UK Access Management Federation for Education and Research

Recommendations for Use of Personal Data

4th July 2014

Version 3.1

ST/AAI/UKF/DOC/002

Contents

1. Introduction

The UK Access Management Federation for Education and Research is designed to protect the privacy of users while giving both Service Providers and User Organisations sufficient assurance that requirements such as licenses and acceptable use policies can be enforced. The architecture chosen for the UK Federation is designed to achieve better user privacy and service provider assurance than common alternative forms of authentication and authorisation; however the measures it provides can only be effective if they are used and respected by User Organisations, Identity Providers and Service Providers.

The basis for the federation is that a user's primary relationship is with their organisation and that personal data should normally be kept within this relationship. Many Service Providers will only need to know that an individual is a recognised user, having a particular status, at a member organisation; where a user makes a series of visits to the same service they may wish to save settings, recent searches etc. between visits. Federated authentication allows this to be done in a privacy-protecting way, without disclosing anything that current UK law considers to be personal data.

For some education and research services there may be a clear benefit to the user, and to the organisation that employs or teaches them, if the service can use additional personal information. This may create some additional risks to the user's privacy and is likely to involve new legal obligations for both the user organisation and the service provider. Organisations will need to satisfy themselves that these risks are justified by the benefits. Service Providers can help by minimising the personal data they request and by behaving in ways that reduce the risk to it. Service Providers should endeavour to provide service, possibly at a reduced level, to users for whom this personal data is not available.

This guide explains the various privacy systems available in the UK Federation and how they can be used to protect the interests of users, user organisations and service providers. The guide first covers the general requirements on UK Federation members and then specific issues relating to the two main areas likely to involve personal data: attributes and logfiles. Some ways of addressing these issues are described as examples of good practice, rather than to be prescriptive. Other approaches that satisfy the legal, contractual and operational requirements may be appropriate in particular circumstances.

1.1 Changes in this Edition

- Section 3.4.2 *Service Provider Categories* updated to include REFEDs R&S Category.

2. Federation Requirements

2.1 Rules of Membership

All members of the UK Federation are required to abide by the Rules of Membership.[1] A condition of the Rules, and therefore of membership, is that members abide by the eight Data Protection Principles set out in the UK's *Data Protection Act 1998* and described in the following section. A breach of these principles may therefore be both a breach of UK law and grounds for exclusion from the UK Federation.

2.2 Legal

Activities of the UK Federation, whether performed by members or the federation operator, are subject to the *Data Protection Act 1998*. This requires that any personal data (defined by the Act as any data that can be associated with an identifiable individual) must be processed (which includes collection and disclosure) according to the eight Data Protection Principles contained in Schedule 1, Part I of the Act:

1. Personal data shall be processed fairly and lawfully;
2. Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes;
3. Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed;
4. Personal data shall be accurate and, where necessary, kept up to date;
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes;
6. Personal data shall be processed in accordance with the rights of data subjects under this Act;
7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data;
8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

[1] http://www.ukfederation.org.uk/library/uploads/Documents/rules-of-membership.pdf

The main requirements these principles impose on User Organisations, Identity Providers and Service Providers are discussed in the following section.

Note.

Additional legal requirements may apply to the handling of information that raises particular privacy risks. For example information relating to racial or ethnic origin, political or religious beliefs, health or offences is classified as sensitive personal data under the *Data Protection Act 1998*; personal information about children may require additional measures to inform responsible adults, to obtain valid consent or to prevent inappropriate use of the data by those handling it. Compliance with these requirements is the responsibility of the User Organisations and Service Providers collecting or using such information, not the UK Federation, and is likely to be ensured by appropriate procedures and contractual arrangements. Guidance on these issues is available from the UK Office of the Information Commissioner.[2] Providers of services to children should also be aware of the Home Office good practice guides for internet services.[3]

2.2.1 Requirements for User Organisations and Identity Providers

Identity Providers have to process personal data about individual users to maintain accounts, authenticate users and investigate problems. All such processing must be fair, lawful, necessary and proportionate to the purpose(s) for which the data is required.

The *Data Protection Act* creates specific duties to inform users, minimise processing, and ensure the security of personal data. Where a User Organisation runs its own Identity Provider, it will be responsible for all of these. In some cases a User Organisation will choose to outsource the operation of its Identity Provider system to a different organisation. The following bullet points suggest one possible way that responsibilities might be shared between the parties. Whatever allocation of responsibilities is agreed, the outsourcing agreement must be clear about these, and about the ways that each party will use and protect the personal information involved (advice on outsourcing agreements is available from the Information Commissioner[4]).

[2] http://www.ico.gov.uk

[3] http://www.education.gov.uk/childrenandyoungpeople/safeguardingchildren/b00222029/child-internet-safety

[4] http://www.ico.org.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/outsourcing_guide_for_smes.ashx

- User Organisations must inform their users at the time personal data is collected of what it will be used for and whether it may be disclosed to other members of access management federations (this is known as a fair processing notice). Using federated services is likely to be a necessary part of the individual's learning, research or employment with the organisation so this information can be provided along with other routine data processing information when they join. If the organisation is planning to collect or disclose optional information in addition to the standard UK Federation attributes described below then, so far as is possible, users should be allowed to opt out of this;

- User Organisations must also ensure that personal data is only released to Service Providers where this is necessary and where they can ensure that the data will not be misused. In many cases this will already be covered by a site licence; otherwise the User Organisation will need to assess the risk of harm based on the type of data being released, the organisation providing the service and the laws and Federation rules that bind it. If the risk appears high, a written agreement may be needed to provide sufficient assurance (guidance on appropriate contracts is available from the Information Commissioner's website referenced above);

- Identity Providers must only release personal data to Service Providers where the User Organisation has determined, as above, that this is necessary and that the data will not be misused. Individual users may be asked to consent to additional personal data being disclosed about them, but consent must be given freely and after the user has been fully informed and understands how their information will be used. Where information is being released based on consent, users must also be provided with a way to change their minds and either alter or prevent future releases of this information;

- User Organisations and Identity Providers must use appropriate technical and organisational measures to protect personal data in their keeping.

2.2.2 Requirements for Service Providers

Some Service Providers may wish to use personal data about individual users, where the benefits to their service justify the additional legal duties this will involve. There are two possible sources from which this personal data may be obtained:

- If the accuracy of the information is critical to the operation of the service (for example information that is used to decide whether or not a particular user should be given access to information), then this should be obtained from the user's Identity Provider as this gives a third party guarantee of accuracy. Note, however, that Identity Providers may only be able to provide the information represented by the UK Federation's core attributes, described in the next section, and will only release it if they are comfortable that the benefits of doing so justify the risk;

- Information whose accuracy does not matter to anyone other than the user (for example information used to personalise a home page) may be obtained direct from the user, since they are free not to provide it; users may welcome a choice, for example to be known by a nickname rather than the formal name on their organisation's official record.

Whichever source personal information is obtained from, the Service Provider must use it in accordance with the law:

- Service Providers must only use personal data for purposes that have been agreed with the User Organisation or user from whom it was obtained;
- Service Providers must only request personal data that is strictly necessary for the stated purpose(s), and must not keep the data for longer than required for the purpose(s). Minimising the amount of personal data requested, and using the least intrusive attributes that will deliver the required function, will reduce the risks to both Service Providers and Identity Providers and make it easier for them to conclude that the remaining risks are justified by the benefits. Note in particular that the federation's Rules of Membership allow problems and misuse to be investigated without the Service Provider needing to know the identity of the individual user;
- Service Providers must use appropriate technical and organisational measures to protect personal data in their keeping;
- If requesting information directly from the user, the Service Provider must first inform them of the purpose(s) for which the data is required, and explain how the user can subsequently change their mind and alter or delete any information disclosed.

2.2.3 What is Personal Data?

Section 1 of the UK *Data Protection Act 1998* defines:

"personal data" means data which relate to a living individual who can be identified–
(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the [holder of the data]

Information with a clear link to an individual person, such as their name or e-mail address, clearly does satisfy this definition, while information that merely identifies a user as being a member of a group (e.g. "one of our students") does not.

The status of opaque identifiers that allow a user to be recognised on their return, but do not permit the identification of the living individual, is less clear. In the hands of the Identity Provider that issues them, such identifiers

normally will be personal data, since Identity Providers that have declared compliance with section 6 of the UK Federation Rules must be able to link the use of a federated service to a known individual user. When used within appropriate technical and organisational frameworks, however, the Information Commissioner's Guidance on Anonymisation[5] suggests that Service Providers may be able to treat them as presenting a very low risk to privacy, or as being effectively anonymous, since it is unlikely that Service Providers will obtain the information needed to link them to the individual user. The Article 29 Working Party's Opinion on the Concept of Personal Data (Opinion 4/2007) appears to support the same approach.

The Information Commissioner recognises that any value that is unique to an individual involves some risk of that individual being linked back to their corresponding value. However there are a number of ways that that risk can be minimised:

- The value is generated and used in a way that effectively conceals any information about the user. The UK Federation recommends that opaque identifiers are generated using state of the art hash cryptographic functions that make discovery of the original information computationally infeasible;

- The value does not allow information from different sources to be cross-linked. The UK Federation recommends that a different value is used for each Service Provider; collusion between services is also prohibited by the UK Federation Rules of Membership;

- The value is disclosed to a limited audience, not published. Within the UK Federation, opaque identifiers are only disclosed to individual Service Providers whose services require them;

- The organisation that knows the link between the value and the person is under a duty not to disclose it. The UK Federation Rules prohibit Identity Providers from disclosing the identity of anonymised users, and the operational procedures are designed so as not to require it.

Since the UK Federation's Rules and recommendations satisfy these requirements for privacy protection, there should be very little risk to Identity Providers and Service Providers in using the Federation's recommended opaque identifier within those Rules.

---

[5] http://www.ico.org.uk/Global/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

3. Attributes

3.1 General

For most access management decisions the actual identity of the user is much less important than their status and other characteristics. Access to a resource will rarely be granted on the basis that someone is called "John Smith"; far more relevant is whether the user is a member of staff or student at an educational organisation, whether they are authorised by the institution to access a particular resource or, in some cases, what subject or class they are studying. Federated access management systems communicate this type of information through attributes. In many cases attributes will not constitute personal data so their use is a significant advance in protecting the privacy of users. An Identity Provider that confirms to a Service Provider only that "this user is a member of the institution" both protects their users much better than one that says "this is John Smith" and also provides the Service Provider with the information that is actually needed when deciding whether the user is entitled to see a particular resource. Attributes that do not reveal personally identifiable information should therefore be used wherever possible. Service Providers should design their services to require only these attributes and Identity Providers should normally expect to release them to Service Providers that are bound by the UK Federation's Rules. Attributes that do contain personal information may involve greater risks and duties for both Identity Providers and Service Providers. These may be justified by the benefits to users and their organisations but Identity Providers should analyse the risks and benefits before deciding to release these attributes. This analysis is discussed further in the section on Attribute Release Policies below.

The UK Federation bases its common attributes on a standard description of a person in education known as the eduPerson schema. Four commonly used attributes are described in the following sections.

3.2 Standard Attributes

A number of standard attributes are defined in the Technical Recommendations for Participants.[6] Since these attributes and their values have agreed definitions across the whole UK Federation, Identity and Service Providers who use them can be confident that they mean the same thing to both parties. The same is generally true when working with members of other federations, but there may be occasional variations.

[6] http://www.ukfederation.org.uk/library/uploads/Documents/technical-recommendations-for-participants.pdf

The implications of these attributes for personal data privacy are described in this section; ways of using them to maximise privacy protection are described in the following sections.

3.2.1 eduPersonScopedAffiliation

The eduPersonScopedAffiliation attribute describes the nature of the user's association with the organisation that knows their identity. A UK-specific interpretation of the eduPerson controlled vocabulary for this attribute is described in the Technical Recommendations for Participants. This has been developed in consultation with representatives of all education sectors and contains all the common relationships between organisations and their members (e.g. "staff", "student", "member"). Note that some values of eduPersonScopedAffiliation may have different meanings in other countries due to differences of language and national education systems.[7] A single user may have more than one eduPersonScopedAffiliation value: for example someone who has the value "staff" or "student" is also likely to have the value "member". eduPersonScopedAffiliation is likely to be sufficient for many access control decisions; because it does not allow a Service Provider to distinguish individual users it also protects privacy. Identity Providers should support eduPersonScopedAffiliation for those they authenticate, releasing the least intrusive value that is required in the particular circumstances (for example if a resource is licensed for all members of the organisation it should not be necessary to disclose whether a particular user is a member of staff or student); Service Providers should design their access control systems to use eduPersonScopedAffiliation wherever possible.

3.2.2 eduPersonTargetedID

For many services the user would like to be able to save information from one session to the next. This may include, for example, personal preferences on how the service should appear or past searches and their results. This requires the Service Provider to be able to recognise the user whenever they return to the service and also to keep the stored information private from other users. However it does not require the Service Provider to know the user's identity.

The eduPersonTargetedID attribute is designed for applications where the Service Provider needs to be able to recognise a returning user. The Identity Provider should set the value of this attribute to be an opaque string, for example a random number unique to each user, containing no information that can be used to identify the person. Different values must be returned for different users. Whenever the same Service Provider requests the eduPersonTargetedID for the same user, the same opaque value should be returned. This persistent but anonymous identifier allows the Service Provider

[7] http://www.terena.org/activities/refeds/docs/ePSAcomparison_0_13.pdf

to retrieve information saved on previous visits. Identity Providers should provide different eduPersonTargetedID values to different Service Providers: this protects the user against collusion by Service Providers to derive or exchange additional information about the user. As discussed above, an attribute constructed in this way represents a very low risk to privacy and may even be considered as anonymous.

Service Providers should design their services to use eduPersonTargetedID for any persistent service.

Wherever a Service Provider retains information about a user, the Service Provider is responsible for ensuring that this information cannot be disclosed to others. This could happen, for example, if the value of a persistent identifier such as eduPersonTargetedID were reused by the Identity Provider, allowing the new holder of the ID to access the previous holder's stored information. To allow Service Providers to protect against this risk the Federation Rules require those Identity Providers that assert that they can account for individual users not to reissue a persistent identifier value to a new user within two years of the last possible use by the previous user. Service Providers can therefore allow an individual account to remain dormant for up to eighteen months before deleting the stored data associated with it. Identity Providers that do not make this assertion are not required to give any such guarantee and Service Providers should therefore be cautious about storing data against identifiers from these Identity Providers.

For many applications a combination of the attributes eduPersonScopedAffiliation and eduPersonTargetedID will be sufficient. Where services require the release of other attributes that may involve personal data, User Organisations should ensure that the benefits to them and their users justify any increased risk.

3.2.3 eduPersonPrincipalName

In order to protect privacy, a user should have a different eduPersonTargetedID attribute value for each service. Where there is a genuine requirement to identify a particular individual across different services or organisations, the eduPersonPrincipalName attribute may be used, as this provides a single identifier (often a login name that gives access to both internal and external services) for an individual. Since eduPersonPrincipalName allows a user's activities to be tracked across both internal and external services its use is likely to involve some privacy risks (one of the Information Commissioner's concerns about pseudonyms is the ability to match data across systems). If a login name is used as a component of the ePPN value then this may also involve risks to the security of both internal and external information systems.

It will often be possible to associate an eduPersonPrincipalName with an individual so it is likely to constitute personal data within the meaning of the

Data Protection Act 1998. This means that the data protection principles in that Act will apply to disclosure or use of eduPersonPrincipalName, in particular the user must be informed that their identity will be disclosed and what this may be used for. Both Identity Provider and Service Provider must take appropriate technical and organisational measures to protect the eduPersonPrincipalName and information stored in association with it. Before disclosing eduPersonPrincipalName the Identity Provider should be satisfied that the risk of misuse is low and justified by the benefits to the user.

As with eduPersonTargetedID above, it is the Service Provider's responsibility to ensure that information stored in association with an eduPersonPrincipalName is not disclosed or otherwise misused. In particular, Service Providers must ensure that their processes take account of the Identity Provider's policy on whether values of eduPersonPrincipalName may be reused.

3.2.4 eduPersonEntitlement

Although most access control decisions will be based simply on the user's status or role within the organisation, for a few services access will only be granted if the individual user satisfies a more complex set of conditions defined by the Service Provider. For example access to medical resources may only be available to users of a certain age and training who have signed a non-disclosure agreement; or access to sensitive cultural artefacts may depend on the age, gender and race of the user. Previously this type of application has involved the Service Provider maintaining a list of logins for authorised individuals: a process that is both hard to maintain and a potential breach of privacy. The eduPerson schema instead provides the eduPersonEntitlement attribute for this purpose: a set of conditions are defined by a Service Provider or other organisation and a unique value (formatted as a Universal Resource Identifier (URI)) chosen for the eduPersonEntitlement to mark those users who satisfy all the conditions. Identity Providers are responsible for ensuring that users who satisfy the particular set of conditions can assert the relevant value of the attribute. Both ease of maintenance and privacy are thereby improved.

In general, eduPersonEntitlement values will not constitute personal data; however where there are only a small number of entitlement holders per organisation it may be possible to identify them as individuals using other information. As the examples above indicate, particular sets of conditions may even contain information classified as "sensitive personal data" by section 2 of the *Data Protection Act 1998*: these may only be stored or disclosed with the explicit permission of the user. Before implementing an eduPersonEntitlement value, therefore, the organisation must consider whether it is likely to constitute non-personal, personal or even sensitive personal data. For values that constitute sensitive personal data, individuals should be asked to consent before the value is assigned to them. In any case, eduPersonEntitlement values

must only be released to Service Providers where they are necessary and relevant to that Service Provider's access terms.

3.3 Other Attributes

As described in the Federation's Technical Recommendations for Participants,[8] individual Identity Providers may choose to provision and release additional attributes if they have specific requirements that cannot be met by the federation's standard attributes described above. These attributes will be covered by the same laws and policies, so comparison with the discussion of the standard attributes above should indicate their privacy and legal implications.

3.4 Attribute Release Policies

Different Service Providers will require different attributes in order to provide service to users. For example

- a Service Provider whose resources are licensed to all members of the organisation should need to know only the "member" value of eduPersonScopedAffiliation and, perhaps, an eduPersonTargetedID to allow each user to store preferences;
- a group of Service Providers that need to link accounts across their services (e.g. to link a VLE to a student record system) may require eduPersonPrincipalName;
- where a Service Provider has defined their own eduPersonEntitlement value, that value should not be released to other SPs.

Disclosing attributes to Service Providers that they do not need creates risks for users, their User Organisation/Identity Provider and the Service Provider, so User Organisations need to manage the attributes their IdPs disclose.

This is normally done by maintaining an Attribute Release Policy for the Identity Provider, listing which attributes will be released to which Service Providers. Service Providers are encouraged to publish a list of which attributes their service requires and, if they involve personal data, what the attributes and any other information will and will not be used for. User Organisations, who are ultimately responsible for protecting their users' privacy, can then quickly identify the required ARP configuration and, in the light of the benefits to users of having access to the service, satisfy themselves

[8] http://www.ukfederation.org.uk/library/uploads/Documents/technical-recommendations-for-participants.pdf

that it represents an acceptably low risk to privacy. Because it separates the authentication and authorisation steps, federated access management will often provide better protection for both privacy and security than individuals setting up user accounts directly with Service Providers.

3.4.1 Default Attribute Release Policies

As noted above, some of the UK Federation's standard attributes are either not personal data at all, or represent a very low threat to privacy. Many sites have therefore concluded that they can safely configure a default Attribute Release Policy that will release these low-risk attributes to any UK Federation member Service Provider where a user attempts to log in, knowing that such Service Providers are bound by the federation's Rules of Membership to handle them in accordance with the *Data Protection Act 1998*. A default ARP of this kind can enable privacy-protecting federated access to many services without any individual configuration.

3.4.2 Service Provider Categories

For services that require the release of attributes that do represent personal data, User Organisations must decide whether the benefit to their users of providing access to the service justifies the risk represented by releasing the attributes. Different service providers will offer a different balance of benefit and risk: one tool that is being developed to help in making this assessment is to define categories of Service Provider that may have similar risk/benefit characteristics.

For example the Research and Education Federations (REFEDs) group, of which Janet is a member, has defined a category for services that are specifically designed to support research and scholarship (R&S), such as research collaboration tools.[9] Services that allow sharing of ideas and resources among a research or study group might be considered to offer particularly high value to User Organisations themselves engaged in research and education. Although these services are likely to require disclosure of a limited amount of additional personal data (principally name and e-mail address to allow group members to recognise one another), the UK Information Commissioner has described "transfer[ring] the academic biographies of its lecturers and research staff to other universities and potential students outside the EEA"[10] as representing a low privacy risk provided individuals were informed and able to opt out. Identity providers might consider that this balance of high benefit with low risk justifies setting a default Attribute Release Policy that offers a wider range of attributes. The

[9] https://refeds.org/category/research-and-scholarship/

[10] http://www.ico.org.uk/for_organisations/data_protection/the_guide/principle_8#assess first example

REFEDs category description recommends which attributes should be offered to R&S services.

The SWAMID (Sweden) and InCommon (USA) federations have previously used their own R&S categories to help users, identity providers, and service providers establish communications quickly and effectively. These federations plan to move to the REFEDs category definition, which is based on their experience.

3.4.3 Personalised Attribute Release Policies

Some Identity Providers may offer their users the ability to change their personal attribute release settings from the default assigned by the Identity Provider to that Service Provider. This may be used either to further restrict the attributes released for a particular user or to provide additional attributes.

Users who are given this facility must be informed of the consequences of using it: restricting attributes may reduce the service available from particular Service Providers, or even prevent access entirely if an attribute was necessary for the Service Provider's authorisation process, while by permitting the release of additional attributes the user may be exposing their personal information unnecessarily. Since releasing additional attributes relies on the user's consent they must be offered the opportunity to change their mind. This may be done by an interface that allows the user to change their attribute release settings either just for the current login session or for an extended period. To ensure that a user's withdrawal of consent is effective, and to ensure they are using up to date information, Service Providers should request fresh attributes from the Identity Provider for each new session rather than storing them locally.

3.4.4 Releasing UK Federation Core Attributes

The following table summarises the questions you should ask when considering releasing each of the UK Federation's standard attributes.

Questions	**Relates to**
Does the service grant access based on organisation membership?	eduPersonScopedAffiliation (relevant values only)
Does the service provide personalised services (e.g. save settings)? Can users refuse to provide personal information? Is the use of information restricted (e.g. by a federation agreement)?	eduPersonTargetedID
Does the service need to link users to other services? Does the benefit of using the service justify the risk?	eduPersonPrincipalName
Has the service defined its own entitlement value? Does the service use a standard entitlement value?	eduPersonEntitlement (relevant value(s) only)

If that value affects users' privacy, have they agreed?	

3.5 Using Attributes

This section gives some examples of how attributes should, and should not, be used in order to protect privacy and satisfy the requirements of the *Data Protection Act 1998*. Note that these do not constitute legal advice on compliance, but merely highlight areas that may require consideration.

3.5.1 Personalisation

Where a Service Provider wishes to personalise the service they offer to each user this should be done using the eduPersonTargetedID attribute, since this provides the required ability to recognise a returning user and recover their stored preferences or other information.

In some cases it may be appropriate for the Service Provider to request additional personal information in order to provide an enhanced service, for example to send e-mail notices of upgrades to the service or information provided, or to greet the user by name or nickname. Note that, as discussed in section 2.2.3 above, doing so is likely to make the eduPersonTargetedID value into personal data so the Service Provider and User Organisation will need to ensure that this risk is acceptable. As discussed in section 2.2.2, the appropriate way to obtain this information – whether from the Identity Provider or the User – will depend on how critical the information is to the provision of the service. Note that in some circumstances the law may allow, or require, that a responsible adult provide information on behalf of a child or other person who cannot themselves give informed consent to its use.

3.5.2 Attribute Sharing

Unless a user has given specific consent, Service Providers must only use attributes obtained from Identity Providers (whether or not they contain personally identifiable data) for the service and purpose for which they were obtained. In particular individual Service Providers must not, without the user's consent, combine information about individual users across different services, and must not share information about individual users with other Service Providers.

3.5.3 Identifying Real World Individuals

For a few types of service, for example a project discussion list, there is a requirement to grant access to a particular real-world individual. It is impossible to establish this type of relationship using only the federated authentication system: the Service Provider cannot link the online identity j.smith with the particular John Smith who is to be granted access. To establish the link the individual and the service manager must exchange a secret that can be used to associate the individual with their online identity.

SWITCH's Group Management Tool[11] demonstrates one way this can be done. The service manager sends the desired group member a unique token by e-mail; the invitee then visits the service registration page, logs in using their home Identity Provider and presents the token. The IdP only needs to release a persistent eduPersonTargeted ID value because the token can be used to link the individual to the new account and authorise their access to the service.

Alternatively the sequence of two steps may be reversed. The service uses the persistent eduPersonTargetedID attribute to establish an anonymous account for a new user, then provides that user with a unique secret. The real-world person can then prove his ownership of the anonymous account by contacting the service manage, for example by telephone, to confirm his identity and knowledge of the secret. The service manager then authorises that account to access the service.

Both approaches protect the privacy of users because no personal information is disclosed until a user actively chooses to join the service. Users who visit the site by mistake or to read documentation cannot be identified.

4. Logfiles

4.1 General

Federation members are expected to keep records of use of their services and by their users. Logfiles may be needed, for example, to identify or trace faults or misuse, to account for use of services or to inform future planning. The same privacy principles apply to logfiles as to other personal data: processes should be designed to ensure that there is no more processing or disclosure of personal information than is strictly necessary and de-personalised information should be used wherever possible.

4.2 Collecting Logfiles

Service Providers may retain logs of the resources used in each user session. If these logs need to be associated with an individual user this should be done by recording the identifier associated with the subject of the Shibboleth SAML (Security Assertion Markup Language) assertion, not any other information purporting to identify the user. Clear information should be provided to users or their responsible adults describing what logs are kept, the purpose(s) they will be used for and the period for which these logs will be retained. Logs must be deleted when they are no longer required for the declared purpose(s).

Identity Providers should retain logs of the authentication decisions they make, linking the subject of the Shibboleth SAML assertion to the local identity that

[11] http://www.switch.ch/aai/support/tools/gmt.html

was authenticated. As above, users must be informed that this personal data is being recorded, the purpose(s) for which it will be used and the period for which the logs will be retained. For fault-finding and tracing misuse logs should be kept for a minimum of three months and a maximum of six; accounting and other purposes may justify longer retention but consideration should be given to removing personal data from the logs if there is no need to account for activity of individual users. Logs must be deleted when they are no longer needed for the declared purpose(s).

4.3 Using Logfiles

Processing of logfiles should be limited to what is strictly necessary. Tracking the particular resources accessed by an individual user is a serious breach of privacy and should only be done when it is necessary to avoid a serious risk of harm.

Individual Service Provider agreements will determine what level of accounting is required for each service. The Service Provider's logs should be sufficient for them to account for use by each subscribing organisation; if more detailed accounting is required then this should be done by the Identity Provider and User Organisation using information provided by the Service Provider. For example the Raptor software suite can be used to generate aggregated statistics.[12] Accounting information for individual users may only be generated if the users have been informed that this will be done. Otherwise accounting records must be de-personalised or aggregated to group together a class or other organisational unit.

Where misuse is suspected, Service Providers should pass relevant sections of logfiles to the User Organisation involved. The User Organisation should then work with its Identity Provider to identify the individuals responsible and ensure that the complaint is dealt with appropriately.

4.4. Disclosing Logfiles

Logfiles containing personal data must not be disclosed to others except with the permission of the individuals concerned or when required by law. In particular Identity Providers must not disclose the identity of individual users to Service Providers or other third parties. Where it is necessary to combine logfiles this should always be done by the User Organisation or Identity Provider to ensure that the privacy of users is protected.

[12] http://iam.cf.ac.uk/trac/RAPTOR

Copyright:
This document is copyright Jisc Collections and Janet Ltd trading as Janet(UK). Parts of it, as appropriate, may be freely copied and incorporated unaltered into another document unless produced for commercial gain, subject to the source being appropriately acknowledged and the copyright preserved. The reproduction of logos without permission is expressly forbidden. Permission should be sought from the Janet Service Desk.

Disclaimer:
The information contained herein is believed to be correct at the time of issue, but no liability can be accepted for any inaccuracies.

The reader is reminded that changes may have taken place since issue, particularly in rapidly changing areas such as internet addressing, and consequently URLs and e-mail addresses should be used with caution.

Jisc Collections and Janet Ltd cannot accept any responsibility for any loss or damage resulting from the use of the material contained herein.

© Jisc Collections and Janet Ltd 2006-14