



---

## INTRODUCTION

DMP training European Commission

September/October 2018

Gwen Franck (FOSTER)



# Training Programme

Recap (Gwen Franck)

RDM concepts and tools (Sarah Jones)

- FAIR data / Lifecycle / What is RDM
- Data formats
- Metadata - standards
- Licensing
- Data repositories
- Persistent identifiers

Reviewing DMPs - what you need to know. (Sarah Jones/Venkataraman Shanmugasundaram)

Break

Exercise on DMPs and feedback

# Recap and introduction

The FOSTER project

What do you know about Research Data Management, Data Management Plans and the EC Policies?



# The FOSTER project

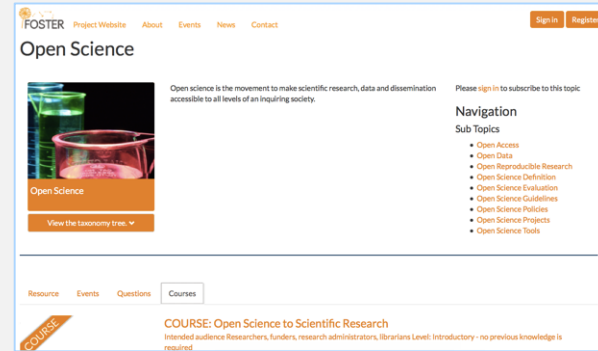
Facilitate Open Science Training for European Research

Phase 1 (2014-2016): Spread the Seeds of Open Science and Open Access

Creation of **Open Science Taxonomy**

**2000+** training materials, categorized in the **FOSTER Portal**

More than **100 f2f training events** in **28 countries** and **25 online courses**, totalling more than **6300 participants**



<http://fosteropenscience.eu>





# The FOSTER project

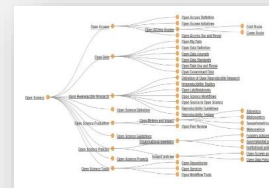
Facilitate Open Science Training for European Research

Phase 2 (2017-2019): ‘Let the Flowers of Open Science Bloom’  
Focus on:

- Training for the **practical implementation** of Open Science (face to face and online) including RDM and Open Data
- Developing intermediate/advanced level/discipline-specific **training resources** in collaboration with three disciplinary communities (and related RIs): Life Sciences (ELIXIR), Social Sciences (CESSDA) and Humanities (DARIAH)
- Update the FOSTER Portal to support moderated learning, badges and gamification

In concrete terms:

- **150** new training resources
- **Over 50 training events** (outcome-oriented, providing participants with tangible skills) and **20 e-learning courses**
- Multi-module **Open Science Toolkit**
- Trainers Network, Open Science Bootcamp, Open Science Training Handbook, and more...



Q1: Why is RDM so important?

Troughout the training: ask a question!



# RDM and research: the primary benefits

RDM helps preserve, protect and proliferate the data behind scientific (research) discoveries and claims - first and foremost it is a **QUALITY** issue...

- When research data is managed actively and responsibly, the evidence that underpins research can be made open for anyone to scrutinise, and attempt to replicate findings. This leads to a more robust scholarly record, and helps discourage and identify academic fraud

A secondary benefit is **PROTECTION**: the rights and legitimate interests of data subjects and IP owners are mindfully protected

- Active and responsible data management reduces the chances of inadvertent data leaks or loss

## Other benefits of RDM

It also has other benefits...

**EFFICIENCY:** Data collection can be funded once, and used many times for a variety of purposes

**ACCESSIBILITY:** Interested third parties can (where appropriate) access and build upon publicly-funded research outputs with minimal barriers to access

**SPEED:** The research process becomes faster

**IMPACT and LONGEVITY:** Data linked to publications receive more citations, over longer periods

**DURABILITY:** Simply put, fewer important datasets will be lost



# Risks of not doing this, or getting this wrong

**LEGAL** - sensitive data is protected by law (and contracts) and needs to be protected

**FINANCIAL** - non-compliance with funder policies can lead to reduced access to income streams

**SCIENTIFIC** - potential discoveries may be hidden away in drawers, on USB sticks or non-networked drives

**OPPORTUNITY COST** - reduced visibility for research > lost opportunities for collaboration

**QUALITY** - the scholarly record becomes less robust

**REPUTATIONAL** - responsible data management is increasingly considered a core element of good scholarly practice in the 21<sup>st</sup> century

# Mentimeter Q2 What is the EC principle when it comes to Open Research Data?

- 'As open as possible, as closed as necessary'
- 'All research data should be made openly available at the moment of publication'
- 'No research data management plan, no funding'

# Mentimeter Q2 What is the EC principle when it comes to Open Research Data?

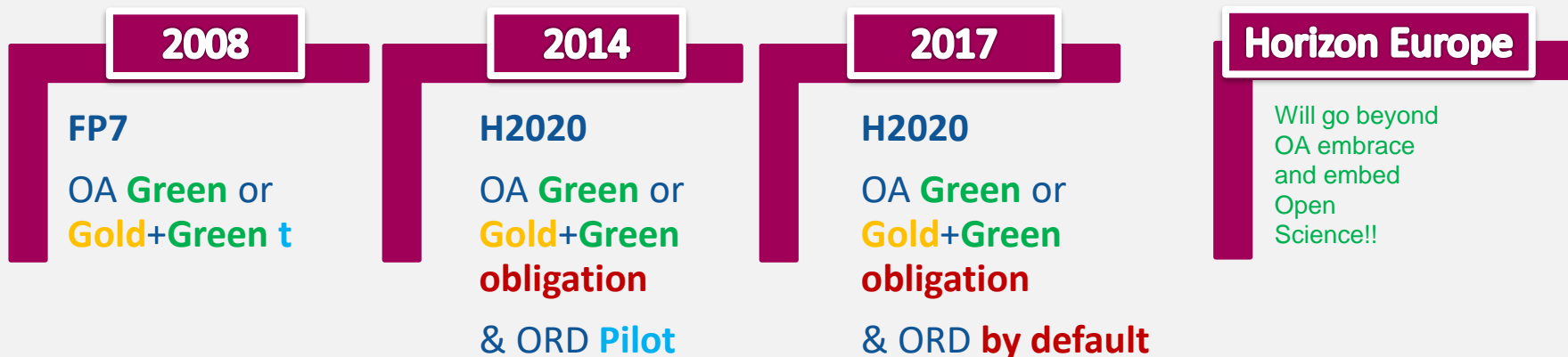
“As of the Work Programme 2017 the Open Research Data pilot is extended to cover all thematic areas of Horizon 2020 per default. However, the Commission recognizes that some research data cannot be made open and applies the principle of 'as open as possible, as closed as necessary'. It is therefore possible to opt out of research data sharing at any stage - before or after the signature of the grant agreement - but reasons have to be given [...] (see General Annex L of the 2017 Work Programme adopted at 25 July 2016).”

[http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm)

## Mentimeter Q3. Which statement is correct?

- ORD was piloted at the start of H2020 and has become the default in 2017
- ORD and RDM are only relevant for data 'underlying' publications
- Participation in the ORD Pilot is a criterium for project evaluation
- DMPs were piloted in FP7 and have become a requirement in H2020

# Mentimeter Q3. Which statement is correct?



*Since participation in the ORD pilot is not an evaluation criterion, the proposal is not expected to contain a fully developed DMP. However, good research data management as such should be addressed under the impact criterion, as relevant to the project. Your application should address the following issues (see [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf))*



# Mentimeter Q3 What kind of data is not required to be discussed in a DMP?

- Fieldwork observations
- **Publication lists**
- Raw data resulting from experiments
- Survey results
- 3D models

# Mentimeter Q4 What kind of data is not required to be discussed in a DMP?

A: Specified "digital research data"

- ‘Digital research data’ is information in digital form (in particular facts or numbers), collected to be examined and used as a basis for reasoning, discussion or calculation; this includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. (H2020 GA)

## Mentimeter Q5. A DMP should NOT include:

- the handling of research data during and after the end of the project
- **an overview of every dataset that will be produced by the project**
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open
- how data will be curated and preserved (including after the end of the project)



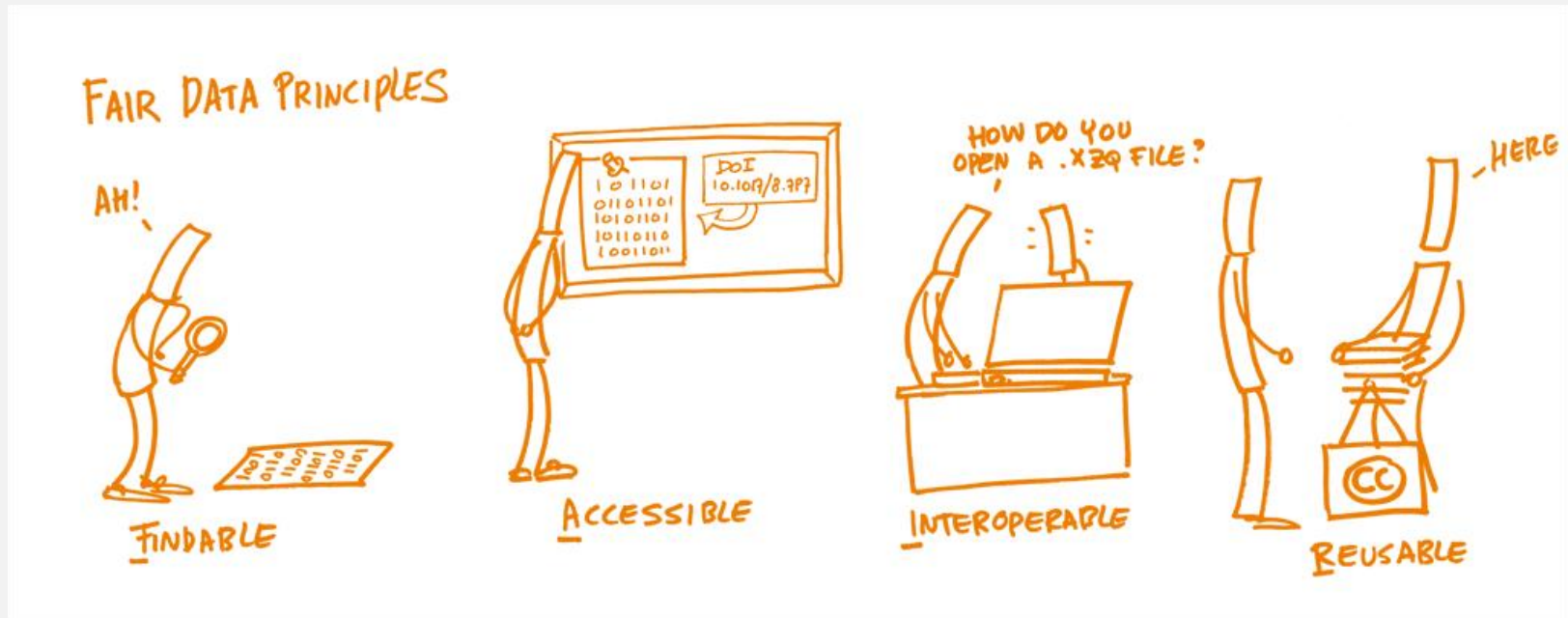
## Mentimeter Q5. A DMP should NOT include:

Ideally, a DMP is short! It should not include a detailed description of every dataset the project plans to produce, but it should identify for each category of dataset the relevant information.

## Mentimeter Q6. What does FAIR stand for?

- Findable, Accountable, Interoperable, Reproducible
- Findable, Accountable, Interoperable, Reusable
- **Findable, Accessible, Interoperable, Reusable**
- Findable, Accessible, Interoperable, Reproducible

# Mentimeter Q6. What does FAIR stand for?



## Mentimeter Q7: Which statement is correct?

- A first version of the DMP is due at M6
- A first version of the DMP needs to be included in the grant proposal
- The DMP needs to be updated at least once a year

# Mentimeter Q7: Which statement is correct?

- A: The DMP **must be updated and completed** (i.e. become more precise) as the project evolves. New versions of the DMP should be created whenever important changes to the project occur, e.g.:
  - New data categories or important new datasets
  - Changes in consortium policies or external factors
  - **Good practise:**
    - DMP outline deliverable M6
    - DMP final version last year of the project
    - M6 deliverable can be reopened for updates

# Mentimeter Q8. What is NOT a good reason to opt out of the ORD Pilot?

- If the project will not generate / collect any data
- **If none of the partners has the technical or financial capacity to participate**
- Conflict with confidentiality/security obligations
- Conflict with rules on protection of personal data
- If the achievement of the action's main objective would be jeopardised

# Mentimeter Q8. What is NOT a good reason to opt out of the ORD Pilot?

- Projects may opt out of the Pilot on Open Research Data in Horizon 2020 in a series of cases (submission stage):
  - If the project will not generate / collect any data
  - Conflict with obligation to protect results
  - Conflict with confidentiality obligations
  - Conflict with security obligations
  - Conflict with rules on protection of personal data
  - If the achievement of the action's main objective would be jeopardised

**Opting out (and opting in) during GAP and even after also possible using the same justifications as during submission stage. (H2020 AGA p. 239 - 240)**

Costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

## AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

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



Top three reasons for opt-out:



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



## 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



## 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)

Data management costs are fully eligible for funding

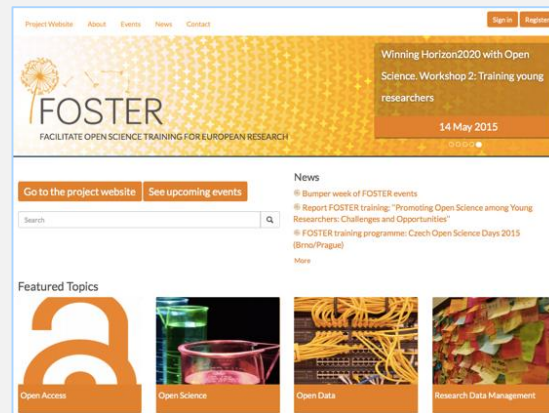
No repository imposed: deposit data where you want



# Contact details

For more information about the FOSTER project:

- Website:  
[www.fosteropenscience.eu](http://www.fosteropenscience.eu)
- Principal investigator: Eloy Rodrigues  
([eloy@sdum.uminho.pt](mailto:eloy@sdum.uminho.pt))
- General enquiries: Gwen Franck  
([gwen.franck@eifl.net](mailto:gwen.franck@eifl.net))
- Events:  
[www.fosteropenscience.eu/events](http://www.fosteropenscience.eu/events)
- Twitter: @fosterscience and #fosteropenscience
- Slides reused from DCC and Victoria Tsoukala.



This work is licensed under the Creative Commons Attribution 2.5 UK: Scotland License.

