

# IRUS-UK best practices for cataloguing identifiers and exposing them in OAI-PMH

---

There are a number of key canonical identifiers, which IRUS-UK needs to capture reliably when ingesting data in order to:

1. Facilitate interoperability with other services
2. Enable consolidation of reporting between repositories, e.g. identifying instances of an article hosted in multiple repositories and reporting combined usage statistics
3. Identify item relationships, e.g. identifying the journal in which an article is published so that we can provide journal reports as well as article reports

Currently, the three main identifiers that help us to achieve these objectives are ISSNs, ISBNs and DOIs. Although mostly relevant for articles and books, they may also be applicable for other item types.

Additionally, ORCIDs (researcher identifiers) are being implemented in a growing number of systems. Cataloguing and exposing ORCIDs will eventually enable IRUS-UK to produce accurate per author statistics.

## Cataloguing identifiers

The internal mechanisms for cataloguing metadata vary across software platforms. Nevertheless, regardless of the specifics of any individual system, it is good practice to catalogue identifiers in their own discrete fields rather in the middle of a (free text) citation, which can cause problems, e.g. an ISSN being confused with date or page number ranges which happen to pass the checksum calculation (the last digit of an ISSN, which may be 0-9 or an X, is a ‘check digit’ whose value is determined by the values of the first seven digits).

## Exposing identifiers in OAI-PMH

IRUS-UK utilises the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to harvest the bibliographic metadata, which describes items that have been downloaded.

There are two OAI-PMH metadata formats relevant to IRUS-UK, namely: `oai_dc` and `rioxx`.

### `oai_dc` metadata format

The OAI interface must support the ‘`oai_dc`’ metadata format, which exposes Dublin Core metadata – this enables IRUS-UK to harvest basic bibliographic metadata that is sufficient for many core operations within the service.

### *DOIs, ISSNs and ISBNs*

We recommend that DOIs, ISSNs and ISBNs should be exposed in their own discrete element in the form of “<namespace>:<value>”, that is the value of the identifier should be prefixed with the appropriate namespace, e.g.

`doi:10.1000/182`

issn:1741-7589

isbn:978-0-8412-3707-0

In most cases DOIs, ISSNs and ISBNs should be exposed in the dc:relation element as they identify items related to the resource in the repository. An exception to this is where a DOI has been minted specifically for the version of the resource in the repository, in which case the dc:identifier element should be used. However, this convention is not adopted by all software platforms so, in practice, the use of either dc:relation or dc:identifier is acceptable.

For DOIs, we recommend that you enter the DOI in its native format, which starts with '10.'. We advise against exposing DOIs as URLs, which can lead to a multitude of problems in accurately matching a DOI across instances of an article hosted in multiple repositories: the same DOI can have different URLs due to differing domains, cut and paste errors, typographical errors, vendor data appended, etc.

### ORCID*s*

We recommend that ORCID*s* should be exposed in their own discrete dc:creator element preferably in the form of "<namespace>:<value>", e.g.

orcid:0000-0002-1825-0097

Note though, it is also acceptable to use the URL form of an ORCID as specified in the ORCID display guidelines:

<http://orcid.org/0000-0002-1825-0097>

### rioxx metadata format

If possible, the OAI interface should also support the 'rioxx' metadata format, which exposes RIOXX-compliant records – this enables IRUS-UK to harvest a much richer metadata set (including data about funders, grant numbers, ORCID*s*, parent publications, etc.) and offer enhanced services to other agencies and organisations such as Jisc, OpenAIRE and Research Councils.

Please follow the 'UK Metadata Guidelines for Open Access Repositories', which can be found on the RIOXX website (<http://www.riox.net/>).