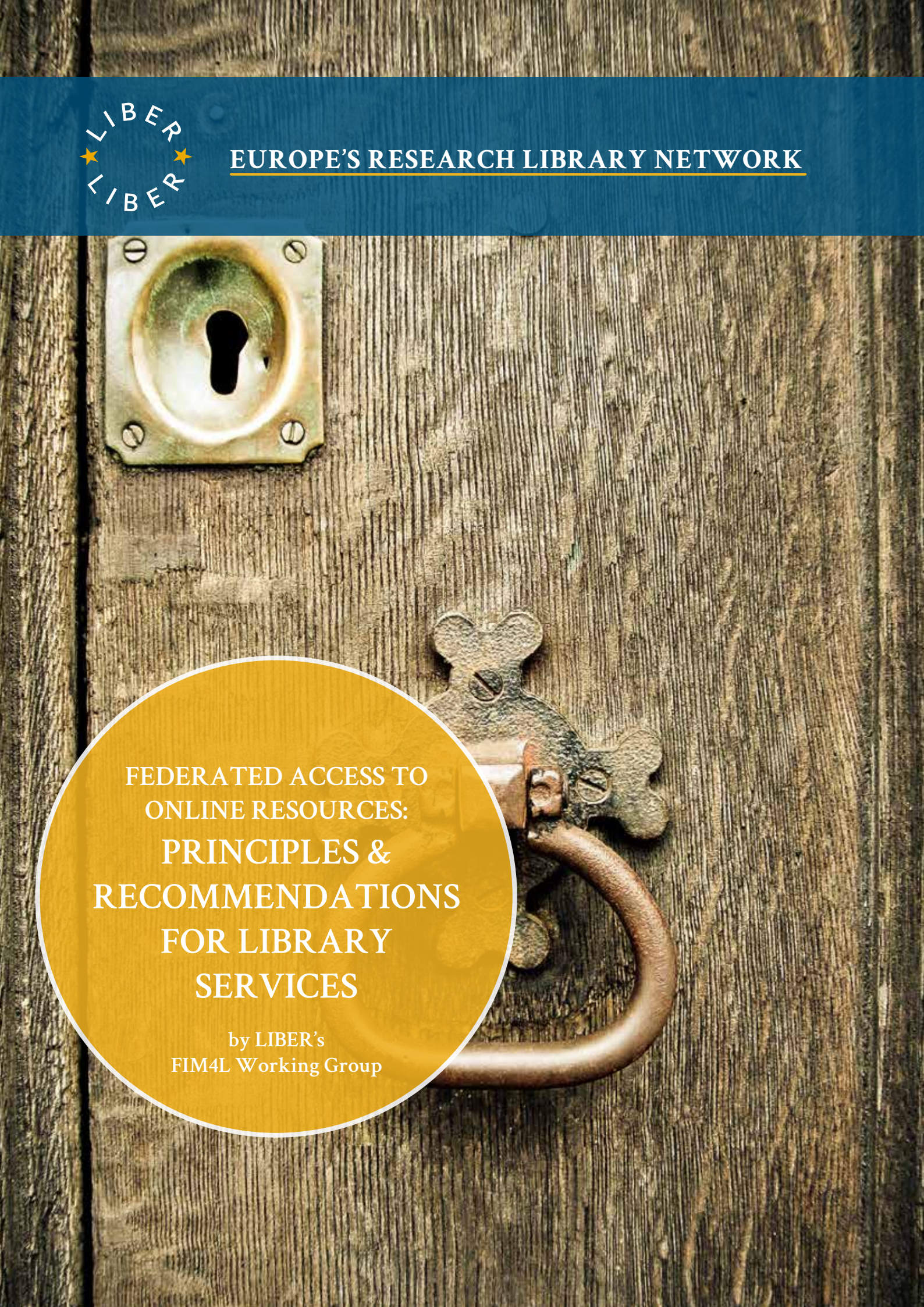
EUROPE'S RESEARCH LIBRARY NETWORK



FEDERATED ACCESS TO ONLINE RESOURCES: PRINCIPLES & RECOMMENDATIONS FOR LIBRARY SERVICES

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 450 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 by LIBER's FIM4L Working Group (2020).

Version 1.0

Table of Contents

INTRODUCTION

<< table of contents

Publishers and suppliers of licensed online resources want to provide authorized 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 and 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.

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

This document aims to function as 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 make a choice whether it will implement principle 4.A or 4.B. This reference is intended to be beneficial for both libraries and publishers.

NOTES:

These 2 terms below are helpful to understand 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 on pages 10&11.

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

SSO IMPLEMENTATION PRINCIPLES

<< table of contents

Principle 1: Legal Compliance

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

Principle 2: Protocol

For access to services based on licensed content, next to the option of access based on IP addresses, it is recommended to use the SAML 2.0 protocol (or its follow-up technology OIDC/OAuth2 if the involved IdPs are able to handle it) to connect and control access.

Principle 3: Federation

eduGAIN has been established as a proper means to interfederate between identity federations, and thus enables service providers to greatly expand their user base. FIM4L encourages publishers to make use of eduGAIN.

Principle 4: Authentication

There are two recommended options for authentication attributes, Transitory Access (4.A) and Personalized Access (4.B).

Both are defined by degree of privacy control. Transitory Access is the most private.

If the purpose of the service is to recognize returning users, so it can present personalized features such as saved searches, profile-based recommendations for reading articles, etc, then Personalized Access is recommended for providing these options to users.

Figure 1 on the following page explains both types of privacy options. Please refer to the table of Terms and Definitions on pages 10&11 if necessary.

Principle 4: Authentication

There are two recommended options for authentication attributes. If the purpose of the service is to recognize returning users, so it can present personalized features such as saved searches, profile-based recommendations for reading articles, etc, then Personalized Access is recommended for providing these options to users.

TRANSITORY ACCESS (4.A)

Holds the highest level of privacy.

The publisher only requires a transient identifier: "privacy star". During a session the user is identified by a transient identifier (NameID) containing a unique alphanumeric string, for a certain Service Provider (SP). If the user logs in again, a new transient identifier will be generated.

This allows for maximum privacy. It doesn't allow the publisher to recognize a returning customer, which makes it impossible to know what resource is downloaded by the same user.

In exceptional cases, for example where misconduct is suspected, 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.

PERSONALIZED ACCESS (4.B)

Maintains a high level of privacy based on a pseudonym, and more user information and tracking can be added.

The publisher requires a persistent but targeted identifier: "personalization and subject tracking possible". A persistent identifier (ID) contains a unique alphanumeric string, such as the transient one, identifying the user for a specific SP, but persisting over multiple sessions. The same ID is then used for the same user on every authentication.

This is an option for services that have a need to recognize returning customers.

It will give the user options for personalized features such as saved searches, profile-based recommendations for reading articles, etc.

For both privacy preferences, the service provider can require extra non-identifiable information.

Figure 1: Privacy preferences infographic

Here is an explanation of both types of privacy options.

4.A TRANSITORY ACCESS

The publisher only requires a transient identifier, “privay star.” During a session, the user is identified by a transient identifier (NameID), containing a unique alphanumeric string, for a certain Service Provider (SP). If the user logs in again, a new transient identifier will be generated. This allows for maximum privacy. It doesn’t allow the publisher to recognize a returning customer, which makes it impossible to know what resource is downloaded by the same user. In exceptional cases, for example where misconduct is suspected, 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.

4.B PERSONALIZED ACCESS

The publisher requires a persistent but targeted identifier: “personalization and subject tracking possible”. A persistent identifier (ID) contains a unique string, like the transient one, identifying the user for a specific SP, but persisting over multiple sessions: on every authentication, for the same user the same ID is used. This is an option for services that have a need to recognize returning customers, for instance so it can present your files, your orders etc. In SAML the Pairwise Subject Identifier is preferred over eduPersonTargetedID (deprecated) and SAML 2.0 persistent NameID. [2]

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 (like the Pairwise Subject Identifier, pairwise-id) is sufficient for the SP to provide personalization features. Sometimes an SP requests more information, like a name and email address. Adding personal information like Name and Email to enrich the user profile should be optional (not mandatory) for the user. Libraries/institutions are advised not to transfer that information during authentication, but 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. Minimize the attribute set provided to the service during the authentication flow.
- Before a service that receives a persistent identifier creates a profile for the user, the service should ask user permission to store and process his/her personal data, for instance via a button “personalize 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 whether or not have a personal profile.

For both privacy preferences, the SP can require extra non-identifiable information. If more information is needed to allow for billing, access control etc., identity providers can supply one or more of the following attributes (from most to least preferred):

- 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
- Usage of schacLocalReportingCode attribute is recommended for statistics purposes once it is well defined. [4]

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

Principle 5: Personal Identifiable Information (PII)

SPs should not require attributes with personal identifiable information (PII). Some publishers state "I need an email address, as my software can't function without it." Publishers with (older) systems that require more attributes for authentication to function should adapt their systems ASAP. Libraries are recommended to stop or not start using services that require more personally identifiable information (PII) than a transient or persistent ID during authentication.

Principle 6: Consent

Apart from generally working according to the GDPR, when requesting information from users, for instance in 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.

Principle 7: Data Processing Agreement

When providing PII to a SP, whether based on consent[6] or not, a respective data processing agreement (DPA) may be needed.

Principle 8: GÉANT Compliance

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

Principle 9: REFEDS 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

Publishers are encouraged to follow the guidelines from the SeamlessAccess.org coalition (formerly known as RA21).

RISKS AND CONCERNS

<< table of contents

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

- Deanonymization: If you provide a targeted ID, as recommended in Principle 4.B, Personalized Access, 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 (and even IP-addresses) are 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 try to collect more information from the user, the SP needs to ask consent via a registration form as recommended for the Personalized Access option.

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 Blue Print 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, use of a username and password.
Authorization	The process of verifying against a set of access controls whether an account is authorized to access a given service or resource.
eduGAIN	eduGAIN enables 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 greatly expand their user base 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 having to contact 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 designed to include widely-used person and organizational 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 of more service providers that the user has been authenticated and may be authorized by the service provider to access relevant resources.

Federated Identity	A digital identity which is asserted by one system (an identity provider) which may be consumed by other systems (service providers) by means of federated authentication.
Federation	A federation is an association of organizations that agree to exchange information as appropriate about their users and resources in order to enable collaborations and transactions such as user authentication.
Identity Provider (IdP)	An organization that manages digital identities and issues authentication assertions and potentially other attributes to Service Providers.
IP address-based Authorization	A method where a SP and a home organization have agreed that every request coming from a range of network/Internet Protocol (IP) addresses associated with the home organization should be authorized for the services provided by the SP. .
REFEDS R&S	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 for the purpose of supporting research and scholarship interaction, collaboration or management, at least in part. Example Service Providers may include collaborative tools and services such as wikis, blogs, project and grant management tools that require some personal information about users to work effectively. This entity category should not be used for access to licensed on-line resources as described in the category definition. For more details see REFEDS documentation.
Service Provider (SP)	An organization that makes online resources available to users based in part 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. This is often achieved by the mechanism of Federated Authentication.
Security Assertion Markup Language (SAML)[7]	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 this argumentation.

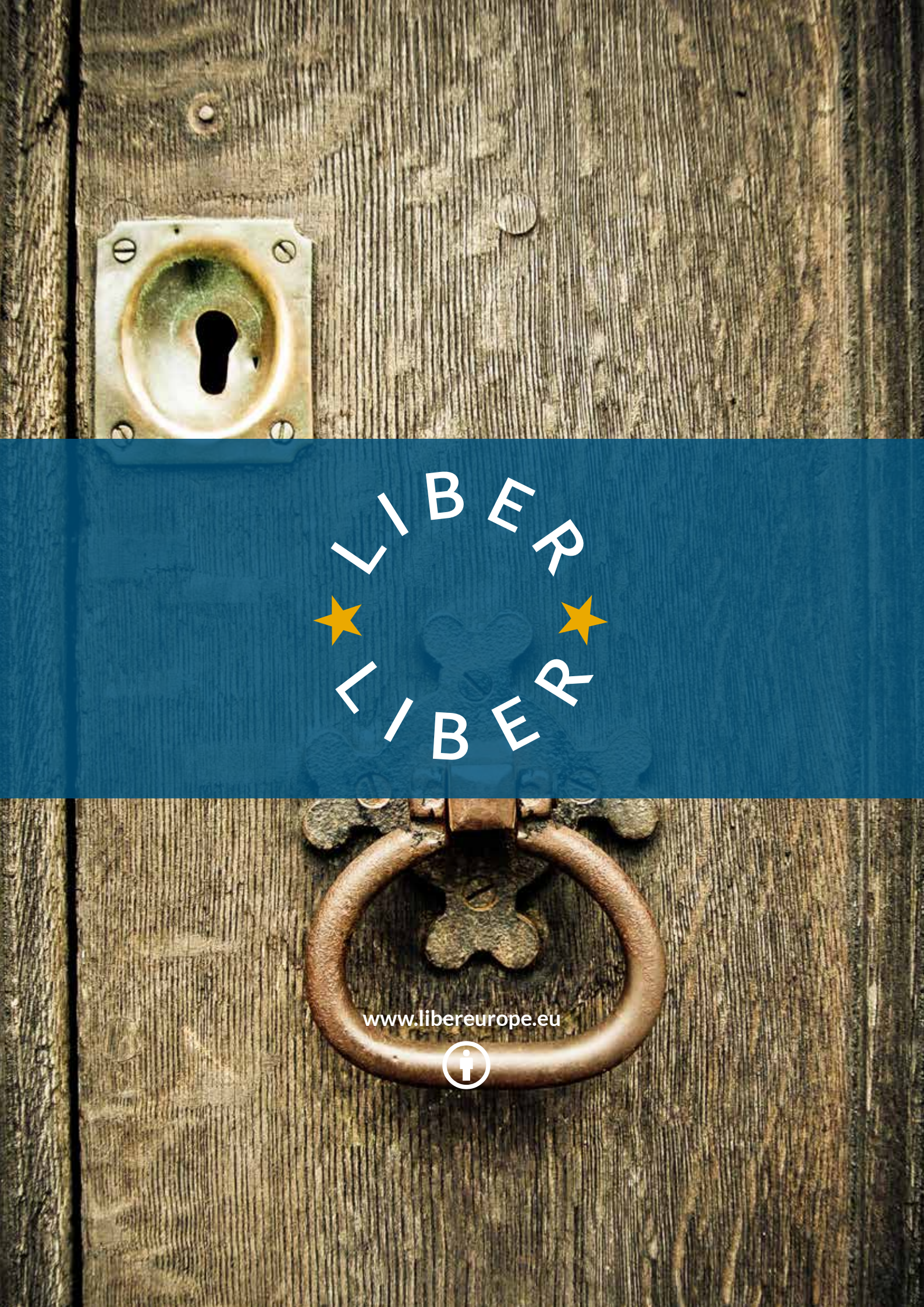
[3.] E.g., "By connecting to this service, I agree that the service provider stores my person related data (ID, affiliation, entitlements sent by my IdP, my IP address sent by my client, and my actions on this platform). Only if I want to receive emails from the service or if I want to be addressed by my name, I will add my email address and name respectively, but this is not needed for any other personalization features like 'point me to the last document and its last page I read', 'my last searches', <include your personalization 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.] We know of and are tracking the internet2 CAR-initiative about consent for optional release of attributes.

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


LIBER
LIBER
www.libereurope.eu