



---

## Introduction to Open Data Policies in H2020

Nancy Pontika, PhD

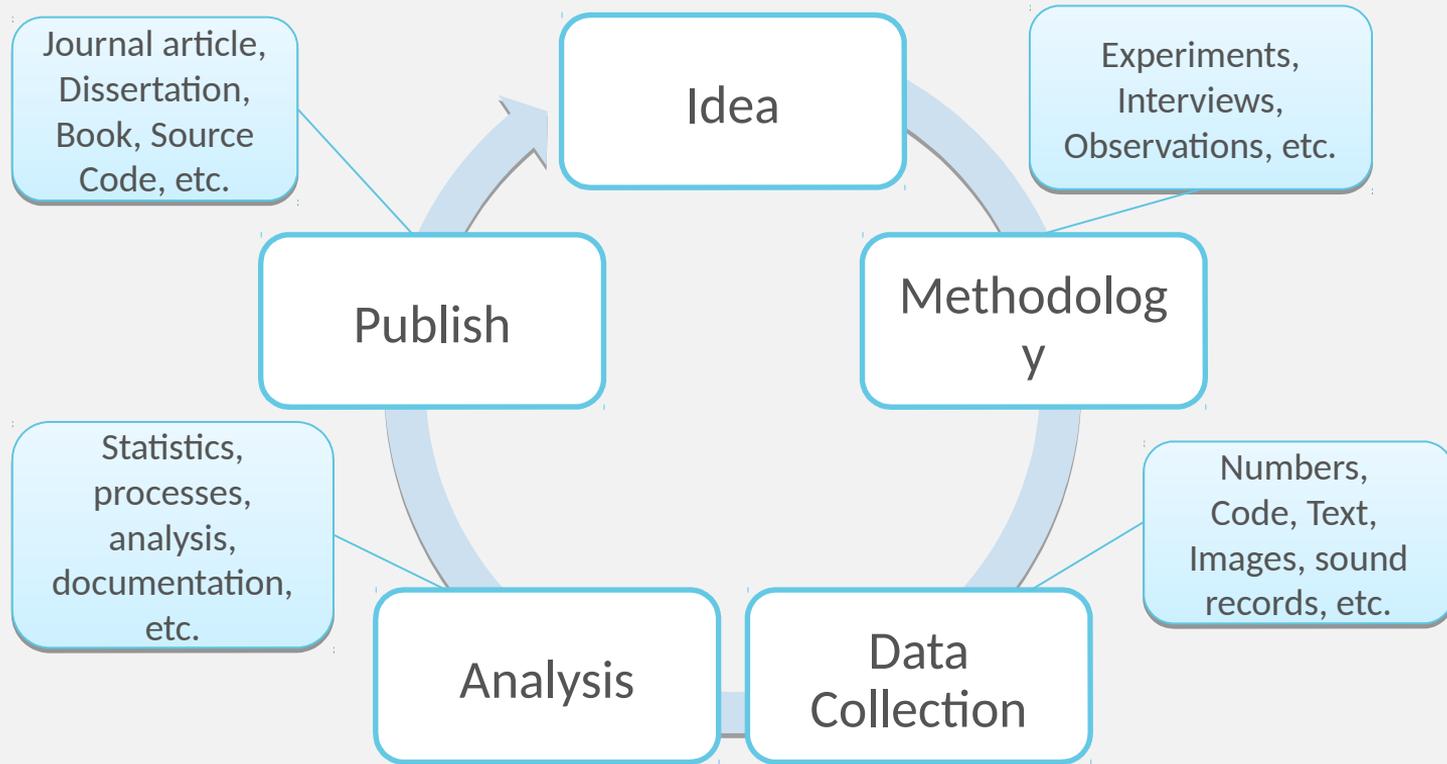
Open Access Aggregation Officer, CORE

The Open University, UK

25th October 2017



# Research Lifecycle



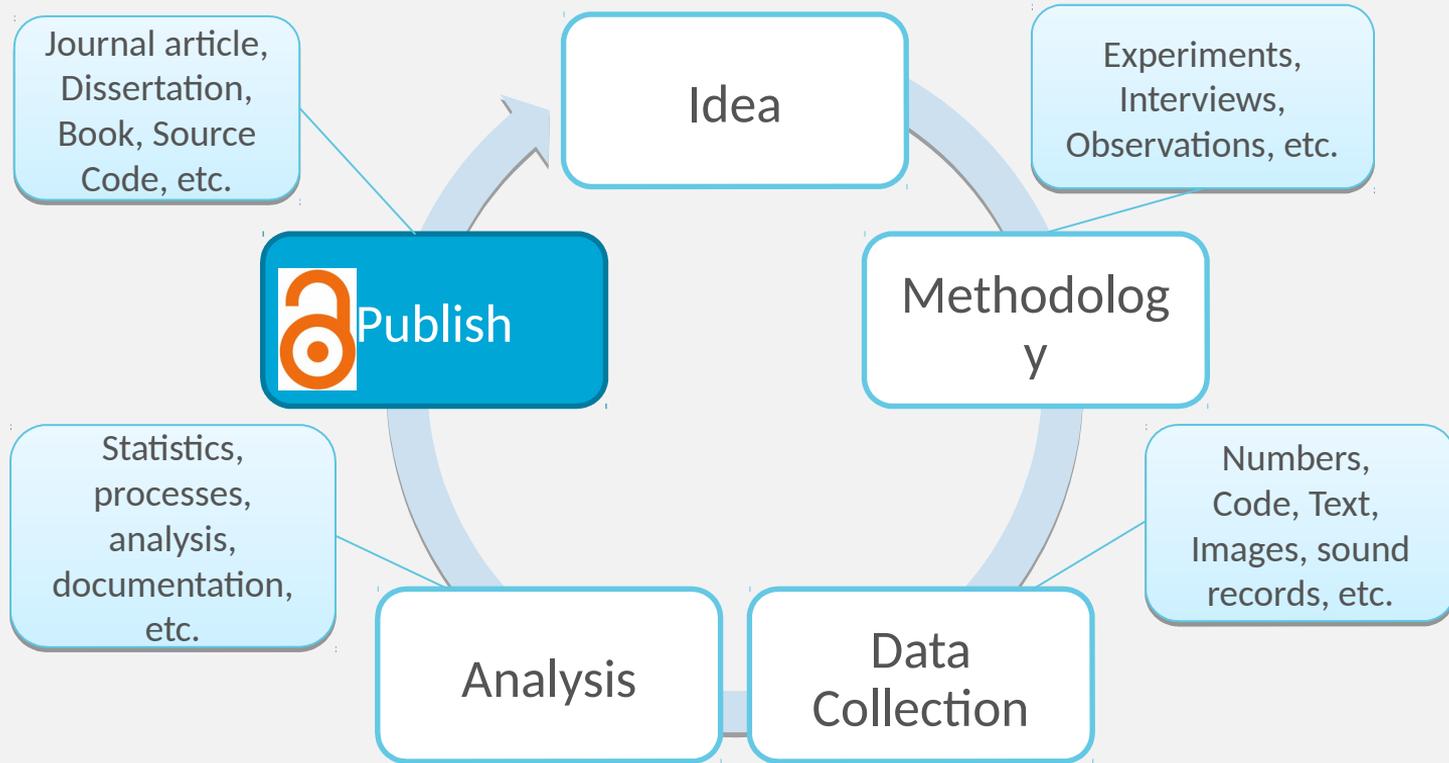
# Open Access routes

Open-access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions. What makes it possible is the internet and the consent of the author or copyright-holder.

(Source: <https://legacy.earlham.edu/~peters/fos/brief.htm>)



# Research Lifecycle: focus on the publications



# Open Access in Horizon2020

*“The European Commission sees open access not as an end in itself but as a tool to facilitate and improve the circulation of information in the European Research Area (ERA) and beyond.”*

(Source: [https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet\\_Open\\_Access.pdf](https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Open_Access.pdf))

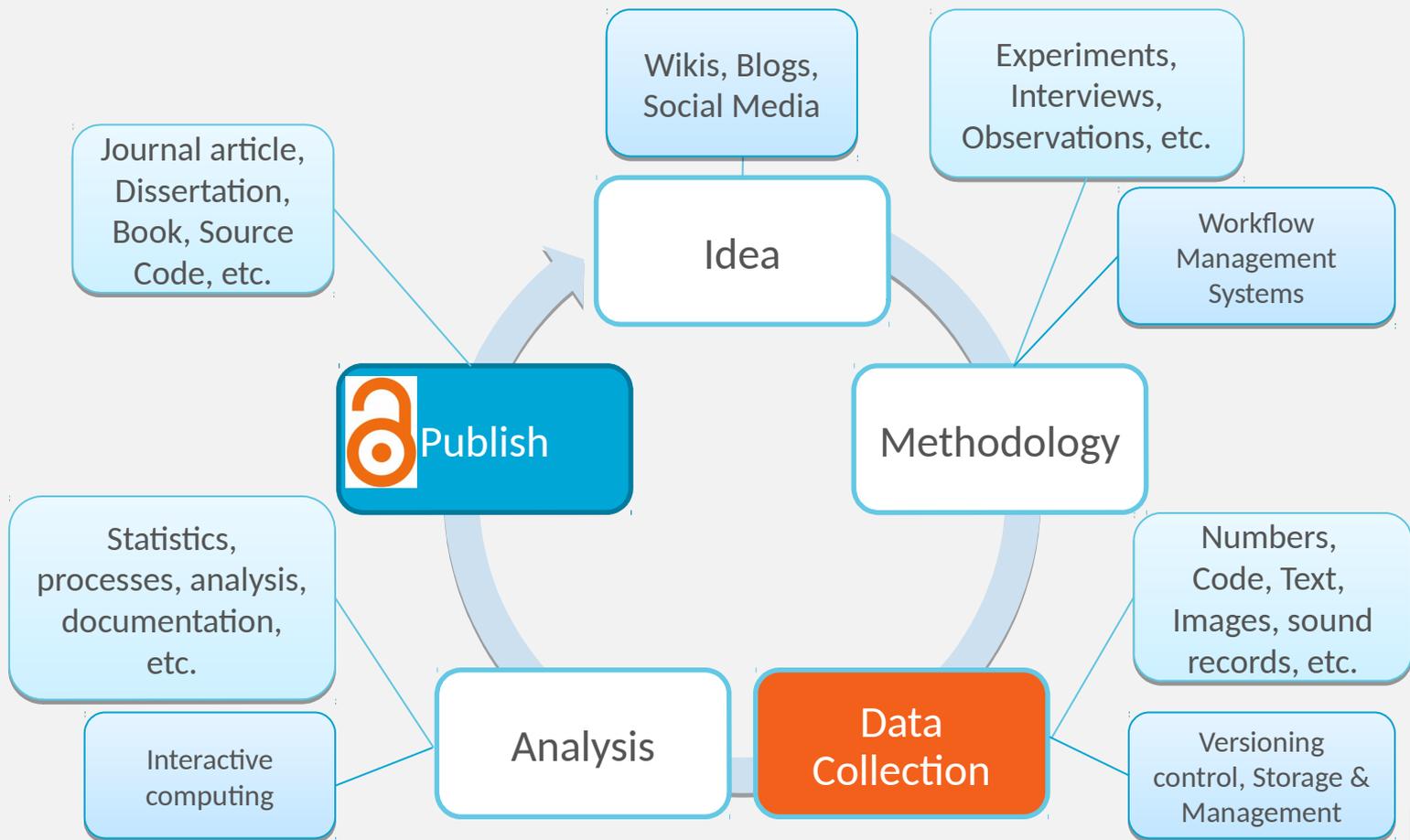
It is **mandated**:

- Immediate deposit of peer reviewed scientific publications
  - i.e. journal articles
- Deposit in a repository (machine readable form and even Gold OA)
  - Respect embargo periods
- If possible use an open license
  - e.g. Creative Commons Attribution, CC-BY
- When possible deposit research data

(Source: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf))



# Research Lifecycle: focus on data



# What constitutes research data?

*'Research data' refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation.*

*In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. **The focus is on research data that is available in digital form.***

(Source: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf))



# H2020 areas participating in pilot (2016-2017)

- Future and Emerging Technologies
- Research Infrastructures
- Leadership in enabling and industrial technologies – Information and Communication Technologies
- Nanotechnologies, Advanced Materials, Advanced Manufacturing and Processing, and Biotechnology
- Societal Challenge: Food security, sustainable agriculture and forestry, marine and maritime and inland water research and the bioeconomy
- Societal Challenge: ‘Climate action, Environment, Resource Efficiency and Raw Materials’ – except raw materials
- Societal Challenge: ‘Europe in a changing world – inclusive, innovative and reflective Societies’
- Science with and for Society
- Cross-cutting activities – focus areas – part Smart and Sustainable Cities.

\* Projects in other areas can participate on a voluntary basis

(Source: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf))



# The scope of participation is growing...

- In 2014-15 work programme, 7 areas participated in the pilot.
- In the 2016 work programme, new topics joined in 3 areas (research infrastructures, nanotechnologies and food security)
- **All calls covered by the 2017 work programme will be part of the pilot.** A move from a pilot to a mandate.



Open Access EC  
@OpenAccessEC

Official! From 2017 no more pilot: [#open](#) [#research](#)  
[#data](#) will become the rule (with opt out) [ec.europa.eu/  
digital-single...](http://ec.europa.eu/digital-single...)

The Commission will make open research data the default option, while ensuring opt-outs, for all new projects of the Horizon 2020 programme. As of 2017

19/04/16 12:17

# H2020 Open Research Data Pilot (ORD)

**OPEN RESEARCH DATA**  
IN HORIZON 2020

**CHALLENGE**

Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results:

- offers better value for EU research funds
- encourages research across scientific fields

a public benefit

essential for solving today's complex societal challenges

**SOLUTION**

Horizon 2020 already mandates open access to **all scientific publications**

From 2017, **research data is open by default**, with possibilities to **opt out**

## Open access to research data

Refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement.

## Why have Open Access to both Publications and Data?

- build on previous research results (improved quality of results)
- encourage collaboration and avoid duplication of effort (greater efficiency)
- speed up innovation (faster progress to market means faster growth)
- involve citizens and society (improved transparency of the scientific process).

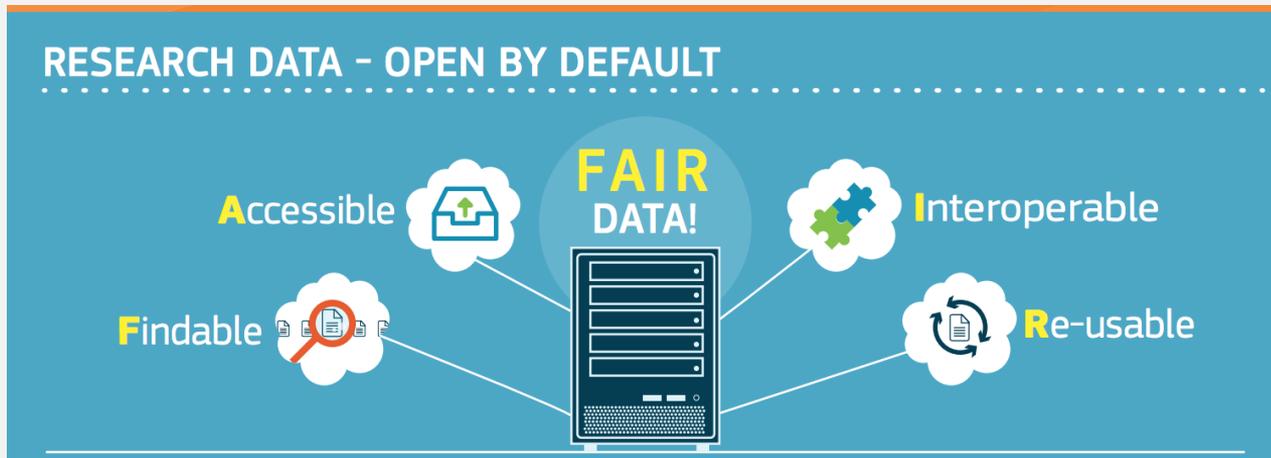
(Source: <http://>

[ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

and [https://ec.europa.eu/research/press/2016/pdf/opaopendata\\_infographic\\_072016.pdf](https://ec.europa.eu/research/press/2016/pdf/opaopendata_infographic_072016.pdf))



# FAIR Data



Search 79,309,134 open access articles

Search

## Aggregating the world's open access research papers

We offer seamless access to millions of open access research papers, enrich the collected data for text-mining and provide unique services to the research community.



# Which data does the ORD pilot apply to?

The data, including associated metadata, needed to validate the results presented in scientific publications

Other curated and/or raw data, including associated metadata, as specified in the data management plan

Doesn't apply to all data (researchers to define as appropriate)

Don't have to share data if inappropriate - exemptions apply

# Key requirements of the ORD pilot

Beneficiaries participating in the ORD pilot will:

- **Deposit** data in a research **data repository**
- Take measures to **enable third parties** to access, mine, exploit, reproduce and disseminate (free of charge for any user) this research data
- Provide information via the chosen repository about **tools and instruments necessary for validating the results** (where possible, provide the tools and instruments themselves)

# Exemptions – reasons for opting out

- If results are expected to be **commercially or industrially** exploited
- If participation is incompatible with the need for **confidentiality** in connection with security issues
- Incompatible with existing rules on the **protection of personal data**
- Would **jeopardise the achievement** of the main aim of the action
- If the **project will not generate / collect** any research data
- If there are other **legitimate reasons** to not take part in the Pilot

Projects can opt out at any stage

Can opt out totally or partially (i.e. for some data only)

Should describe issues in the project DMP

# Data Management Plans

Projects participating in the pilot will be required to develop a Data Management plan (DMP), in which they will specify what data will be open.

Note that the Commission does NOT require applicants to submit a DMP at the proposal stage.

A DMP is therefore NOT part of the evaluation.

DMPs are a deliverable.

# Information on RDM: what and when

## PROPOSAL STAGE

Where relevant\*, H2020 proposals can include a section on data management which is evaluated under the criterion 'Impact'

- What types of data will the project generate/collect?
- What standards will be used?
- How will this data be shared/made available? If not, why?
- How will this data be curated and preserved?

\* For "Research and Innovation actions" and "Innovation Actions"

## IN PROJECT

DMPs are a project deliverable for those participating in the open data pilot.

Not a fixed document – should evolve and gain precision

- Deliver first version within initial 6 months of project
- More elaborate versions whenever important changes to the project occur. At least at the mid-term and final review.

# How can researchers make data open?



<https://okfn.org>

## 1. Choose the dataset(s) to share

- What can be made open? This step may need to be revisited if problems are encountered later.

## 2. Apply an open license

- Determine what IP exists. Apply a suitable licence e.g. CC-BY

## 3. Make the data available

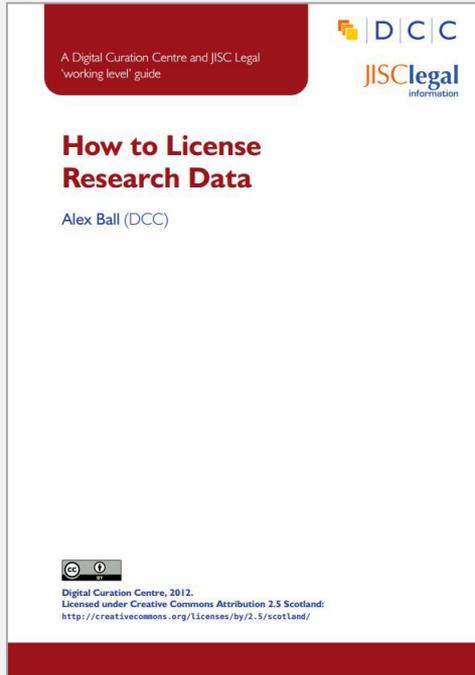
- Provide the data in a suitable format. Use repositories.

## 4. Make it discoverable

- Post on the web, get a unique ID, register in catalogues...



# Licensing research data openly



This DCC guide outlines the pros and cons of each approach and gives practical advice on how to implement a data licence

Source:

<http://www.dcc.ac.uk/resources/how-guides/licence-research-data>

# Which licenses are appropriate?

Creative Commons clauses that limit sharing



NC NonCommercial



**What counts as commercial?**

ND NoDerivatives

**Severely restricts use**

**These clauses are not open licenses**

Horizon 2020 Open Access guidelines point to:

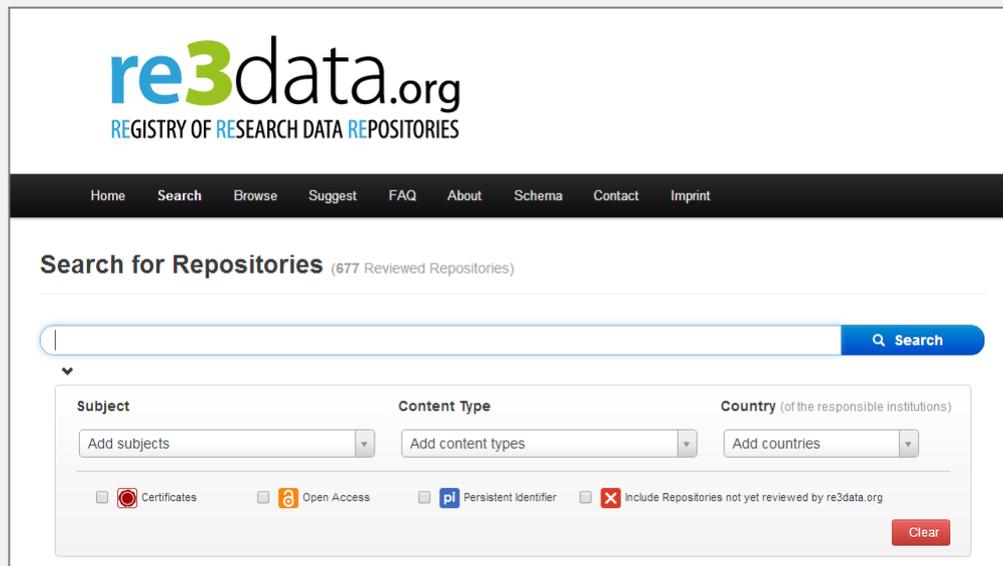


or



# Deposit in research data repositories

The EC guidelines point to Re3data as one of the registries that can be searched to find a home for data



The screenshot shows the re3data.org website interface. At the top, the logo "re3data.org" is displayed in blue and green, with the tagline "REGISTRY OF RESEARCH DATA REPOSITORIES" below it. A navigation bar contains links for Home, Search, Browse, Suggest, FAQ, About, Schema, Contact, and Imprint. The main section is titled "Search for Repositories (677 Reviewed Repositories)". It features a search input field with a magnifying glass icon and a "Search" button. Below the search field are three dropdown menus for "Subject", "Content Type", and "Country (of the responsible institutions)". At the bottom, there are four checkboxes: "Certificates" (unchecked), "Open Access" (checked), "Persistent Identifier" (unchecked), and "Include Repositories not yet reviewed by re3data.org" (checked). A "Clear" button is located at the bottom right of the filter section.

Source: <http://service.re3data.org/search>

# Zenodo

Zenodo is a multi-disciplinary repository that can be used for the long-tail of research data

- Multidisciplinary repository accepting
  - Multiple data types
  - Publications
  - Software
- Assigns a Digital Object Identifier (DOI)
- Links funding, publications, data & software



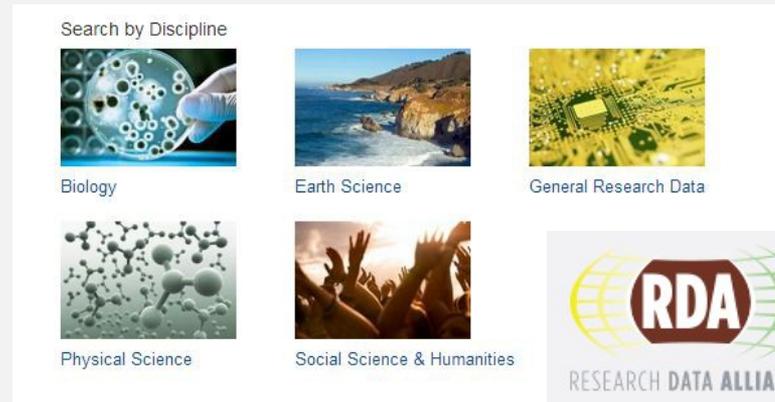
(Source: [www.zenodo.org](http://www.zenodo.org))

# Metadata and documentation

Metadata is basic descriptive information to help others identify and understand the structure of the data e.g. title, author...

Documentation provides the wider context e.g. the methodology / workflow, software, tools and any information needed to understand and reuse the data

Relevant standards should be used for interoperability – check out the Metadata Standards Directory from the Research Data Alliance



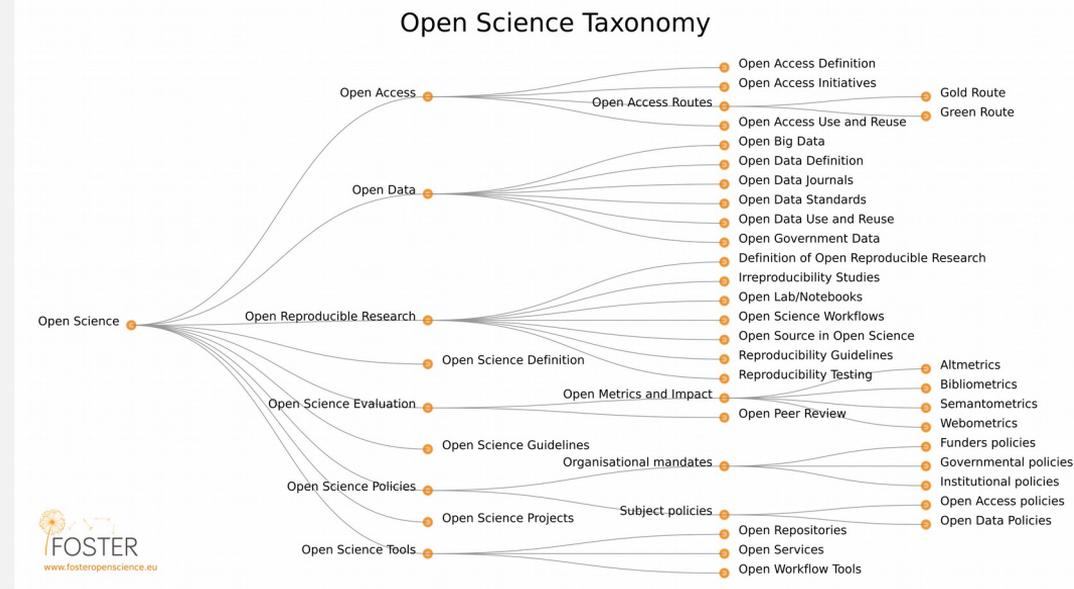
(Source: <http://rd-alliance.github.io/metadata-directory>)



# FOSTER project

Facilitate Open Science Training for European Research

- Network of open access trainers
- Programme of open science courses
- Portal to training materials
- E-learning courses on open access and open data



(Source: [www.fosteropenscience.eu](http://www.fosteropenscience.eu))



# FOSTER Course

## The Horizon 2020 Open Research Data pilot

A course for researchers, research support staff and project officers.

Learning objectives:

1. Understand what is required of participants in the H2020 Open Research Data pilot
2. Learn about the concepts of open data, metadata, licensing and repositories
3. Identify key resources and services that can help you to comply with requirements
4. Undertake short tests to check your understanding

You are not logged in, in order to take the course you need to [login](#)

[Go to the Course forum](#)

## Full details

Level of knowledge: Introductory: no previous knowledge is required

## Topics



## Slides attribution:

Slides 12-24 are based on "The Horizon2020 Open Data Pilot" by Sara Jones, Digital Curation Center [https://  
www.slideshare.net/sjDCC/h2020-open-data-pilot](https://www.slideshare.net/sjDCC/h2020-open-data-pilot)

Thank you!