for the transfer of personal data to third countries.

EC, 2012

European Commission. 2012/0011. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

van Eecke, P. Truyens, M., 2011

Data Protection analysis eduGAIN project. Memorandum. 29 June 2011 <https://refeds.terena.org/images/6/65/DLAPiperMemo.pdf>

EU, 1995

Directive 95/46/EC of the European Parliament and of the Council. On the protection of individuals with regard to the processing of personal data and on the free movement of such data.

InCommon, 2013

InCommon. Research and Scholarship Category. < https://spaces.internet2.edu/display/InCFederation/Research+and+Scholarship+Category> [Accessed 13 August 2013]

Liberty, 2005

Liberty Alliance Project. Circles of Trust: The Implications of EU Data Protection and Privacy Law for Establishing a Legal Framework for Identity Federation. February, 2005

Linden, M., 2005

Organising Federated Identity in Finnish Higher Education. Computational Methods in Science and Technology, 11(2), pp.109–118

Linden, M., 2011

How to satisfy the data protection regulations in federated identity management? TERENA Networking Conference, 2011. <https://tnc2011.terena.org/core/presentation/40>

REFEDS, 2013

REFEDS Federation Survey. TERENA. <https://refeds.terena.org/index.php/Federations> [Accessed 3 March 2013]

REFEDS, 2013b

Data protection Code of Conduct. <https://refeds.terena.org/index.php/Data_protection_coc> [Accessed 3 March 2013]

REFEDS, 2013c

Data protection Code of Conduct. Pilot report. <https://refeds.terena.org/index.php/CocPilotReport> [Accessed 13 April 2013]

RENATER, 2013

RENATER. List of the federation "Fédération éducation et recherche" Identity Providers and resources . <https://services-federation.renater.fr/liste?action=view_all&type=sp&federation=renater&lang=en>

[Accessed 3 March 2013]

Simonsen, D. Madsen, J., 2009

Trusted third party based ID federation, lowering the bar for connecting and enhancing privacy. TERENA Networking Conference, 2009. <http://tnc2009.terena.org/schedule/presentations/show20b7.html?pres_id=29>

SWAMID, 2013

SWAMID. Entity Categories. <https://portal.nordu.net/display/SWAMID/Entity+Categories> [Accessed 3 March 2013]

TERENA, 2012

Metadata Explorer Tool. TERENA. <http://beta.terena-met.yaco.es/> [Accessed 3 March 2013]

Wegdam, M. van der Harst, E. Janssen, R., 2011

The user perspective on consent for identity federations. TERENA Networking Conference 2011. <https://tnc2011.terena.org/core/presentation/71>

WP29, 2011.

Article 29 Working Party. Opinion 15/2011 on the definition of consent. <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf>

Biographies

Mikael Linden works at CSC – the IT Center for Science Ltd and has been operating the Haka federation of the Finnish research and higher education since 2005. He received his doctoral degree in information security from Tampere University of Technology in 2009. He has chaired the subtask for the policy development in the eduGAIN task (SA3 T3) of the GN3 project.

Steven Carmody is an IT Architect at Brown University. He is a member of Internet2's Middleware Architecture Committee for Education (MACE), and the InCommon Federation Technical Advisory Committee. He also chairs the REFEDS attribute release workgroup. He was the Project Manager for the Shibboleth project for ten years.