Open Research Data & DMPs

Sarah Jones

Digital Curation Centre, Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC

"Open Science for Successful Research Grants" workshop, University of Latvia, 24 October 2018

H2020 Open Research Data Pilot

H2020 open data pilot

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

Pilot in H2020

Expanded to open data by default

Provision of support

?
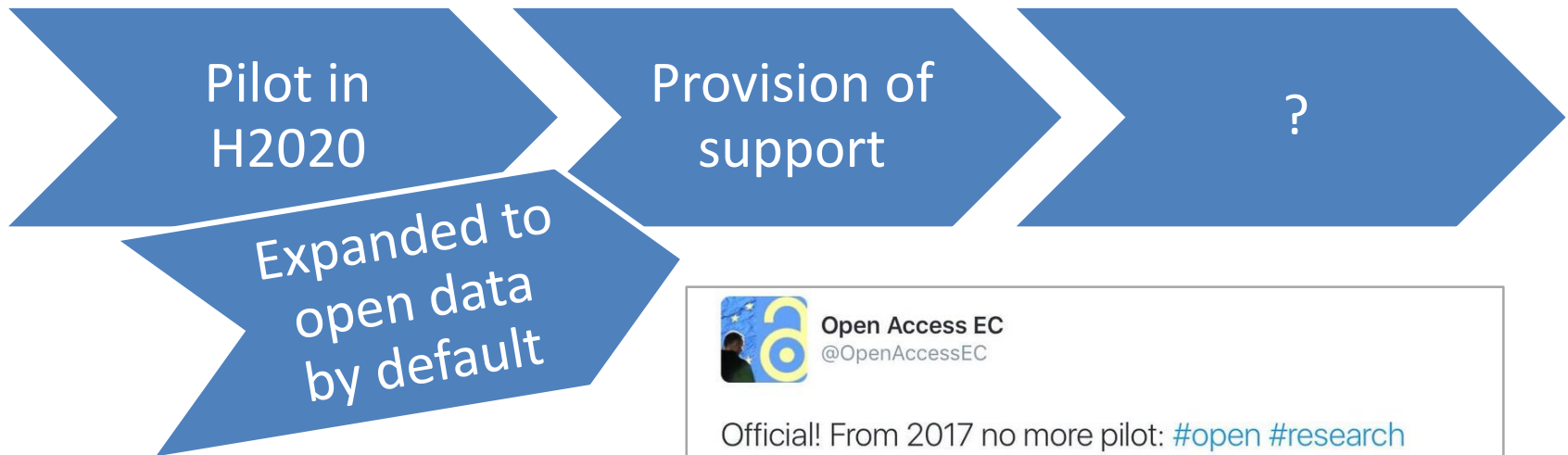
Open Access EC
@OpenAccessEC

Official! From 2017 no more pilot: #open #research #data will become the rule (with opt out) ec.europa.eu/digital-single...
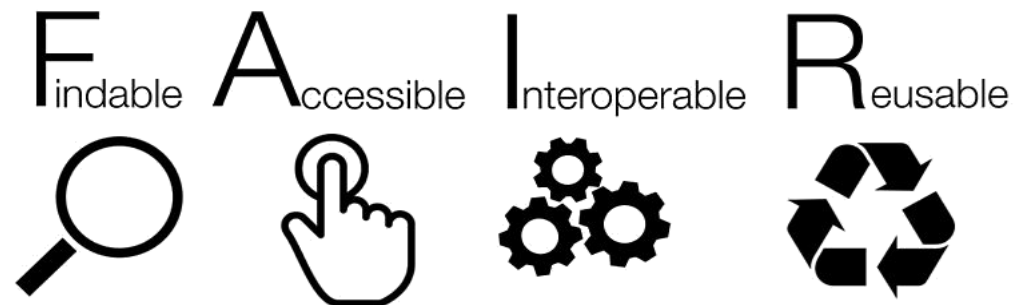
The Commission will make open research data the default option, while ensuring opt-outs, for all new projects of the Horizon 2020 programme. | As of 2017
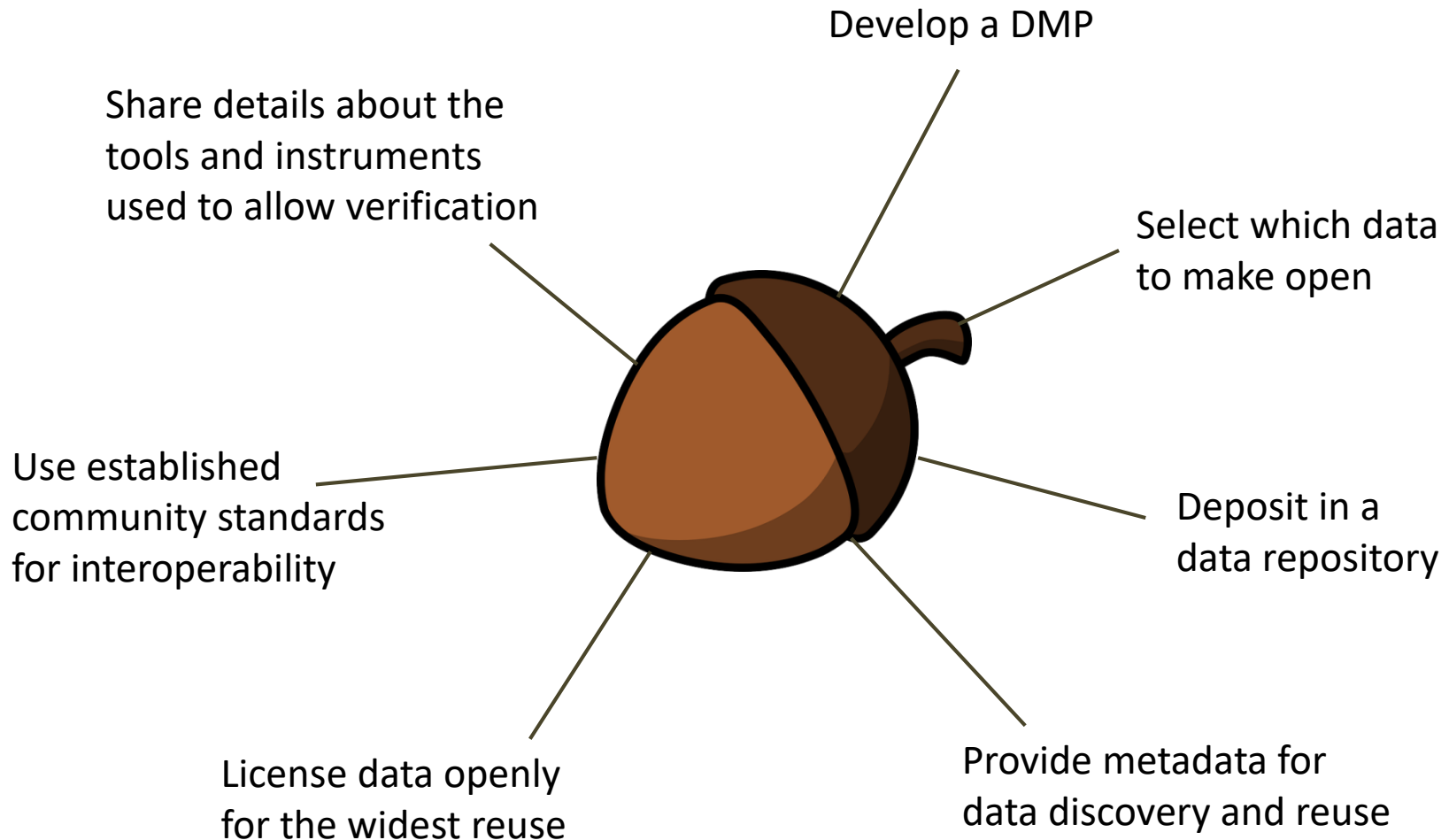
19/04/16 12:17

What is FAIR?

A set of principles that describe the attributes data need to have to enable and enhance reuse, by humans and machines

Findable   Accessible   Interoperable   Reusable

Approach:

as open as possible, as closed as necessary

Requirements in a nutshell

Develop a DMP

Share details about the tools and instruments used to allow verification

Select which data to make open

Use established community standards for interoperability

Deposit in a data repository

License data openly for the widest reuse

Provide metadata for data discovery and reuse

RESEARCH DATA – OPEN BY DEFAULT

Projects must have

DATA MANAGEMENT PLAN (DMP)

Provides information on:

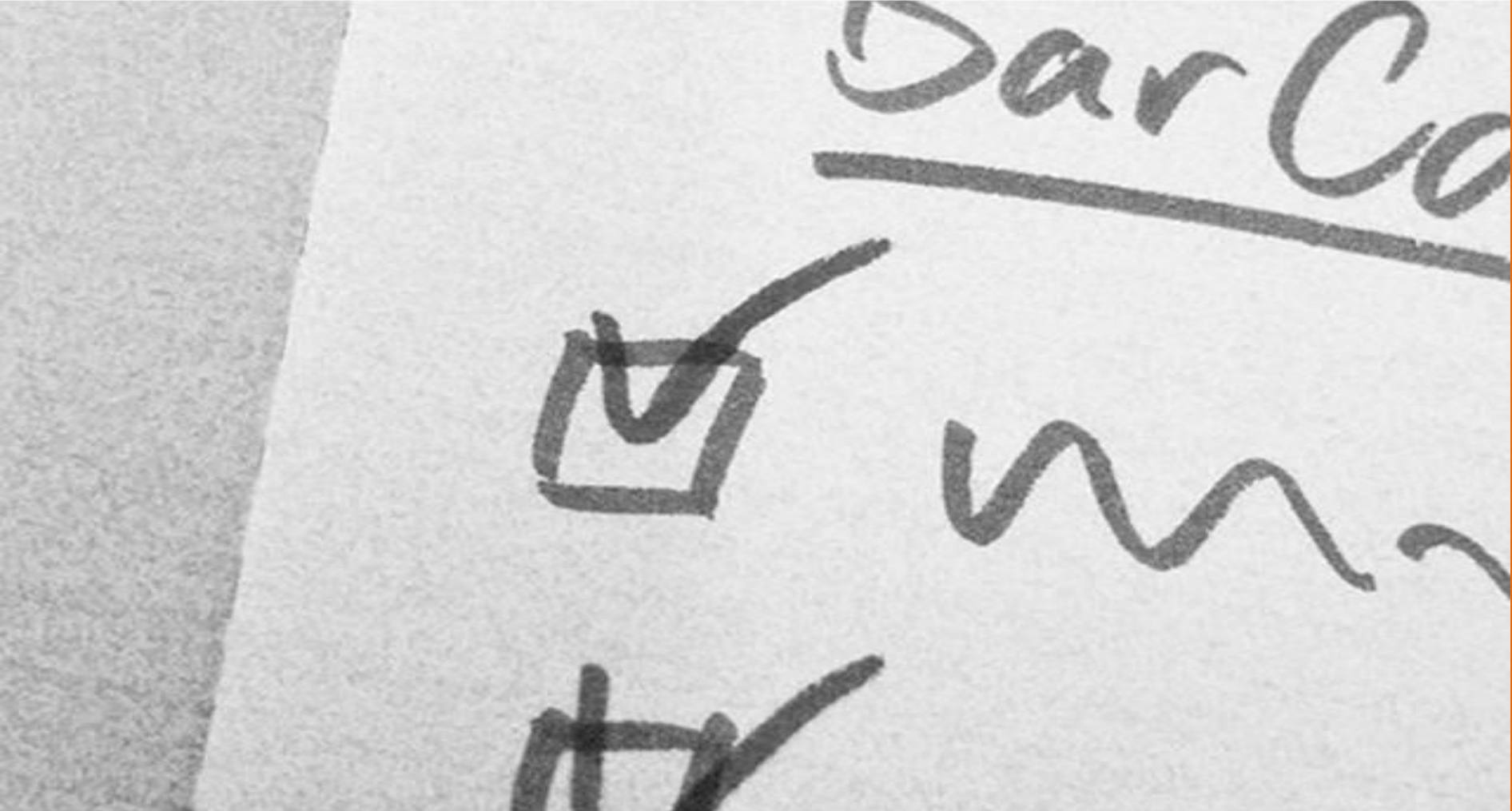the data the research will generate

how to ensure its curation, preservation and sustainability

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

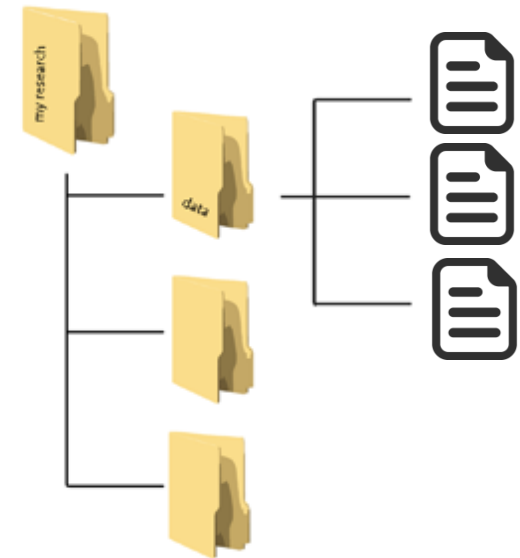**Research and Innovation**

European Commission

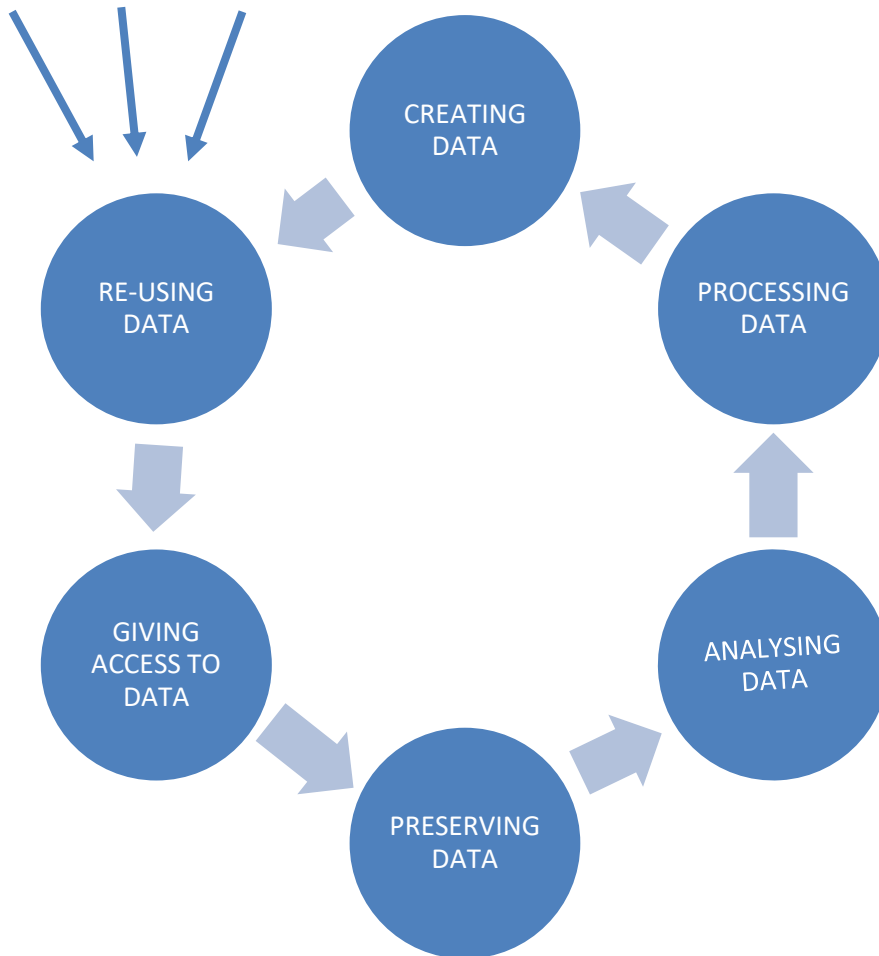# Data Management Planning

What is a DMP?

A DMP is a brief plan to define:

- how the data will be created

- how it will be documented

- who will be able to access it

- where it will be stored

- who will back it up

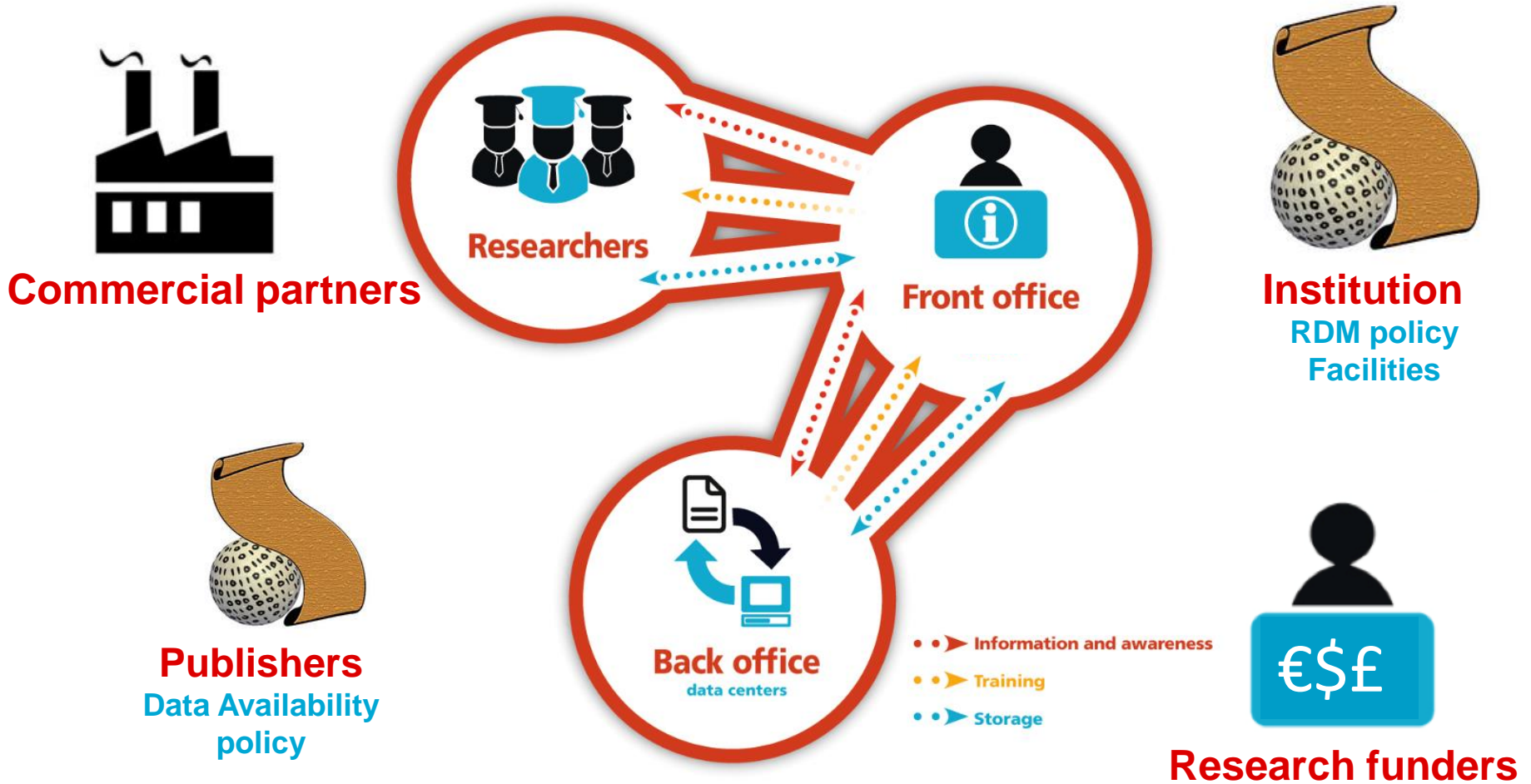- whether (and how) it will be shared & preserved

DMPs are often submitted as part of grant applications, but are useful whenever researchers are creating data.

Planning trick 1: think backwards
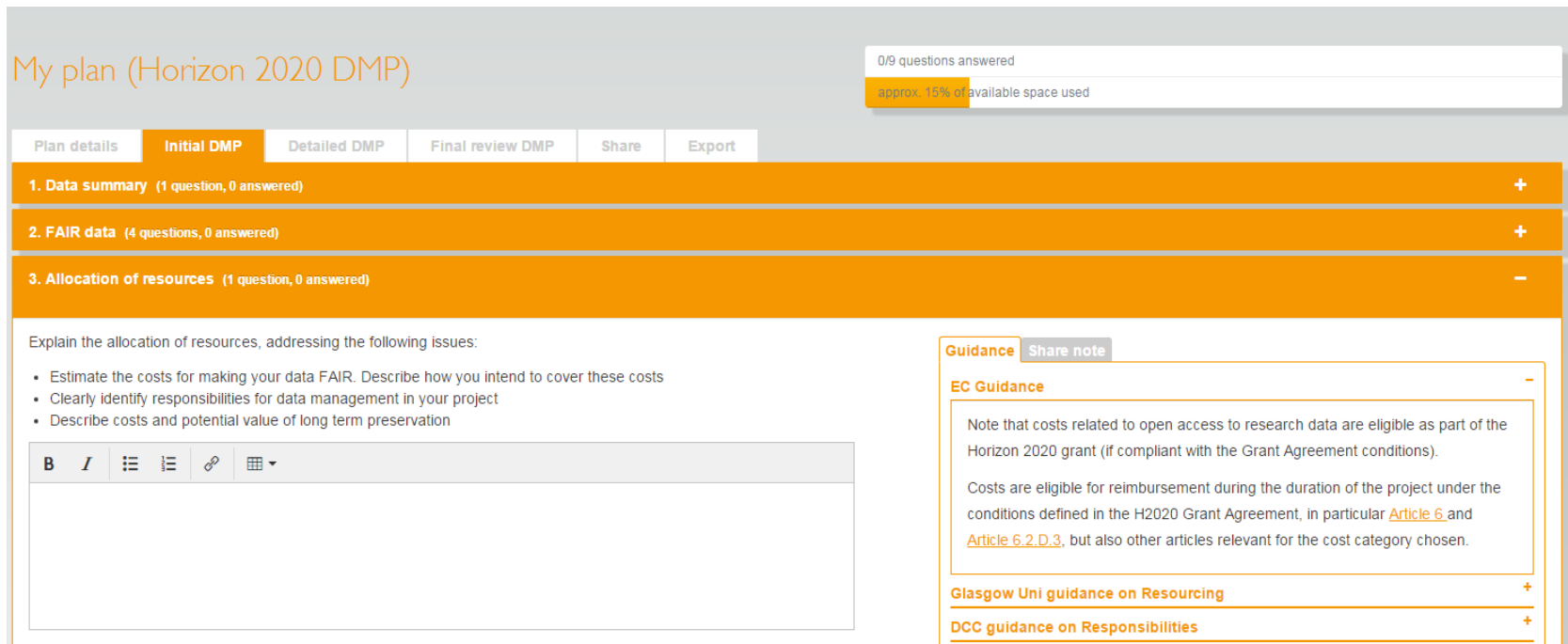
What data organisation would a re-user like?

Planning trick 2: include RDM stakeholders

Commercial partners

Researchers

Front office

Institution
RDM policy
Facilities

Back office
data centers

- Information and awareness
- Training
- Storage

Publishers
Data Availability policy

Research funders

www.openaire.eu/briefpaper-rdm-infonoads

Planning trick 3: ground your plan in reality

Base plans on available skills, support and good practice for the field – show it's feasible to implement

Different services will be applicable depending on the stage of the research and nature and scale of your data….

What is DMPonline?

A web-based tool to help researchers write DMPs

Includes a template for Horizon 2020

https://dmponline.dcc.ac.uk

Options for unis to customise DMPonline

Organisations can:
- Add their own template(s)
- Customise existing funder templates
- Provide example and suggested answers
- Local guidance with links to support and services
- Include their own logo and text in a banner
- Review basic statistics
- …

Key differences in H2020

- The Commission does NOT require applicants to submit a DMP at the proposal stage. It's a deliverable (due by month 6).

- A DMP is therefore NOT part of the evaluation

- Optional section on data management in proposal is worth doing, especially to help justify costs

- A DMP is a living or "active" document that should be updated whenever important changes occur (or at review times)

- DMPs available online are typically quite long, but don't necessarily have to be

H2020 template

1. Data summary

2. FAIR data

   2.1 Making data findable, including provisions for metadata

   2.2 Making data openly accessible

   2.3 Making data interoperable

   2.4 Increase data re-use (through clarifying licences)

3. Allocation of resources

4. Data security

5. Ethical aspects

6. Other issues

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Reviewing DMPs in H2020

- DMPs are a deliverable, checked primarily by project officers and in some cases external reviewers too

- Guidelines are being developed to give reviewers pointers on what to check. These are based on the template.

- The reviewer has access to the full project documentation

- Process is only just evolving and this is a pilot so feedback may be variable initially

European Commission

Making data open

How to make data open?

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 guidelines point to:

CC BY

or

CC ZERO

CREATIVE COMMONS LICENSES

	COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
PUBLIC DOMAIN	✓	✗	✓	✓	✓
CC BY	✓	✓	✓	✓	✓
CC BY-SA	✓	✓	✓	✓	✗
CC BY-ND	✓	✓	✓	✗	✗
CC BY-NC	✓	✓	✗	✓	✓
CC BY-NC-SA	✓	✓	✗	✓	✗
CC BY-NC-ND	✓	✓	✗	✗	✗

✓ You can redistribute (copy, publish, display, communicate, etc.)

✓ You have to attribute the original work

✓ You can use the work commercially

✓ You can modify and adapt the original work

✓ You can choose license type for your adaptations of the work.

EUDAT licensing tool

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

Do you own copyright and similar rights in your dataset and all its constitutive parts?

Yes   No

Do you allow others to make commercial use of you data?

Yes   No

Creative Commons Attribution (CC-BY)
This is the standard creative commons license that gives others maximum freedom to do what they want with your work.
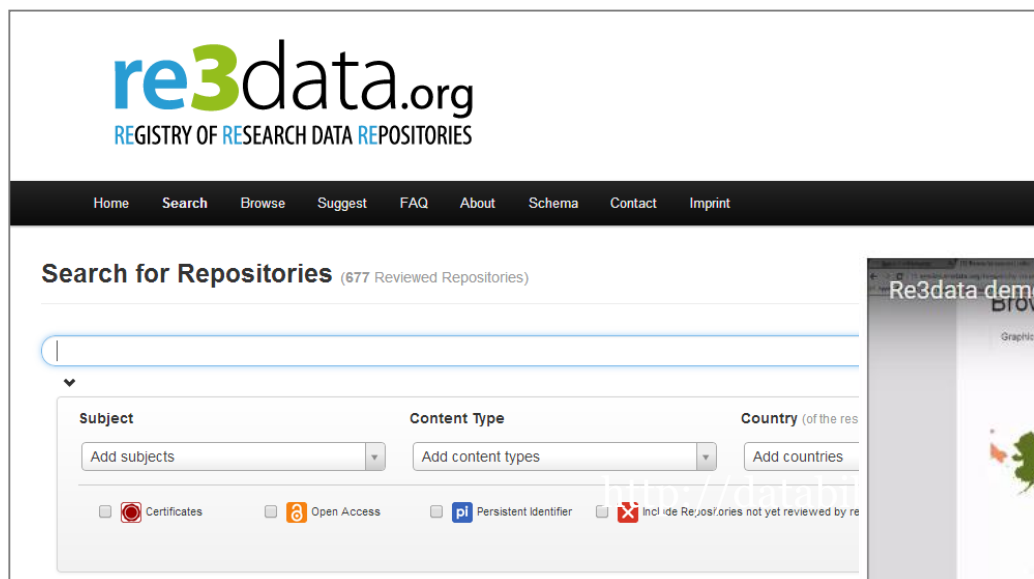
Public Domain Dedication (CC Zero)
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

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

FAIRsharing

A portal of data standards, databases, and policies

Focused on life, environmental and biomedical sciences

https://rdamsc.dcc.ac.uk

https://fairsharing.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

Planning for sharing

Consent and ethics

Copyright

Documenting your data

Formatting your data

Storing your data

Strategies for centres

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

FOSTER and 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

- Focus on disciplinary materials in FOSTER+

www.fosteropenscience.eu

FOSTER toolkit

What is Open Science?
This introductory module will help you to understand what open science is and why it is something you should care about.

Best Practice
This module introduces policies and other environmental factors that influence good practice in open research.

Open Peer Review (OPR)
This module will introduce you to OPR and let you know how you can get started with it.

Data Protection and Ethics
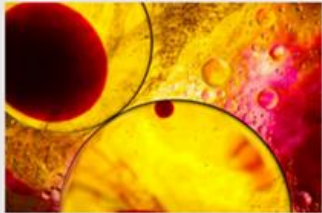This module helps you to get to grips with responsible data sharing.

Licensing
This module helps you to find the best license for your open research outputs.

Open Data
In this module, you'll focus on which data you can share and how you can go about doing this most effectively.

OSS and Workflows
This module introduces Open Source Software (OSS) and workflows as an emerging but critical component of Open Science.
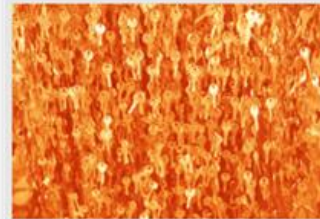
Open Innovation
This module will show you how Responsible Research and Innovation is accelerated through Open Science.

Open Access Publishing
This module will help you become skilled in Open Access publication in the wider context of Open Science.

Preprints
This module introduces the practice of sharing preprints and helps you to see how it can support your research.

https://www.fosteropenscience.eu/toolkit

Thanks for listening

DCC resources on Data Management

www.dcc.ac.uk/resources

Follow us on twitter:

@digitalcuration and #ukdcc