Schwerpunktinitiative „Digitale Information“

Recommendations on methods for controlling access to scientific information resources

A joint paper by Deutscher Bibliotheksverband e.V. (dbv) and the Digital Information Priority Initiative of the Alliance of Science Organisations in Germany, 29.11.2019

Preamble

Access to digital content is changing for scientists. In Germany, all science organisations and their infrastructure facilities are pursuing a common objective: establishing open access as the standard publication model for scientific information. This complies with the BMBF open access strategy, which specifies the establishment of open access as the standard for scientific publishing as a guiding principle. Free access without technical, legal or financial barriers guarantees the optimum usability of information and safeguards the rights of the scientists both in their role as authors and in the subsequent use of the published information.

However, subscription-based scientific information is still subject to access restrictions that are controlled via authentication and authorization procedures. Traditionally, the main form of access authorization is the IP-based procedure. The single sign-on procedure (e.g. Shibboleth) has also existed alongside this for years, but has not been used on a broad basis in the publishing environment till now. There have been further developments in access control in recent years by commercial providers, driven amongst other things by a desire for simplification, as well as the aim of facilitating personalized offers. This is accompanied by a fundamental risk that there will be more of a focus on collecting personal data, passing it on and using it to analyze the usage behaviour of scientists.

Free access to information and privacy protection are essential concepts for the libraries and information sector and fundamental prerequisites for exercising human rights to freedom of expression and intellectual freedom. The IFLA Code of Conduct therefore requires all information service providers to "take steps to limit the collection of personal information about users and the services they use"[1].

The information infrastructure facilities in Germany are also obliged to follow this principle. They therefore commit to the protection of their users' personal data. The way in which this data is used and saved must be transparent and with the permission of the users. The following recommendations can act as orientation in this sense and set out the framework for the development and introduction of new technologies for access to electronic information resources in Germany.

[1] International Federation of Library Associations Statement on Libraries and Intellectual Freedom, https://www.ifla.org/DE/publications/node/8767.

Case example: Project RA21 – Resource Access in the 21st Century [2]

"Resource Access for the 21st Century" (RA21) was a joint project by the International Association of Scientific, Technical and Medical Publishers and the National Information Standards Organization (NISO), with the fundamental aim of implementing SAML-based procedures (e.g. single sign-on) as the standard technology for access to electronic scientific information resources, thereby replacing the current widespread use of IP addresses for access control.

The aim of RA21 was to develop a standardized and simplified single sign-on procedure (SSO) to access electronic resources. This procedure involves authorization for the requesting users between the service provider and identity provider. First, the identity provider for the requesting party must be established via a service (discovery service[3]). A best-practice reference implementation for improved discovery services was developed as part of RA21 that can be used by all participating service providers. Once the respective identity provider has been selected, the information is saved in the browser so that a discovery service compatible with SeamlessAccess.org can read this automatically. This means that users no longer have to search again for the identity provider, which has often been perceived as inconvenient in the current SSO procedure in use.

The next step is to forward the requesting users to their identity provider (IdP) for authentication. The identity provider sends the service provider (SP) the data on the basis of which access is granted or refused (authorization). The identity provider decides which data is to be transferred to the service provider on a technical level. The infrastructure intended by RA21 is not part of the processing of this data.

Evaluation and recommendations

Web-based single sign-on solutions such as those developed in RA21 and implemented by SeamlessAccess.org are suitable for managing access to licensed, access-restricted contents and services. They allow anonymous or pseudonymous use of such resources and personalised individual user registration. Compared to IP-based access control, their main benefit is to make it easier for users to use services outside their facility. There are also administrative benefits; individual sessions can be better identified technically and user activities can be tracked on a granular level even with anonymous access, enabling you to track misuse, for example. Personalized services can also be based on these technical options.

These benefits are potentially countered by the risk that service providers will determine the transfer of personal information by the institutional identify provider above and beyond the required extent as a prerequisite for use as SSO develops further. From the point of view of information infrastructure facilities, determination using only SSO-based procedures is not desirable, as this could put facilities that do not use their own identity provider at a disadvantage. Basically, IP-based access control should always be offered as an alternative. When implementing web-based SSO procedures, it must be ensured that data protection can also be implemented to the full extent in terms of privacy by design.

[2] The RA21 project officially ended on 30 June 2019. At this point, the recommendations from the project were transferred to a beta phase service led by the Coalition for Seamless Access (www.seamlessaccess.org).

[3] "Discovery service" in this context means a technical service that researches and selects the identity provider, not the research system that uses the same name as standard.

This general evaluation leads to the following concrete recommendations:

1. Access to information resources must fundamentally be possible without passing on personal data. Facilities are called upon to inform users and take a critical look at their responsibility for technical framework conditions. Libraries should consult with their relevant IT infrastructure facilities in this regard (e.g. computer centres).
2. In general, only the data required for the respective service or content being used should be transferred to the service provider by the identity provider. Authorization for access to information resources in itself does not require any personalized attributes in the SSO procedure. The attributes that have already been in use for a long time (eduPersonEntitlement and/or eduPerson-ScopedAffiliation) are sufficient.
3. Where services require personalization, the aim should be national and international agreement on attribute approval that is compliant with data protection regulations. The regulation of the transfer of attributes should be transparent and specific to the service groups. Pseudonymous SP-specific identifiers or attributes should be used for personalization, so that user tracking across multiple service providers is no longer possible.
4. Technical recommendations should be generated for the configuration of contents when transmitting attributes for identity providers. Consideration should be given here to existing guidelines such as the GEANT Data Protection Code of Conduct or the recommendations of the DFN Legal Research Centre.
5. As the intermediary discovery service can also collect and process personal data for the selection of the identity provider, this service should be run by trustworthy and neutral facilities.
6. In terms of implementation under licensing law, determination only on single sign-on technologies should not be accepted for access control to contents and services. The possibility of using an established alternative procedure – such as IP control – must always be provided as an additional option.
7. Single sign-on technologies can simplify access to licensed contents. However, the open access transformation of the scientific publication market should primarily be pursued in order to ensure access to and scientifically adequate subsequent use of information resources while making the most of digital opportunities. Open access to scientific contents is the most suitable way to solve access problems.