Horizon 2020 and the Open Research Data pilot

Sarah Jones
Digital Curation Centre, Glasgow
sarah.jones@glasgow.ac.uk
Twitter: @sjDCC

Data Management Plan (DMP) workshop, e-infrastructures Austria, Vienna, 17 November 2016

The EC Open Research Data pilot

Key sources of information

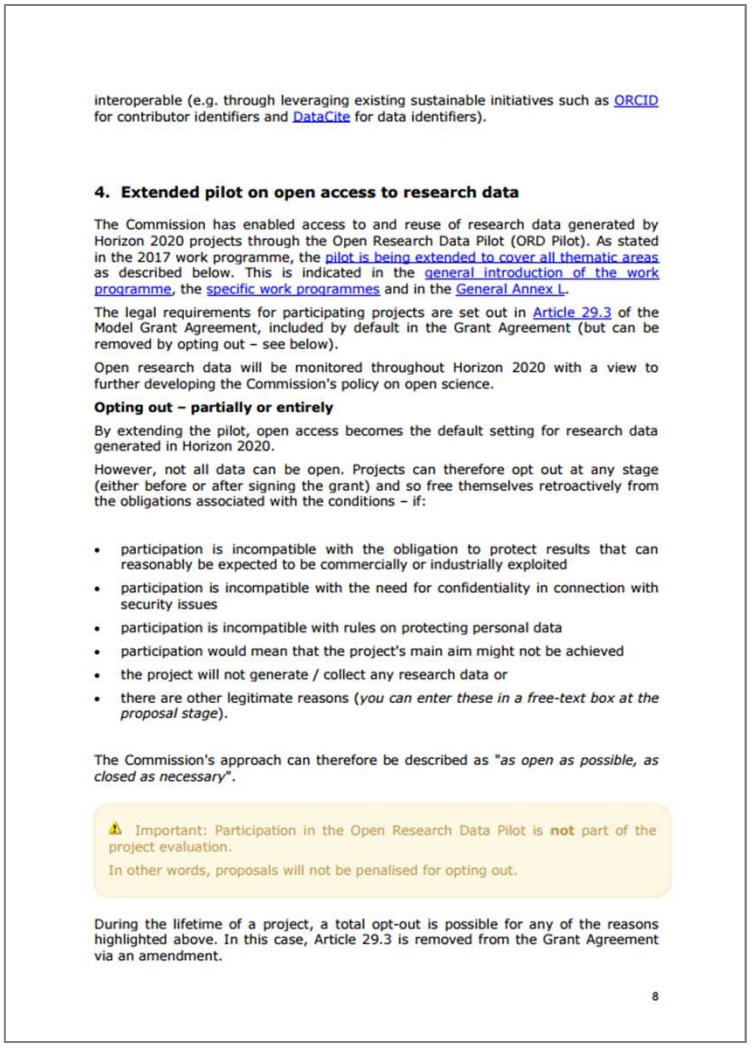
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
  http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

- Guidelines on FAIR Data Management in Horizon 2020
  http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

- Annotated model grant agreement, clause 29.3
  http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

- New infographic summarising key policy points
  http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf

Guidelines on open access



interoperable (e.g. through leveraging existing sustainable initiatives such as ORCID for contributor identifiers and DataCite for data identifiers).

4. Extended pilot on open access to research data

The Commission has enabled access to and reuse of research data generated by Horizon 2020 projects through the Open Research Data Pilot (ORD Pilot). As stated in the 2017 work programme, the pilot is being extended to cover all thematic areas as described below. This is indicated in the general introduction of the work programme, the specific work programmes and in the General Annex L.

The legal requirements for participating projects are set out in Article 29.3 of the Model Grant Agreement, included by default in the Grant Agreement (but can be removed by opting out – see below).

Open research data will be monitored throughout Horizon 2020 with a view to further developing the Commission's policy on open science.

Opting out – partially or entirely

By extending the pilot, open access becomes the default setting for research data generated in Horizon 2020.

However, not all data can be open. Projects can therefore opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions – if:

- participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
- participation is incompatible with the need for confidentiality in connection with security issues
- participation is incompatible with rules on protecting personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate / collect any research data or
- there are other legitimate reasons (*you can enter these in a free-text box at the proposal stage*).

The Commission's approach can therefore be described as "*as open as possible, as closed as necessary*".

⚠ Important: Participation in the Open Research Data Pilot is **not** part of the project evaluation.

In other words, proposals will not be penalised for opting out.

During the lifetime of a project, a total opt-out is possible for any of the reasons highlighted above. In this case, Article 29.3 is removed from the Grant Agreement via an amendment.

8

- Includes a section on the pilot
- Explains opting out
- Notes exceptions (funding instruments that aren’t included)
- Explains types of data
- Specifies requirements and support via EUDAT & OpenAIRE
- Defines research data and open access routes (section 2)

FAIR Data Management guidelines

- Notes the extension of the pilot
- Clarifies concept of FAIR data
- Explains what a DMP is and when they should be updated
- Notes what happens at proposal, submission and evaluation
- Explains costs are eligible
- Provides a DMP template

Model grant agreement

29.3 Open access to research data

[OPTION for actions participating in the open Research Data Pilot: Regarding the digital research data generated in the action ('data'), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

(ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

(b) Provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves)

The following European Commission branded slides come from the EC's open access team and provide an overview to the key points. Content from Jean-Francois Dechamp and colleagues.

Mail: RTD-open-access@ec.europa.eu

Web: http://ec.europa.eu/research/openscience/index.cfm

Twitter: @OpenAccessEC

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

a public benefit

encourages research across scientific fields

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**

RESEARCH DATA - OPEN BY DEFAULT

Research and Innovation

European Commission

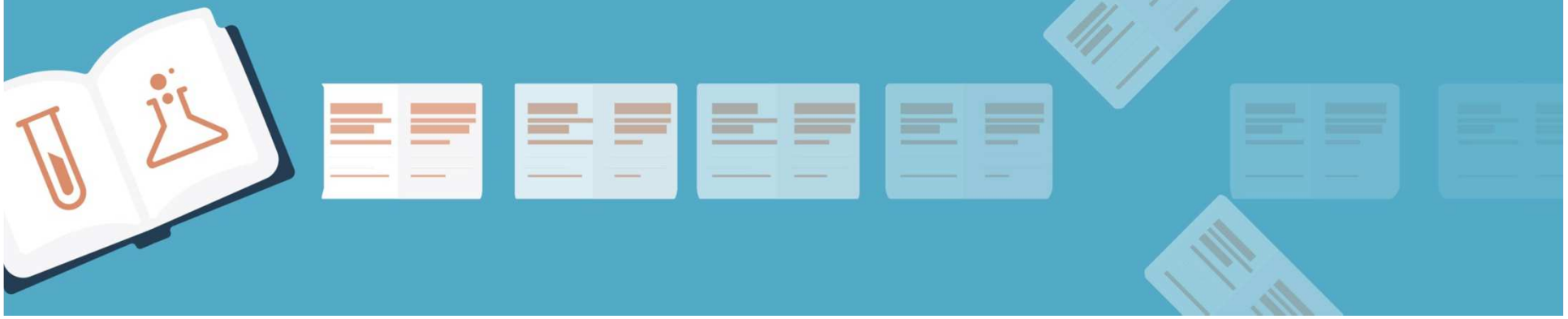
RESEARCH DATA - OPEN BY DEFAULT

Horizon 2020 grantees are required

take measures to ensure open access to the **data underlying their scientific publications**

provide open access to **any other research data of their choice**

Horizon 2020 grantees are **encouraged to also share datasets** beyond publication



Research and Innovation

European Commission

RESEARCH DATA - OPEN BY DEFAULT

Projects must have

Provides information on:

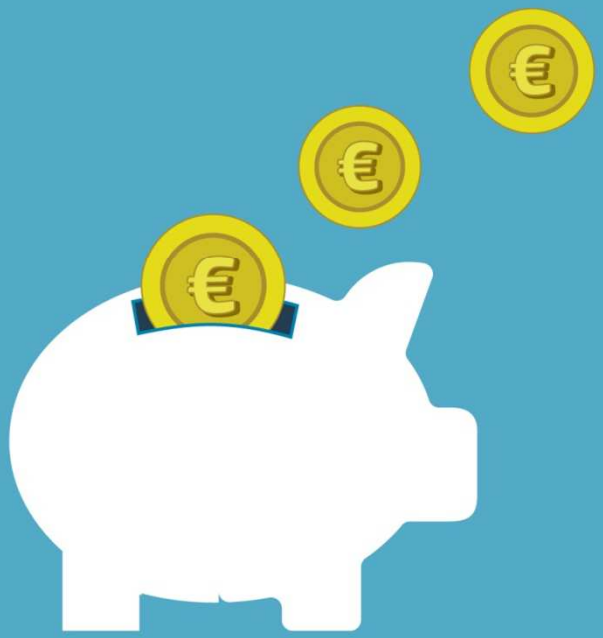
the data the research will generate

how to ensure its curation, preservation and sustainability

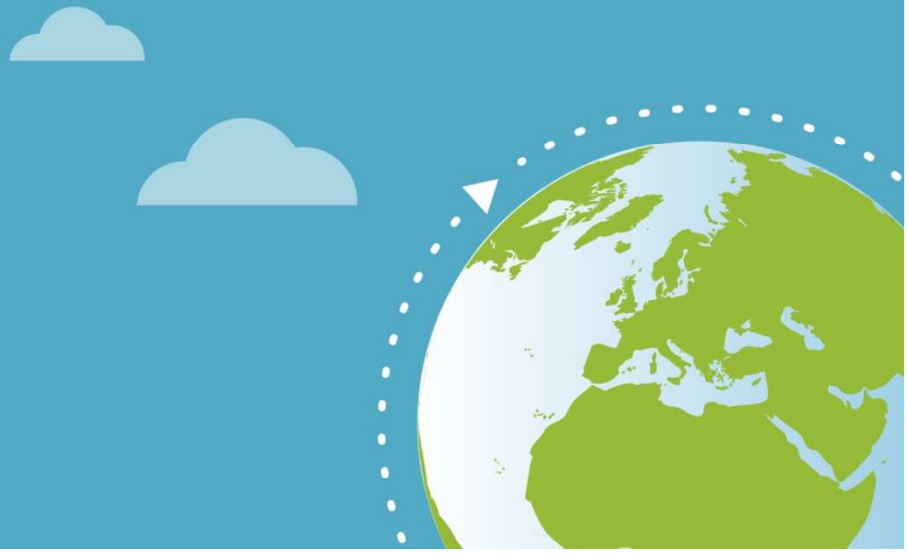
what parts of that data will be open (and how)

RESEARCH DATA - OPEN BY DEFAULT

Data management costs are fully eligible for funding



No repository imposed: deposit data where you want



AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

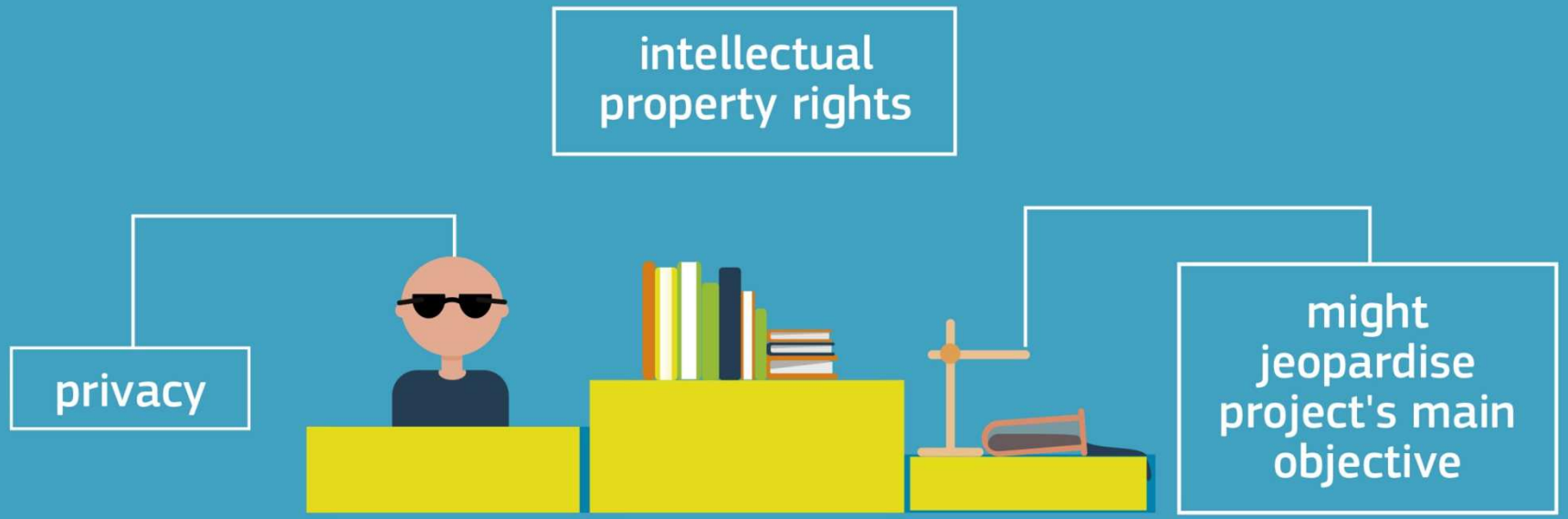
Grantees have the right to **opt-out**, but need to say **why**

Research and Innovation

European Commission

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Top three reasons for **opt-out**:



AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

The approach has been tested during a Horizon 2020 pilot action

2015

of 431 signed projects

opted to share data

from 2017

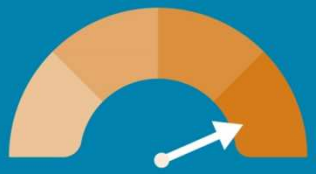
the current
Open Research Data Pilot
expands to cover all areas
of Horizon 2020,
with the same rules

HOW IT WORKS


Get funding
Write DMP
Gather data during research
Finish research
Publish results
Choose repository and deposit data
Inform OpenAIRE the EU-funded Open Access portal

BE PART OF THE NEW ERA OF OPEN SCIENCE

reach more people, have greater impact

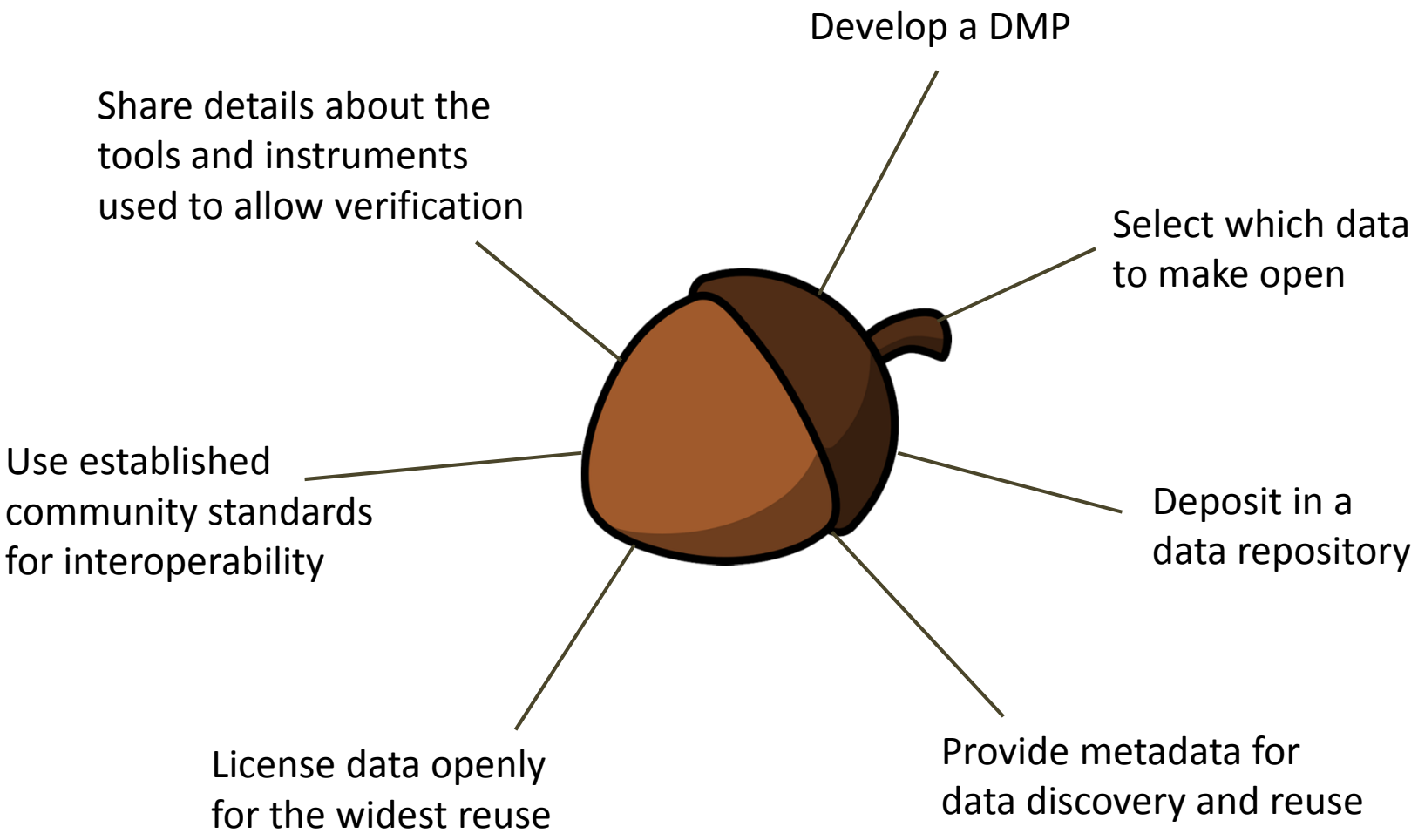


avoid duplication of efforts

preserve data for future researchers

simplify final Horizon 2020 reporting thanks to an up-to-date DMP

Requirements in a nutshell

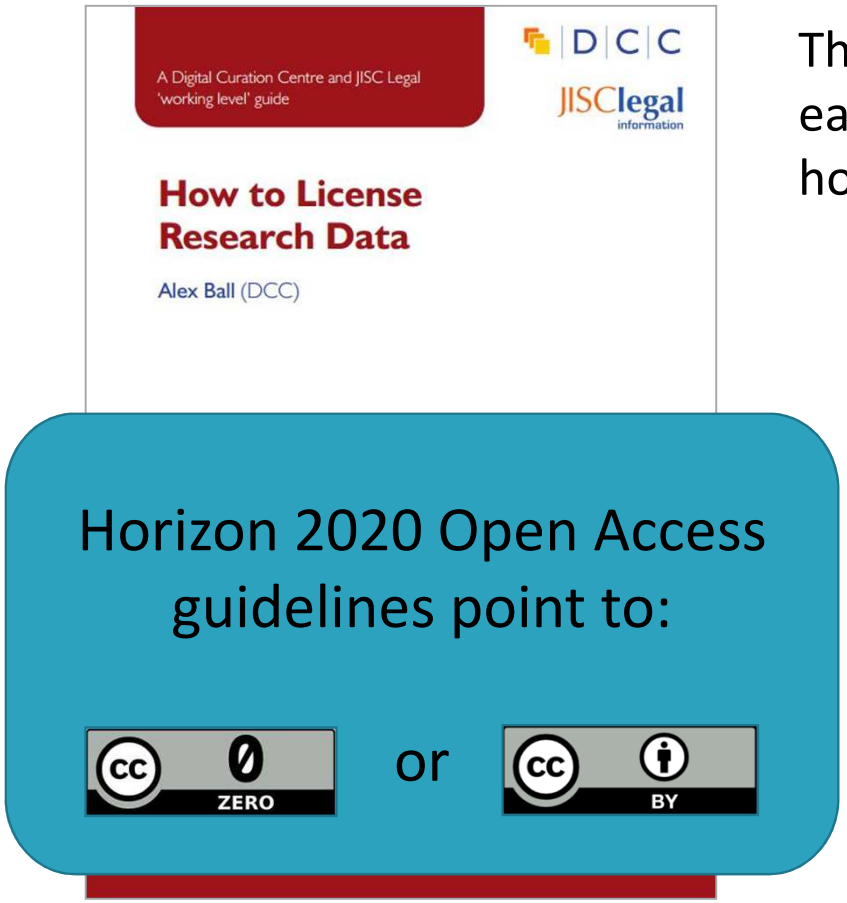


How to make data open?

https://okfn.org

1. Choose your dataset(s)
   - What can you may open? You may need to revisit this step if you encounter problems 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, register in catalogues…

License research data openly



Horizon 2020 Open Access guidelines point to:

CC ZERO or CC BY

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

CREATIVE COMMONS LIMITATIONS

NC Non-Commercial
What counts as commercial?

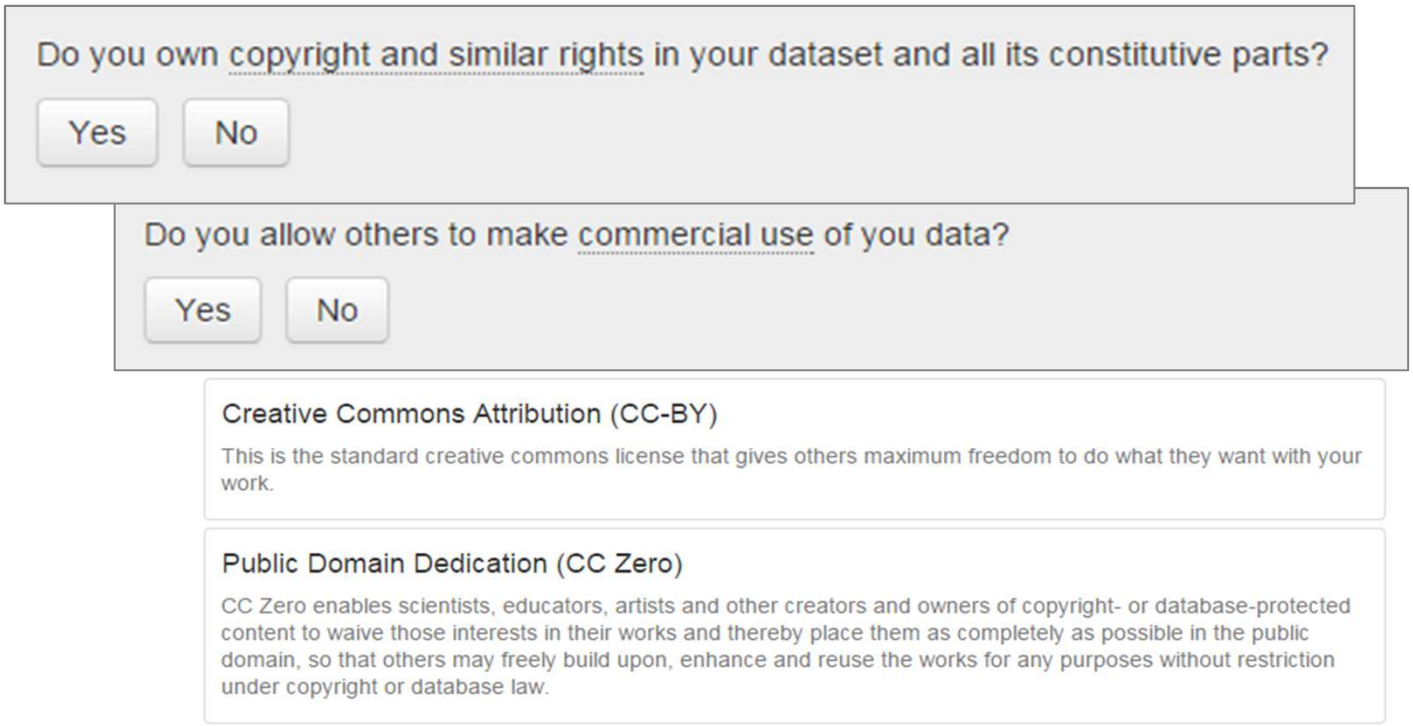
ND No Derivatives
Severely restricts use

These clauses are not open licenses

www.dcc.ac.uk/resources/how-guides/license-research-data

EUDAT licensing tool

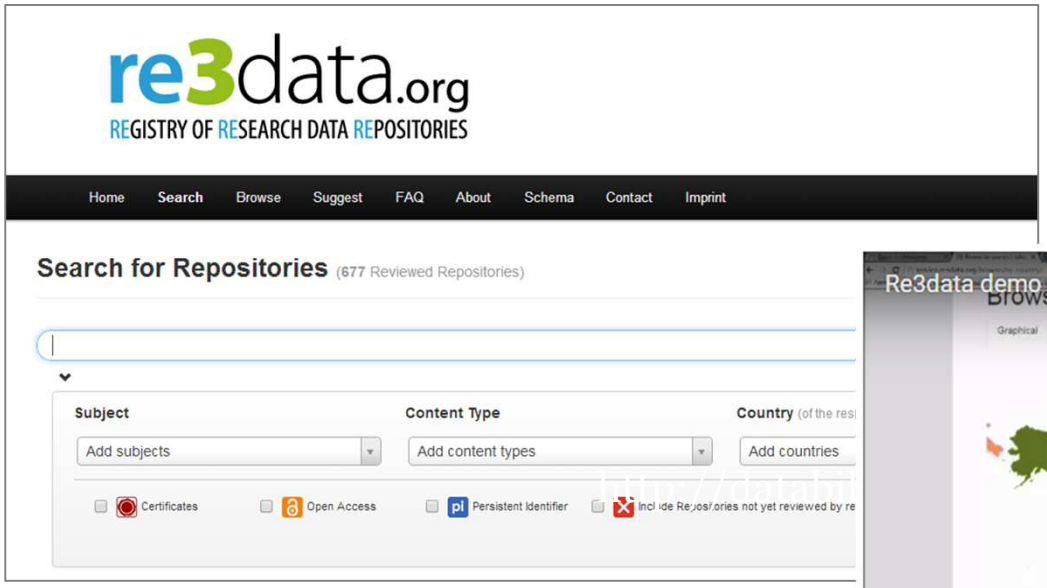
Answer questions to determine which licence(s) are appropriate to use



http://ufal.github.io/lindat-license-selector

Deposit in a data repository

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



http://service.re3data.org/search

www.fosteropenscience.eu/content/re3data-demo

How to select a repository?

- Look for provision from your community, university, publisher, funder etc
- Check they match your particular data needs: e.g. formats accepted; mixture of Open and Restricted Access.
- See if they provide guidance on how to cite the deposited data.
- Do they assign a persistent & globally unique identifier for sustainable citations and to links back to particular researchers and grants?
- Look for certification as a '*Trustworthy Digital Repository*' with an explicit ambition to keep the data available in long term.

www.openaire.eu/opendatapilot-repository

Zenodo

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

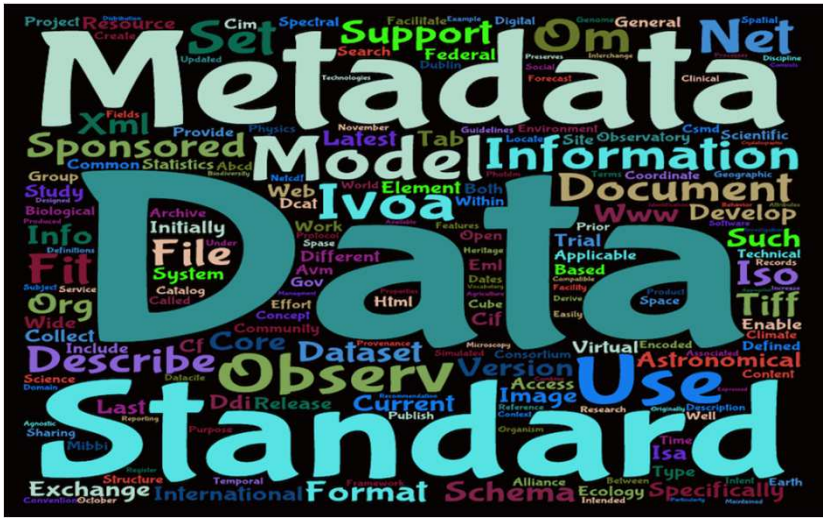
- An OpenAIRE-CERN joint effort
- Multidisciplinary repository accepting
  - Multiple data types
  - Publications
  - Software
- Assigns a Digital Object Identifier (DOI)
- Links funding, publications, data & software

www.zenodo.org

Use metadata standards

Metadata Standards Directory

Broad, disciplinary listing of standards and tools. Maintained by RDA group



http://rd-alliance.github.io/metadata-directory

Biosharing

A portal of data standards, databases, and policies

Focused on life, environmental and biomedical sciences

https://biosharing.org

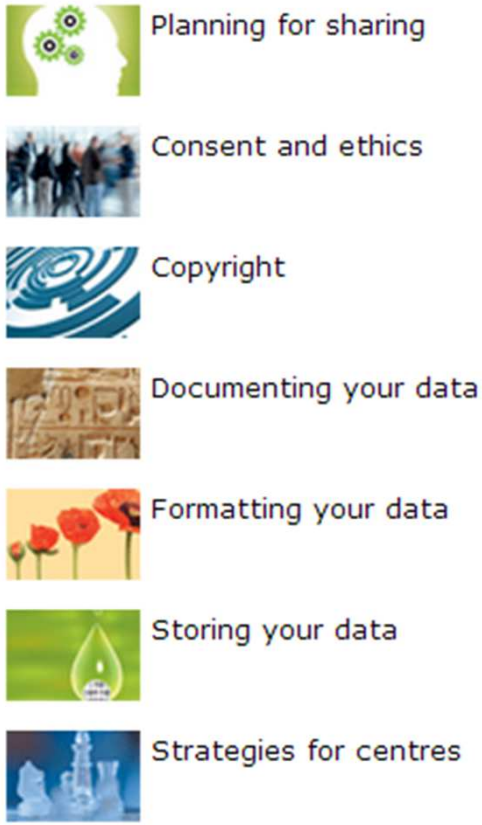
Choose appropriate file formats

If you want your data to be re-used and sustainable in the long-term, you typically want to opt for open, non-proprietary formats.

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

Further examples:
www.data-archive.ac.uk/create-manage/format/formats-table

Managing and sharing data: a best practice guide



http://data-archive.ac.uk/media/2894/managingsharing.pdf

FOSTER

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, open data, open science

www.fosteropenscience.eu

OpenAIRE

Open Access Infrastructure for Research in Europe

- aggregates data on OA outputs
- mines & enriches it content by linking thing together
- provides services & APIs e.g. to generate publication lists or support EC reporting
- lots of guidelines on H2020 Open Data pilot and DMPs

www.openaire.eu

http://vimeo.com/108790101

EUDAT services

EUDAT offers a pan-European solution, providing a generic set of services to ensure minimum level of interoperability

Building common data services in close collaboration with 25+ communities

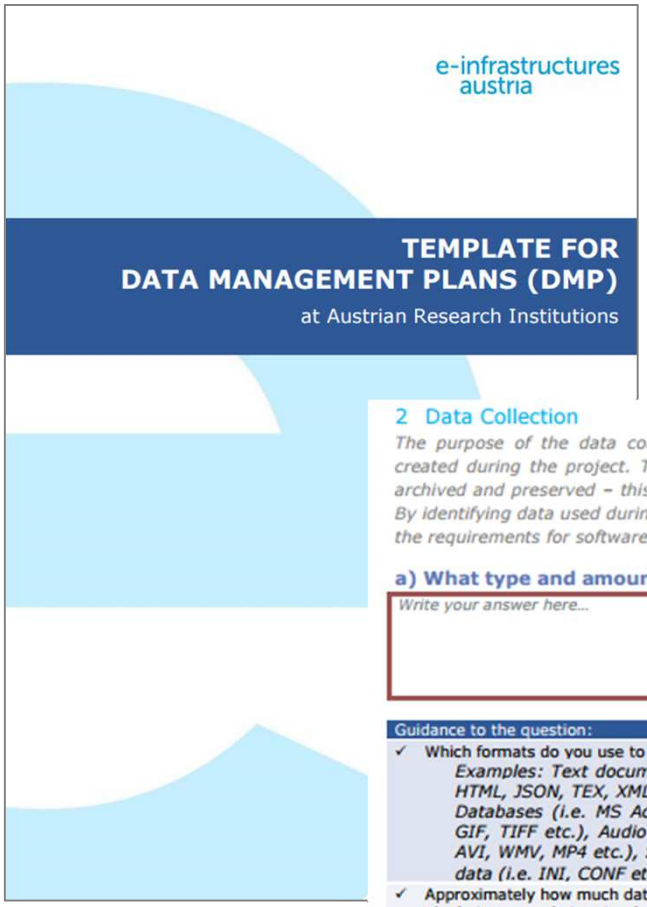
www.eudat.eu

EUDAT

Discipline-specific infrastructure

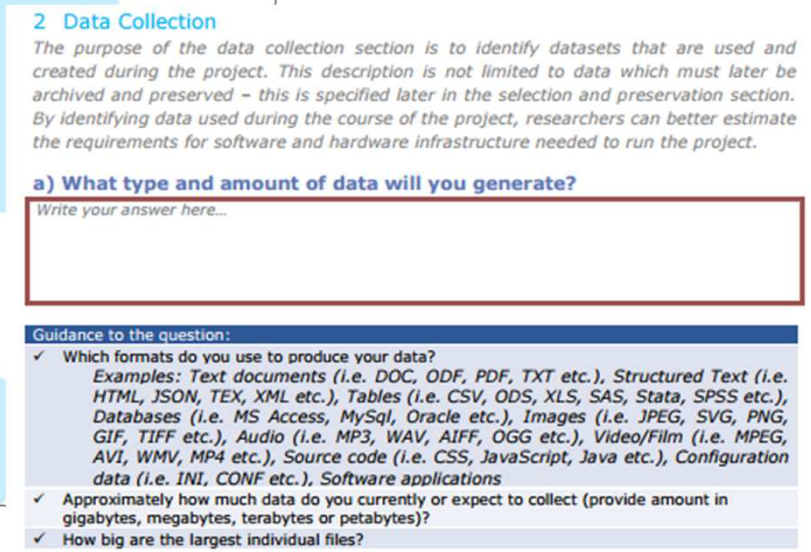
- ELIXIR in life sciences
- BBMRI-ERIC (biobank and biomedical resources)
- Euro Argo ERIC (oceanography)
- CLARIN for language resources
- DARIAH-EU for arts and humanities

Also look for national & local support!



phaidra.org

OpenAIRE


2 Data Collection

The purpose of the data collection section is to identify datasets that are used and created during the project. This description is not limited to data which must later be archived and preserved – this is specified later in the selection and preservation section. By identifying data used during the course of the project, researchers can better estimate the requirements for software and hardware infrastructure needed to run the project.

a) What type and amount of data will you generate?

Write your answer here…

Guidance to the question:

- ✓ Which formats do you use to produce your data?
  Examples: Text documents (i.e. DOC, ODF, PDF, TXT etc.), Structured Text (i.e. HTML, JSON, TEX, XML etc.), Tables (i.e. CSV, ODS, XLS, SAS, Stata, SPSS etc.), Databases (i.e. MS Access, MySql, Oracle etc.), Images (i.e. JPEG, SVG, PNG, GIF, TIFF etc.), Audio (i.e. MP3, WAV, AIFF, OGG etc.), Video/Film (i.e. MPEG, AVI, WMV, MP4 etc.), Source code (i.e. CSS, JavaScript, Java etc.), Configuration data (i.e. INI, CONF etc.), Software applications
- ✓ Approximately how much data do you currently or expect to collect (provide amount in gigabytes, megabytes, terabytes or petabytes)?
- ✓ How big are the largest individual files?

www.openaire.eu/contact-noads

http://e-infrastructures.at/en/the-project/deliverables

Thanks for listening

DCC resources on Data Management

www.dcc.ac.uk/resources

Follow us on twitter:

@digitalcuration and #ukdcc