OA tracker

The purpose of this document is to outline the data sources and expected inputs and outputs for the OA tracker. The goal of the app is to allow users to enter an ORCID number (more input formats shall be added later) to retrieve a list of DOIs of journal articles, determine what the publishing journal allows in term of OA, determine what OA versions are already available on the Web, and identify whether one of these available OA versions is in a specified location (e.g., Dalspace, the Dalhousie institutional repository).

# Summary

The goal of the OA tracker is to check for the OA status of a list of publications.

## ORCID API <https://info.orcid.org/documentation/features/public-api/>

ORCID (Open Researcher and Contributor ID) allows you to download research publications of a person. Ultimately this part of the code should receive as input the ORCID of a researcher and return a list of DOIs.

## CrossRef API <https://www.crossref.org/documentation/retrieve-metadata/rest-api/>

The CrossRef API can be used to retrieve useful metadata (e.g., title, journal, publisher) from a doi. It will be most likely be useful to retrieve the journal information that we will need to request OA policies from the Sherpa Romeo API.

## Sherpa Romeo API <https://v2.sherpa.ac.uk/api/>

Sherpa Romeo is a directory of journal policies. For each publication retrieved from ORCID, we want to identify the best possible OA version permitted according to the journal’s policy. Policies usually have three variables:

* Location
* Article version
* Embargo period

This essentially tells us **what** papers **can** be made available in OA, **what version** of the paper **can** be made available, **where** they **can** be made available, and **when** they **can** be made available.

## Unpaywall API <https://unpaywall.org/products/api>

Unpaywall.org is a service that tracks the different locations where an article with a DOI can be found. So it tells us **what** papers **are** available in OA, **what version** of the papers **are** available in OA, and **where** these OA versions of the paper **are** located.

# Outputs

## Data

As a first step we need to be able to retrieve information about the allowed OA and available OA for each doi. This can be achieved by retrieving two tables from Sherpa Romeo and Unpaywall. The former should have the allowed OA location and types for each publication (doi), and the former should provide a list of current OA locations where the publication (doi) is currently available.

Table 1. journal\_policy

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| doi | String | Doi of the publication |
| Location | String | Location (or types of location) where green OA is allowed |
| Version | String | Version of the article that can be made available at the location. |
| Embargo | Smallint | Embargo period before the version of the article can be made available at the location |

Table 2. oa\_location

See: <https://unpaywall.org/data-format>

At this point we can retrieve the entire OA Location object.

## User interface (upcoming)

Once we are able to collect the data for a list of DOIs, we will determine what kind of user interface will be optimal to inform the users about the OA status of their publications, provide information about their best options for OA, and facilitate self-archiving in Dalspace (the institutional repository of Dalhousie University) if the article can, but is not, there.