UK Access Management Federation for Education and Research

Recommendations for Use of Personal Data

6 October 2006

V0.13
ND/ATG/AAI/DOC/023

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 Shibboleth architecture chosen for the federation is designed to protect user privacy; however, the measures it provides can only be effective if they are used and respected by 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 not normally be disclosed outside this relationship. Service providers should normally only need to know that an individual is a recognised user, having a particular status, at a member organisation. Where service providers need to obtain personal data from individual users they should do so directly, after informing the user what the data will be used for and what benefit the user will receive. Service providers should endeavour to provide service, possibly at a reduced level, to users who do not provide their personal data.

This guide explains the various privacy systems available in the federation and how they should be used to protect the interests of users. The guide covers first the requirements on federation members, then specific issues relating to the two main areas likely to contain personal data: attributes and logfiles.

2. Federation Requirements

2.1 Policy

All members of the federation are required to abide by the Federation Policy. A condition of the Policy, 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 federation.

2.2 Legal

Activities of the federation, whether performed by members or the federation operator, are subject to the Data Protection Act 1998. This requires that any personal data (defined 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.

The main requirements these principles impose on identity providers and service providers may by summarised as follows.

2.2.1 Requirements for Identity Providers

Identity providers will need to process personal data about individual users. All such processing must be fair, lawful, necessary and proportionate to the purpose(s) for which the data is required.

- identity providers 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 the UK Access Management Federation for Education and Research (this is known as a fair processing notice); so far as is possible, users must be allowed to refuse permission for their personal data to be processed;
- identity providers must only release personal data to service providers where this is necessary and where they can ensure that the data will not be misused. In most cases, and particularly where the service provider is located outside the European Economic Area, this will require a contract between the identity provider and the service provider (guidance on appropriate contracts is available from the Information Commissioner's website). Individual users may be asked for positive consent to their personal data being disclosed, but this must be given freely and after the user has been fully informed of how their information will be used. Wherever possible, only non-personal data should be released to service providers;
- 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, even though this imposes considerable additional legal responsibilities on the service provider. Personal data may be obtained from the user's identity provider or directly from the user.

- Where information is obtained directly from the user, the user must be informed at the time of collection of the purpose(s) for which the data is required;
- service providers must only use personal data for purposes that have been notified to the identity provider 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);
- service providers must use appropriate technical and organisational measures to protect personal data in their keeping.

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, and whether they are authorised by the institution to access a particular resource or, in some cases, what subject or class they are studying. The Shibboleth architecture communicates 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 release only these attributes. Requesting or disclosing attributes that contain personal data imposes significant additional legal burdens on both service provider and identity provider so should be regarded as highly unusual and only done following careful investigation of the objective requirement.

The 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. Since these attributes and their values have agreed definitions across the whole federation, Identity and service providers who use them can be confident that they mean the same thing to both parties. The implications of these attributes for personal data privacy are described in this section; the use of these attributes is 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 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’). 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 represents the least intrusion into the user’s privacy and is likely to be sufficient for many access control decisions. identity providers should support eduPersonScopedAffiliation for those they authenticate, releasing the least intrusive value that is required in the particular circumstances; 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 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 to satisfy 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 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.

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 cannot be disclosed to others. This may happen, for example, if the value of a persistent identifier such as eduPersonTargettedID is reused by the identity provider, allowing the new holder of the ID to access the previous holder’s stored information. identity providers that assert they can account for individual users are required not to reissue a persistent identifier value to a new user within two years; other identity providers are not required to give any such guarantee.

For most applications a combination of the attributes eduPersonScopedAffiliation and eduPersonTargetedID will be sufficient. A requirement to provide other attributes should be regarded as exceptional by both Identity and service providers and will involve considerable additional responsibilities for both.

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 will involve significant privacy issues. If a login name is used 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 must be assumed that eduPersonPrincipalName constitutes 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, imposing significant operational and legal burdens on both identity provider and service provider. 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 information stored in association with the eduPersonPrincipalName.

A requirement to supply eduPersonPrincipalName or other personally identifiable attributes should therefore be regarded as exceptional.

As with eduPersonTargettedID above, it is the service provider's responsibility to ensure that information stored in association with an eduPersonTargettedID is not disclosed or otherwise misused.

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 set 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 URI (Universal Resource Identifier)) 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 assigning an eduPersonEntitlement value to an individual, therefore, the organisation must consider whether it is necessary to obtain the individual's consent. In any case, eduPersonEntitlement values must only be released to service providers where they are necessary and relevant to the service provider's access terms.

3.3 Attribute Release Policies

Different service providers will require different Attributes in order to provide service to users. A service provider whose resources are licenced to all members of the organisation needs to know only the 'member' value of eduPersonScopedAffiliation and, perhaps, an eduPersonTargetedID to allow the user to store preferences. Disclosing unnecessary attributes to service providers could breach the privacy of users and must therefore be prevented.

Every identity provider should maintain an Attribute Release Policy (ARP) listing which attributes and values may be released to which service provider: this Policy should explicitly list those attributes and attribute values that can be released and block release of all others. No attributes should be released to service providers that are not in the ARP. The attribute release settings for each service provider should be a matter of negotiation: service providers are recommended to publish which attributes they require so that identity providers who wish to allow their users to access the service can quickly identify the correct ARP configuration. However, identity providers should be cautious about releasing attributes and should challenge any service provider requirement that is not plainly necessary. The identity provider is responsible for protecting the privacy of its users.

Some identity providers may give their users the ability to change their personal ARP settings from the default provided by the identity 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, while by permitting the release of additional attributes the user may be exposing their personal information unnecessarily.

3.4 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.

3.4.1 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. This information may be obtained in two different ways:

- Where the user is able to give both reliable information and informed consent to its use, the information should be obtained directly from the user after displaying the required fair processing notice. This also allows the user to select the most appropriate information for the particular situation.
- If the information or consent cannot be reliably obtained from the user then it may be possible for the identity provider to disclose it to the service provider. This should only be done where the parties have a contract that states how the information will be used and protected, and where the user has been informed that the information will be disclosed. In some circumstances the law may allow consent to be given by a responsible adult on behalf of a child or other person who cannot themselves give informed consent.

3.4.2 Attribute Sharing

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 combine information about individual users across different services, and must not share information about individual users with other service providers.

3.4.3 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 through solely online methods: the service provider cannot link the online identity j.smith with the particular John Smith who is to be granted access. To make this link the individual (or their identity provider) and the service provider must exchange

offline a secret that can be used to associate the individual with their online identity.

One possibility is for the individual to tell the service provider the username that will appear as their eduPersonPrincipalName. However, this requires the identity provider to disclose personal information about a particular individual to the service provider, while not disclosing the same personal information about users who do not use the service. This therefore requires a per-user Attribute Release Policy, either managed by the identity provider or made available as a service to its users.

It may be simpler for this type of service to use the persistent eduPersonTargettedID attribute to establish an anonymous account for a new user, then provide this user with a unique secret which the real-world person can use to prove his ownership of the anonymous account, for example by telephoning the service provider to confirm his identity and knowledge of the secret. This is similar in effect to the two-stage sign-up process used by many mailing lists and allows the individual to retain control of their personal data.

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 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 then they should record the identifier associated with the subject of the Shibboleth SAML assertion, not any other information purporting to identify the user. Users should be informed 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 was authenticated. Users must be informed that this personal data is being recorded, and of 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 seldom be necessary.

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 using information provided by the service provider. identity providers may only generate accounting information for individual users 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 identity provider involved. The identity provider should then 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 identity provider to ensure that the privacy of users is protected.