Antónia Correia
Universidade do Minho

Data Management Plans

DESIR WINTER SCHOOL
FCSH UNL – 12th december 2019

OpenAIRE

@openaire_eu

Agenda

Research Data Management

1

Research Data Policies in H2020

2

Data Management Plans

3

Tools for developing data management plans

4

OpenAIRE

Learning objectives

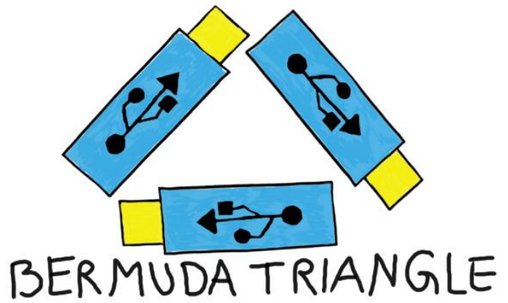
- Understand the importance of research data management;
- Discover how a Data Management Plan (DMP) can help you be more efficient in your research;
- Be aware of the European Comission's requirements on research data;
- Be able to start your own research data management plan.

1

Research Data Management

Why manage data?

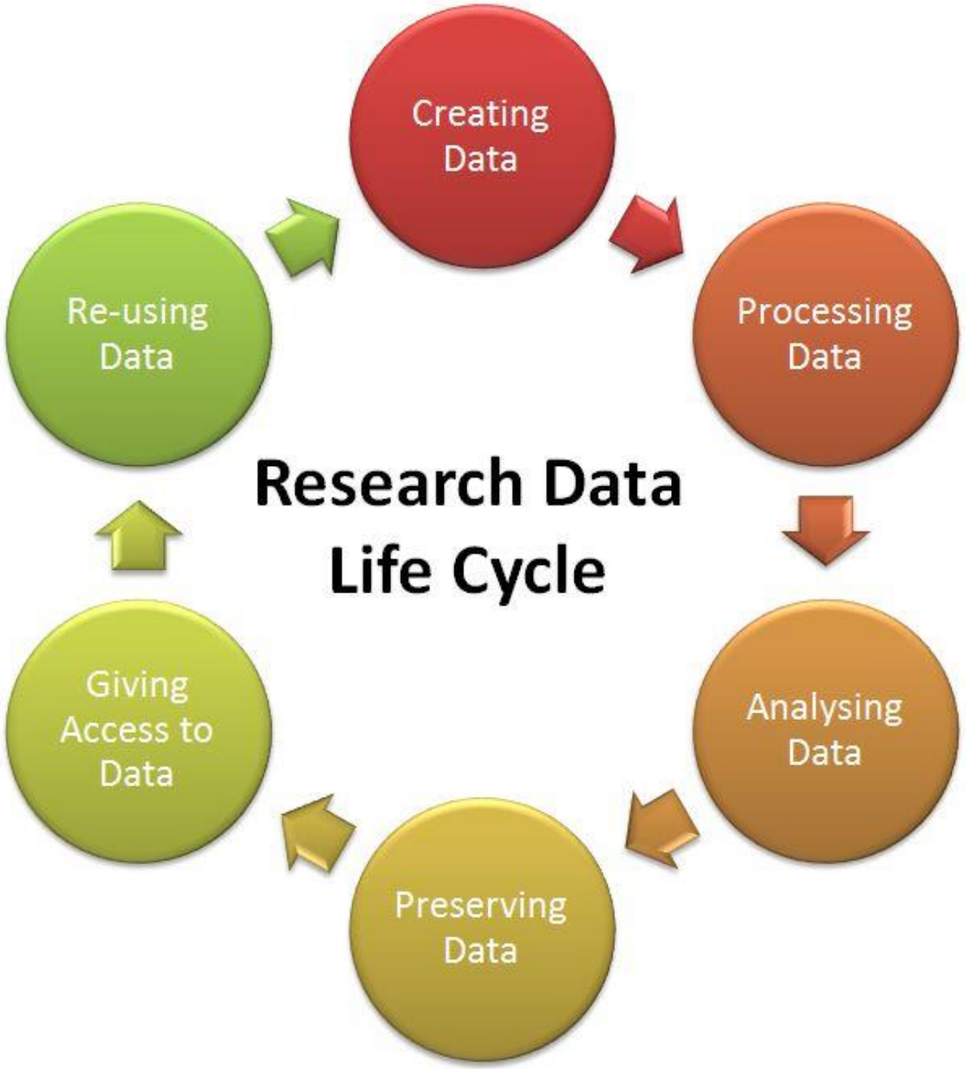
- Make your research easier
- Save data for later
- Avoid accusations of fraud or bad science
- Share your data for re-use
- Get credit for it
- Meet funder/institution requirements
- Prevent data loss



Data loss – frequent causes

- Natural disaster
- Facilities infrastructure/ storage failure
- Server hardware/software failure
- Format obsolescence
- Legal encumbrance
- Human error/ Malicious attack
- Loss of staffing competencies
- Loss of institutional commitment
- Loss of financial stability
- Changes in user expectations

Research Data lifecycle



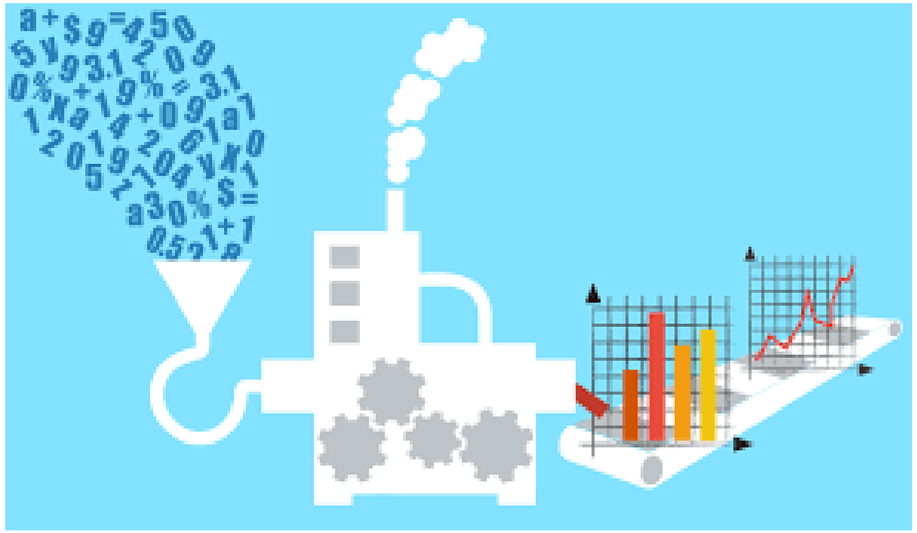
http://ukdataservice.ac.uk/manage-data/lifecycle.aspx

Creating data

- antecipate what data will be created, involving the whole research team
- develop methodology for data collection, assuring consistency
- choose software and formats that allow reproducibility and long time preservation
- plan consent for sharing – both from human subjects and among the project partners
- locate existing data
- collect data (experiment, observe, measure…)
- capture and create metadata

Processing data

- enter data, digitise, transcribe, translate…
- Document the methodology to allow reproducibility
- check, validate, clean data
- anonymise data where necessary
- describe data
- manage and store data



Analysing data

- interpret data
- derive data
- produce research outputs
- author publications
- prepare data for preservation

Preserving data

- migrate data to the best format
- migrate data to a suitable medium
- back-up and store data: more than one copy, in different media, one offsite
- create metadata and documentation
- archive data in a repository

Recommended formats for preservation

File formats extensions for reusability/preservation:

Type of data	APPROPRIATE	ACCEPTABLE	NOT SUITABLE
Tabular data with extensive metadata	.csv - .hdf5	.txt - .html - .tex - .por	
Tabular data with minimal metadata	.csv - .tab - .ods - SQL	.xml if appropriate DTD - .xlsx	.xls - .xlsb
Textual data	.pdf - .txt - .odt - .odm - .tex - .md -.htm - .xml	.pptx - PDF with embedded forms - .rtf	.doc - .ppt
Code	.m - .R - .py - .iypnb - .rstudio - .rmd – NetCDF	.sdd	.mat - .rdata
Digital image data	.tif - .png - .svg - .jpeg	jpg - .jp2 - .tif - .tiff - .pdf – GIF - BMP	.indd - .ait - .psd
Digital audio data	.flac - .wav - .ogg	.mp3 - .mp4 - .aif	
Digital video data	.mp4 - .mj2 - .avi - .mkv	.ogm - .webm	.wmv - .mov
Geospatial data	NetCDF, tabular GIS attribute data, .shp - .shx - .dbf - .prj - .sbx - .sbn – PostGIS - .tif - .tfw -GeoJSON	/.mdb/.mif/	
CAD/vector and raster data	.dwg - .dxf - .x3d - .x3dv - .x3db - .pdf - PDF3D		
Generic data	.xml - .json - .rdf		

EPFL Fast Guides on Research Data Management

Giving access to data

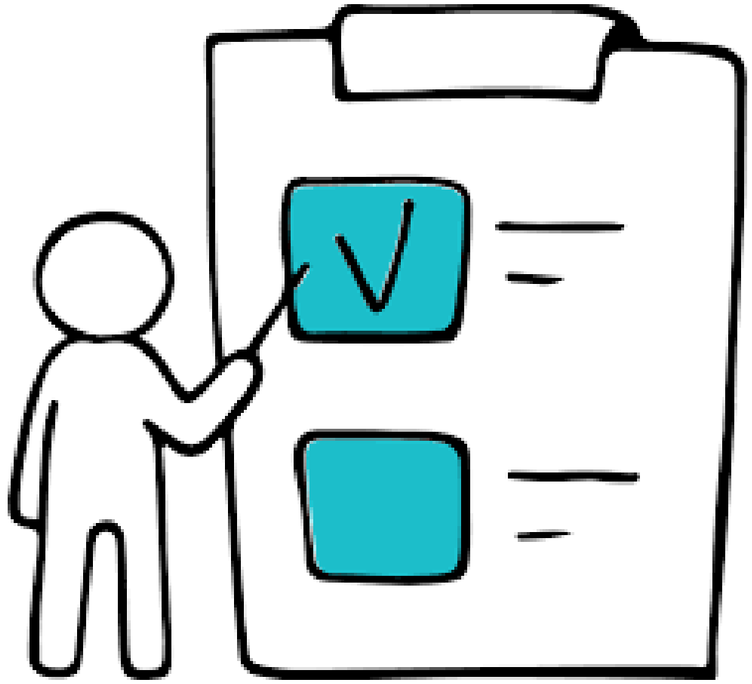
- Different stages = different access options for data (read/ edit)
- Myth: all public or all closed - Some datasets can be public, others should remain embargoed (ex: guest books dataverse)
- establish copyright
- Promote data (eg. social media)

Consider who else has a say about sharing data:

- Collaborators
- Research participants
- Commercial partners
- Data repository

Plan for sharing

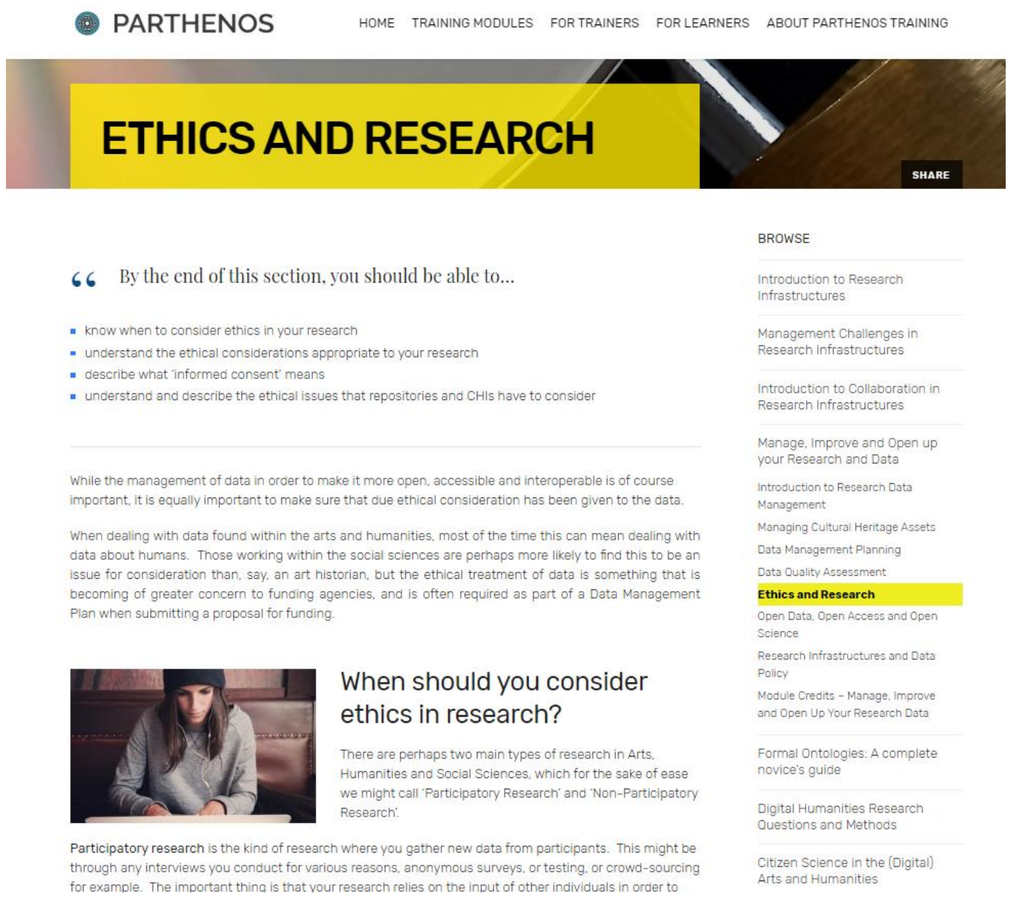
- Desirebly at data creation
- Identify which bits of personal data will be collected
- Do you really need to collect personal data? (eg. irrelevant questions in questionnaires)
- Consider how will anonymization costs be covered



Ethics and Research

The kind of research data that would usually be subject to ethical approval includes (but is not necessarily limited to:

- Any recorded interviews (either video or audio)
- Surveys or questionnaires that collect personal information such as date/place of birth or anything else that could identify the participant
- Research where the participant is asked to reveal or reflect on instances from their past (e.g. oral histories, psychology experiments)
- Anything that involves the participation of minors (additional ethical requirements may be in place for such instances)
- Anything in which the participant is asked to reveal something that might cause them or others physical or mental harm or embarrassment if it were to be made public.
- Any research in which the participant is asked to complete tests, or test-like scenarios.



https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/ethics-and-research/

Informed consent

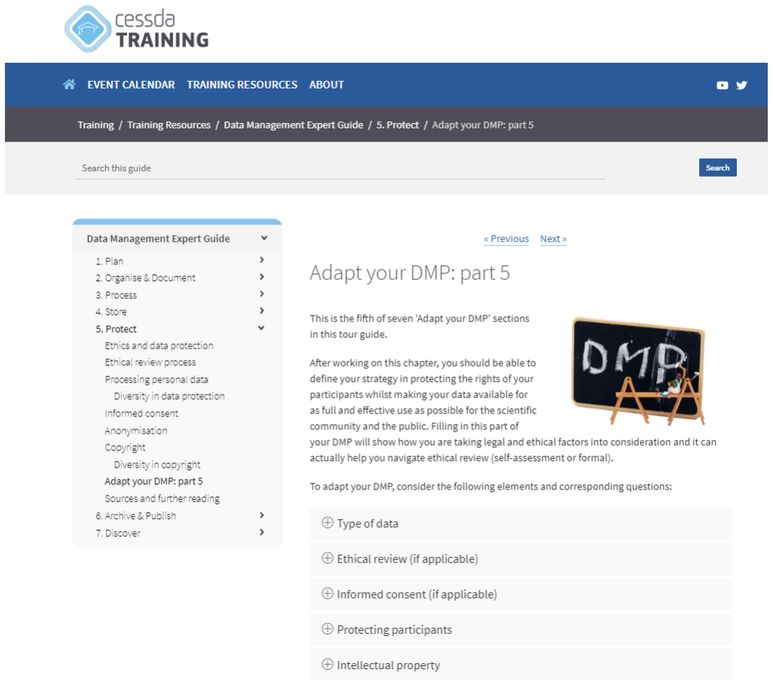
"Informed consent is the process by which a researcher discloses appropriate information about the research so that a participant may make a voluntary, informed choice to accept or refuse to cooperate." (as defined in the CESSDA Expert Tour Guide RDM)

- When creating consent forms, researchers should make sure to:
  - inform participants about their rights
  - introduce relevant aspects of the research in an understandable, transparent, and precise way
  - explain data protection measures that will be taken
  - be clear about plans for data sharing in the consent form - both during the life of your project and after it ends

Examples: https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms

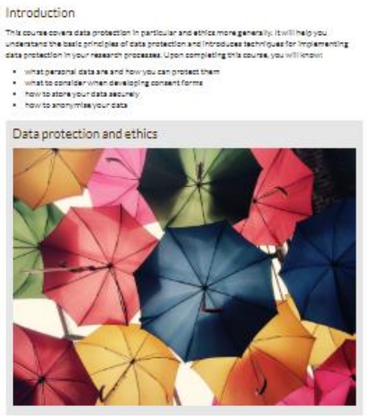
More Resources

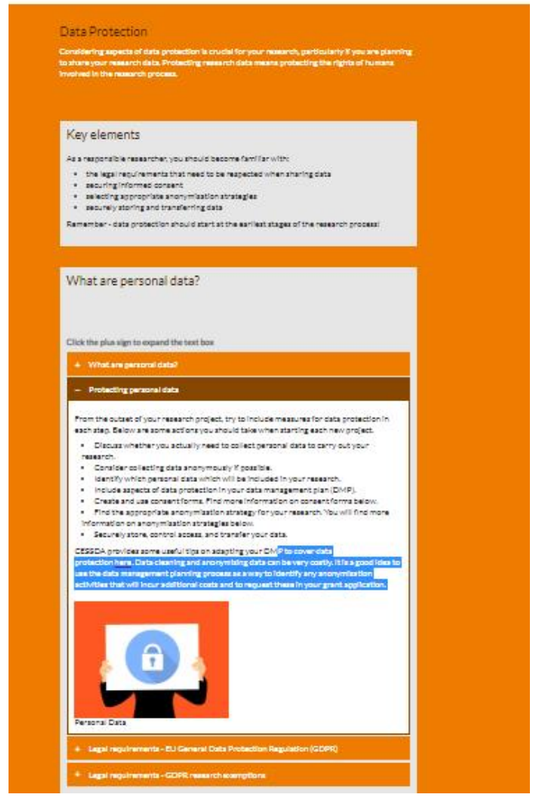
- CESSDA training



https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect

- FOSTER course On Data Protection and Ethics





www.fosteropenscience.eu/learning/data-protection-and-ethics

Re-using data

- Your own or other's, you need context! DMPs are a great help
- follow-up research/ do new research
- FAIR data maximizes reuse of data

Data selection

- Five steps to follow
  1. Could this data be re-used
  2. Must it be kept as evidence or for legal reasons
  3. Should it be kept for its potential value
  4. Consider costs – do benefits outweigh cost?
  5. Evaluate criteria to decide what to keep

5 steps to decide what data to keep

www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

2

Research Data Policies in H2020

Open Science

Results

Publications
Research data
Software
…

Processes
Tools
Infrastructures
…

Methods

Open Science – Why?

Accelerate research

Get credit and citations

Safeguard data

Avoid malpractice or unethical behaviour

Share data and allow reuse

Fulfill funder’s mandates

OpenAIRE

European Commission

Open Science: Scientific progress - Emergency science

OXFORD UNIVERSITY PRESS | Journals

About Us | Careers | Contact Us | Help | 中文 | عربي

View Basket | Sign In / Register

Search articles... Search Advanced Search

Journals A-Z | Arts & Humanities | Law | Medicine & Health | Science & Mathematics | Social Sciences | Academic Books & Online Resources

You are here: Home » Our Journals » Medicine & Health » Free access to OUP resources on *Aedes aegypti* and Zika virus research

Special Collection

on Aedes Aegypti and Zika Virus research

Free access to OUP resources on *Aedes aegypti* and Zika virus research

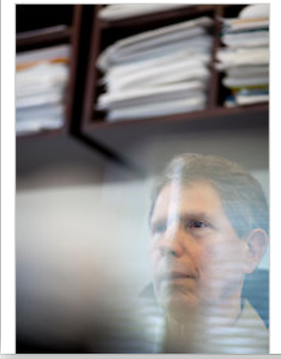
Browse all Medicine & Health titles

Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA
Published: August 12, 2010

In 2003, a group of scientists and executives from the National Institutes of Health, the Food and Drug Administration, the drug and medical-imaging industries, universities and nonprofit groups joined in a project that experts say had no precedent: a collaborative effort to find the biological markers that show the progression of Alzheimer's disease in the human brain.

Enlarge This Image



Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.

And the collaboration is already serving as a model for similar efforts against Parkinson's disease. A $40 million project to lo
by the Micha
study subjec

World Health Organization

عربي 中文 Eng

centre | Publications | Countries | Programmes | Governance | About WHO

Bulletin of the World Health Organization

Zika Open

These papers are posted in the c
International Concern declared b
Organization 1 February 2016.

The data in these papers are free
reproduction in any medium, prov
indicated by the Creative Comm
license (CC BY IGO 3.0).

If you wish to submit a manuscri

EDITORIAL

Data sharing in public health e
- Christopher Dye, Kidist Bartolo
Posted: 4 February 2016
http://dx.doi.org/10.2471/BLT.16.

Science AAAS

Home | News | Journals | Topics | Careers

Latest News | ScienceInsider | ScienceShots | Sifter | From the Magazine | About News | Quizzes

SHARE

A ministry of health official visits a family in Guatemala City to help control the mosquitoes that transmit Zika virus. Conrad Guatemala (CC BY-NC-ND 2.0)

A plea for open science on Zika

PM

www.nytimes.com/2010/08/13/health/research/13alzheimer.html?pagewanted=all&_r=0

OpenAIRE

Open Science: Validate, Correct Results, Combat fraud

The error that could subvert George Osborne's austerity programme

The theories on which the chancellor based his cuts policies have been shown to be based on an embarrassing mistake

Charles Arthur and Phillip Inman
The Guardian, Thursday 18 April 2013 21.10 BST

George Osborne says that Ken Rogoff, the man whose economic error has been uncovered, has strongly influenced his thinking. Photograph: Stefan Wermuth/PA

nature

International weekly journal of science

nature news home | news archive | specials | opinion | features | news blog | natu

comments on this story

Stories by subject
- Brain and behaviour
- Lab life

Stories by keywords
- Deiderik Stapel
- Tilburg University
- Academic fraud
- Retractions
- Social psychology

This article elsewhere
- Blogs linking to this article
- Add to Digg
- Add to Facebook
- Add to Newsvine
- Add to Del.icio.us
- Add to Twitter

Published online 1 November 2011 | *Nature* **479**, 15 (2011) | doi:10.1038/479015a
Updated online: 1 November 2011
Updated online: 8 December 2011

News

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

Ewen Callaway

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report (go.nature.com/tqmp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.

Dutch psychologist Diederik Stapel.
Persbureau van Eijndhoven

"We have some 30 papers in peer-reviewed journals where we are actually sure that they are fake, and there are more to come," says Pim Levelt, chair of the committee that investigated Stapel's work at the university.

Stapel's eye-catching studies on aspects of social behaviour such as power and stereotyping garnered wide press coverage. For example, in a recent *Science* paper (which the investigation has not identified as fraudulent), Stapel reported that untidy environments encouraged discrimination (*Science* **332,** 251-253; 2011)

Retraction Watch

Tracking retractions as a w

Raw files help fix 2003 figure by heart researcher accused of fraud

without comments

A researcher accused of misconduct by an anonymous Japanese blogger has corrected a 2003 paper in *Circulation Research*, after providing a university investigation with the original source files.

Allegations of fraud have dogged Shokei Kim-Mitsuyama for years, and even caused him to step down from his position as editor in chief at another journal. However, Kim-Mitsuyama and his colleagues call the latest correction a "mistake," which didn't affect any of the paper's conclusions.

We've unearthed a total of five publications co-authored by Kim-Mitsuyama that have earned corrections, the latest of which cites an investigation by the university:

Read the rest of this entry »

Share this:

Email | Facebook 5 | Twitter

Written by Shannon Palus
April 21st, 2016 at 2:00 pm

Posted in am j physio heart circ phys,American Heart Association,cardiology retractions,cardiovascular research,circulation research,corrections,erroneous data,hypertension research,japan retractions,misconduct investigations,nature publishing group,plos,plos one,scientific reports,society journal retractions

Authors retract, replace highly cited JAMA Psych paper for "pervasive errors"

with 4 comments

Authors have retracted a highly cited *JAMA Psychiatry* study about depression after failing to account for some patient recoveries, among other mistakes.

JAMA Psychiatry

It's a somewhat unusual notice — it explains that the paper has been retracted and replaced with a new, corrected version.

The study, which included 452 adults with major depressive disorder, concluded that cognitive therapy plus medication works better to treat depression than pills alone. But after it was published, a reader pointed out that some of the numbers in a table were incorrect. The authors reviewed the data and redid their analysis, and discovered "a number of pervasive errors."

OpenAIRE

European Commission

www.guardian.co.uk/politics/2013/apr/18/uncovered-error-george-osborne-austerity

Open Science: improve and promote reproducibility

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Aut

Archive > Volume 533 > Issue 7604 > News Feature > Article

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

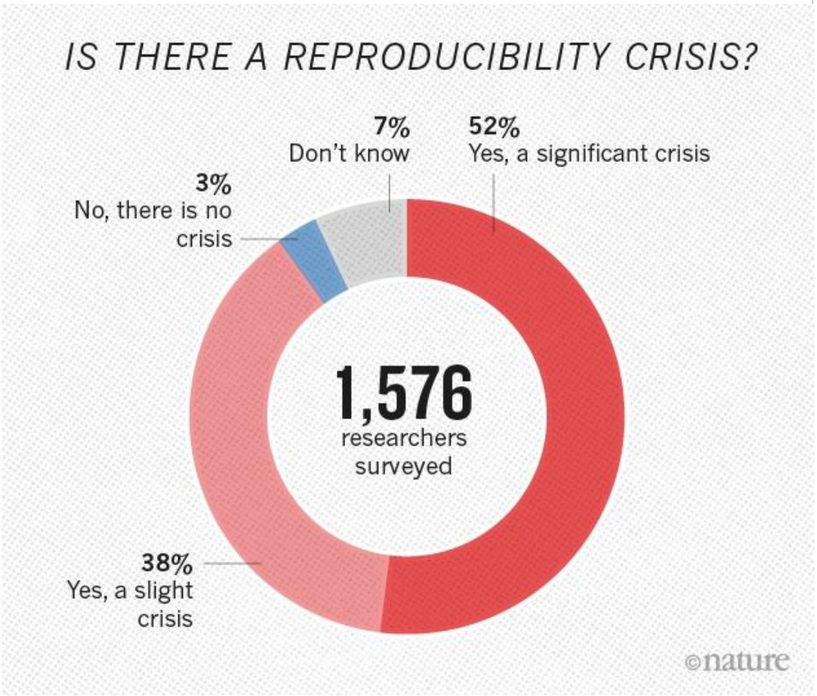


Table 1: Reproducibility of research findings Preclinical research generates many secondary publications, even when results cannot be reproduced.

From: Raise standards for preclinical cancer research

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

*Source of citations: Google Scholar, May 2011

nature International weekly journal of science

NATURE | NEWS

Over half of psychology studies fail reproducibility test

Largest replication study to date casts doubt on many published positive results.

Monya Baker

27 August 2015

Bee business

The bitter battle over the world's most popular insecticides

Begley, C. G. & Ellis, L. M. *Nature* **483**, 531-533 (2012).

The backgroud: evolution of the EU funding programmes for R&I

FP7

Open Access **Pilot**

Deposit and open access

H2020

OA **Mandatory**

Deposit and open access

& Open Research Data / DMP **Pilot**

H2020

OA **Mandatory**

Deposit and open access

& ORD/DMP **by default (opt-out)**

Horizon Europe

- OA **Mandatory**
- Deposit and open access
- DMP + FAIR data **Mandatory**
- OD **by default (opt-out)**
- & Open Science **embedded**

OpenAIRE

European Commission

Open Science is here to stay: the sooner we embrace its principles, the better.

Lennart Martens, Paola Masuzzo (2017)

“The question is no longer „if“ we should have open access. The question is about „how“ we should develop it further and promote it.”

Neelie Kroes (2011)

Open Innovation, Open Science, Open to the World

Carlos Moedas (2015)

OpenAIRE

H2020 Open Research Data Pilot: aims

To make the research data generated by Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access.

Information already paid for by the public should not be paid for again. Open data is data that is free to access and reuse

EC

OpenAIRE

European Commission

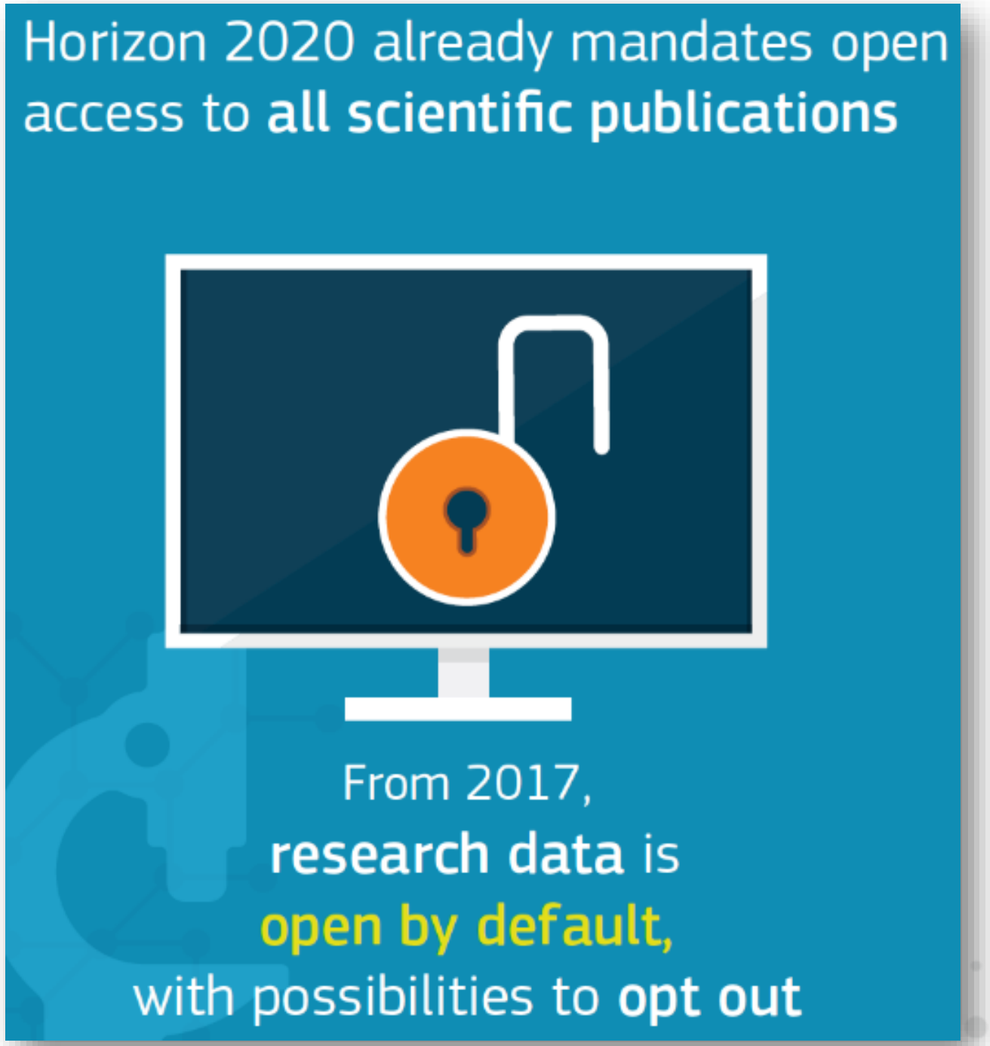
From January 2017 research data is open by default

Pilot
2014-2016

Open by default
2017-2020

DMP +
FAIR data
Mandatory

OD **by default**
(opt-out)



Multi-beneficiary General Model Grant Agreement

29.2 Open access to scientific publications

29.3 Open access to research data

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

From the Model Grant Agreement

29.3 Open Access to research data

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 I);

(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).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex I, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.]

***[OPTION:** not applicable]*

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

Open Research Data policy requirements

DATA, including metadata, needed to validate the results in scientific publications.

Other data, including metadata, as specified in the Data Management Plan.

Horizon 2020 grantees are encouraged to also share datasets beyond publication

OpenAIRE

European Commission

STEP 1

WRITE A DMP

Update at

- 6 months
- Periodic evaluation
- Final review

STEP 2

FIND REPOSITORY

Matches data needs

Data Repositories

- discipline/institutional
- www.re3data.org
- Zenodo

STEP 3

DEPOSIT DATA

(Open) Data

Metadata

Other tools

- Standard File Formats
- Standards metadata schema
- (Open) Licences

SUPPORT

Supporting infrastructure and information

Open Research Data Pilot

- EC guidelines
- OpenAIRE.eu
- peers

Where to find a repository?

More information:

https://www.openaire.eu/opendatapilot-repository

Zenodo: http://www.zenodo.org

Re3data.org: http://www.re3data.org

OpenAIRE

European Commission

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

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

intellectual property rights

privacy

might jeopardise project's main objective

Reasons for total or partial opting out

- ✓ Incompatible with the Horizon 2020 obligation to protect results if they can reasonably be expected to be commercially or industrially exploited;
- ✓ Incompatible with the need for confidentiality in connection with security issues;
- ✓ Incompatible with existing rules concerning the protection of personal data;
- ✓ If the project will not generate / collect any research data;
- ✓ If there are other legitimate reasons to not take part in the Pilot

50 shades of “No”

- Too expensive
- There’s no business case
- There’s no commercial value
- It’s private
- It’s secret
- It's our data
- We have invested a lot of money in this
- Link enough data and one will arrive at
- sensitive private information
- It's not data, it's information
- It will never work
- We don't know how to do this
- We don't have the right people to do this
- We need the money
- It’s not ours, and we don’t know who’s data it is
- No idea what the quality of the data is
- We don’t know where to find it
- It’s not our job
- It isn’t in the right format
- I am not authorized
- Who is going to use this anyway
- People are going to misuse it
- (…)

Source http://philarcher.org/diary/2015/50shadesofno/

OpenAIRE

Exercice 1: Open Data Excuse Bingo

Terrorists will use it	People may misinterpret the data	We'll get spam	I don't mind, but someone else might
We will get too many enquiries	It's not very interesting	Lawyers want a custom License	Data Protection
Thieves will use it	We might want to use it in a paper	There's already a project to...	It's too big
What if we want to sell it later	Poor Quality	It's too complicated	There's no API

Source http://tinyurl.com/cvkba2h

OpenAIRE

European Commission

Open Data Excuse Bingo

The "Open Data Excuse" Bingo lists common arguments used by researchers when explaining why they can't share their data

- **Role playing: in groups of 4, for 10 minutes the groups discuss these excuses - half the participants give reasons for sharing, half give reasons not to**
- **Then individually choose your favorite excuse – or the one that applies to you – and explain.**

3

Data management plans

What is a Data Management Plan?

A data management plan is a document outlining **how the research data collected or generated will be handled during a research project**, and after it is completed, describing **what data will be collected** / generated and following **what methodology and standards**, **whether and how this data will be shared and/or made open**, and **how it will be curated and preserved** (...) The use of a detailed data management plan covering individual datasets is required for funded projects participating in the Open Research Data Pilot.

Benefits of DMPs

DMPs help researchers to:

- Make informed decisions to anticipate & avoid problems
- Avoid duplication, data loss and security breaches
- Develop procedures early on for consistency
- Ensure data are accurate, complete, reliable and secure

Create a DMP

Handling of data during and after project

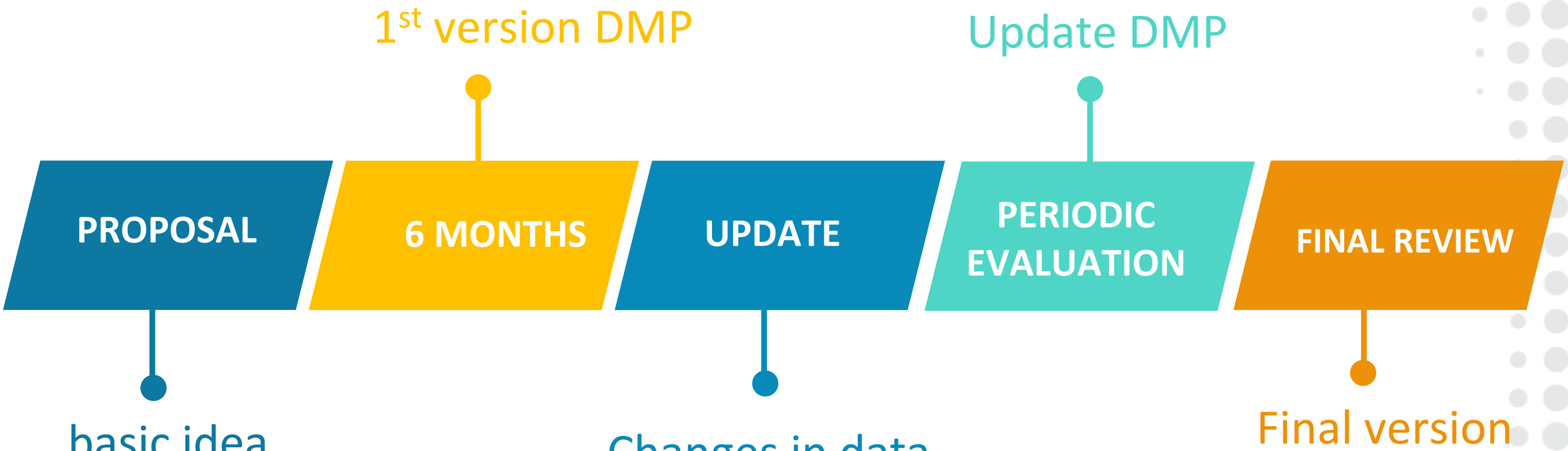
Living document: update

Reflects on curation, preservation, sustainability and security

What parts will be open and how?

OpenAIRE

Timeline



OpenAIRE

RESEARCH COUNCILS UK

Accessibility | Media Enquiries | Contact

search Search

Home | Funding | Research | Innovation | Skills | Public Engagement | News, Events and Publications | About

RCUK Common Principles on Data Policy

Home > Research > RCUK Common Principles on Data Policy

Making research data available to users is a core part of the Research Councils' remit and is undertaken in a variety of ways. We are committed to transparency and to a coherent approach across the research base. These RCUK common principles on data policy provide an overarching framework for individual Research Council policies on data policy.

Principles

- Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.
- Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.
- To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council
may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of th
varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.
- In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data sho
acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.
- It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which car
from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

THE ROYAL SOCIETY

Home | Fellows | Events | Grants, Schemes & Awards | Topics & policy | Journals | Collections

Data sharing and mining

Open data policy

To allow others to verify and build on the work published in Royal Society journals, it is a condition of publication that authors make available the data, code and research materials supporting the results in the article.

Datasets and code should be deposited in an appropriate, recognised, publicly available repository. Where no data-specific repository exists, authors should deposit their datasets in a general repository such as Dryad or Figshare.

To encourage best practice in data sharing, *Biology Letters*, *Proceedings B* and *Royal Society Open Science* have Dryad data deposition integrated into the journal submission system. For all its science journals, the Society will cover the cost of depositing up to 20GB of data with Dryad. In addition, we deposit all supplementary material into the Figshare repository on the author's behalf.

Exceptions to the sharing of data, code and materials may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species. Authors must disclose upon submission of the manuscript any restrictions on the availability of data, code and research

e article.

d in **both**

NIH National Institutes of Health
Office of Extramural Research

Grants & Funding
NIH's Central Resource for Grants and Funding Information

Entire Site | Search this S

eRA | Glossary &

HOME | ABOUT GRANTS | FUNDING | POLICY & COMPLIANCE | NEWS & EVENTS

Home » Policy & Compliance » Policy & Guidance » NIH Data Sharing Policy and Implementation Guidance

Policy & Compliance

- NIH Grants Policy Statement
- Notices of Policy Changes
- Compliance & Oversight
- Select Policy Topics +

NIH Data Sharing Policy and Implementation Guidance

(Updated: March 5, 2003)

This guidance provides the National Institutes of Health (NIH) policy statement on data sharing and additional information or this policy.

- Goals of Data Sharing
- Applicability
- Implementation
  - Timeliness of Data Sharing
  - Human Subjects and Privacy Issues
  - Proprietary Data
  - Methods for Data Sharing
  - Data Documentation
  - Funds for Data Sharing
  - Review Considerations

wellcome

Funding | What we do | About us | News

Scheme finder | Managing a grant

Policy on data, software and materials management and sharing

As a charity, Wellcome works to ensure that the results of the research we fund are applied for the public good. This includes creating an environment that enables and incentivises researchers to maximise the value of their research outputs, including data, software and materials.

We expect our researchers to manage research outputs in a way that will achieve the greatest health benefit. This may involve making outputs widely available or using intellectual property (IP) as a tool to help protect and commercialise an original idea, product or technology.

There is international consensus on the need to share and preserve research datasets in a way that maximises their long-term value. Key documents such as the UK concordat on open research data (2016) articulate this.

Which other funders require a DMP?

● Full Coverage ◐ Partial Coverage ○ No Coverage

	Policy Coverage		Policy Stipulations					Support Provided			
Research Funders	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

www.dcc.ac.uk/resources/policy-and-legal/ overview-funders-data-policies

Data Management Plans Requirements - general

- Description of data to be collected / created (i.e. content, type, format, volume…)
- Standards / methodologies for data collection & management
- Ethics and Intellectual Property (highlight restrictions on data sharing e.g. embargoes, confidentiality)
- Plans for data sharing and access (i.e. how, when, to whom)
- Strategy for long-term preservation

Elements required by H2020 grant

1. Dataset description

- How data will be collected and, or how existing data with be reused
- What kinds of data (ex. types, formats, volume) will be collected or produced

2. Documentation and data quality

- Which metadata and documentation will accompany the data (ex. Data collection methodology)
- Which control, quality measures will be put in place?

Elements required by H2020 grant

3. Backup and storage

- How will metadata be stored during the research process?
- How will data and sesitive data be secured during research?

4. Ethical requirements and code of conduct

- How will the personal and sensitive data legislation be addressed?
- How will intellectual property rights be addressed?
- How will other ethical issues be addressed? Will there be a code of conduct?

Elements required by H2020 grant

5. Data sharing and long time preservation

- How /will datasets be shared in open access? Will there be embargo periods or restrictions?
- How will datasets be selected for long time preservation? Which repositories will be chosen?
- Which software tools will be necessary to access and reuse data?
- How will persistent IDs be assigned?

6. Responsabilities and resources in data management

- Who (eg. Role, position adnd institution) will be responsible fpr data management?
- Which resources (eg. financial, time) will ne dedicated to research data management and to make efforts to render them FAIR?

H2020 template

1. Data summary
2. FAIR data
    1. Making data findable, including provisions for metadata
    2. Making data openly accessible
    3. Making data interoperable
    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

RESEARCH DATA – OPEN BY DEFAULT

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 stage
- ✓ Explains costs are eligible
- ✓ Provides a DMP template

OpenAIRE

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

FAIR DATA PRINCIPLES

Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;
F2. data are described with rich metadata;
F3. metadata clearly and explicitly include the identifier of the data it describes;
F4. (meta)data are registered or indexed in a searchable resource;

Interoperable:

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;

Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol;
A1.1 the protocol is open, free, and universally implementable;
A1.2. the protocol allows for an authentication and authorization procedure, where necessary;
A2. metadata are accessible, even when the data are no longer available;

Reusable:

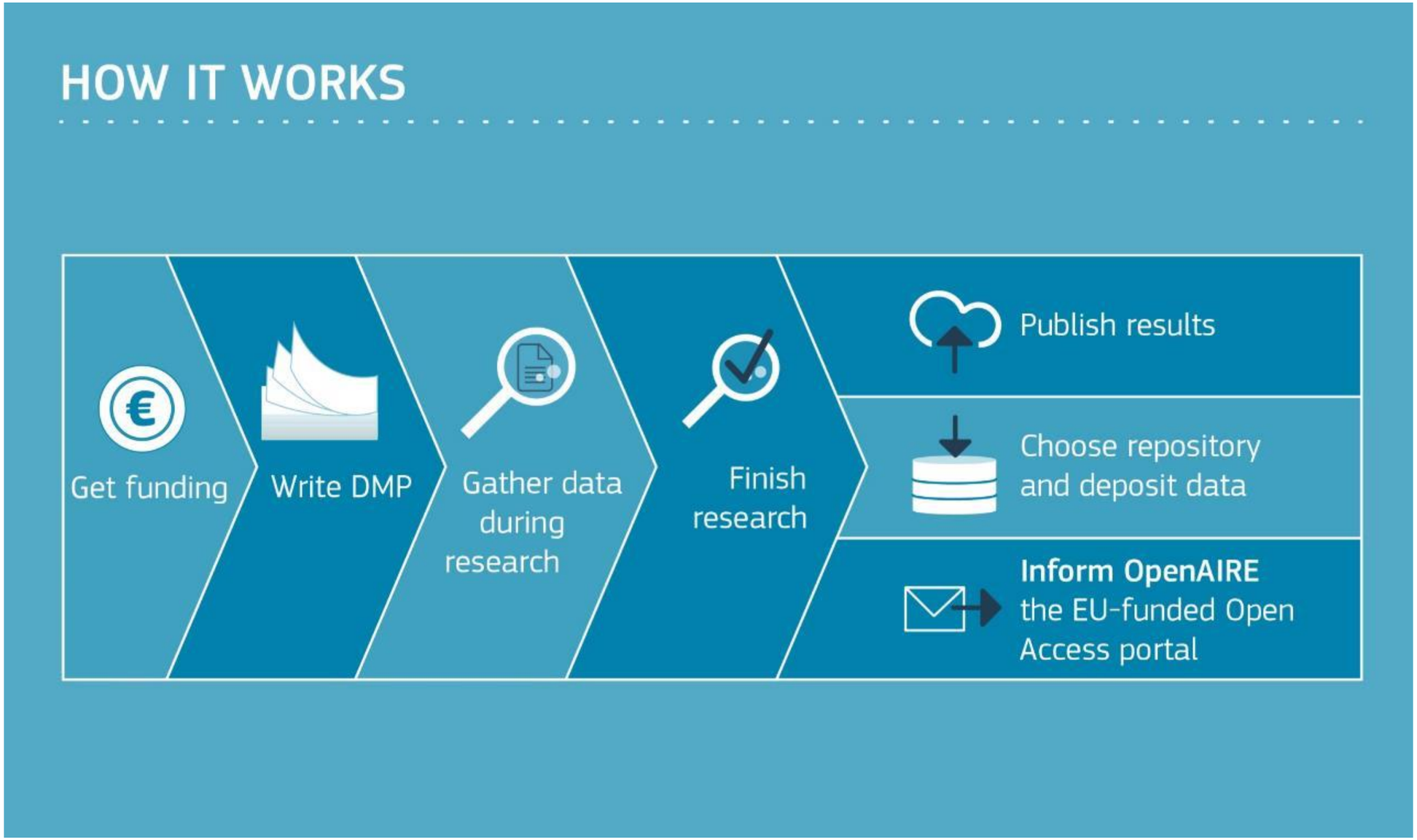
R1. meta(data) are richly described with a plurality of accurate and relevant attributes;
R1.1. (meta)data are released with a clear and accessible data usage license;
R1.2. (meta)data are associated with detailed provenance;
R1.3. (meta)data meet domain-relevant community standards;

OpenAIRE

European Commission

www.force11.org/group/fairgroup/fairprinciples

Open Access to Research Data



What makes a good DMP?

- Does the plan show data and Open RDM awareness?
  - Is the focus on the data or on publications?
  - A publication that describes the data =/= depositing the data.
  - Is Openness on the author's mind?
  - Does the plan distinguish between storing and archiving?
- Does the plan show that data management has many stakeholders? IT department, ethics committee, long-term repository…
- Overall approach to DMPlanning
  - Is the plan specific, or mainly intentional?
  - "should", "possibly", "where suitable/ appropriate/ relevant"…
  - Is the whole project team involved?
  - (When) will the DMP be evaluated and updated?

Why is this a good DMP? Marjan Grootveld and Ellen Leenarts, DANS (TUD Seminar "Towards cultural change in data management – data stewardship in practice", May 24th, 2018

OpenAIRE

European Commission

Tips to share: writing DMPs

- Keep it simple, short and specific
- Seek advice - consult and collaborate
- Base plans on available skills and support
- Make sure implementation is feasible
- Justify any resources or restrictions needed

Also see: http://www.youtube.com/watch?v=7OJtiA53-Fk

From: DCC's RDM for librarians

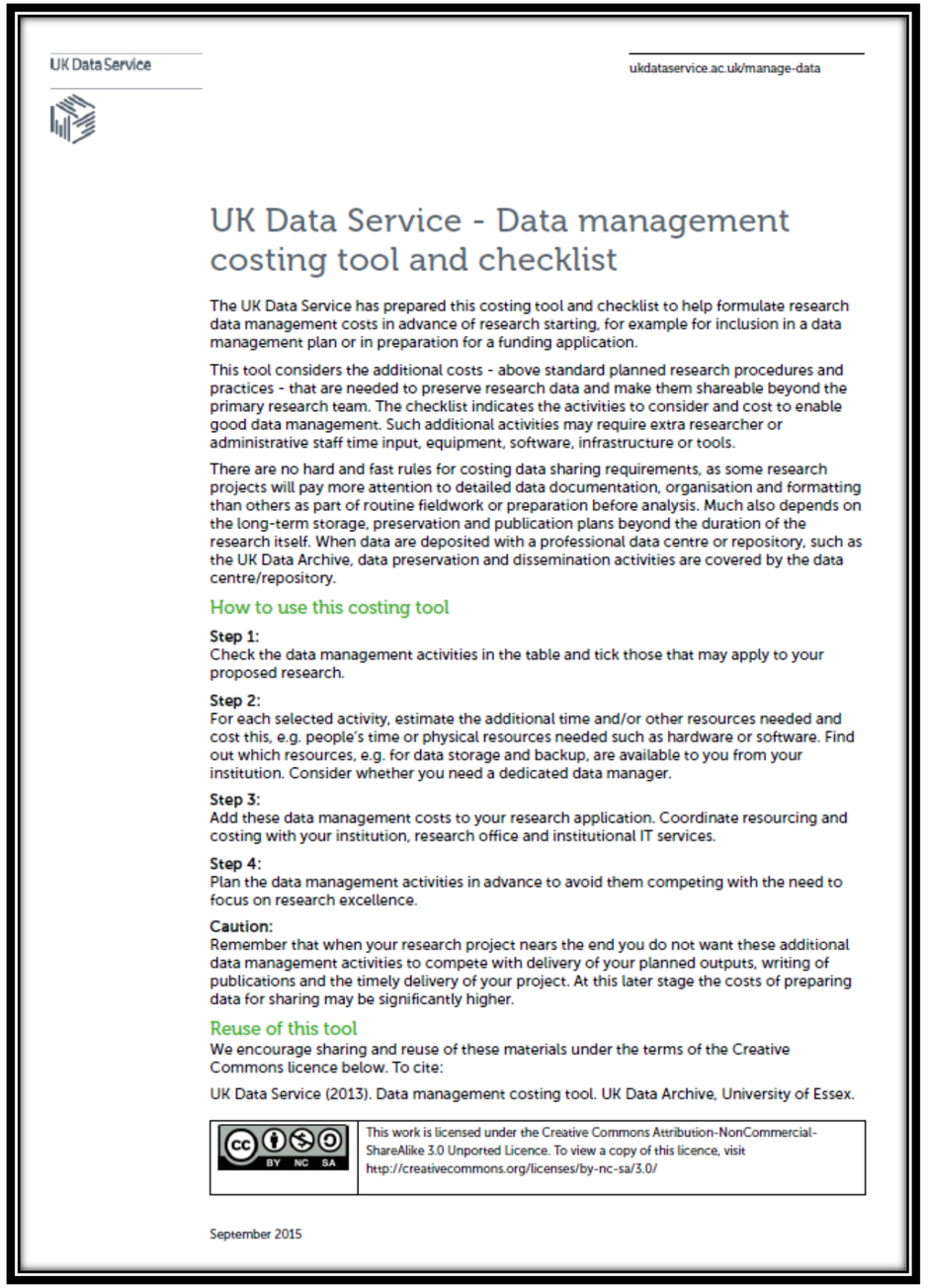
DMPs revision

- **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 so feedback may be variable initially.

UK Data Service

Data management costing tool and checklist

www.ukdataservice.ac.uk/media/622368/costingtool.pdf



UK Data Service

ukdataservice.ac.uk/manage-data

UK Data Service - Data management costing tool and checklist

The UK Data Service has prepared this costing tool and checklist to help formulate research data management costs in advance of research starting, for example for inclusion in a data management plan or in preparation for a funding application.

This tool considers the additional costs - above standard planned research procedures and practices - that are needed to preserve research data and make them shareable beyond the primary research team. The checklist indicates the activities to consider and cost to enable good data management. Such additional activities may require extra researcher or administrative staff time input, equipment, software, infrastructure or tools.

There are no hard and fast rules for costing data sharing requirements, as some research projects will pay more attention to detailed data documentation, organisation and formatting than others as part of routine fieldwork or preparation before analysis. Much also depends on the long-term storage, preservation and publication plans beyond the duration of the research itself. When data are deposited with a professional data centre or repository, such as the UK Data Archive, data preservation and dissemination activities are covered by the data centre/repository.

How to use this costing tool

Step 1:
Check the data management activities in the table and tick those that may apply to your proposed research.

Step 2:
For each selected activity, estimate the additional time and/or other resources needed and cost this, e.g. people's time or physical resources needed such as hardware or software. Find out which resources, e.g. for data storage and backup, are available to you from your institution. Consider whether you need a dedicated data manager.

Step 3:
Add these data management costs to your research application. Coordinate resourcing and costing with your institution, research office and institutional IT services.

Step 4:
Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.

Caution:
Remember that when your research project nears the end you do not want these additional data management activities to compete with delivery of your planned outputs, writing of publications and the timely delivery of your project. At this later stage the costs of preparing data for sharing may be significantly higher.

Reuse of this tool

We encourage sharing and reuse of these materials under the terms of the Creative Commons licence below. To cite:

UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex.

September 2015

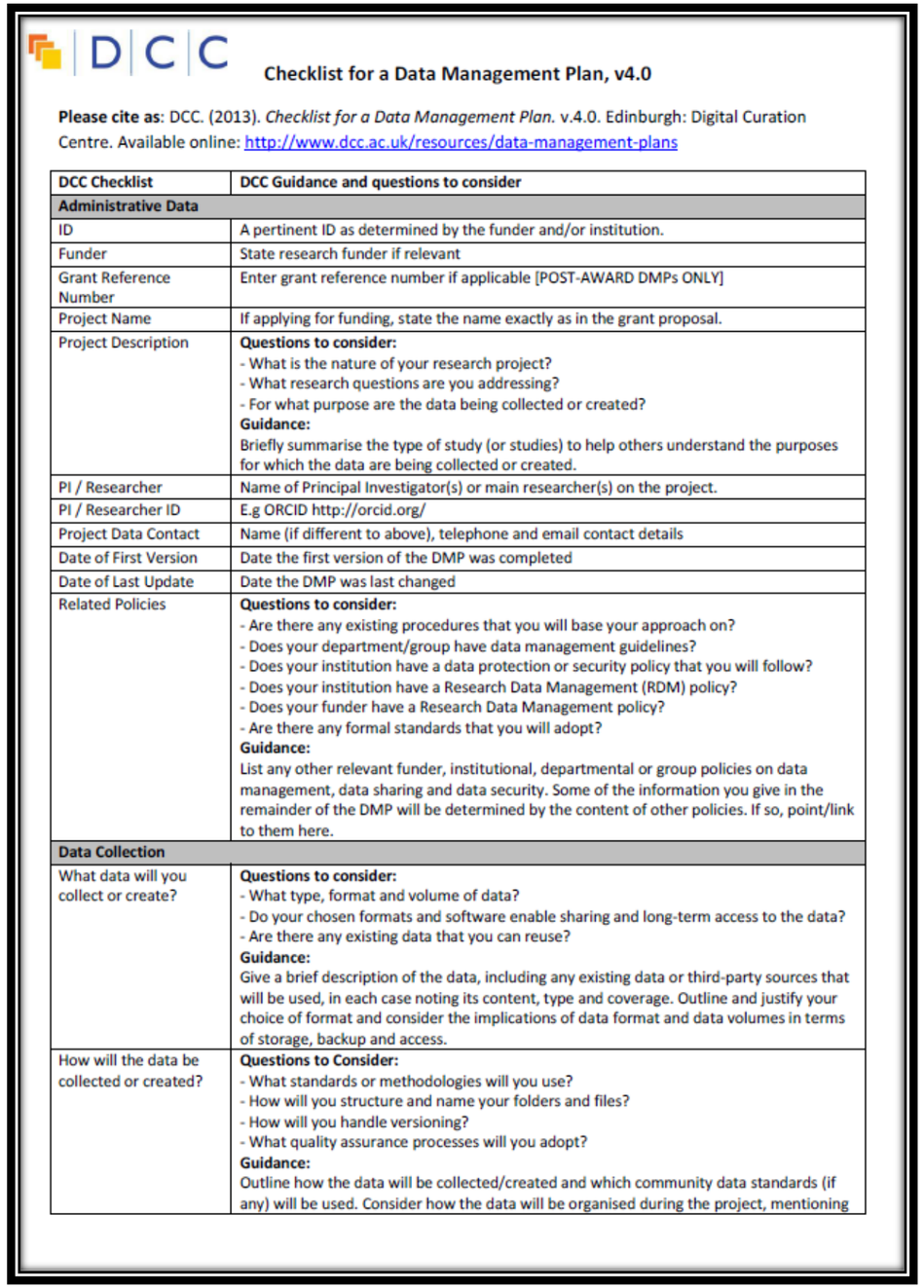
OpenAIRE

European Commission

Digital Curation Center

Checklist for a Data Management Plan

http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf



DCC

Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan.* v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines? - Does your institution have a data protection or security policy that you will follow? - Does your institution have a Research Data Management (RDM) policy? - Does your funder have a Research Data Management policy? - Are there any formal standards that you will adopt? **Guidance:** List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.
Data Collection	
What data will you collect or create?	**Questions to consider:** - What type, format and volume of data? - Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse? **Guidance:** Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.
How will the data be collected or created?	**Questions to Consider:** - What standards or methodologies will you use? - How will you structure and name your folders and files? - How will you handle versioning? - What quality assurance processes will you adopt? **Guidance:** Outline how the data will be collected/created and which community data standards (if any) will be used. Consider how the data will be organised during the project, mentioning

OpenAIRE

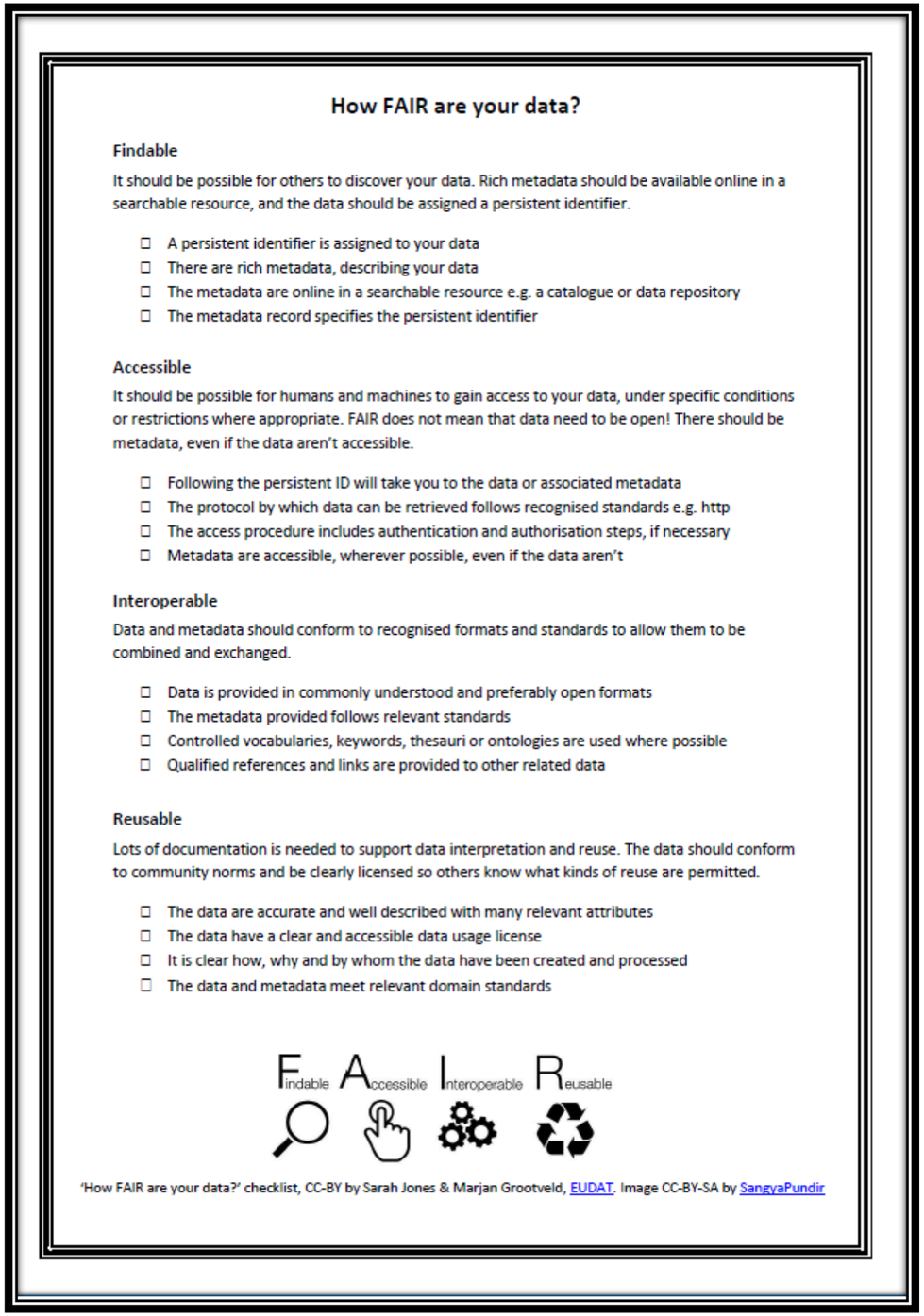
European Commission

FAIR data checklist

- **How FAIR is your data? Checklist**

  https://zenodo.org/record/1065991

- **FAIR data does not equal open data**
  - Data can remain closed
  - Open access data may not be FAIR (e.g. if they lack documentation of don't have a license explaining conditions for reuse.



How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

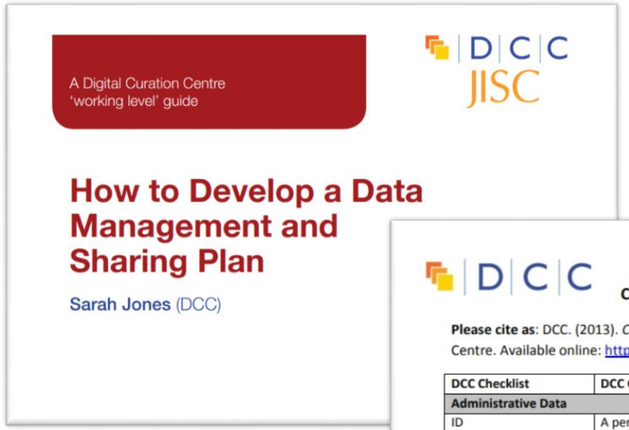
Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

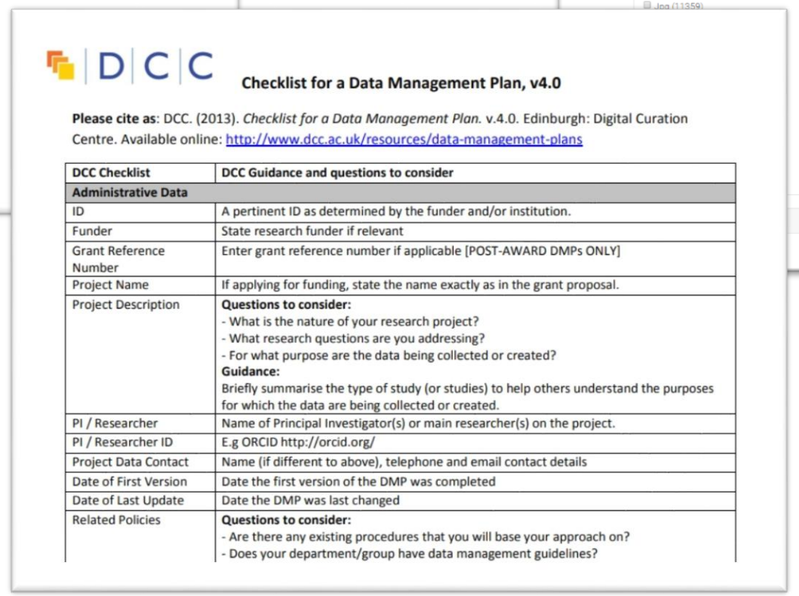
- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

OpenAIRE

European Commission

Guidance



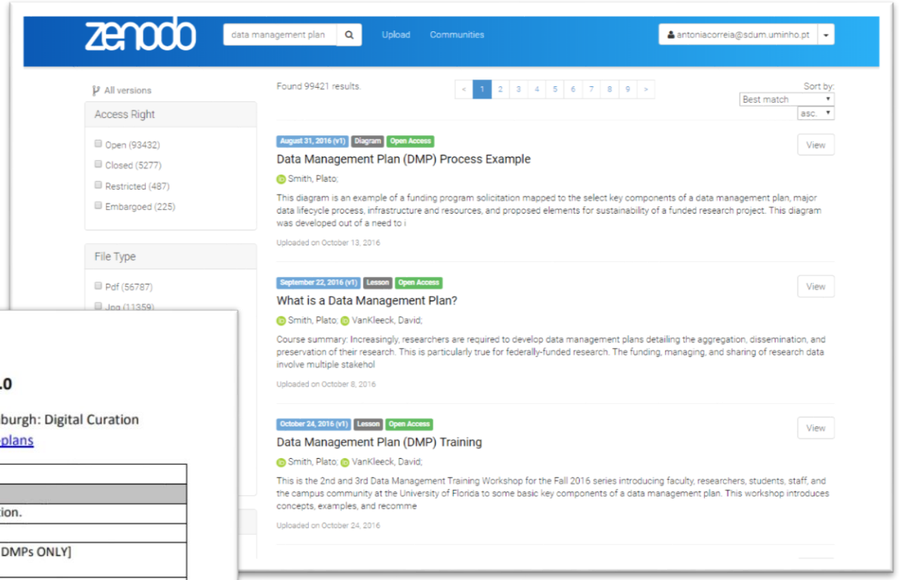


D C C

Checklist for a Data Management Plan, v4.0

Please cite as: DCC. (2013). *Checklist for a Data Management Plan.* v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution.
Funder	State research funder if relevant
Grant Reference Number	Enter grant reference number if applicable [POST-AWARD DMPs ONLY]
Project Name	If applying for funding, state the name exactly as in the grant proposal.
Project Description	**Questions to consider:** - What is the nature of your research project? - What research questions are you addressing? - For what purpose are the data being collected or created? **Guidance:** Briefly summarise the type of study (or studies) to help others understand the purposes for which the data are being collected or created.
PI / Researcher	Name of Principal Investigator(s) or main researcher(s) on the project.
PI / Researcher ID	E.g ORCID http://orcid.org/
Project Data Contact	Name (if different to above), telephone and email contact details
Date of First Version	Date the first version of the DMP was completed
Date of Last Update	Date the DMP was last changed
Related Policies	**Questions to consider:** - Are there any existing procedures that you will base your approach on? - Does your department/group have data management guidelines?



https://zenodo.org/

http://www.dcc.ac.uk/resources/data-management-plans

4

Tools for developing data management plans

Tools – DMP Online

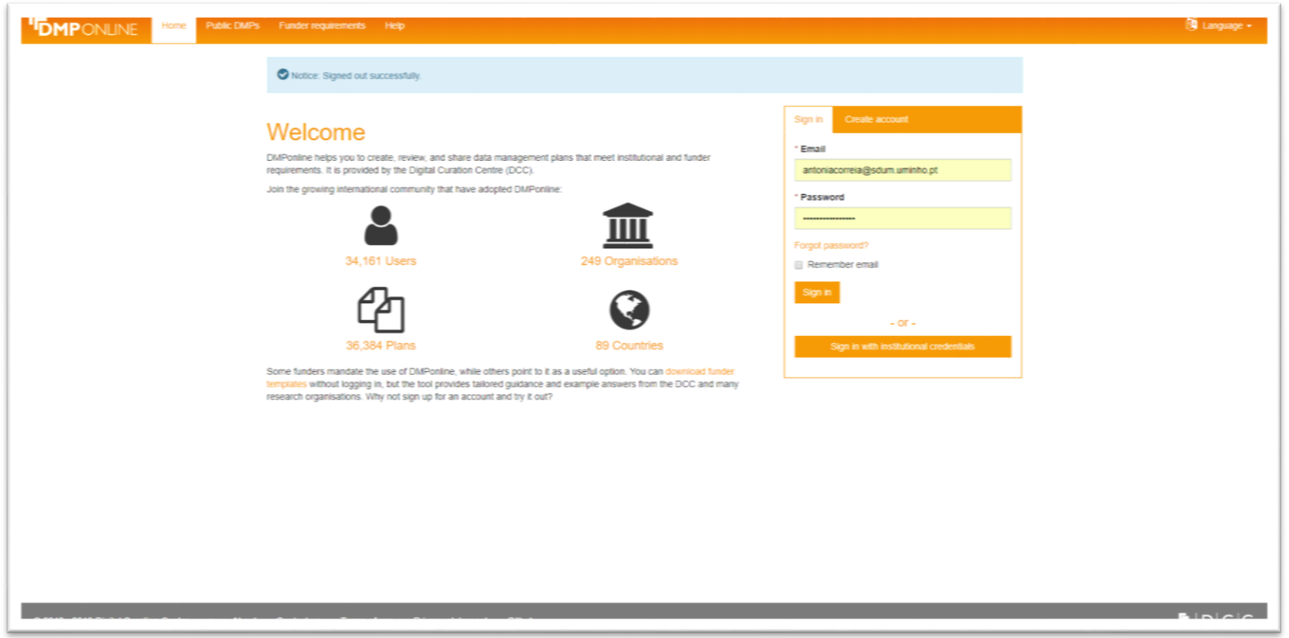
- DMP Online is a free tool developed by the Digital Curation Centre
- Relies on templates based on funding agencies for describing the data
- Generic DMP template for those not in a project
- Guidance
- Different versions for different stages of the project
- Allows sharing and exporting

https://dmponline.dcc.ac.uk

Public DMPs

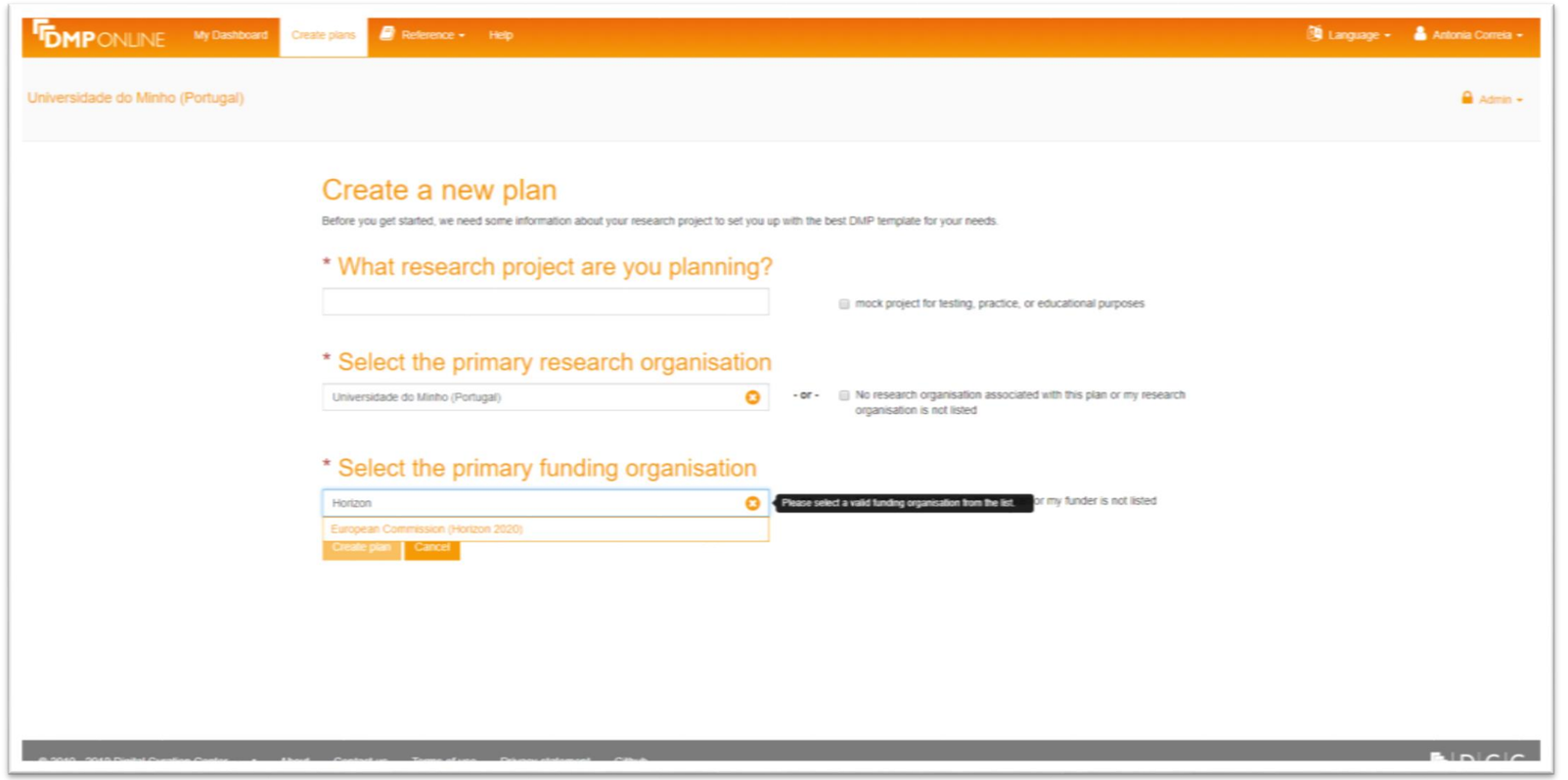
https://dmponline.dcc.ac.uk/public_plans

DMP Online register

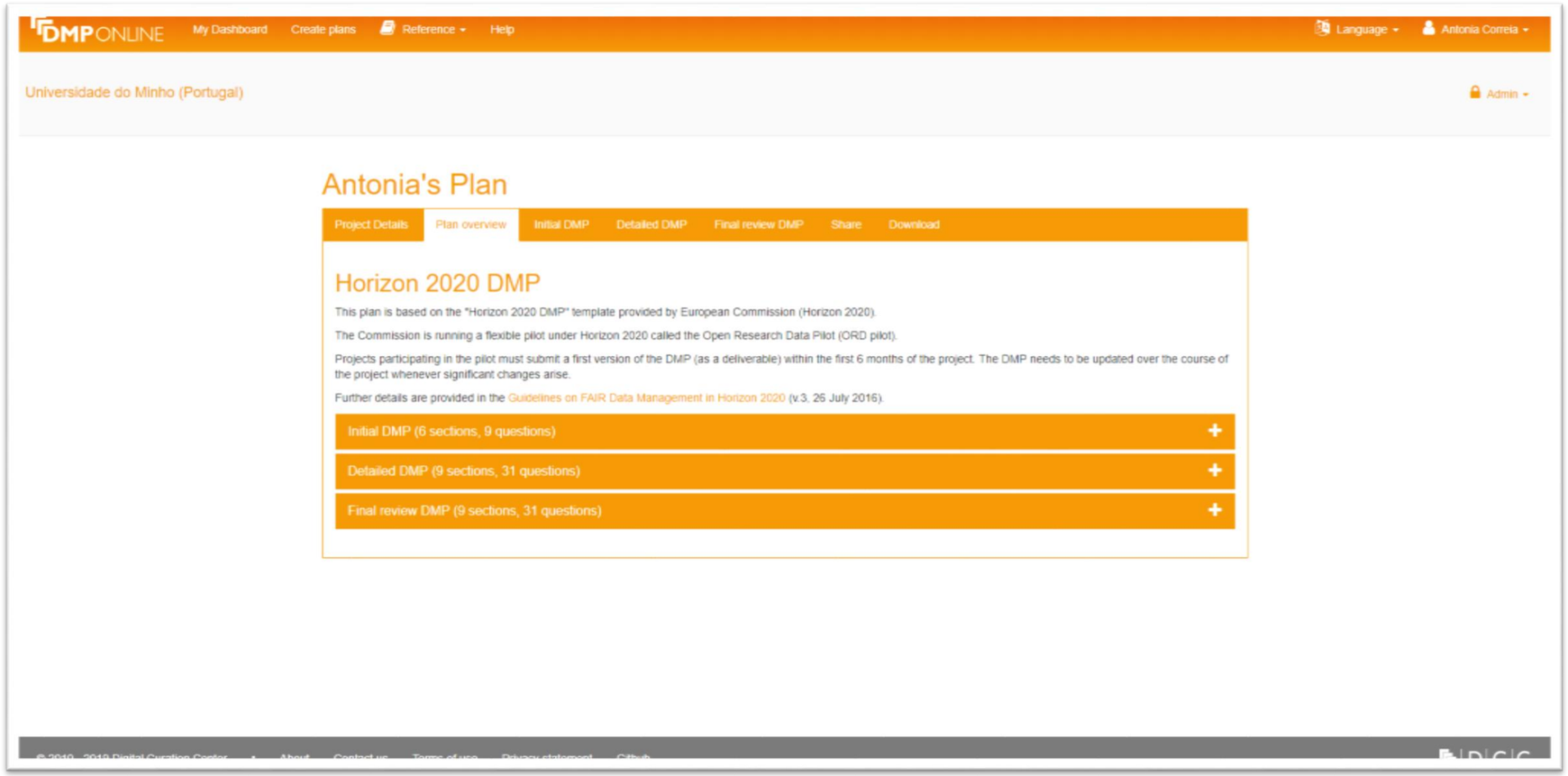


https://dmponline.dcc.ac.uk

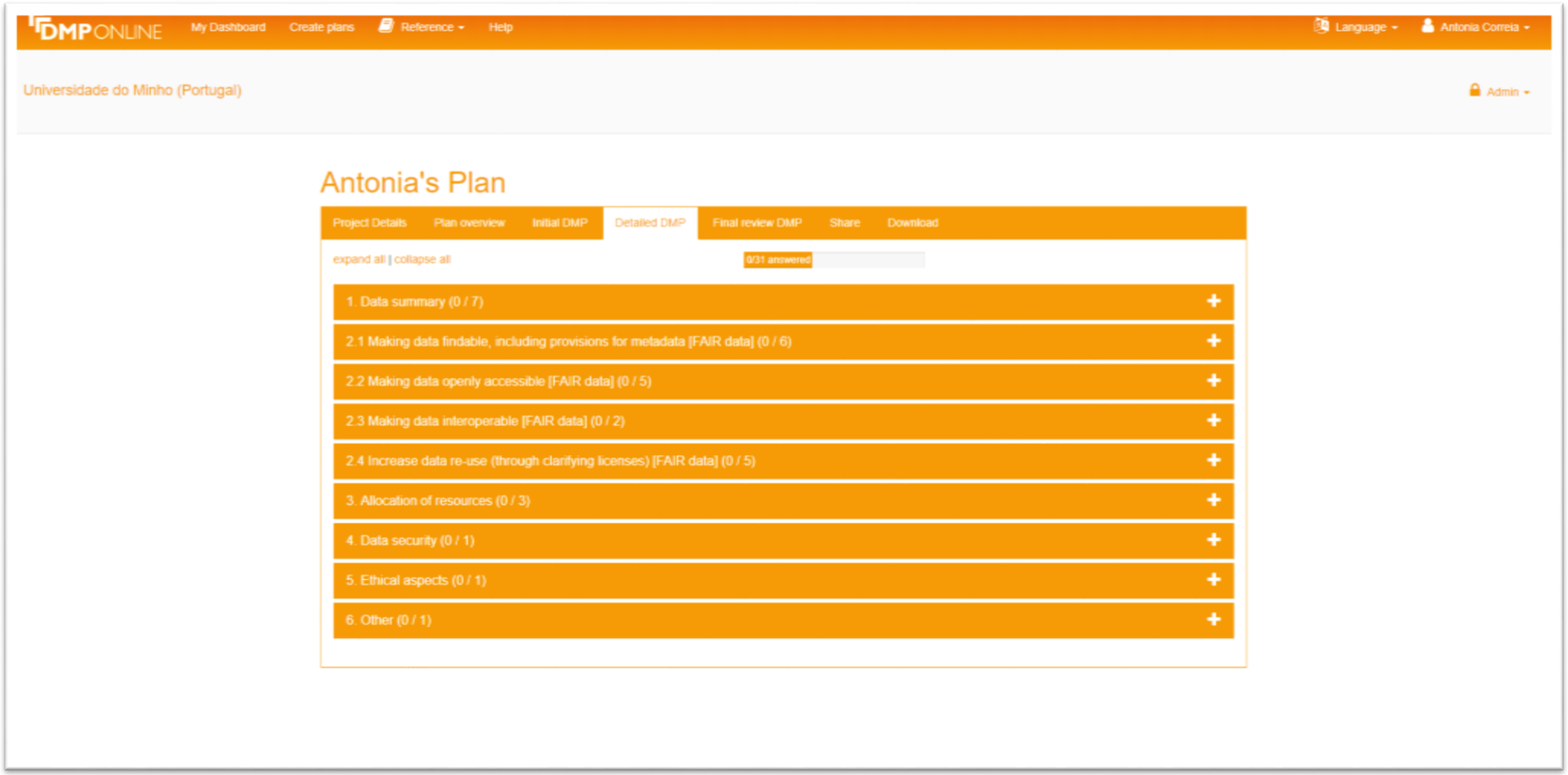
DMP Online template H2020 - create



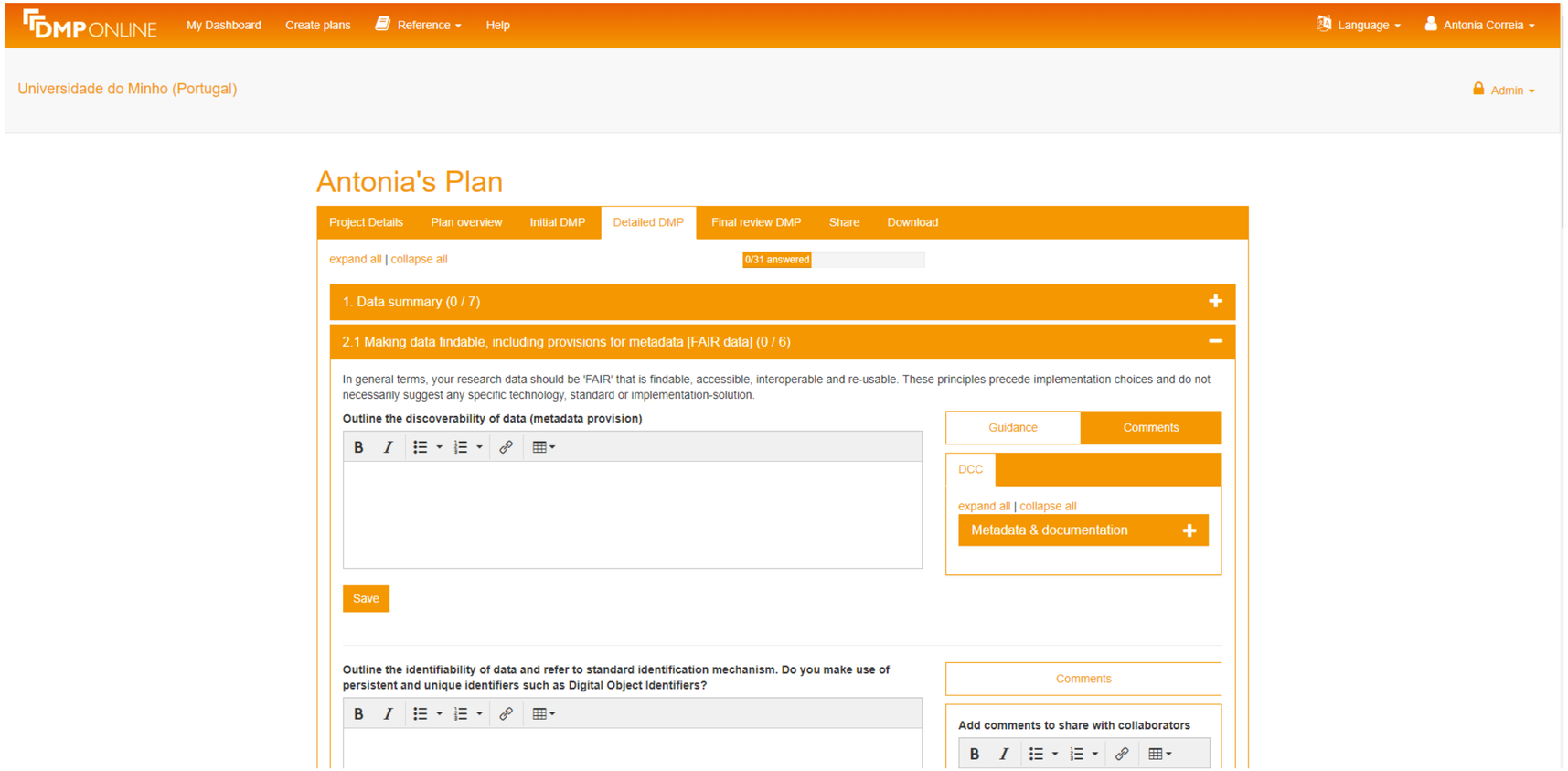
DMP Online template H2020 - versions



DMP Online template H2020 – detailed plan



DMP Online template H2020 - FAIR



Activity 2

- Start elaborating a Data Management Plan using the DCC checklist for a Data Management Plan and DMP Online
- Check the public DMPs for inspiration
- Point out your 3 main difficulties

ARGOS

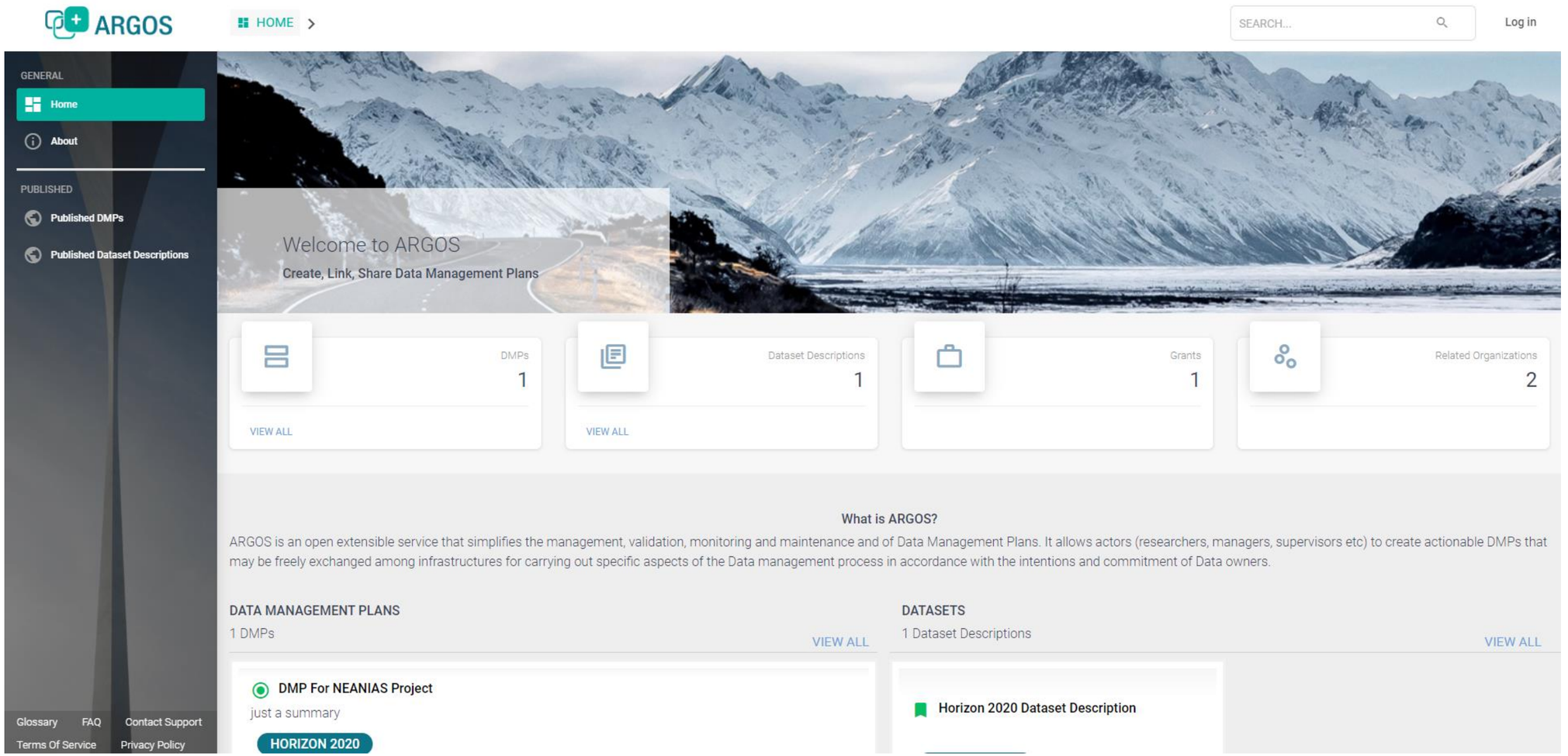
Create, Link, Share Data Management Plans

argos.openaire.eu

Functionalities:

- Collaborative creation of DMPs
- Specific or adaptative DMP templates for funders or institutions
- Dataset-based description
- DMP export to several formats
- REST web services to allow integration with research infrastructures
- Public roadmap for implementation tracking here: https://trello.com/b/x49lylnK/argos

OpenAIRE



argos.openaire.eu

OpenAIRE

References

- CESSDA Data Management Expert Guide
- Digital Curation Center
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
- Guidelines on Data Management in Horizon 2020
- Parthenos training
- Research Data MANTRA – Edina
- UK Data Archive

Thank you!

www.openaire.eu

@openaire_eu

facebook.com/groups/openaire

antoniacorreia@sdum.uminho.pt