

INFLIBNET Access Management Federation

Recommendations for use of Personal Data

November 2011

Version 1.0

INFED: Recommendations for use of Personal Data

1.0 Introduction

The INFLIBNET Access Management Federation (INFED) 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 INFED is designed to protect user privacy; 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. This involves no personal data being disclosed. Where Service Providers need to obtain additional personal data about individual users they may either request it from the appropriate User Organisation (this will usually need to be covered by a legal agreement), or ask the individual user to provide it, seeking free and informed consent by 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 for whom personal data is not available.

This guide explains the various privacy systems available in the INFED and how they can be used to protect the interests of users. The guide first covers the general requirements on INFED members and then specific issues relating to the two main areas likely to contain 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.

2.0 Federation Requirements

2.1. Rules of Membership

All members of the INFED are required to abide by the Rules of Membership available at http://parichay.inflibnet.ac.in/documents/rules_for_membership.pdf . A condition of the Rules, and therefore of membership, is that members abide by the eight Data Protection Principles and described in the following section. A breach of these principles may therefore be both a breach of law and grounds for exclusion from the INFED.

2.2. Legal

Activities of the INFED, whether performed by members or the federation operator which requires that any personal data (i.e. 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 as follows

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 and only for the purposes for which data has been procured;
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 India 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 are summarised 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. 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 those collecting or using such information, not the INFED, and is likely to be ensured by appropriate procedures and contractual arrangements.

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. In some cases a User Organisation will wish its Identity Provider system to be run by a different organisation. Since this will involve the transfer of personal data between the two organisations there should be a detailed agreement between them on how the information will be used and how it will be protected. The following bullet points suggest one possible way that the agreement might assign responsibilities between the parties.

Where the User Organisation runs its own Identity Provider the single organisation bears all the duties and responsibilities.

- 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 the INFLIBNET Access Management Federation (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 collected, processed or disclosed;
- Identity Providers must only release personal data to Service Providers where the User organisation has determined that 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 India, this will require a contract between the User Organisation and the Service Provider that also describes what information the Identity Provider will release. 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 and understood how their information will be used. Wherever possible, only non-personal data should be released to Service Providers;
- 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, even though this imposes considerable additional legal responsibilities on the Service Provider. Personal data may be obtained either from the user's Identity Provider or directly from the user. Personal data whose accuracy is essential for the provision of the service should normally be obtained from the user's Identity Provider as this is likely to raise fewer compliance issues and have a third-party guarantee of accuracy. Personal data that is optional for the service (for example to personalise a home page) may be obtained from the user, since they can choose not to provide it without affecting the delivery of the service.

- 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); note 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 information is requested directly from the user then the user must be informed before the information is collected of the purpose(s) for which the data is required, and must have the option to refuse to provide it.

2.2.3. What is Personal Data?

“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 email 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. However, the status of 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 federation Rules must be able to link the use of a federated service to an individual user; however it is possible that a Service Provider receiving such an identifier may be able to treat it as nonpersonal data.

The non-personal status of identifiers in the hands of Service Providers may be achieved if the following conditions are met:

1. The identifier is generated and used in a way that effectively conceals any information about the user;
2. The Identity Provider undertakes not to disclose any information to the Service Provider that might help them to identify the user;
3. The Service Provider undertakes not to obtain or use any additional information (whether obtained from the Identity Provider, the user, or elsewhere) likely to identify the user.

It is clear that all of these measures help to preserve the privacy of the individual user, they can be seen as reducing (or even eliminating) the likelihood of additional linking information coming into the possession of the Service Provider. The federation’s Rules and recommendations support these three conditions. For example, pseudonymous identifiers should be generated using one-way hash functions, and any investigation of misuse should be carried out by the Identity Provider, not the Service Provider (see section 4 of this document). These Recommendations are therefore written on the basis that privacy preserving pseudonymous identifiers (and in particular the eduPersonTargetedID attribute) can, if used properly, be treated by Service Providers as non-personal data.

3.0 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 “Jagdish Arora”; 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. 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 Jagdish Arora” 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 INFED 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

Identity and Service Providers who use following attributes 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 INFED specific interpretation of the eduPerson controlled vocabulary for this attribute is described in the Technical Recommendations for Participants. This 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 eduPersonTargetedID is 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 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. This means that

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. Where this is required, the user organisation must ensure that users are notified and that the personal data is protected both by themselves and the Service Provider.

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 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 nondisclosure 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": 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. 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 the user to store preferences; a Service Provider with more fine grained authentication or logging requirements may require more attributes. 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 list those attributes and attribute values that can be released and block release of all others. Attributes should only be released if this is permitted by a rule 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 User Organisations and Identity Providers who wish to allow their users to access the service can quickly identify the correct ARP configuration. However User Organisations should be cautious about releasing attributes and challenge any Service Provider requirement that is not plainly necessary. The User Organisation is ultimately 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 . 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. 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 Identity Provider will need to ensure that this is covered by their legal agreement. 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

3.4.2. 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.

3.4.3. 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.arora with the particular Jagdish Arora 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 wish to use the service but may visit it accidentally. 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 eduPersonTargetedID 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

4.0 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 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. 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.

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. The INFLIBNET Centre cannot accept any responsibility for any loss or damage resulting from the use of the material contained herein.

Acknowledgement:

This document has been adapted from “Recommendations for use of personal data” a copyright document of The JISC Content Procurement Company Limited trading as JISC Collections available at <http://www.ukfederation.org.uk/library/uploads/Documents/recommendations-for-use-of-personal-data.pdf> (Last accessed on 15th November 2011) and modified to suit the requirement of the INFLIBNET Access Management Federation (INFED).