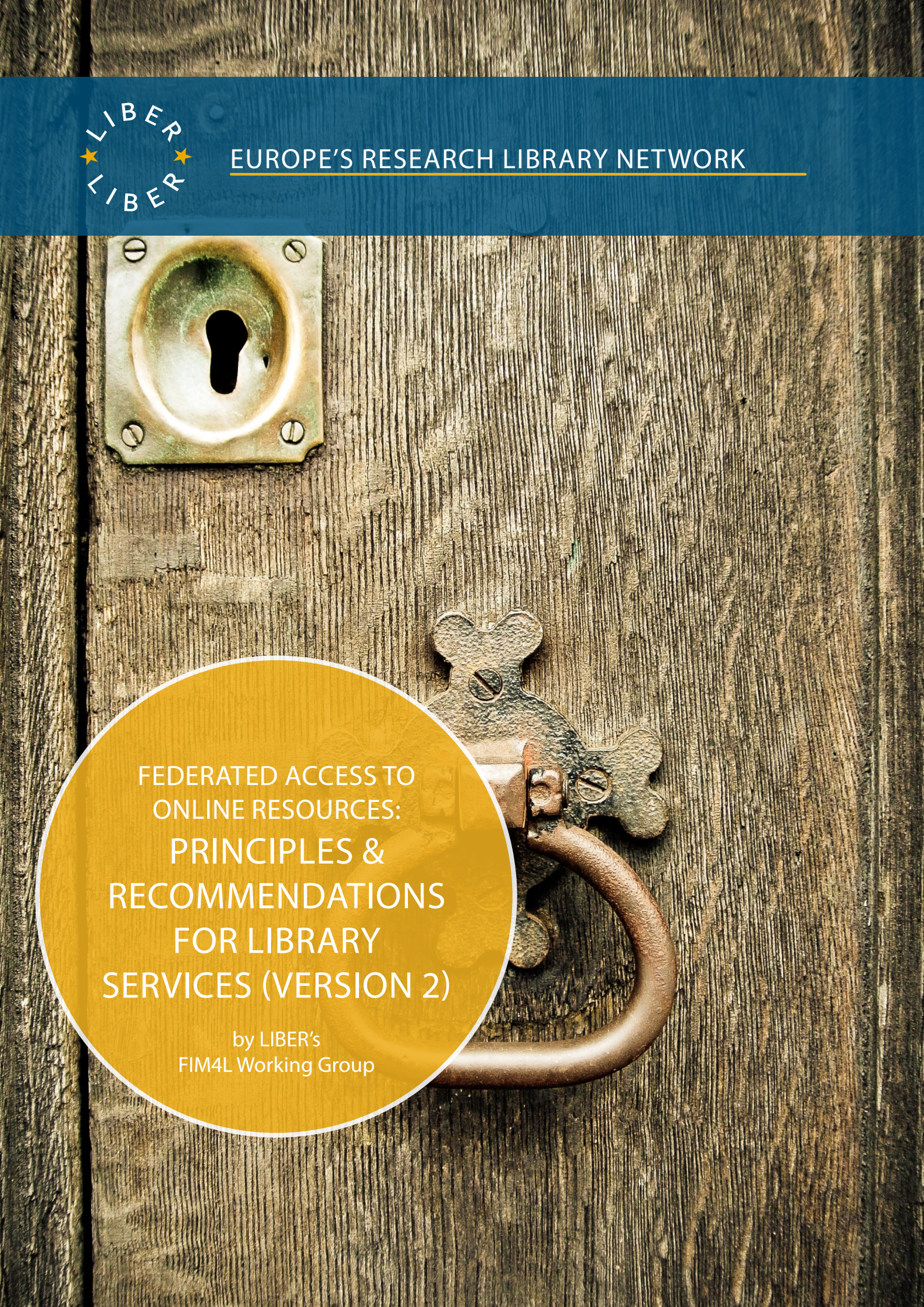
LIBER
LIBER
EUROPE'S RESEARCH LIBRARY NETWORK
FEDERATED ACCESS TO ONLINE RESOURCES:
PRINCIPLES & RECOMMENDATIONS FOR LIBRARY SERVICES (VERSION 2)
by LIBER's FIM4L Working Group

Colophon

ABOUT LIBER

LIBER *(Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries)* is the main network for research libraries in Europe. Founded in 1971, LIBER has grown steadily to include some 400 national, university and special libraries.

Together we work to represent the interests of European research libraries, their universities and their researchers.

HOW TO CITE THIS DOCUMENT

Westerbeke, J. & Gietz, P. & Pavlik, J. (Eds.) Federated Access to Online Resources: Principles & Recommendations for Library Services (Version 2) by LIBER's FIM4L Working Group (2022).

Version 2.0

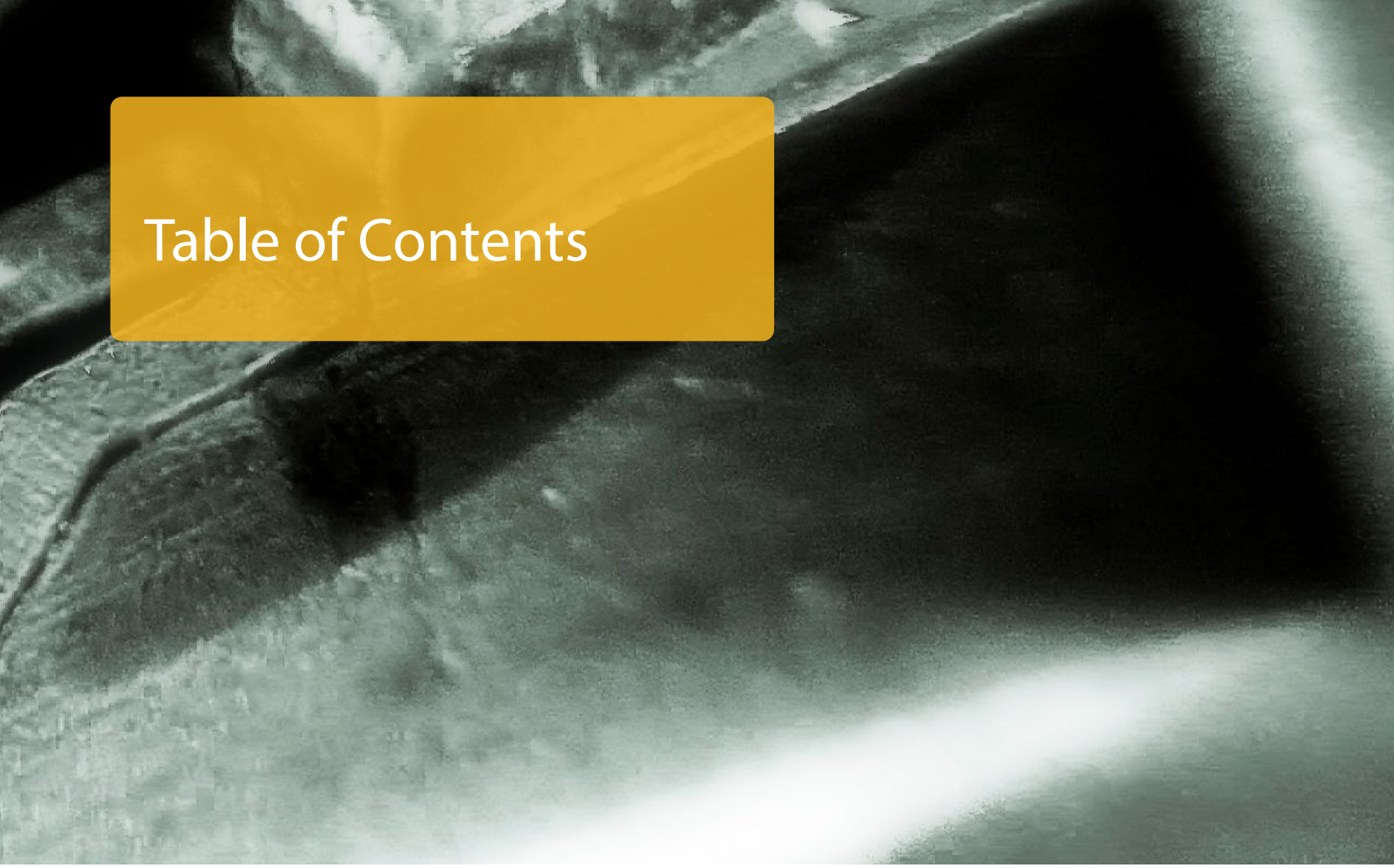

Table of Contents

INTRODUCTION

<< table of contents

Publishers and suppliers of licensed online resources want to provide authorised users of institutions for higher education and research with access to their services in a controlled way. The commonly used access method based on IP address has limits when users want access from anywhere and any device at any time. Solutions based on federated authentication, also called Single Sign-On (SSO), are viable alternatives, as long as attention is paid to how these connections are configured. Libraries should protect the privacy of their users, who, in turn, should have control over their privacy.

To make the configuration and management of federated authentication easier for both libraries and publishers, scholarly libraries from around the world have agreed on the following guidelines to control access to services based on licensed content.

This document is a reference for libraries and publishers who want to set up an SSO connection. Principle 4 is the core principle for this action. The library has to choose whether it will implement principle 4.A. or 4.B. This reference is intended to benefit both libraries and publishers.

Notes:

These two terms below are helpful for understanding the content of this document.

- Publishers are Service Providers (SP)
- Institutions/libraries are Identity Providers (IdP).

Please refer to the table of Terms and Definitions further on in this document.

ABOUT THE FIM4L WORKING GROUP

<< table of contents

The Federated Identity Management for Libraries (FIM4L) is an international activity initiated and headed by libraries. The recommendations in this document have been defined by library representatives from around the world.

The FIM4L Working Group operates as part of LIBER's Strategic Direction on Research Infrastructure, one of the pillars of LIBER's 2018-2022 Strategy.

The group aims to develop a library policy for federated authentication that is broadly supported and implemented by libraries and publishers. The authors firmly believe it is important to protect the privacy of library users by keeping the handling of research library user information within the library administration.

Visit our website: FIM4L.org.
Or visit the LIBER FIM4L webpage.

SSO IMPLEMENTATION PRINCIPLES

<< table of contents

Principle 1: Legal Compliance

The configuration and solution have to be in line with data protection regulations, particularly the General Data Protection Regulation (EU GDPR).[1]

Principle 2: Protocol

For federated access, it is recommended to use the SAML 2.0 protocol to connect and control access.

Principle 3: Federation

eduGAIN has been established as a proper means to interfederate between identity federations, including InCommon and UKfederation and thus enables service providers to expand their user base greatly. FIM4L encourages publishers to make use of eduGAIN; SPs are recommended to have their attribute requirements up-to-date in the eduGAIN metadata.

Principle 4: Authentication and Attribute Exchange

There are two recommended options for federated authentication and attributes: Anonymous Access (4.A.) and Pseudonymous Access (4.B.). Both are defined by the degree of privacy control and a minimal disclosure policy. Anonymous Access is the most private.

If the purpose of the service is to recognise returning users so it can present personalised features such as saved searches, profile-based recommendations for reading articles, etc., then Pseudonymous Access (4.B.) is recommended for providing these options to users.

On the right, you will find an explanation of both types of privacy options.

Principle 4: Authentication & Attribution Exchange

ANONYMOUS ACCESS (4.A)

This Access Method Holds The Highest Level Of Privacy.

The publisher cannot recognise returning customers across visits. During a session, the user session could be identified by a transient identifier containing a unique alphanumeric string for a certain Service Provider (SP). A new transient identifier will be generated if the user logs in again. This allows for maximum privacy.

It doesn't allow the publisher to recognise a returning customer, which makes it impossible to know what resource is downloaded by the same user.

It doesn't allow the library to trace a user in case of misconduct. In exceptional cases, however, users could be identified if libraries (IdPs) have configured their systems to allow for a thorough investigation of log files and if libraries are willing to carry out this investigation.

When choosing 'Anonymous Access' it is recommended to support REFEDS Anonymous Authorisation entity category in the SP metadata.

PSEUDONYMOUS ACCESS (4.B)

Maintains a high level of privacy based on a pseudonym and makes personalisation possible.

The publisher requires a persistent but targeted identifier. A persistent identifier (ID) contains a unique alphanumeric string, like the anonymous one, that identifies the user for a specific SP, but persists over multiple sessions. The same ID is then used for the same user on every authentication. This is an option for services that need to recognise returning users for personalised features.

In SAML the samlPairwiseID is preferred over eduPersonTargetedID (deprecated) and SAML 2.0 persistent NameID.[2]

The IdP should release the samlPairwiseID attribute according to organisational privacy preferences or user consent.

When opting for a persistent ID, consider the following:

- A persistent ID allows the library (not the publisher) to translate the ID to a patron in case of misconduct.
- It is possible to lock down access for a particular user in case of misconduct.
- A persistent ID (samlPairwiseID) is sufficient for the SP to provide personalisation features. Sometimes an SP requests more information, like a name and email address. Adding personal information like name and mail to enrich the user profile should be optional (not mandatory) for the user. Libraries/institutions are advised not to transfer that information during authentication. Still, they have the SP offer the user a profile page in their service, where users provide consent and can voluntarily provide name, email or other information. Minimise the attribute set provided to the service by the IdP during the authentication flow.
- Before a service that receives a persistent identifier creates a profile for the user, the service should ask for user permission to store and process their personal data, for instance, via a button "personalise account" or at least be informed by a message on data privacy. [3] In no way should the permission request be mandatory or seemingly mandatory for the user; the user must be free to choose whether or not to have a personal profile.

When choosing 'Pseudonymous Access' it is recommended to support REFEDS Pseudonymous Authorisation entity category in the SP metadata.

Figure 1: Privacy preferences infographic

Principle 5: Attributes for authorisation, personalised access, and analytics

For both privacy preferences, the SP can require extra non-identifiable information. Suppose more information is needed to allow for billing, access control etc. In that case, identity providers can supply one or more of the following attributes:

- eduPersonEntitlement, with the specific value urn:mace:dir:entitlement:common-lib-terms
- eduPersonScopedAffiliation
- eduPersonEntitlement, with other values, representing group or role memberships in alignment with AARC Guidelines on expressing group membership and role information
- Organisation, via either eduPersonScopedAffiliation, eduPersonOrgDN or schacHomeOrganization, when necessary to confirm which organisation the authenticated user is from.
- eduPersonAnalyticsTag, which is recommended for statistical purposes once it is well defined. [4]

Any combination of extra attributes like these needs to be agreed upon between the SP and the federation or in bilateral agreements with the IdP.

If an SP needs to identify a user and requires PII attributes like email address from the IdP, then it is recommended to support the REFEDS Personalized Access entity category in the SP metadata. In this case, the user cannot access the resource and stay anonymous.

The SP should not store local passwords of library patrons for personalised access. It is recommended to offer 'account linking' to other IdPs. This provides account portability between institutions.

Principle 6: Consent

Apart from generally working according to the GDPR, when requesting information from users, for instance, on a profile page, publishers have to adhere to the most recent EU "Guidelines on Consent"[5] to make sure that free consent is given in compliance with the GDPR.

Although we recommend that the Service provider request personal data directly from the user, there might be some cases where the IdP should send such data to the Service Provider. If that is the case, we strongly advise to have a consent module implemented at the IdP side. A very flexible, user-friendly tool is CAR, which is centrally managed and can be used by other provisioning processes besides the IdP. CAR puts the user in control, so privacy-focused users can choose to release as little information as possible. In contrast, convenience-focused users can opt to release more information.

Principle 7: Data Processing Agreement

When providing PII to an SP, whether based on consent or not, a respective data processing agreement (DPA) may be needed.

Principle 8: GÉANT Data Protection Code of Conduct Compliance

Publishers are encouraged to declare compliance with the GÉANT Data Protection Code of Conduct version 2.0.

Principle 9: REFEDS Sirtfi framework Compliance

Publishers are encouraged to declare compliance with the assertions of the REFEDS Sirtfi framework (Research and Education FEDerations group, Security Incident Response Trust Framework for Federated Identity).

Principle 10: Seamless Access

Apart from the attributes, SPs are encouraged to follow the guidelines from the SeamlessAccess.org coalition (formerly RA21). 'Standard' or 'Advanced' flavour integration with Seamless Access Button and WAYF are recommended.

RISKS&CONCERNS

The privacy recommendations impact some risks, which we want to make explicit.

- **Deanonymization: If you provide a targeted ID, as recommended in Principle 4, Part B above, you have to be aware that other data already collected by the SP, could be linked to this ID.**
- **Apart from the fact that for GDPR pseudonymous IDs are PII (just like static IP addresses may constitute PII), normally, users would see a consent or information screen when accessing an SP for the first time and would see which attribute release policy the IdP has selected. If the SP wants to collect more information from the user, the SP needs to ask for consent via a registration form.**

TERMS AND DEFINITIONS

<< table of contents

AARC	Authentication and Authorization for Research Collaborations, Project funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreements 653965 and 730941. AARC was successful in establishing a Blueprint Architecture for the deployment of FIM technologies in research infrastructures, as well as in establishing guidelines on respective technical and policy matters.
Authentication	The process of verifying the identity of a user, process or device; the ability of a user to access an account, often, but by no means exclusively, by use of a username and password.
Authorization	The process of verifying (against a set of access controls) whether an account is authorised to access a given service or resource.
eduGAIN	eduGAIN enables the trustworthy exchange of identity information between federations without many bilateral agreements, reduces the costs of developing and operating services, improves the security and end-user experience of services, enables service providers to expand their user base greatly and enables identity providers to increase the number of services available to their users. Regarding costs of operating services, when a resource provider is updating its metadata, it is easier to send it to just one federation and then propagate it to eduGAIN instead of contacting many national federations separately. On the federation side, getting updated metadata from eduGAIN has no maintenance costs is undoubtedly an advantage. See AARC and eduGAIN: expanding access to online resources for students, teachers and researchers, How to reach global customers with Federated Identity Management and How to Join eduGAIN as Service Provider for more details.
eduPerson schema	Lightweight Directory Access Protocol (LDAP) schema is designed to include widely used person and organisational attributes in higher education. More info: https://wiki.refeds.org/display/STAN/eduPerson
Federated Authentication	The mechanism by which an identity provider, such as a home organization, indicates to one or more service providers that the user has been authenticated and may be authorised by the service provider to access relevant resources.

Federated Identity	A digital identity asserted by one system (an identity provider) that may be consumed by other systems (service providers) through federated authentication.
Federation	A federation is an association of organisations that agree to exchange information as appropriate about their users and resources to enable collaborations and transactions such as user authentication.
Identity Provider (IdP)	An organisation that manages digital identities and issues authentication assertions and potentially other attributes to Service Providers.
IP address-based Authorization	A method where an SP and a home organisation have agreed that every request from a range of network/Internet Protocol (IP) addresses associated with the home organisation should be authorised for the services provided by the SP.
REFEDS	Research and Education FEDerations group, see refeds.org. The REFEDS Research and Scholarship Entity Category (R&S) has been designed as a simple and scalable way for Identity Providers to release minimal amounts of required personal data to Service Providers serving the Research and Scholarship Community. Candidates for the Research and Scholarship (R&S) category are Service Providers that are operated to support research and scholarship interaction, collaboration or management, at least in part. Example Service Providers may include collaborative tools and services such as wikis, blogs, and project and grant management tools that require some personal information about users to work effectively. This entity category should not be used to access licensed online resources as described in the category definition. For more details, see REFEDS documentation.
Service Provider (SP)	An organisation that makes online resources available to users based partly on information, in particular authentication assertions, from IdPs.
Single Sign On (SSO)	The ability of a user to access multiple discrete systems or sets of resources with a single set of access credentials. The mechanism of Federated Authentication often achieves this.
Security Assertion Markup Language (SAML)	A standards-based approach to federated or single sign-on (SSO) authentication. Many interoperable open source and commercial implementations of SAML are available.

ENDNOTES

<< table of contents

[1.] Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016; cor. OJ L 127, 23.5.2018, https://gdpr-info.eu.

[2.] This is in line with the argumentation found in the wiki of the Austrian national research federation on eduPersonTargetedID, see https://wiki.univie.ac.at/display/federation/eduPersonTargetedID.

[3.] E.g., "By connecting to this service, I agree that the service provider stores my person-related data (ID, affiliation, entitlements sent by my IdP, my IP address sent by my client, and my actions on this platform). Only if I want to receive emails from the service or if I want to be addressed by my name I will add my email address and name, respectively. However, this is not needed for any other personalisation features like 'point me to the last document and its last page I read', 'my last searches', <include your personalisation feature here>, etc. Whenever I wish to do so, I may request to see and to have deleted all data stored about me."

[4.] Please note that this attribute is not available in many federations and IdPs, so if the SP would like to receive that attribute, it will take specific communication between SP and IdP and possibly the federation.

[5.] Guidelines on Consent under Regulation 2016/679, https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051.

[6.] https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language.


LIBER
LIBER
www.libereurope.eu