

equipment.data

Aggregating equipment data
from across UK Higher Education

THE OPD

ORGANISATION PROFILE DOCUMENT

UK NATIONAL EQUIPMENT & FACILITIES PORTAL

EPSRC

In partnership with **Jisc**



CONTENTS

2

THE LANDSCAPE

3 Importance of the OPD

4

WHY CREATE AN OPD?

5 Data Profiles

6 Ownership

7

ROLES AND RESPONSABILITIES

8

WHO? WHERE? WHAT?

8 Define your identity

9 What are you publishing


11 Maximise your publicity

12 About your business

13

AUTO-DISCOVERY





With over 1300 providers the UK Higher Education (HE) sector has hundreds of reporting obligations, many statutory, often leading to duplication in data capture and management, and in many instances, the use of a range of different information systems generates interoperability challenges.

Establishing harmonised vocabularies and interoperability in the data management landscape is becoming an increasing requirement. There is a need to deliver increased efficiency in management and reporting along with greater benefits to the wider community from published data.

Initiatives such as the Higher Education Data and Information Improvement Programme (HEDIIP), funded by HEFCE, aim to promote a new data landscape harmonising data profiles and promoting more efficient data management practice. However, there is still a need to develop and promote emerging technologies and standards underpinning how we make this data discoverable, and therefore easy to share.

The Organisation Profile Document (OPD) is one such technology. This simple piece of rdf (a machine readable file) supports a linked open data based architecture now well established in the HE sector. Initially developed to support delivery of the National Equipment Portal, equipment.data.gov.uk, it also supports cross-institution cross-sector data autodiscovery and sharing, which in the future could also deliver reporting requirements.

IMPORTANCE OF THE OPD



The Organisation Profile Document (OPD), is a simple script hosted on your website, with an associated embedded link in the home page, that acts as the key enabler to the process of data autodiscovery.

The OPD provides key information of who, what, where, when and how by enabling discovery of data, describing the organisation, and stating what is published (the catalogue of datasets). It also provides essential organisational information such as the organisation ID, official name, organisation type, official logo and geographical location.

For the equipment.data project to reach its goal of a fully sustainable system, we needed a method of updating sources as efficiently as possible. To encourage adoption of a sustainable method of contribution we established a compliance rating system with gold, silver and bronze ratings to indicate to what level each contributing institution's data input is sustainable. Changes to data can be as small as an institution altering its logo, to moving its data source from one system to another, but they still need to be captured in an efficient way.

The ideal is that all institutions will be operating the gold compliance rating for equipment.data, using a fully autodiscoverable OPD. Therefore no human intervention is required from either the contributing institution or the equipment.data service in updating information as it will be automatically identified by the OPD.

In the future we may see the OPD used for research data management (RDM), highlighting compliance e.g. published insurance policies, ethics policy, as well as research governance strategies or data management planning strategies and in enabling the discovery of a much wider range of business related datasets e.g. capability and expertise (e.g. via ORCID).



¹ <http://orcid.org/>





WHY CREATE AN OPD

OPEN DATA ARCHITECTURE

- In the world of open data
the OPD is rapidly becoming the standard
for data discovery

DEFINE YOUR IDENTITY

- Defines the official name of the organisation
- Provides your organisation's unique ID number
 - Publishes current up to date logo
- Makes your organisation's contact details e.g.
current phone number, email address and postal
address available

GET DISCOVERED

- Published data will automatically become
discoverable to new sources and markets
 - Describes your business
- Up to date RSS feeds can be published
allowing them to be discovered
- Lists official social media accounts



DATA PROFILES

Data aggregations offering function and value would sadly not be possible without standards – agreed “data profiles”. For published data, made discoverable via an OPD, to have real value to end users and aggregators it should follow an agreed profile. Those currently published or discoverable include “facilities and equipment”, “research outputs” and webpages included within the “Linking you” vocabulary.

equipment.data would not have been possible without establishing of UNIQUIP Data Publishing Specification which is now RCUK’s preferred standard for the capture and publishing of equipment data. This standard is the culmination of HE sector engagement through the UNIQUIP Project and the subsequent development work through equipment.data, including early discussion with CASRAI to establish an entry in the CASRAI Dictionary.

The data profiles for these datasets have been established with the consensus of the HE community and have been adopted by system vendors and institutions in enabling system interoperability and standardised reporting. For example, agreement to use Dublin Core was driven by the need for a standard way of describing publications (and outputs) resulting in the development of the OAI-PMH as the chosen mechanism for making publication metadata information accessible to aggregation and publication services.

With growing interest in data sharing and system interoperability we have seen development strands formed in the Research Data Alliance to explore metadata standards for research outputs and the organisation CASRAI formed, with the aim to create a dictionary of community agreed and managed data profiles established to support “use cases” e.g. Research Data Management.

¹ <http://casrai.org/>

OWNERSHIP

As wider use of OPDs increases the challenge will be to establish appropriate ownership and governance of the OPD within your organisation.

It may be logical for this to be your marketing and communications department who typically will be responsible for your website (home page) and could therefore manage the OPD content and/or link to the OPD.

To date, due to the focus on research equipment data, the equipment.data project team have mainly worked with teams from research support offices and IT departments although as more data links are established, and OPDs have a wider use, ownership could change.

Governance of the OPD could be reasonably defined in your Data Management Planning Strategy. However, to enable such decisions the sector will require greater confidence in this emerging technology through standards development and increased application.

These discussions will invariably emerge as a greater emphasis is placed on the use of the OPD in data discovery services.

IF YOU ARE PUBLISHING CONTENT (non-technical)

If you are responsible for the publishing and management of your organisation's OPD you will need to consider its governance i.e. long term management, maintaining discoverability and the core information is current.

Any governance will need to ensure links to published datasets are maintained i.e. if the location changes, either web page or end point, this is amended on the OPD. The following pages provide examples of the content from the "Human readable" output on the OPD Checker service and further information describing the type of content published within each section.

IF YOU ARE WRITING AN OPD (technical)

Writing an OPD is straightforward and hopefully you will have been provided with the location of the relevant data feeds and datasets to be published.


An OPD is written in rdf and typically will be around one page of code.

Examples of the code structure required for the "Core" fields, facilities and equipment, news feeds and "Linking you" are provided in the following pages. For more detailed information visit <http://opd.data.ac.uk/docs/core>

WHO? WHERE?

Define your identity.

The “Core” fields will describe your organisation (name location, web home page, organisation i.d. e.g. ISNI¹) and any parent or sub organisations.

HUMAN READABLE VERSION OF OPD	RESOURCE DESCRIPTION FRAMEWORK RDF
Self assigned ID: http://id.example.ac.uk	
Type: An Academic Institution Formal Organisation (charity, business etc.) An Organisation (general)	
Preferred Label: University of Example	
Hidden Label: Example University	
Label: University of Example	
Homepage: http://www.example.ac.uk/	
Phone (foaf): +442380123456	
Email: info@example.ac.uk	
Logo: 	
SPARQL: http://sparql.data.example.ac.uk/	
End point is the same as: http://dbpedia.org/resource/University_of_Example http://education.data.gov.uk/id/school/123456 http://id.learning-provider.data.ac.uk/ukprn/101234	<pre><http://id.example.ac.uk/> a org:FormalOrganization ; skos:prefLabel "The University of Example"; skos:hiddenLabel "Example University"; vcard: sortLabel "Example, University of"; foaf: homepage <http://www.example.ac.uk/> ; foaf:phone <tel:+441234567890> ; foaf:mbox <mailto:info@example.ac.uk> ; foaf: logo <http://www.example.ac.uk/ example-logo.png> ; owl: sameAs <http://dbpedia.org/ resource/University_of_Example> ; owl: sameAs <http://education.data.gov. uk/id/school/123456> ; owl: sameAs <http://id.learning-provider. data.ac.uk/ukprn/101234> .</pre>
Located: long/lat: 50°56'08"N 1°23'46"E (50.935557, 50.935557) OS Grid Reference: SU 425 152 (442537,115296)	

¹ <http://www.isni.org/>

WHAT are you publishing?

Open Data Architecture for your Facilities and Equipment

To ensure compatibility with international data aggregations we recommend data made available under these sections also conforms to an internationally recognised data profile e.g. UNIQUIP Data Publishing Specification.¹

HUMAN READABLE VERSION OF OPD	RESOURCE DESCRIPTION FRAMEWORK RDF
Equipment Dataset URL: http://id.example.ac.uk/equipment.csv	
Organization URI: http://id.example.ac.uk/	< http://id.example.ac.uk/equipment.csv >
Contact: facshare@example.ac.uk	oo:organization < http://id.example.ac.uk/ > ; oo:corrections < mailto:facshare@example.ac.uk > ;
Corrections Contact: facshare@example.ac.uk	oo:contact < mailto:facshare@example.ac.uk > ; dcterm:subject < http://purl.org/openorg/theme/Equipment > ;
Subject: Equipment Dataset	dcterm:conformsTo < http://equipment.data.ac.uk/uniquep > ; dcterm:license < http://creativecommons.org/publicdomain/zero/1.0/ > .
Conforms to: Uniquip Data Publishing Specification	
License: OGL - The (UK) Open Government License for Public Sector Information	

¹ <http://equipment.data.ac.uk/uniquep>

Research outputs

Publishing data availability under this section will make your outputs metadata accessible to research data aggregation services. We have chosen OAI-PMH as it is an established protocol for the exchange of Metadata from insitutional repositories.

HUMAN READABLE VERSION OF OPD	RESOURCE DESCRIPTION FRAMEWORK RDF
End point URL: http://eprints.example.ac.uk/cgi/oai2	
Organization URI: http://id.example.ac.uk/	<code><http://eprints.example.ac.uk/cgi/oai2</code>
Contact: eprints@example.ac.uk	<code>oo:organization <http://id.example.ac.uk/> ; oo:contact <mailto:eprints@example.ac.uk> ; oo:corrections <mailto:eprints@example.ac.uk> ; dcterms:subject <http://purl.org/openorg/theme/ResearchOutputs> ;</code>
Corrections Contact: eprints@example.ac.uk	<code>dcterms:license <http://www.nationalarchives.gov.uk/doc/open-government-licence/> ; dcterms:conformsTo <http://www.openarchives.org/OAI/openarchivesprotocol.html> .</code>
Subject: Research Outputs	
License: OGL - The (UK) Open Government License for Public Sector Information	
Conforms to: OAI-PMH	

MAXIMISE YOUR PUBLICITY

Get Discovered

A news “feed” for an organisation is generally one of the following:

- Events related to the organisation.
- News items from and about the organisation e.g. Vice Chancellor opens new building (this does not include sites which provide an actual news service)
- Notices from the organisation. e.g. Building 23 will be closed on 4th May for essential maintenance.

NB. iCalendar data is probably not going to be news or notices, but please still mark it as events for easy discovery.

FEED-TYPE is one of: RSS or Atom or iCalendar

THEME is one of: news or events or notices

HUMAN READABLE VERSION OF OPD

RESOURCE DESCRIPTION FRAMEWORK RDF

RSS News Feed URL:

<http://www.example.ac.uk/news.xml>

Type: RSS Feed

Organization URI:

<http://id.example.ac.uk/>

Subject: News Feed

License: Creative Commons CC0

```
<http://www.example.ac.uk/news.xml>
a <http://purl.org/xtypes/Document-RSS>
oo:organization <http://id.example.ac.uk/> ;
dcterms:subject <http://purl.org/openorg/theme/news>;
dcterms:license <http://creativecommons.org/
publicdomain/zero/1.0/> .
```

HUMAN READABLE VERSION OF OPD

RESOURCE DESCRIPTION FRAMEWORK RDF

Org has social Twitter Account:

@uniexample

Type: Online Account

Account Name: uniexample

Account Service Homepage: Twitter

```
<http://id.example.ac.uk/>
foaf:account <https://twitter.com/uniexample> .
<https://twitter.com/uniexample>
a <http://xmlns.com/foaf/0.1/OnlineAccount> ;
foaf:accountName "uniexample";
foaf:accountServiceHomepage <https://twitter.com/> .
```

ABOUT YOUR BUSINESS

This is an agreed vocabulary¹ of common URLs for organisational webpages such as “contact” and “about”.

Publishing these pages through your OPD can enable the discovery of business critical pages such as your “Enterprise” landing page or availability of “Business incubation” facilities.

In cooperation with the international Research Data Management (RDM) community, the Digital Curation Centre (DCC) has defined a list of 15 core components for RDM infrastructure - both technical elements and soft infrastructure components such as policies, support and guidance. The intention is to establish these components as additional items on the “Linking you” data profile with the aim to support early developments in research integrity. It is hoped the basic RDM profile, currently defined, can be progressed as a CASRAI standard.

HUMAN READABLE VERSION OF OPD	RESOURCE DESCRIPTION FRAMEWORK RDF
Events Page: http://www.example.ac.uk/events/	
About Us Page: http://www.example.ac.uk/about/	< http://id.example.ac.uk/ > lyou:events < http://www.example.ac.uk/events/ > ; lyou:about < http://www.example.ac.uk/about/ > ;
Business Page: http://www.example.ac.uk/business/	lyou:business < http://www.example.ac.uk/business/ > ; lyou:news < http://www.example.ac.uk/news/ > .
News Page: http://www.example.ac.uk/news/	

¹ <http://openorg.ecs.example.ac.uk/linkingyou/>

AUTODISCOVERY

Sustainable linked data solutions can be adopted enabling the discovery of a range of data profiles by using the infrastructure the equipment.data project has developed. This approach avoids the need for prescribing defined systems, instead allowing data to be captured from a range of sources.

Essentially all that is required are 4 key components:

- An authoritative list of institutions you wish to capture the data from (e.g. UKPRN)
- Auto-discovery programming hosted by the requesting organisation (the equipment.data programming is open source)
- A requirement that all reporting organisations host an OPD pointing to their data profiles,
- The data is managed/collated to required standardised profiles e.g. UNQUIP, ORCID, Research outcomes profile (Gateway to Research) ideally managed through a standards organisation e.g. CASRAI

If it sounds simple, it is. equipment.data is using these very simple components in the discovery of research equipment data across HE.

Future data profiles could include student reporting information, course information (including CPD), Research Data Management (RDM) and research data formalising the Digital Curation Centre (DDC) UK RDDS metadata profile.



equipment.data

Equipment Data
University of Southampton,
Highfield Campus SO17 1BJ,
023 8059 4456
info@data.ac.uk
equipment.data.ac.uk