

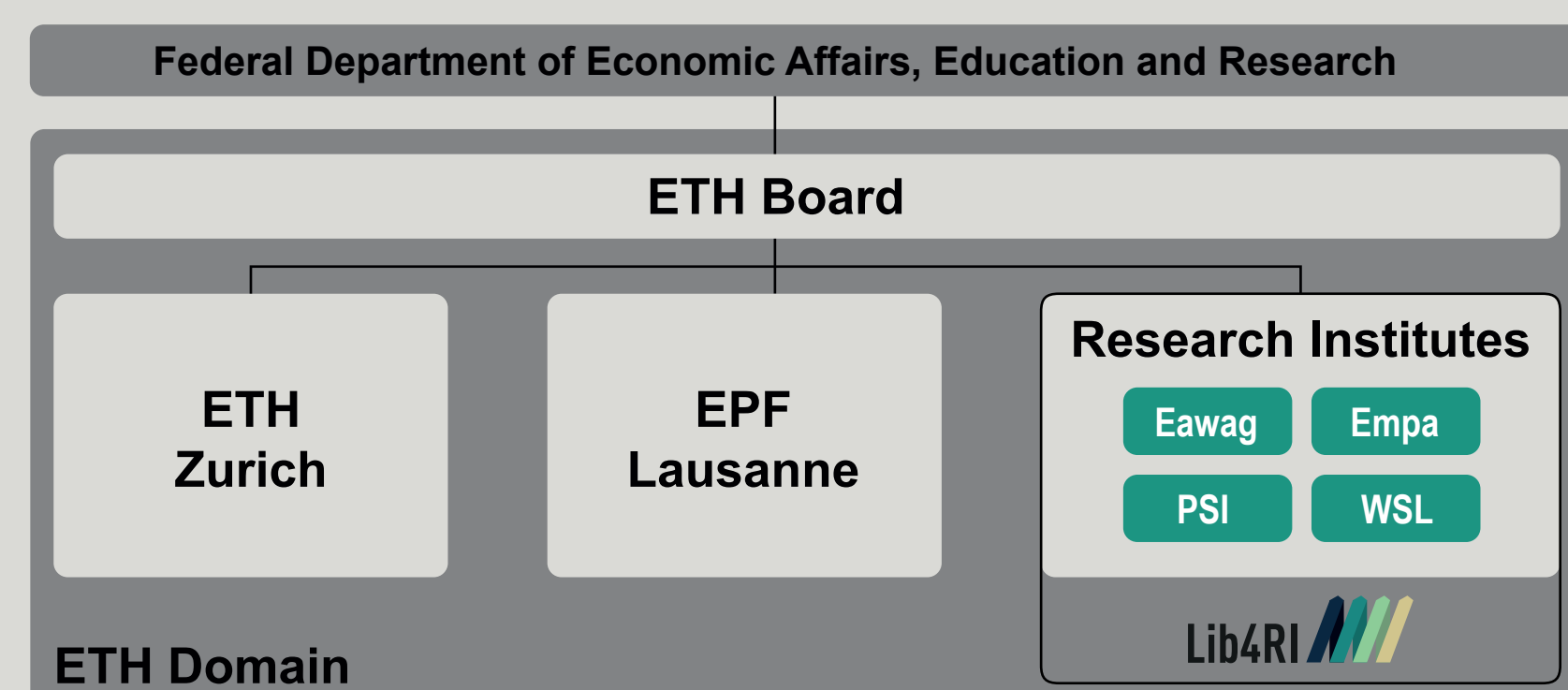
Highly Automated Import of Metadata into an Institutional Repository: A PHP Tool

Federico Cantini, Laura A. Konstantaki, Lothar Nunnenmacher

Lib4RI - Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL
Überlandstrasse 133, 8600 Dübendorf, Switzerland

Lib4RI

Lib4RI is the joint «Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL». As a scientific library, Lib4RI provides all affiliates of the four institutes with comprehensive access to information resources. After the shift from print to online resources in recent years and the increasing importance of Open Access, we are expecting services to play a much greater role for the library than collections in the next years. <http://www.lib4ri.ch/>



Introduction

DORA is the bibliography and institutional repository for all publications affiliated with the four Research Institutes, launched in 2016 and now containing 60.000+ items, of which 16.000+ are Open Access. It is based on Drupal/Islandora with a high degree of customization.

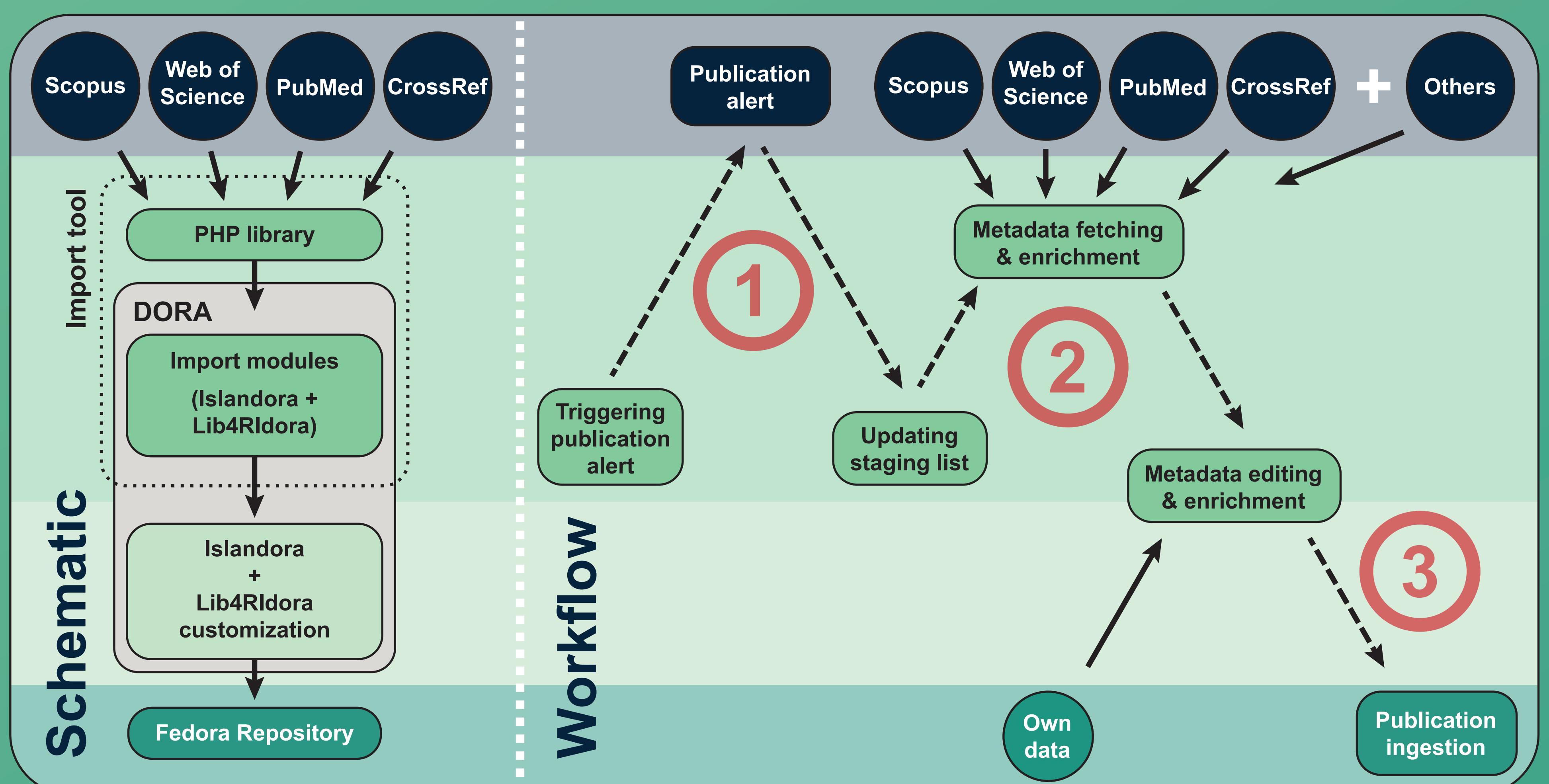
To relieve researchers of a time-consuming administrative task, the software imports publications directly from other databases instead of asking for manual submissions.

Import Tool

A pure PHP library abstracts the interaction with metadata providers allowing it to fetch metadata from several services. The PHP library also provides functionality for metadata transformation based on XSLT documents or call-back functions. This software library can be used in any system using PHP code to interact with metadata providers.

Two Drupal import modules use the PHP library to implement publication alerts, metadata fetching, integration, enrichment and ingestion. The first module implements functionalities common to a standard Islandora installation and can be used in any Islandora site. The second module implements functionalities that are specific for our Islandora customization and is therefore fit for use on our system only.

The PHP library and the two import modules will be available on GitHub soon. <https://github.com/lib4ri>



Publication Alert

Several services offer APIs for querying for publications (e.g. Scopus, PubMed, Web of Science, CrossRef). A restriction on authors affiliated with the relevant institution and for a relevant time period (e.g. the last two years) is possible. The results, containing only a few relevant IDs (as DOI and e.g. Scopus ID), are stored in a staging list and displayed to the library team. A duplicate check is done and publications already in our repository or in the staging list are discarded. The fetching of the IDs for the relevant publications can be triggered manually or scheduled to happen regularly. Publications that should not be ingested (i.e. including authors erroneously considered affiliated to our institutes) can be "blacklisted" and excluded from the staging list.

ID	DOI	SCOPUS ID	FETCHING
1	10.1038/441535-019-0146-3	2+2.0-85061359944	Monday, A
2	10.1038/441598-018-36336-w	2+2.0-85060124707	Monday, A
3	10.1038/441598-019-38556-0	2+2.0-85061586097	Monday, A
4	10.1038/441467-019-08763-4	2+2.0-85062019233	Monday, A

Metadata Fetching, Integration & Enrichment

For each publication, the full metadata is fetched from several sources and then integrated and displayed when opening the form for editing. Further enrichment of metadata is performed semi-automatically with tools helping the library team to add information stored in our system such as funding programs and the full internal affiliation.

Publication Ingestion

The publication is ingested into the repository and made available to website visitors. The library team triggers a notification e-mail to the researcher, asking for a check of the metadata and for the accepted version in case the published version is not open access.

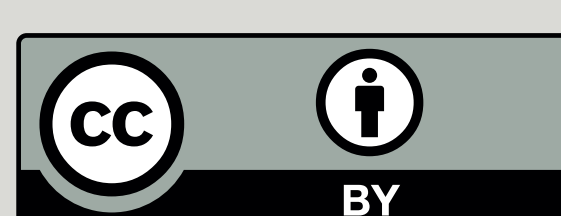
Take home ...

- The library actively adds new publications to the institutional repository. This is done efficiently and with high quality metadata (PHP tool will be available soon).
- After providing such a service to the researchers we can ask for their contributions – especially to add the accepted versions of their publications, increasing the quantity of articles publicly available via green Open Access.



Federico Cantini
federico.cantini@lib4ri.ch

Open Repositories
Hamburg, June 10-13, 2019



This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>.