**Discovering research outputs – a model for sustainable discovery**

Whilst the aggregation of metadata is straightforward the challenge for long term success and resilience of an aggregation service is dependent on appropriate mechanisms for sustainability.

Introducing such a system raises some fundamental questions: Are procedures for improving quality of the research output metadata in place? Institutionally this will include guidance in Data Management Planning (DMP); what is seen as a minimum acceptable metadata to enable successful identification of a research output? we may even see a policy intervention by the funders to improve the capture of the minimum metadata to enable identification of outputs.

Discovery of research metadata, fundamentally via OAI-PMH end points, will require appropriate management to ensure ongoing publishing to a required standard and a sustainable mechanism for discovery. The equipment.data service autodiscovers published research equipment through the use of the [Organisation Profile Document](http://opd.data.ac.uk) (OPD). The OPD is a simple script that can be hosted anywhere on an institution’s website, with an associated embedded link in the home page, and acts as the key enabler to the process of data autodiscovery. It provides key information of who, what, where and how by enabling discovery of data, describing the organisation, and stating what is published (the catalogue of datasets) and to what format it conforms. It also provides essential organisational information such as the organisation ID, official name, organisation type, official logo and geographical location. To date a number of institutions have included their OAI-PMH end point on their OPD enabling its autodiscovery using the same process as equipment.data.

Lessons from equipment.data highlight the need to ensure there is identified role responsible for content on the OPD. Although a model with either manual reporting of data end points or autodiscovered will have a requirement to indicate the location of an OAI-PMH end point, using the OPD ensures this is managed overtime alongside the catalogue of data the OPD provides and that other data aggregators can also discover the end point.

Using this established process of autodiscovery provides added value in that this infrastructure is exploitable by others wishing to aggregate data for other applications. This already includes the NCUB who are ingesting the equipment.data and OPD datasets in .csv form in complimenting data searchable through the Intelligent Brokerage Tool, [Konfer](https://konfer.online).

What the OPD offers is a mechanism for discovering both end point and details of the contact responsible for the that dataset (e.g. the repository) This enables the data aggregator (the service) to set up automated emails for problems such as a break in the data feed, suspected duplicate records or mandated data field omissions e.g. author. Such a feature is fundamental in an automated service and one which places an emphasis on the institutions to ensure curation and management of their data feed.

Establishing a workflow that embeds these simple steps within an institution utilising the OPD enables greater ownership and responsibility for the published data placing an emphasis on quality and fundamentally data provision.