FIM4L workshop

Managing Open Access with FIM

&

Educational material

Peter Gietz
Co-founder and CEO of DAASI International, Germany

Jos Westerbeke
Library IT specialist / manager at Erasmus University Rotterdam, Netherlands

More contact details at LIBER website

November 2023

Workshop agenda

- **Introduction to FIM4L and Federated Access (10 Min.)**
- **Presentation of current FIM4L work (15 Min.)**
- **Open Access, Open Science and FIM**
  - **Introduction (10 Min.)**
  - **Workshop (40 Min.)**
- **Educational material for library access to e-resources**
  - **Introduction (10 Min.)**
  - **Workshop (20 Min.)**

Introduction to FIM4L

- FIM4L (Federated Identity Management for Libraries) is initiated and headed by libraries. With the ending of the AARC project in which LIBER was involved, FIM4L started in 2019. (http://fim4l.org)
- Involved people from related organizations including, besides a number of research related libraries, AARC, GÉANT, REFEDS and NRENs.
- FIM4L is a LIBER working group since October 2019.
- The group also operates an international email list for communication and consultation.
- Published recommendations.
- Recommendations are endorsed by: LIBER, CARL, CAUL and RLUK.
- FIM4L cooperates with FIM4R, the research branch.

Federated Access to Online Resources: Principles & Recommendations for Library Services.

Published on Zenodo: https://zenodo.org/record/7313371

A libraries' "online collection" demands managing access:
- Identification,
- authentication and
- authorization

for licensed resources. But note that in other cases, like some Open Access providers, they want the user to sign in to get data for analytical purposes e.g.

Libraries need to manage their users (identities) and their privacy.

Libraries should always provide the ability to read articles and books anonymously.

Emerging role of libraries: Identity Management. Guiding users and preserving privacy by technical and contractual means.

Libraries could play an important role in Open Science managing access to online resources as a trusted partner. Note that Open Access is sometimes managed access to open materials.

"We protect each library user's right to privacy"
(ALA Code of Ethics)

"Library users shall have the right to personal privacy and anonymity. Librarians and other library staff shall not disclose the identity of users or the materials they use to a third party."
(IFLA Statement on Libraries and Intellectual Freedom)

Privacy supports freedom

Federated Identity Management for Libraries (FIM4L.org) - A LIBER working group

Authentication

- IP address-based for on-campus. For off-campus-use extended with
    - VPN (Virtual Private Network)
    - EZproxy (URL rewriting proxyserver)
- Increasing usage of federated access: Single Sign-On (SSO)

"State of the art"

Sign in, please.

Sign in via institution: federated SSO
- Not disclose any private information
- Only persistent pseudonymous identifier
- More data could be asked *optionally* by publisher

According to FIM4L.org recommendations.

Help library users protect themselves online

- LIBER Working Group "Federated Identity Management for Libraries (FIM4L)" published second and current version of the FIM4L Recommendations on Zenodo.org

- FIM4L cooperated with Seamless Access creating a Model License Agreement.

- Volunteers at FIM4L are there to help librarians.

- The recommendations are useful for both libraries and publishers.

LIBER
EUROPE'S RESEARCH LIBRARY NETWORK

FEDERATED ACCESS TO ONLINE RESOURCES:

PRINCIPLES & RECOMMENDATIONS FOR LIBRARY SERVICES (VERSION 2)

by LIBER's
FIM4L Working Group

FIM4L.org

LIBER

FIM4L

Published on Zenodo:
https://zenodo.org/record/7313371

What's in the recommendations:

https://zenodo.org/records/7313371

Principle 4: Authentication & Attributes 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.

Federated Identity Management

(c) RUBENKING, NEIL J.: Securing web services
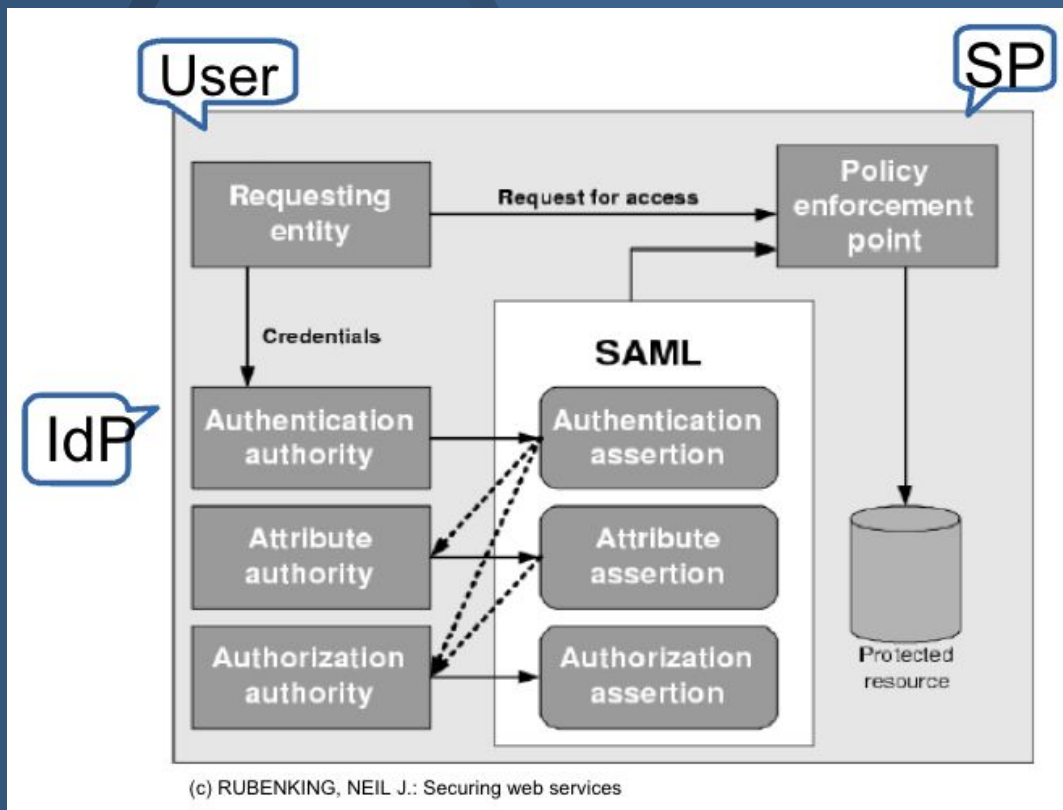
SAML: Security Assertion Markup Language
   SP: Service Provider
   IdP: Identity Provider

Attribute Assertions potentially contain privacy related information (name, email, etc.)

Working on Educational Material

Second part of the workshop

Library Access Guide

With privacy in mind

But let's start first with an Open Access journey

First part of the workshop:
Open Science and managing access

The goal of this session is "to receive feedback from the participants in order to create an informed strategy for the FIM4L working group."

Agenda:
- What are we speaking about: OA, Open Science, etc.
- Questions: consulting you.
- Creating input for strategy.
- What role can FIM4L WG play?

Finding new FIM4L WG strategy

- **Supporting community owned and governed publishing infrastructure.**
- **Supporting integration of ecosystems.**
- **Working towards recommendations from the library field for FIM, with special attention for using (SAML) attributes.**

What is Open Access?

A publication is defined 'open access' when there are no

- financial,
- legal
- or technical

barriers to accessing it. (**OpenAccess.nl**)

Open Access is the free, immediate, online availability of research articles coupled with the rights to use these articles fully in the digital environment. Open Access ensures that anyone can access and use these results. (**SPARC**)

The author(s) ... grant(s) to all users a free, irrevocable, worldwide, right of access ..." (**Berlin Declaration**)

Open Science Concepts

- **Make research outputs (and data sets) freely available**
- **Multiple policy goals, like**
  - **Equity**
  - **Accelerate discovery**
  - **Enable citizen science**
- **Taxpayer Euro's fund the research; taxpayers should be able to access it**

Drivers of Open Science

- **In the EU, key open access (OA) policy requirements instituted several years ago, now rolling out**
    - **Broad definition of science, including social sciences and humanities**
    - **Usual EU per nation variances and allowances**
- **In the US, per agency initiatives**
    - **Catalyzed by the Nelson memo**
    - **Each agency, sigh, is creating its own plan https://open.science.gov/**
    - **C Level orgs also embrace - https://sparcopen.org/**

Open Science infrastructure and authentication

- Diamond OA needs a publication infrastructure managed by the university library.
- Relevant infrastructure providers and projects:
  - EOSC
    - Federation of onboarded EU (data)services
    - Indexing and discovery of content and projects
  - OpenAIRE
    - Authentication & Authorization Infrastructure (AAI) - IdP federation
  - CRAFT-OA (2023-2025)
    - Creating a Robust Accessible Federated Technology for Open Access. Specifically for self-publishing/Diamond OA.

More European Initiatives on Federated Identity Management (FIM)

- **GÉANT**
  - The collaboration of European National Research and Education Networks (NRENs).
  - Provides Trust & Identity Services and runs the world-wide interfederation eduGAIN.
  - Member of the EOSC Association.
- **REFEDS**
  - The Research and Education FEDerations group which is to be the voice that articulates the mutual needs of research and education identity federations worldwide.
  - Creates specifications for e.g. attribute release.
- **AARC** (I, II and III)
  - Authentication and Authorisation for Research and Collaboration.
  - EU Projects researching the future of FIM in the context of research infrastructures.
- **FIM4R**
  - Research infrastructure integration of R&E federations.
- And more initiatives

Complex set of issues at national and campus levels

- Time and geo embargoes, e.g.:
  - licences content until 70 years until the death of an author
  - national license allows OA only from within the country
- Privacy of researchers and data subjects
- Existing economy of research publication
- Institutional infrastructure
- Measuring impacts

Managing Access before publishing as OA

Before something is finally published as Open Access, some use cases need access control.

- Control access while creating Open Access material (control who has access before publication)
- Infrastructure for discussing Open Access materials
  - You want to know who you are discussing with you
  - Authorship of reviewers
    - Reviews can be open access, but need to be authenticated by the authors

Publication infrastructure landscape of Open Source software.

(From: Mind The Gap, John Maxwell, 2019, p.15, pubpub.org)
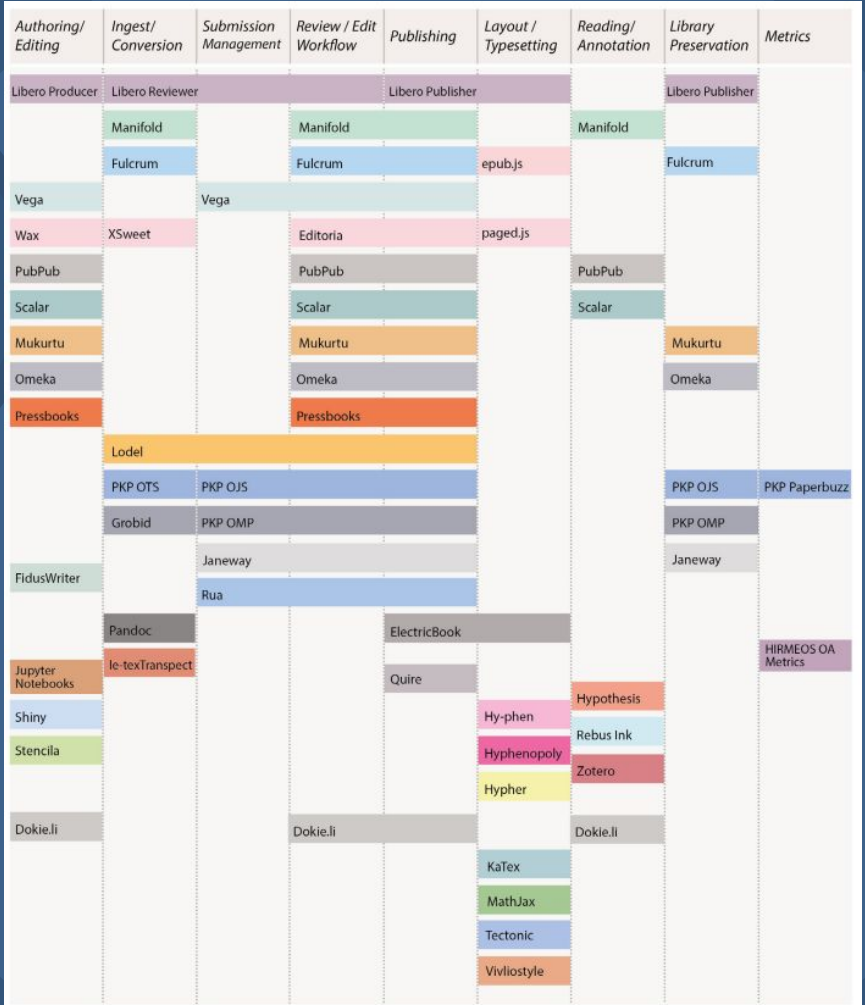
Figure 3: Software development across hypothetical workflow stages

Controlled/Managed access

When a publication is not open, as OA advocates, its access should be managed, when publishing to certain groups or persons is needed. Access then needs to be controlled, but should adhere to the FAIR principle "Accessible".

- Find use cases in this workshop, like:
  - Special collections
  - Datasets

Find use cases for:
The pre-publishing road to OA

1. What steps in the process are needed before something can be published as OA in a library repository or university press?
   ○ Author commitment, licence CC-BY etc. (And other non-technical requirements.)
   ○ Pre-print stage
   ○ Peer-to-peer reviews
   ○ Access restriction because of embargo's
   ○ ?

2. Who needs to take action and how?
   ○ ?

Find use cases for:
The pre-publishing road to OA

1.

Find use cases for:
Restricted (managed) access

1. What in the library collection **could not be published as OA?** (Which articles, journals, ebooks, datasets, etc.)
   - ?
2. Why?
   - Special collection restrictions
   - gated communities
   - Subject privacy
   - Geo location (Country restriction)
   - Community based content with access restrictions
   - Age
   - ?

Find use cases for:
Restricted (managed) access

1.

Discussion

Infrastructure for enabling discussions/reviews

- How would ideal systems look like to discuss OA materials
- Who should be in charge of running such systems
  - the libraries?
  - big research infrastructures
  - commercial providers such as researchgate, academia.edu
  - What role can ORCID play for authenticating authors?
  - Computing centers
  - …

Publishing infrastructure and metering

- How do we measure the usage of OA?
  - COUNTER alike metrics possible?
  - Controlled publication?

Discussion

Second part of the workshop:
Educational Material

Creating a guide

"Guiding library patrons (and librarians) with privacy-preserving access to online resources"

Agenda

- What do we have? (Showcase)
- What do you think we need?
- Putting ideas into action: Creating a guide

Why an access guide?

The library should be a safe place online:
- **Providing safe access**
- **Protecting its patrons' privacy**
- **Teaching and guiding patrons**

A trusted safe place with privacy

Library Access Guide

Workshop:

Creating a library access guide