The Digital Research Object Portal

Opening research data and publishing open access articles have become a requirement for many publicly-funded research projects. This is the case of research projects funded through EU Framework Programmes.

The ultimate goal of opening data is to increase their visibility, use and re-use. There is evidence that studies that make their data available are shared more frequently, hence receive more citations than similar studies that do not (Piwowar and Vision, 2013)[1]; re-use of data has produced five times more scientific publications than the initial exploitation of the data by the research groups that have produced it (Gaillard, 2014)[2]; technological progress has moved science from the analysis of a molecule in a test tube to high-throughput approaches that allow the collection of enormous amount of data (genomics, transcriptomics, proteomics, metabolomics, etc.) that could benefit to the scientific community.

Facts and observations

The number of infrastructures that host open research data has increased to adapt to different types of data and scientific journals are adapting to the open access policy.

The multiplication of open data/open access infrastructures has to be seen positively as they contribute to the visibility, use and re-use of scientific information. Yet, the large number of diverse infrastructures has the counter-effect to scatter information which may be more difficult to reach, or which is 'hidden'.

Open data means:

- Data is open all over the world
- Many people from different disciplines open data
- Different types of data is available
- Different formats of data are used
- Different infrastructures to host data exist
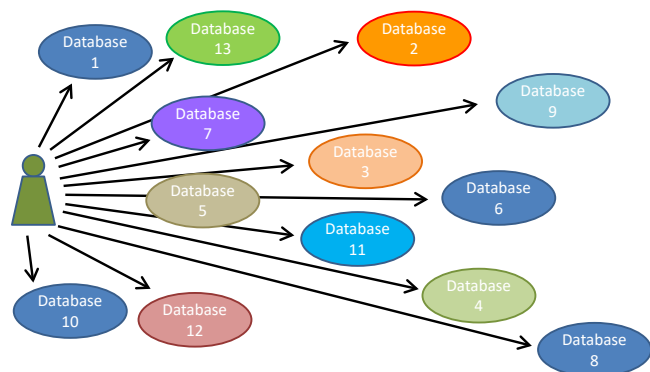

The Digital Research Object Portal (DROP)

To address the abovementioned limitations associated to open data approaches and policies, it was proposed to build a node. The opportunity to develop a prototype to focus on open data and open access documents was given by the XF-ACTORS project.

---

[1] Piwowar H.A. and Vision T.J. Data reuse and the open data citation advantage. PeerJ 1:e175 (2013).

[2] Gaillard R. De l'Open data à l'Open research data: quelle(s) politique(s) pour les données de recherche? ENSSIB (2014).

The online publicly accessible Digital Research Object Portal (DROP) works as a node, a unique entry point that facilitates retrieval of digital research objects (data and documents) on *Xylella fastidiosa*. Priority was given to open data and open access documents produced in the framework of the XF-ACTORS project, but digital objects from other sources were also referenced, when relevant.
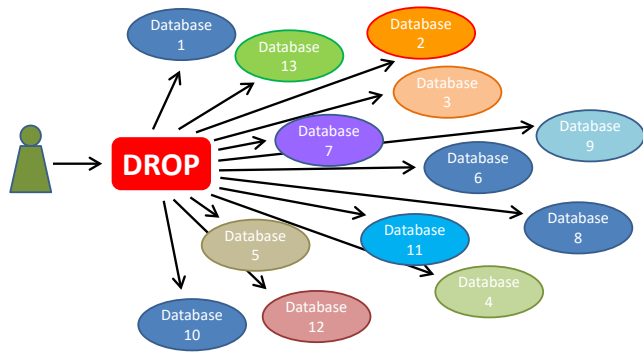
Why a portal, rather than a database?

- Because a database implies that information is centralised in one infrastructure, but scientists want to keep the freedom of choosing where to put their data. Moreover, in an international context, research funders have different policies for opening data, that the producers of the data have to comply with
- Because there are many types of data, each one with its own requirements in terms of format, metadata, *etc*. As an example, on the OpenDOAR Directory of Open Access Repositories, 5325 databases are referenced (as of 2020-03-18), but the number of databases worldwide is larger, if we consider e.g. small and local (e.g. university based) infrastructures
- Because different data requires different levels of checking or curating
- Because it is less resource consuming to maintain a portal, than a database

How to use the Portal

The Portal is accessible on the url https://drop.euphresco.net/

Each digital research object referenced in the Portal is described using the following set of metadata element derived from the Dublin Core metadata element set:

Title, the name given to the resource.

Creator, the entity primarily responsible for making the resource. In DROP, creators and contributors (see below) are the same, i.e. the scientists that produced the data or the document.

Contributor, the entity responsible for making contributions to the resource. In DROP, contributors and creators (see above) are the same, i.e. the scientists that produced the data or the document.

Subject, the topic of the resource. Typically, the subject will be represented using keywords, key phrases, or classification codes. DROP uses three types of keywords: the categories, the tags and the organisms.

Categories use a controlled vocabulary, in order to better structure the information. Categories relate to research fields, such as biochemistry, molecular biology, but also diagnostics, pest risk assessment, surveillance.

Tags are free-text keywords that allow to describe more precisely and finely the resource.

Organisms is the EPPO code associated to the organism covered by the resource, the pest and/or the host.

Description, the account of the resource, the abstract.

Publisher, the entity responsible for making the resource available Examples of a Publisher include a person, an organization, or a service.

Funding, the funders, grant information, *etc*.

Date, a point or period of time associated with an event in the lifecycle of the resource. Generally, the date of the publication of the resource.

Type, the nature or genre of the resource. The digital objects referred to in the portal are all files.

Format, the file format, physical medium, or dimensions (size) of the resource.

Identifier, an unambiguous reference to the resource within a given context. Generally, the DOI.

Source, a related resource from which the described resource is derived.

Language, the language of the resource. By default, this is English.

Relation, a related resource.

Rights, Information about rights held in and over the resource.

An example of page associated to a digital object is presented below (accessible here)

The metadata allow easy retrieval of the digital research objects by:

▪ Browsing the digital objects. The categorization of the digital objects is visible from the Portal homepage and the user can select one criterium for the search. Criteria are:

a. Category: the research field covered by the digital research object
b. Tags: as identified by the owners/authors of the data/document
c. Organism: the pests (based on the EPPO codes)
d. Author: the owner of the data or author of the document
e. Publication year: date of publication of the digital research object
f. Data type: collection, dataset, event, image, interactive resource, service, software, sound, text (scientific article/review article, protocol/guideline, standard, opinion)

Browse into Data

By category	quick filter
Agriculture	18
Agrobiodiversity	1
Bacteriology	175
Biochemistry	187
Bioinformatics, biocomputing, DNA and molecular computation	2
Biological control	71

By tag	quick filter
1,2-dichloroethane	1
16s	20
16s-23s	1
16s-23s intergenic	1
16s-23s intergenic spacer	2
16s-23s rdna	1
16S/ITS sequencing	6

By organism	quick filter
Acer rubrum	1
Acer sp.	1
Acetobacter	1
Acetobacter aceti subsp. xylinus	2
Acidovorax citrulli	3
Acinetobacter	1
Acinetobacter baumannii	2

By author	quick filter
Abou Kubaa, Raied	4
Abrahams, BR	1
ABRAHAMS, BR	2
Abrahantes, Jose Cortinas	1
Abreu, FA	1
Abreu, IL	1
Acciani, Giuseppe	1

By publication year	
empty	3
2020	69
2019	82
2018	75
2017	45
2016	53
2015	43

By Data type	
Collection	3
Dataset	59
Event	2
Text	5
Text - opinion	11
Text - protocol/guideline	1
Text - scientific article/review article	715

- Using the DROP multi-criteria search. It is possible to search for digital objects by combining different criteria (category, tags, organism, author, publication year, data type). It is possible to use a combination of criteria (e.g. organism and data type) or the same criteria multiple times (e.g. different tags).

Advanced search

	Filter by	Value
	title or keywords ▼	🔍 Search in title or keywords
X	Category ▼	-- select -- ▼
X	Data type ▼	All ▼
X	Publication date ▼	All ▼
X	Author ▼	👤 Search for author
X	Organism ▼	🜊 Select an organism to add ▼

[Add filter] [Apply filters]

Search results - 645 item(s) found

Status of the Portal

Currently, the portal references documents and data produced in the framework of Euphresco activities.

The future of the Digital Research Object Portal

Discussions are in progress on the services that DROP can provide to the phytosanitary community that use data and documents produced in the framework of research activities. There is an interest to extend the scope of DROP to cover all regulated and emerging pests and to reference important plant health data (data that is unique, data that is expensive to produce, data that can be re-used, *etc.*) and documents.

What are the benefits of DROP?

- Today, two types of databases are available. Multi-thematic databases (like Zenodo), that host data from any discipline. These multi-thematic databases are often used by scientists that have not found a specialised database, thus they are a 'stuff all'. Specialised databases, they can be many (e.g. the databases focussing on biochemistry or on genomics) and not all known widely. Different databases have different keywords, in general added by the depositors. This is linked to the view of the depositor and creates diversity and inconsistency of metadata that complicate retrieval of data. By linking databases, DROP reduces fragmentation; by referencing data using agreed metadata DROP smoothens the diversity associated to different databases.
- DROP can contribute to harmonise practices for opening data in the phytosanitary community. Having a node is the first step towards agreeing on the data that is important to open, agreeing on the infrastructures to use, agreeing on the ontology of metadata. DROP can be the catalyser for discussions on all these points.
- DROP can shape operations: perennity of data is ensured, data is more visible and can be exploited more efficiently; cross-sectoral collaboration is incentivised; adoption of best practices is favoured.