Dissemin: an open source information system for open access policies

Antonin Delpeuch[1,2] and Marie Farge[1,2]

[1] École Normale Supérieure
[2] Committee for the Accessibility of Publications in Sciences and Humanities (CAPSH)

Librarians around the world struggle to fill their institutional repositories, while researchers upload millions of papers to academic social networks. This is a serious issue for the open access movement, because many social networks are actually not open enough. But institutional repositories have limitations too. They do not yet form the interconnected and decentralized network we dream of. Moreover, much too often only metadata are available, not the full text, which leads to the present bad habit of quoting papers without having read them.

Why should the number of publications deposited in your own institutional archive be an indicator of the success of your open access policy? Publications can also be deposited in subject repositories, or in the institutional repositories of your researcher's colleagues. What actually matters is the number of publications which end up in some well-indexed, metadata-rich and permanent database, not necessarily in yours.

Dissemin (http://dissem.in) is an open source platform measuring just that. We analyze the publications of any researcher and spot which ones are missing from open repositories. This analysis is summarized by a simple chart, where publications are split into 5 categories (as shown by the example figure on the right).

- available from the publisher website,
- available from an open repository,
- not available but the publisher allows deposit,
- not available but the publisher's authorization to deposit is unknown,
- not available since publisher forbids deposit.

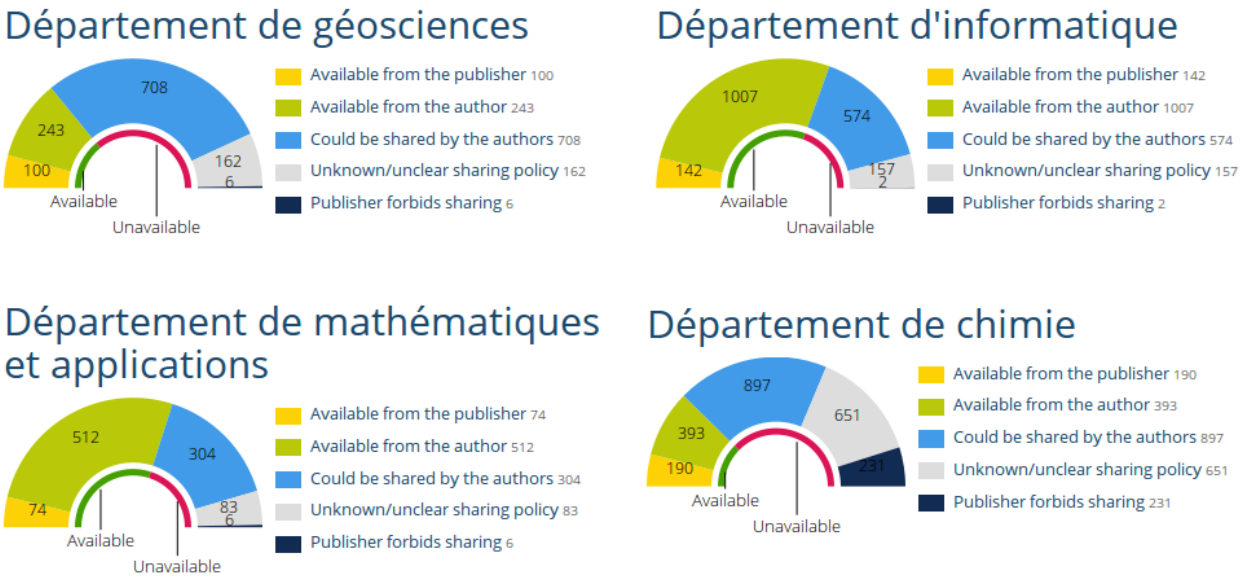
Dissemin provides a detailed list of all publications available in open access for a given researcher, and each one is linked to the publisher and to the institutional repositories from where it can be downloaded. The publisher policy is retrieved from SHERPA/RoMEO, and classified into simple categories (fully open access, allows some kind of green open access, unclear policy, or forbids any kind of deposit).

Dissemin also provides a simple interface allowing researchers to upload their publications which are not in open access to Zenodo (the open repository of OpenAIRE hosted at CERN), without having to register or to fill long metadata forms. The metadata stored in Dissemin are bundled with the PDF provided by the user and sent to Zenodo via its REST API. Adding similar interfaces to deposit in other repositories, for instance via the SWORD protocol, is made easy by the modular structure of the source code.

Dissemin harvests publications from various metadata sources, such as CrossRef and the Bielefeld Academic Search Engine (BASE). It complies with ORCID to identify the authors, and performs

some automatic disambiguation for papers without ORCID identifiers. Researchers who use Dissemin are encouraged to log in via ORCID, or to sign up for ORCID if they do not have a profile yet.

We are also working with institutions to provide them a birds' eye view on the availability in open access of their research output. A pilot experiment has been carried out at École Normale Supérieure (ENS Paris). The administration has sent us a list of more than one thousand researchers and from Dissemin's results we have computed department-wise statistics, that can be seen at http://dissem.in/institution/1/ . This analysis will then be the basis for a university-wide open access policy.



This Dissemin-ENS platform is only a prototype, and we would be interested to streamline this process to make it easy for other universities to do their repository. We are therefore actively looking for funding and partners. Our project is run by the non-profit organization CAPSH and we believe the scientific community can benefit from an open source, publisher-independent CRIS-like system, whose core purpose is to foster the use of open repositories.