

Technology choices for API

Gareth Murphy

@garethcmurphy

@essneutron

- OAI-PMH (Open Access Initiative - Protocol for Metadata Harvesting)
- ResourceSync (successor to OAI-PMH, also by OAI)
- OpenAire (interoperability service)
- B2find (discovery service)
- EOSC hub

- One endpoint, 6 verbs
- **GetRecord** retrieve an individual metadata record from a repository.
- **Identify** retrieve information about a repository.
- **ListIdentifiers** an abbreviated form of ListRecords which retrieves only headers rather than records.
- **ListMetadataFormats** retrieve the metadata formats available from a repository, or optionally the formats available for a specific item.
- **ListRecords** harvest records from a repository. ListSets retrieve the set structure of a repository



Harvesting Metadata

Contents

- [Harvesting channels](#)
 - [OAI-PMH](#)
 - [JSON-API](#)
 - [CSW2.0](#)
- [Initial uptake of a new data provider](#)
- [Synchronous and operational ingestion](#)

Harvesting channels

Harvesting is the process of automatically fetching remote metadata. This section describes how B2FIND harvests metadata records from data provider sites. While OAI-PMH, as the de facto standard for metadata harvesting, is preferred, B2FIND also supports other APIs, as described in the section [Harvesting channels](#) . Once one of these transfer methods has been successfully implemented, B2FIND first takes up a few test samples to analyse their content, as described in the section [Initial uptake of a new data provider](#). As soon as the harvesting and mapping has been consolidated and the data provider gives their consent, the metadata are published in the B2FIND database, and an operational and stable ingestion process is established (see section [Synchronous and operational ingestion](#)).

OAI-PMH

OAI-PMH is B2FIND's preferred metadata harvesting protocol. It can be used to fetch metadata directly from the data providers within research communities. The simplicity of the protocol allows a controlled and easy-to-manage transfer of metadata. Very little information must be provided to enable B2FIND to perform the harvesting process using this protocol:

- **OAI endpoint:** This is the URL of the OAI provider server on data provider site, which must be open for OAI-PMH read requests.
- **OAI mdprefix:** This is the OAI acronym for the metadata schema in which the provided XML records are coded in.
- **OAI sets (optional):** It is recommended to group your records in subsets, because this simplifies the controlled harvesting.

Example:

- Provides data discovery service
- Based on OAI-PMH
- Need to register at re3data.org
- OpenAire Advance

- Improved successor to OAI-PMH
- Faster, allows change lists
- Based on sitemap technology
- Maybe supported in OpenAire Advance

- We can use b2find to send our data to EOSC hub