

DHARMA Kickoff-Meeting,  
16 September 2019, Berlin



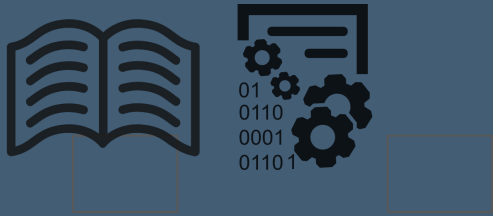
# ERC Rules on Open Access Publication and the Open Data Research Pilot

Emmanuelle Morlock (CNRS, HiSoMA Research Centre)





Making research results openly available is **not an afterthought** when the project is over, but, rather, it is **an effective way of doing the research itself.**



---

## Outline

1. Context: OA policies in Europe
2. The Horizon 2020 OA mandate and ERC specificities
3. Key notions:  
FAIR, DMP, datapapers...

2007 - FP7

OA **pilot for publications**

ERC guidelines  
recommends OA for **data**  
from the beginning



2014 - Horizon 2020

OA **obligation** for publications + **OA data pilot**



2017 - Horizon 2020

OA **by default** for publication and data

2018 - EOSC

**Zooming** on Open Science



sept 2018

Science Europe initiative to  
**accelerate the transition**  
to full Open Access



2021 - Horizon Europe

Open Science **modus operandi**  
(Open data by default)

Evolution of policies in  
the EU funding programs  
for R&I

EVERYONE ON THE ROAD TO OPEN SCIENCE



**Open Science** is the practice of science in such a way that others can **collaborate and contribute**, where research data, lab notes and other research processes are **freely available**, under terms that enable reuse, **redistribution and reproduction** of the research and its underlying data and methods.

**Open research beyond academia:  
innovation + citizen science**

Source: [FOSTER e-learning platform](#)

On the road to Open Science (By Patrick Hochstenbach CC BY 4.0)

# Benefits of Open Science

## EFFICIENCY

Increase research efficiency  
(avoid duplication of research,  
more results with the same  
data, more participation in the  
work.

## KNOWLEDGE

Enables the creation of  
new research questions.

## QUALITY & INTEGRITY

wider evaluation and scrutiny  
by the scientific community.

## COLLABORATION

International, better  
understanding of challenges  
requiring coordinate efforts at  
a large scale.

## VISIBILITY

Share your work and  
make it more visible.

## ECONOMIC BENEFITS

Reduce publishing costs,  
value creation through  
innovation.

# What is Open Access (OA)?

H2020 => Online access to scientific information that is:

- **free of charge** to the end-user
- **reusable**

Scientific information:

- peer-reviewed scientific research **articles**
- research **data**



The EU Framework Programme  
for Research and Innovation

**HORIZON 2020**



**H2020 Programme**

**AGA – Annotated Model Grant Agreement**

Version 5.2  
26 June 2019

**Disclaimer**

This guide is aimed at assisting beneficiaries. It is provided for information purposes only and is to replace consultation of any applicable legal sources. Neither the Commission nor the Executive Agency (any person acting on their behalf) can be held responsible for the use made of this guidance.



**European Research Council**  
Executive Agency

Established by the European Commission



**Horizon 2020**  
European Union Funding  
for Research & Innovation

**European Research Council (ERC)**

**Multi-Beneficiary Model Grant Agreement**

**ERC Starting Grants, Consolidator Grants, Advanced Grants and Synergy Grants**

**(H2020 ERC MGA – Multi)**

Version 5.0  
18 October 2017

**Disclaimer**

This document is aimed at assisting applicants for Horizon 2020 funding. It shows the full range of provisions that may be applied to this type of grant agreement, and is provided for information purposes only. The legally binding grant agreement will be that which is signed by the parties for each action.

# Multi-beneficiaries General Model Grant Agreement

- **29.2.** Open Access to scientific publications
- **29.3.** Open Access to research data



## H2020 OA mandate for publications

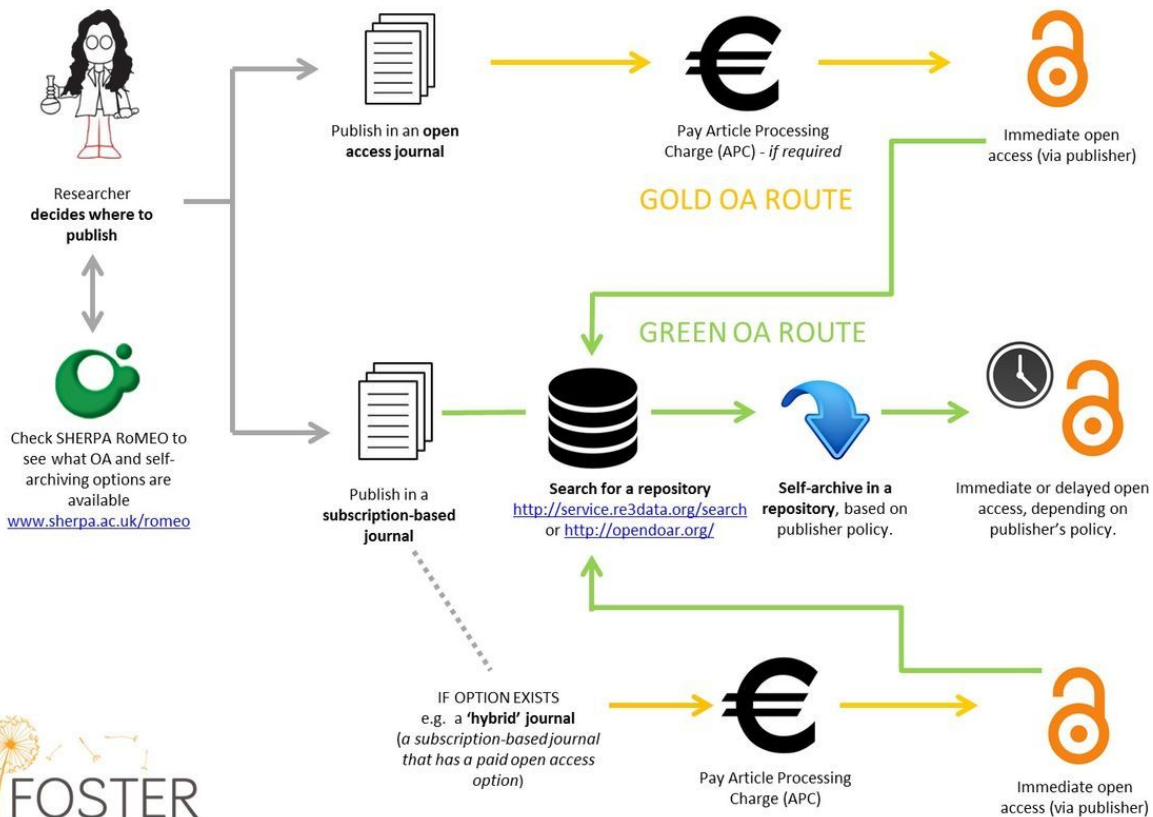


**All peer-reviewed** scientific publications related to results (including monographs and book chapters) must be:

- deposited in a **repository** (even for if in a OA journal already)
- in the form of a **machine readable electronic copy**
- the published **version** or the final peer-reviewed manuscript accepted for publication
- as **soon** as possible
- with identifying **metadata**

Open access costs eligible for reimbursement when the project runs.

# Publications: routes to Open Access



Maximum embargo 6 months (12 for SSH)

## H2020 OA mandate for data



- Limited pilot launched in 2014, in 2017
  - participating is the **default option** for all projects (opt out possible)
- Develop a **Data Management Plan (DMP)**
- Provide open data (if possible)
  - for **(meta)data needed to validate results**
  - other (meta)data curated or raw data specified in the DMP

**As open as possible,  
as closed as necessary!**



## ERC specificities for publications



- ERC rules closely aligned to the H2020 requirements (cf. MGA 29.2)  
expected: **data underlying Publications**
- Disciplinary or institutional repositories recommended
- **Embargo** extended to 12 months if pub. linked to SSH (upon request)
- **Metadata**: lighter requirements for bibliographic metadata
  - but PID mandatory (DOI)
  - Good practice: ERC grant number indicated in metadata
- **Open license** encouraged (CC BY)

## ERC specificities for research data



Slightly different 29.3 (data): opt-in to the ORD pilot optional

- Data management must follow the **FAIR principles**
- Submit the first version of a **DMP within 6 months** specific template for ERC
- Deposit data mentioned in the DMP in **a trusted (certified) repository**
- **Exceptions acceptable** if it would jeopardise any objective (not just main)
- Opt-out: no justification needed
- Recommended: Information and tools used for the validation of the results



Ultimately, it is for the individual investigator to **decide** which data merit conservation and/or sharing.

# Non compliance with OA requirements

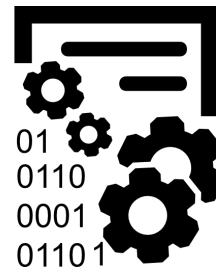
- Grant may be reduced (art. 43)
- Other measures describes in chapter 6 of the MGA





## Publications

- All publications including monographs and book chapters
- Always deposit in Open repository
- A PID is required



## Data

- Degrees of data sharing
- Needs Research Data Management compliant with the FAIR principles
- Don't forget to: think about tools/infos needed to validate results





PUBLICATIONS AND DATA

Data is increasingly becoming a legitimate citable research output.

**DATA?**

# What is data?

Definitions vary based on the research community or discipline, legal context, data governance policies...

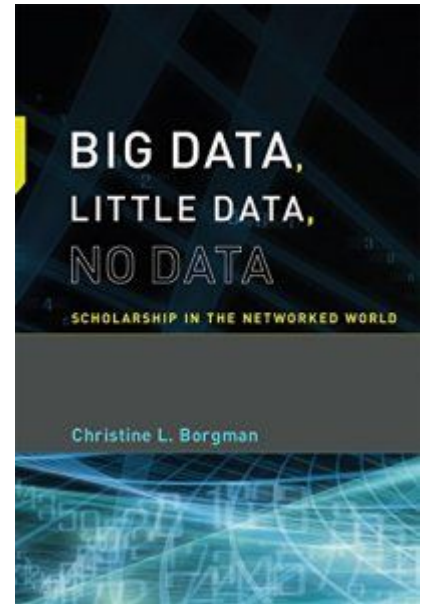
**It's impossible to frame them as a fixed ontological category... especially in the Humanities.**

=> Everything that is produced in a research process (but not administrative or pedagogical outputs)

# Publications <-> Data

Publications are arguments made by authors,  
Data are evidence used to support the arguments.

C.L. Borgman





A digital, selectively constructed,  
**machine-actionable abstraction**  
representing some aspects of a given  
object of humanistic inquiry.

—Schöch, Christoph. “Big? Smart? Clean? Messy?” Data in the Humanities”.  
Journal of Digital Humanities. Vol. 2, n°3, 2013 2018



**European Research Council**  
Scientific Council

Established by the European Commission

## **Open Research Data and Data Management Plans**

Information for ERC grantees

*by the ERC Scientific Council*

**FAIR?**



**F**

## FINDABLE

To aid automatic discovery of relevant datasets



**A**

## ACCESSIBLE

Explicit limitations on the use of data, protocols for querying, downloading or copying data are made explicit for both humans and machines



**I**

## INTEROPERABLE

(Meta)data should use standards, have references to other (meta)data and be machine actionable



**R**

## REUSABLE

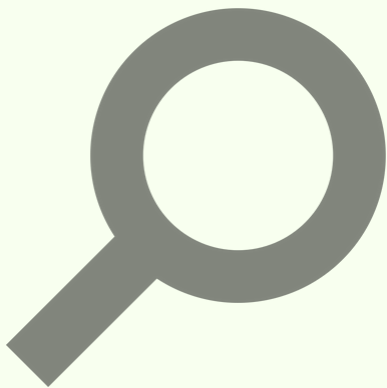
(Meta)data are sufficiently well described for humans and computers to be able to understand them and have a clear and accessible data usage licence



# FINDABLE

RICH DESCRIPTIVE  
(META)DATA

---



- How will they be discovered?
- What is necessary to make explicit so they can be understood outside the team?

- F1. (Meta)data are assigned a globally unique and **persistent identifier**
- F2. Data are described with **rich metadata** (defined by R1 below)
- F3. Metadata clearly and explicitly **include the identifier** of the data they describe
- F4. (Meta)data are registered or **indexed in a searchable resource**

# ACCESSIBLE

EXPLICIT ACCESS PROTOCOLS  
FOR BOTH HUMANS AND  
MACHINES

---

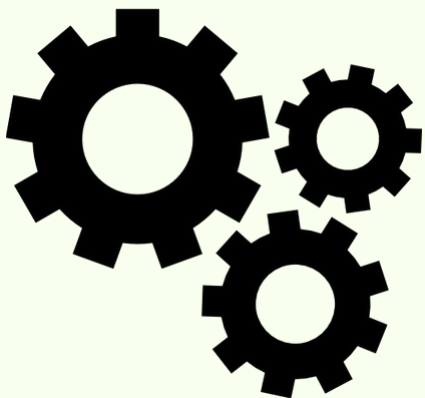


- Where the data will be found?
- Are they openly and freely accessible or are the limitations clear?
- A1. (Meta)data are retrievable by their identifier using a **standardised communications protocol**
  - A1.1 The **protocol is open**, free, and universally implementable
  - A1.2 The protocol allows for an **authentication and authorisation procedure, where necessary**
- A2. **Metadata are accessible**, even when the data are no longer available

# INTEROPERABLE

INTEGRATED WITH OTHER  
DATA & INTEROPERATE WITH  
APPLICATIONS

---



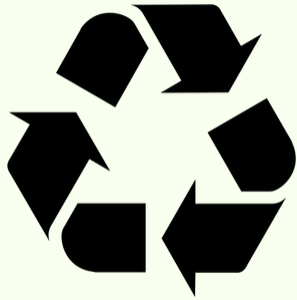
- What are the formal languages in use?
- Is the data intelinked to other (meta)data?
- Is the data machine-actionable?

- I1. (Meta)data use a **formal**, accessible, shared, and broadly applicable **language for knowledge representation**.
- I2. (Meta)data use **vocabularies** that follow FAIR principles
- I3. (Meta)data include qualified **references** to other (meta)data

# REUSABLE

## OPTIMISE THE REUSE OF THE DATA

---



- How will they be discovered?
  - Are reuse conditions clear?
  - What is necessary to make explicit so they can be understood outside the team?
- 
- R1. Meta(data) are **richly described** with a plurality of accurate and relevant attributes
    - R1.1. (Meta)data are released with a clear and accessible **data usage license**
    - R1.2. (Meta)data are associated with detailed **provenance**
    - R1.3. (Meta)data meet **domain-relevant community standards**



**F**

**FINDABLE**

Metadata  
Persistent identifier  
Naming  
conventions  
Keywords  
Versioning



**A**

**ACCESSIBLE**

Software  
Documentation  
Access  
modalities  
Data repository



**I**

**INTEROPERABLE**

Standards  
Vocabularies  
Machine  
actionable  
(Meta)data



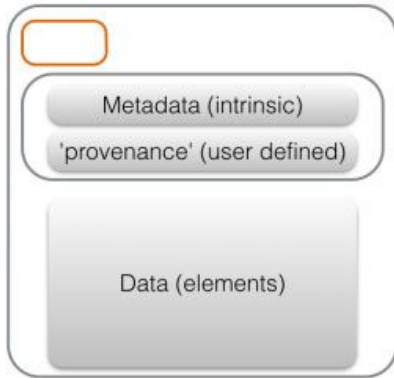
**R**

**REUSABLE**

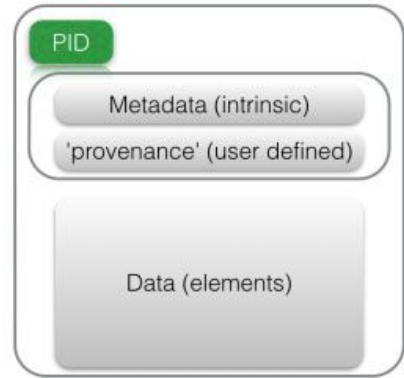
Licensing  
Context  
Provenance info  
Methodologies

# Data as increasingly FAIR Digital Objects

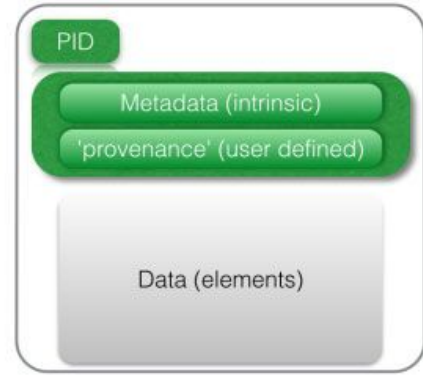
Totally UNFAIR



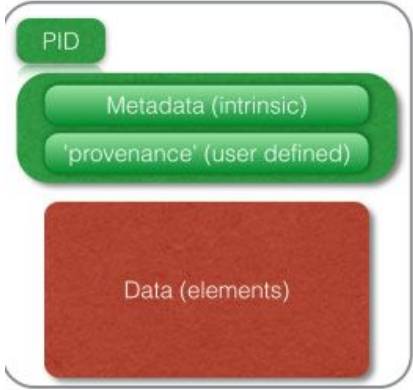
Findable  
Usable for Humans



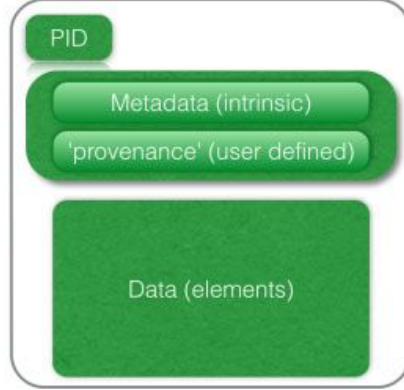
FAIR metadata



FAIR data-  
restricted access



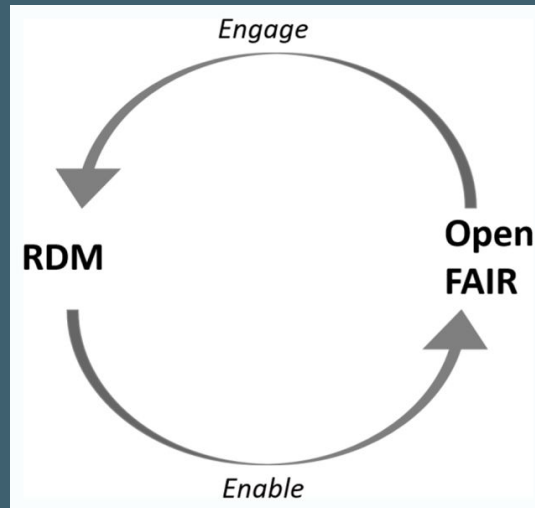
FAIR data-  
Open Access



FAIR data-  
Open Access/Functionally Linke



# FAIR vs Open vs RDM



Source : Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights*, 32(1), 18. DOI: <http://doi.org/10.1629/uksg.468>

## How to comply with the ORD Pilot?

1. Write a **DMP**
2. Find a **repository** matching data needs
3. **Deposit** (meta)data and other tools



**ARCHAIDE**

GRANT AGREEMENT NUMBER: **693548**

PROJECT ACRONYM: **ArchAIDE**

PROJECT TITLE: **Archaeological Automatic Interpretation and Documentation of eDramas**

FUNDING SCHEME: **H2020-REFLECTIVE-6-2015**

PROJECT COORDINATOR: **Prof Maria Letizia Gualandri, UNipi**

TEL: **+39 05022 15817**

E-MAIL: **maria.letizia.gualandri@unipi.it**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N.693548

Doc Title: **ORDP- Open Research Data Pilot D N° 10.2**

version: **0.2**

Revision: **Final Version**

Work Package: **10**

Lead Author (Org): **Tiziana Evans (Unipi)**

Contributing Author(s) (Org): **MS**

Due Date: **30<sup>th</sup> November 2016**

*Mobilising the Past for a Digital Future: The Potential of Digital Archaeology*  
**Data Management Plan**

**Types of Data**

The data produced by this proposed project will include presentations given by participants and data generated from moderated discussions in the proposed workshop. Presentations will consist of image presentations, video recording, and text files documenting a multidisciplinary two-day workshop on establishing best practices and protocols for mobile computing in digital archaeology. Dialogue and discussion generated during and after the workshop is an additional source of data. A website will be generated by the project team to publicize the event; post-text, video, and graphic images from the event; post best practice and protocol documents generated after the event; and continue to post discussion generated after the workshop. A project white paper and final report will be produced for the NHT.

**Data Standards and Capture**

The Creighton Digital Repository (CDR) has committed to providing a stable and secure digital archive indefinitely in order to store, preserve, and make accessible the data generated from the proposed workshop using DSpace. Data for this project consists of documents related to the organization, running, and follow up on the two-day workshop. Project directors in coordination with the CDR will upload event promotional material, and practical information for attendees and participants, as well as the presentations (text and PowerPoint), video, and discussion from the workshop itself. After the workshop, continuing discussion, and best practice and protocol documents will be uploaded by project directors and participants. Wakeborough Institute of Technology has committed to providing audio and video recording of workshop events; these will be captured as uncompressed audio or video but will be converted and maintained in the form of widely-adopted open formats such as MP3 that are suitable for events will be live-streamed. Effort will be made to use what is likely to remain to be performed to normalize

g items with the Creighton  
 rm and Dublin Core as its

all property rights in material  
 id for the conference under  
 to both Creighton and  
 participants and the workshop  
 (and, in similar agreement,  
 video presentations to be  
 or captured for this project.

University Libraries and the  
 led by Creighton University

th.pdf

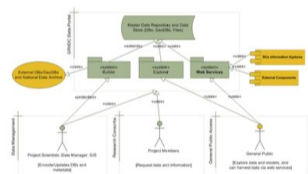


Figure 4.4. GRIDC Data Portal conceptual use case.

**5. DATA PROVENANCE AND METADATA**

The advent of the internet and ability of researchers to move data easily, and facilities to extract data from documents or databases, has introduced a new problem when defining the origin of the data. In a multi-disciplinary/multi-institutional scenario, it is often difficult to determine the origins of the data collected, compiled and archived in a central repository. To facilitate the process of tracing the origin of data, sensors or process used and transformations (if any) and allow the validation of scientific databases, the use of metadata standards will be imposed. Data provenance is central in the validation of data and as such, no data will be archived in RDs and served to the public unless data provenance is addressed. The RDs will maintain its integrity and reliability when the process used to create them are reproducible and analyzable for defects.

# Step 1: What is a DMP?

A DMP is a brief plan, formalised in a document, 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**

## Step 1: DMP Timeline

The DMP is a living document:

- initial DMP within first 6 months
- updated DMP
  - for significant changes
  - or periodical reviews
- final DMP





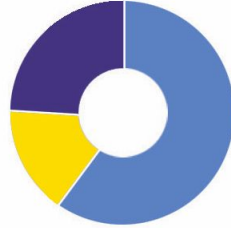
**Writing a DMP (...) should be part of the research process from the outset.**

# DMPs can be useful...

## OpenAire & FAIR data expert group survey (2017)

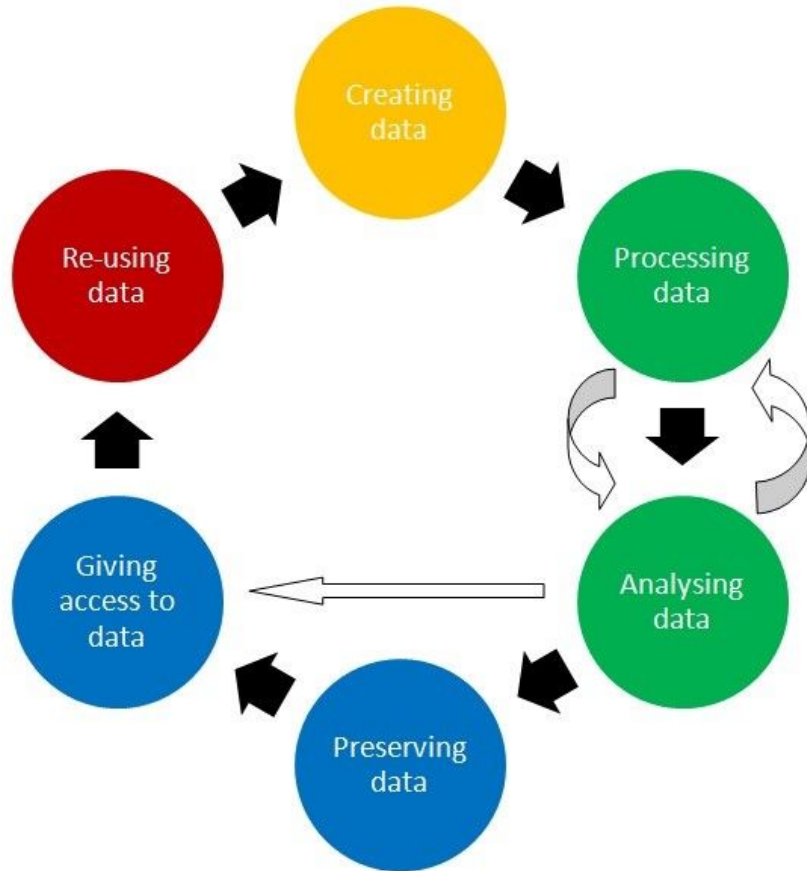
Report : <http://doi.org/10.5281/zenodo.1120245>

overall experience



of 189 respondents

- positive 60%
- not applicable 24%
- negative 16%



## Benefits of writing a DMP

- Guides the decision making
- Centralises information about the data
- Optimizes data management
- Help building dissemination strategies

# Evaluation criteria for DMPs

Performance Criteria	Performance levels		
	Detailed	Addressed but incomplete / unsatisfactory	Not addressed
1. Summary of digital outputs and technologies	<p>Plan provides a clear description of the digital output or digital technology and the type of access envisaged e.g. 'freely available online'</p> <p>The summary explains how this will be achieved technically so reviewers can assess whether plans are appropriate.</p>	<p>There's a partial description for example an explanation of the data but not the access plans, or too little information given the scale and importance of the digital output. There is insufficient information to assess whether the overall plans are sound.</p>	<p>It is unclear what will be created or how.</p> <p>The proposal indicates that significant digital outputs will be created but the technical plan doesn't match this.</p>
2a. Standards and formats	<p>Relevant statistics are provided to explain the size, quantity and duration of the data.</p> <p>An explanation is given about which formats and standards will be used and why. This demonstrates an understanding about which formats are appropriate to each purpose e.g. to provide access or to archive data.</p>	<p>A data volume is stated, but it's difficult to interpret how many files that pertains to or how large and complex the dataset is.</p> <p>Formats and standards are named but it is not clear that these are understood and that choices have been made for valid reasons.</p>	<p>The description misses important information and doesn't give a clear sense of sounds decisions regarding the technical methodology.</p>
2b. Hardware and software	<p>It is clear what (if any) additional hardware and software will be required. The additional kit is clearly necessary and aligns with statements in the justification of resources.</p> <p>If nothing is required, the plan should clearly state this is not applicable.</p>	<p>The suggested resources seem unconvincing (e.g. items that should already provided by the institution or software that doesn't seem necessary for the proposed digital output).</p>	<p>The plan does not state that this section is not applicable.</p>
2c. Technical methodology	<p>There is a full description of the development process from data capture to delivery of a digital resource. Key issues such as quality control and backup are covered for all the research environments / circumstances e.g. in-house procedures and fieldwork.</p> <p>There is a clear timetable for technical delivery that relates to the overall project milestones in the case for support.</p>	<p>Certain information is missing e.g. how quality will be checked if a number of assistants are transcribing and documenting resources.</p> <p>Backup is not explained for all circumstances in which data will be collected or stored.</p> <p>A timetable is missing, so it is unclear how the technical work aligns with the research activity.</p>	<p>There is no clear technical methodology or plan for how the data will be acquired, processed and used.</p>

## Step 2: Select a suitable repository

1. Domain specific repository
2. Institutional repository
3. General purpose repository such as Zenodo



# Repositories are not created equal!





## Step 3: Deposit Data

### WHAT?

- Data that validate the results presented in publications
- Other data: everything needed to find, assess, understand, reuse data:
  - “raw” data
  - derived data
  - code
  - **Methods**

## Step 3: Deposit Data

### PLANNED PREPARATION (with the DMP)

- Standard file formats
- Standards metadata schemas
- Documentation: what do others need to understand your data?
- (Open) Licences

# Preparing data for sharing

## Embedded documentation

- code, field and label description
- descriptive headers or summaries
- information recorded in the file properties (IPTC images, word doc)

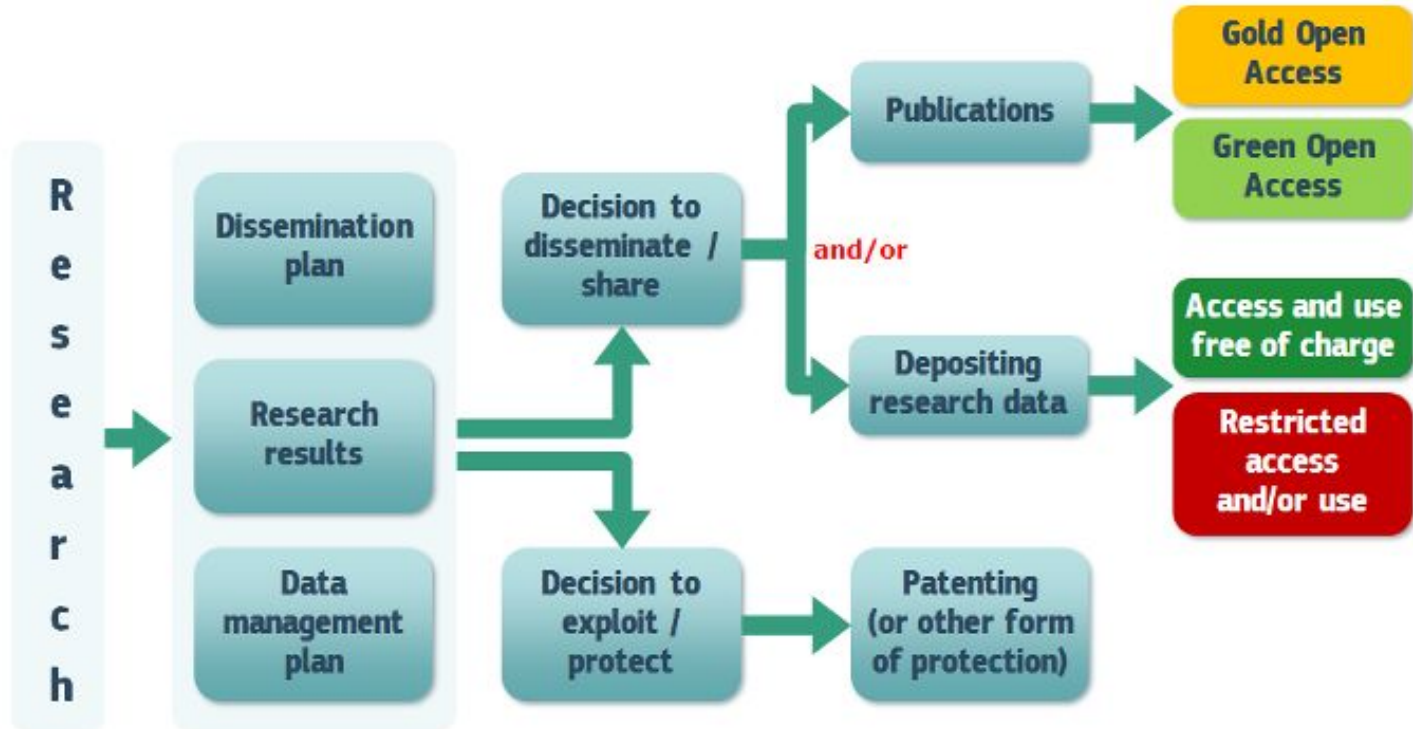
## Supporting documentation

- Working papers, blog posts, field notebooks
- Questionnaires, interview guides
- Final project reports
- Catalogue metadata

**Opening and sharing your data is not just  
about compliance:  
they are strategic decisions.**

The ERC lets you decide what to share and how, but expects you to provide the necessary metadata and documentation that will make it most useful!

# Decide early and record decisions in the DMP



# Pragmatic advice

## Prioritize

Data needed to validate results presented in publications

Data that can be re-used

**Plan for sharing**

## Promote

Give visibility to the project with Open Data

Commit to produce **data papers** for important dataset deposits (and retro-plan from that objective)

## Collaborate

The DMP helps you think about **automatic data stewardship procedures**

Set up collective work sessions to select data for sharing/preservation  
define quality control steps

# Datapapers Journal example

The screenshot displays the homepage of the Journal of Open Archaeology Data. The top navigation bar includes links for Home, About, Contact, Content, and Research Integrity, along with a search bar, Log in, and Register options. The main header features the journal's logo and two prominent orange buttons: 'Start Submission' and 'Become a Reviewer'. Below the header, a breadcrumb trail reads 'Reading: Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Easte...'. A row of social sharing icons (Facebook, Twitter, Google+, LinkedIn) and utility icons (Download, font size controls, citation, comments, print, email) is visible. The article content is organized into two columns. The left column contains the title 'Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Eastern Solent, England', the author 'Julian Whitewright' with an email icon, and an abstract paragraph. The right column has a 'JUMP TO' menu with a 'DISCUSSIONS' sub-menu, listing links for Abstract, Context, Methods, Dataset Description, Reuse Potential, Acknowledgements, and References. The abstract text discusses the potential future reuse of an archived dataset from an archaeological survey of the shipwreck remains of the Flower of Ugie, a wooden sailing barque wrecked in the Eastern Solent in 1852. The keywords listed are aggregate, archaeology, Flower of Ugie, maritime, and shipwreck.

Home About Contact Content Research Integrity Search... Log in Register

Journal of open archaeology data Start Submission Become a Reviewer

Reading: Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Easte... Share: Download A- A+ Alt. Display

**Data papers**

## Archaeological Survey and Monitoring Data from the Flower of Ugie, Wrecked 1852 in the Eastern Solent, England

**Author:** [Julian Whitewright](#)

### Abstract

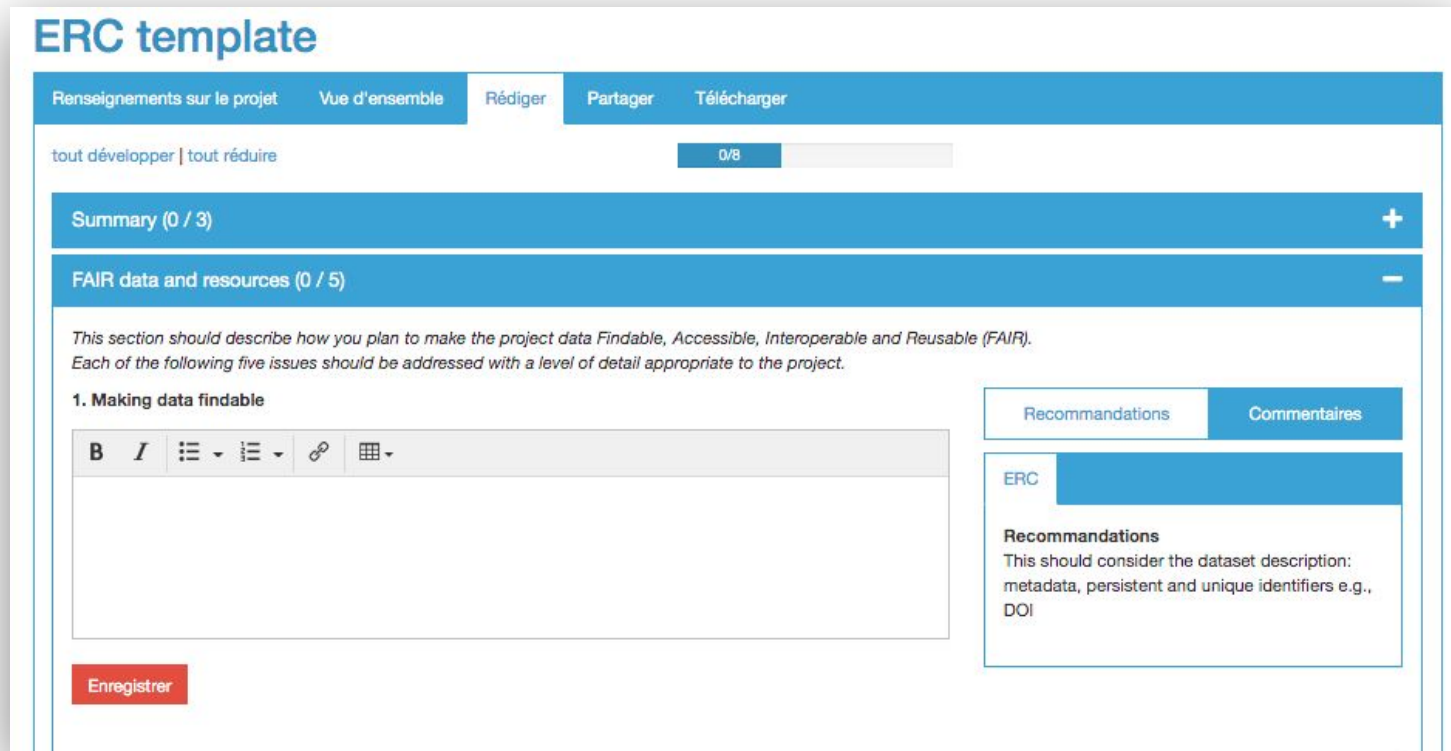
This paper discusses the potential future reuse of the archived dataset resulting from the archaeological survey of the shipwreck remains of the Flower of Ugie, a wooden sailing barque wrecked in the Eastern Solent in 1852. Seven years of archaeological survey are represented in the site plans which record the structural details of the vessel along with many of its material characteristics. This record serves to establish the baseline condition of the site, against which future monitoring of the site will be conducted allowing informed management decisions to be made.

**Keywords:** [aggregate](#) , [archaeology](#) , [Flower of Ugie](#) , [maritime](#) , [shipwreck](#)

**JUMP TO** **DISCUSSIONS**

- [Abstract](#)
- [Context](#)
- [Methods](#)
- [Dataset Description](#)
- [Reuse Potential](#)
- [Acknowledgements](#)
- [References](#)

# A collaborative tool for writing DMPs

A screenshot of a web-based interface for writing a Data Management Plan (DMP) using an ERC template. The interface is in French and has a blue and white color scheme. At the top, there's a navigation bar with tabs: "Renseignements sur le projet", "Vue d'ensemble", "Rédiger" (active), "Partager", and "Télécharger". Below the navigation bar, there's a progress indicator showing "tout développer | tout réduire" and a progress bar at "0/8". The main content area is titled "ERC template" and contains several sections. The first section is "Summary (0 / 3)" with a plus sign icon. The second section is "FAIR data and resources (0 / 5)" with a minus sign icon. Below this section, there's a paragraph of text: "This section should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues should be addressed with a level of detail appropriate to the project." The third section is "1. Making data findable", which contains a rich text editor with a toolbar (bold, italic, list, link, table) and a large empty text area. At the bottom left of the main content area, there is a red button labeled "Enregistrer". On the right side of the interface, there are two tabs: "Recommandations" (active) and "Commentaires". Below these tabs, there's a section titled "ERC" with a blue header, and a sub-section titled "Recommandations" containing the text: "This should consider the dataset description: metadata, persistent and unique identifiers e.g., DOI".





Thanks!

Questions?



# Credits

Presentation template by [Slidesgo](#)