



Open Access e FAIR data

Elena Giglia



Candiolo, 27 maggio 2019
elena.giglia@unito.it

Open Access

Open Access significa
accesso aperto, immediato
e libero da ogni restrizione
ai risultati e ai dati della ricerca scientifica



Open Access

Berlin Declaration

1. The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.

Open scholarly

10 Myths around Open Scholarly Publishing

Myth 1

Preprints will get your research 'scooped'

Preprints typically provide a time-stamp and a DOI, therefore establishing priority of discovery

Myth 6

Copyright transfer is required to publish and protect authors

Copyright transfer procedures do not protect authors nor contribute to the advancement of scientific progress

Myth 2

JIF and journal branding are measures of quality for researchers

The JIF is a flawed metrics that was never meant to be used for evaluation of research and researchers

Myth 7

Gold Open Access is synonymous with the APC business model

Most DOAJ-indexed journals do not have APCs and are funded from other sources, such as research institutes and grants

Myth 3

Approval by peer review proves that you can trust a research article

The current peer review system is prone to a number of flaws including corruption, human bias and ghostwriting

Myth 8

Embargo periods on 'green' OA are needed to sustain publishers

Traditional journals can peacefully coexist with zero-embargo self-archiving policies on author manuscripts

Myth 4

Without journal peer review, the quality of science suffers

Researchers are more than responsible and competent enough to ensure their own quality control as part of intrinsic scientific integrity

Myth 9

Web of Science and Scopus are global databases of knowledge

Neither represent the sum of current global research knowledge including Africa, Latin America and Southeast Asia

Myth 5

Open Access has created predatory publishers

Predatory journals have been around for a long time before the recent push towards Open Access publishing

Myth 10

Publishers add no value to the scholarly communication process

Publishers are responsible for quite some key functions, from peer-review management to production and archiving of final version articles

Preprint

NOT PEER-REVIEWED March 11, 2019

"PeerJ Preprints" is a venue for early communication of research before peer review. Data may be used for research purposes only. Learn more about preprints or browse peer-reviewed articles instead.

View 34 tweets

Ten myths around open scholarly publishing

Literature review Science and Medical Education Science Policy

Si — può — fare!!!!!!

SI PUÒ FARE OPEN SCIENCE
CONTEMPORANEAMENTE
A VQR, ASN, SUA-RD...

NON SONO INCOMPATIBILI!!!

Green road - deposito

81% degli editori internazionali lo consente
(Elsevier, Wiley, Springer...), elenco:



...ovviamente, con limiti:

- quasi mai pdf editoriale, spesso «**postprint**»
- possibile **embargo**

[=mesi in cui articolo pur depositato non è visibile]

Author's Pre-print:	✓ author can archive pre-print (ie pre-refereeing)
Author's Post-print:	✓ subject to Restrictions below, author can archive post-print (ie final draft post-refereeing)
Restrictions:	• 12 months embargo
Publisher's Version/PDF:	✗ author cannot archive publisher's version/PDF
General Conditions:	• Some journals have separate policies • On author's personal website • Author's pre-print may not be used for other purposes • Author's pre-print must acknowledge the journal • Non-Commercial • Publisher's version/PDF cannot be used for other purposes • Publisher source must be acknowledged • Must link to publisher version • If OnlineOpen is available, Biomed Central must be used • If OnlineOpen is available, BioRxiv must be used

Attenzione...

l'editore ha voce in capitolo
SOLO SE avete firmato CONTRATTO
con CESSIONE DI TUTTI I DIRITTI



Arsenate toxicity on the apices of *Pisum sativum* L. seedling roots: Effects on mitotic activity, chromatin integrity and microtubules

Stefania Dho, Wanda Camusso, Marco Mucciarelli, Anna Fusconi

Abstract

Arsenic (As) is one of the most toxic plant growth. Despite the growing of this element on meristem activity study, short-term experiments with whether plant growth impairment was studied by evaluating apical fragmentation and microtubule or that arsenate, at the lowest concentrations, whilst the other concentrations, mitotic and labelling index (after bromo-deoxyuridine administration) (through immunofluorescence). Mitotic metaphases increased, as did the mitotic spindles, which closely resemble those induced by colchicine. Chromosome breaks and anaphase/telophase bridges were virtually absent. These data point to a poor clastogenic activity of As and implicate that microtubules are one of the main targets of As.

Keywords

Pea; Arsenic; Apical meristems; Aberrations; Immunofluorescence; TUNEL test

1. Introduction

Arsenic (As) is a toxic element, frequently found in soils and water. A main natural source of As is the erosion of mother rock, even though a consistent part of As environmental pollution comes from human activities (Meharg and Hartley-Whitaker, 2002 and Patra et al., 2004). The As in unpolluted fresh water is usually in the range 1–10 µg/l. According to EPA and WHO, the maximum permissible As concentration in drinking water is 50 µg/l (Mandal and Suzuki, 2002).

Arsenic is a well-established human carcinogen (Qin et al., 2008a) and has been shown to be genotoxic in a variety of *in vitro* studies (Hughes, 2002). In plants, it severely affects growth and development, and its toxicity is strongly dependent on the concentration, exposure time and physiological state of the plant (Singh et al., 2007). However, plants vary in their sensitivity to As, and a wide range of species have been identified in As-contaminated soils (Meharg and Hartley-Whitaker, 2002). Besides, hyperaccumulators such as *Pteris vittata*, which tolerate high internal As content, may also use this As to defence themselves against herbivore attack (Mathews et al., 2009).

Higher plants take up As mainly as arsenate (V), the dominant form of phytoavailable As in aerobic soils. According to Meharg and Hartley-Whitaker (2002), As competes with phosphate for plant phosphate transporters. Upon absorption, most arsenate is rapidly reduced to arsenite (III), due to an arsenate reductase activity (Xu et al., 2007), hence, the arsenate cytoplasmic concentration is generally not high enough to exert toxicity (Meharg and Hartley-Whitaker, 2002). Both As species interfere with various metabolic pathways: arsenate, as an analogous chemical to phosphate, may replace phosphate in the ATP and in various



UNIVERSITÀ DEGLI STUDI DI TORINO

This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process – such as editing, corrections, structural formatting, and other quality control mechanisms – may not be reflected in this version of the text. The definitive version of the text was subsequently published in ENVIRONMENTAL AND EXPERIMENTAL BOTANY, 69(1), 2010, 10.1016/j.envexpbot.2010.02.010.

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

- (1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>), 10.1016/j.envexpbot.2010.02.010

The definitive version is available at:
<http://linkinghub.elsevier.com/retrieve/pii/S0098847210000353>



Contents lists available at ScienceDirect

Environmental and Experimental Botany

journal homepage: www.elsevier.com/locate/envexpbot



Arsenate toxicity on the apices of *Pisum sativum* L. seedling roots: Effects on mitotic activity, chromatin integrity and microtubules

Stefania Dho^a, Wanda Camusso^a, Marco Mucciarelli^b, Anna Fusconi^{a,*}

^a Dipartimento di Biologia Vegetale, CEBIOVEM, Viale Mattioli 25, I-10125 Torino, Italy

^b Dipartimento di Morfologia Vegetale, Via Leonardo da Vinci 44, I-10095 Grugliasco (To), Italy

ARTICLE INFO

Article history:

Received 20 July 2009

Received in revised form 9 February 2010

Accepted 14 February 2010

Keywords:

Pea
Arsenic
Apical meristems
Aberrations
Immunofluorescence
TUNEL test

ABSTRACT

Arsenic (As) is one of the most toxic pollutants in the environment, where it severely affects both animal and plant growth. Despite the growing literature data on As effects on plant development, alterations induced by this element on meristem activity of the root have not been explored to any great extent. In the present study, short-term experiments with arsenate have been conducted on *Pisum sativum* L. seedlings to assess whether plant growth impairment is due to DNA/chromosome or mitotic microtubule damages. Root growth was studied by evaluating apical meristem activity and cell elongation. Mitotic aberrations, DNA fragmentation and microtubule organization of the apical cells were also analyzed. The results have shown that arsenate, at the lowest concentration (0.25 µM), slightly increases root growth and some related parameters, whilst the other concentrations have a dose-dependent negative effect on root growth, on the mitotic and labelling index (after bromo-deoxyuridine administration), and on the mitotic arrays of microtubule (through immunofluorescence). The main effects on mitosis occurred for 25 µM As. The percentage of metaphases increased, as did the irregular metaphases and c-mitoses. This was related to alterations in the mitotic spindles, which closely resemble those induced by colchicine. Chromosome breaks and anaphase/telophase bridges were virtually absent, whilst DNA fragmentation only increased from 25 µM arsenate onwards. These data point to a poor clastogenic activity of As and implicate that microtubules are one of the main targets of As.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Arsenic (As) is a toxic element, frequently found in soils and water. A main natural source of As is the erosion of mother rock, even though a consistent part of As environmental pollution comes from human activities (Meharg and Hartley-Whitaker, 2002; Patra et al., 2004). The As in unpolluted fresh water is usually in the range 1–10 µg/l. According to EPA and WHO, the maximum permissible As concentration in drinking water is 50 µg/l (Mandal and Suzuki, 2002).

Arsenic is a well-established human carcinogen (Qin et al., 2008a) and has been shown to be genotoxic in a variety of *in vitro* studies (Hughes, 2002). In plants, it severely affects growth and development, and its toxicity is strongly dependent on the concentration, exposure time and physiological state of the plant (Singh et al., 2007). However, plants vary in their sensitivity to As, and a wide range of species have been identified in As-contaminated soils (Meharg and Hartley-Whitaker, 2002). Besides, hyperaccumulators such as *Pteris vittata*, which tolerate high internal As content,

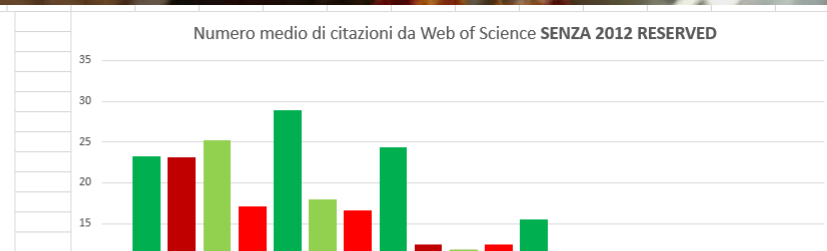
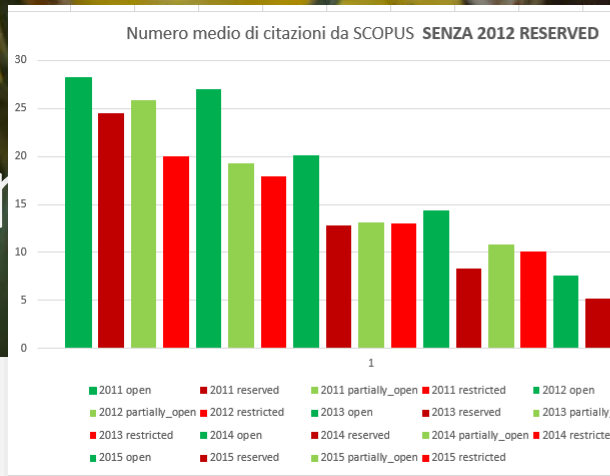
may also use this As to defence themselves against herbivore attack (Mathews et al., 2009).

Higher plants take up As mainly as arsenate (V), the dominant form of phytoavailable As in aerobic soils. According to Meharg and Hartley-Whitaker (2002), As competes with phosphate for plant phosphate transporters. Upon absorption, most arsenate is rapidly reduced to arsenite (III), due to an arsenate reductase activity (Xu et al., 2007), hence, the arsenate cytoplasmic concentration is generally not high enough to exert toxicity (Meharg and Hartley-Whitaker, 2002). Both As species interfere with various metabolic pathways: arsenate, as an analogous chemical to phosphate, may replace phosphate in the ATP and in various phosphorylation reactions, leading to the disruption of the energy flow in cells. The toxicity of arsenite is mainly ascribed to its reaction with sulphhydryl groups of proteins that interfere with their functions (Meharg and Hartley-Whitaker, 2002; Patra et al., 2004).

Exposure to high concentrations of As induces the production of reactive oxygen species (ROS) (Singh et al., 2007; Wang et al., 2007; Lin et al., 2008; Shri et al., 2009) and the conversion of arsenate to arsenite is regarded as one of the causes of ROS generation (Wang et al., 2007). Oxidative stress induced by As can damage cells, mainly through lipid peroxidation of membranes (Singh et al., 2007) and DNA fragmentation, as has been demonstrated in leaves and roots

* Corresponding author. Tel.: +39 011 6705968; fax: +39 011 6705962.
E-mail address: anna.fusconi@unito.it (A. Fusconi).

I var



Demonstrating Achievements

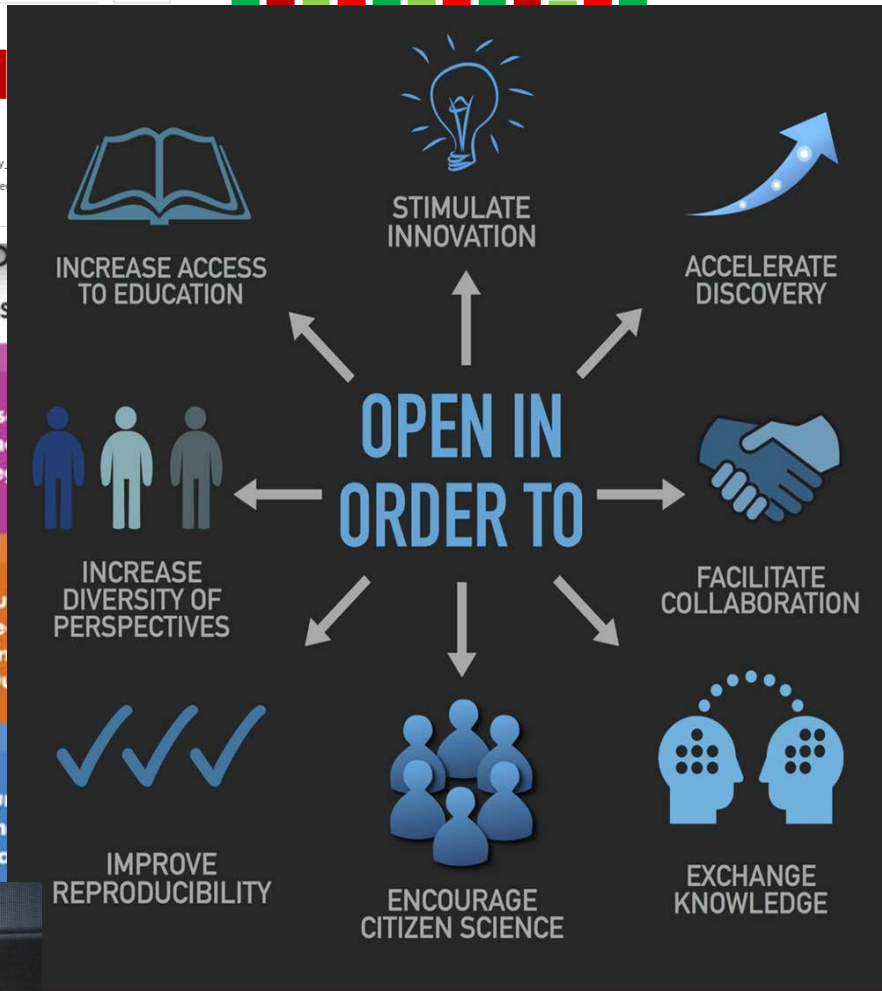
Open access for Institutions



bepress™

Reputation

Open access



whyopenresearch.org

#OAweek

GOLD OA - In cosa le riviste Open Access sono diverse?

PUBBLICANO I **DATI** INSIEME ALL' ARTICOLO

- **TRASPARENZA**
- **RIPRODUCIBILITÀ**

PUBBLICANO LE **REVISIONI** INSIEME ALL' ARTICOLO

- **TRASPARENZA**
- **CONOSCENZA**

PUBBLICANO CON **LICENZE** CREATIVE COMMONS E
NON CHIEDONO CESSIONE DEI DIRITTI

- **RIUSO**
- **TEXT E DATA MINING**

PUBBLICANO RAPIDAMENTE

(spesso) PUBBLICANO IN FORMATI MACHINE-READABLE

- **TEXT E DATA MINING**

La «red road»

la Gold road **non** è la «Open Choice»
degli editori tradizionali
(Elsevier, Springer, Wiley...)

pagando 3000 \$,
UN SINGOLO articolo viene reso Open Access,
mentre la rivista resta IN ABBONAMENTO

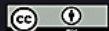
di fatto, paghiamo due volte...

serve solo se l'ente finanziatore
stabilisce embargo massimo inferiore a
quello stabilito dall'editore
(es. Horizon 2020, che però rimborsa)

[DA EVITARE SE POSSIBILE, perché auto-
archiviando ottengo stesso effetto, gratis!!!]

COME RENDERE LA VOSTRA RICERCA OPEN ACCESS

LEGALMENTE E GRATUITAMENTE



Jon Tennant and Lisa Matthias
Translated by Elena Giglia

... in pratica

...c'è sempre un modo
per fare Open Access,
anche gratis
(e questo non rovina la
vostra carriera)

Be Happy
Sevilla

...parliamo di dati (di nuovo)

We
need
Stagg



Perché occuparci dei dati?

1995 WaveLab and Reproducible Research

Jonathan B. Buckheit and David L. Donoho

Stanford University, Stanford CA 94305, USA

*An article about computational science in a scientific publication is **not** the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

UN ARTICOLO SENZA I
DATI È SOLO LA
PUBBLICITÀ DELLA
RICERCA



No data?

Is withholding your data simply bad science, or should it fall under scientific misconduct?



A recent study sent data requests to 200 researchers and 100 journals, stating 'data available upon request'. Most researchers think about those withholding data as misconduct? **Nicole Janz** argues that if you have professional standards in research, and you classify data secrecy as misconduct



Alastair Dunning

@alastairdunning

Following

To me, data are like footnotes. I might not always read them, but I get suspicious if they are not there.

Traduci dalla lingua originale: inglese

12:49 - 27 feb 2018

<https://twitter.com/alastairdunning/status/968453078218395648>

2 Retweet 8 Mi piace



Jon Tennant

@Protohedgehog

Following

My first talk of the year! Message is going to be that the opposite of 'open science' isn't 'closed science' - it's bad science.

**Gold Standard
Research Integrity**

**Questionable Research
Practices**

**Scientific
Misconduct**



Data secrecy

Open data
Open code
Pre-registration
Version control

P-hacking
Sloppy statistics
Peer review abuse
Inappropriate research design
Not answering to replicators
Lying about authorships

Fabrication
Falsification
Plagiarism

Parliamo di dati

«pezzi»
di conoscenza osservabili

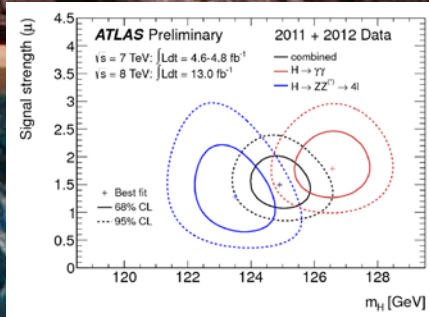


Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample from the cohort.

OTU	A01_TP1	A01_TP2	A01_TP3	A03_TP1	A03_TP3	A04_TP1	A04_TP2	A04_TP3	A05_TP1
OTU_1	261	76	1206	523	2131	25707	64473	60665	
OTU_2	49	52	117	43035	206	119	1152	539	
OTU_9	148	162176					22858	1898	
OTU_6	21	17					1457	29	
OTU_1	19	85					546	214	
OTU_2	292	37					18	170	
OTU_6	6	4							



Wilma van Wezenbeek

@wvanwezenbeek

Following

#osc2018 Wolfram Horstmann wants us to talk about datadiversity, like we do with biodiversity #openscience

Traduci il Tweet

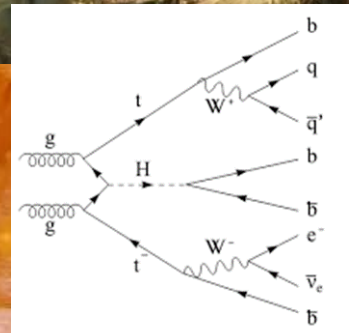
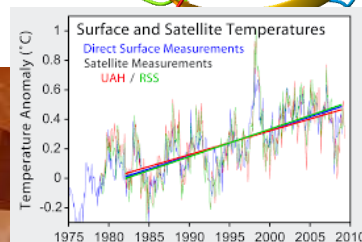
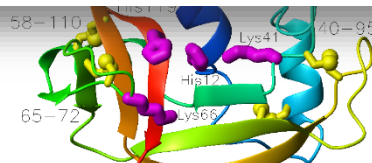
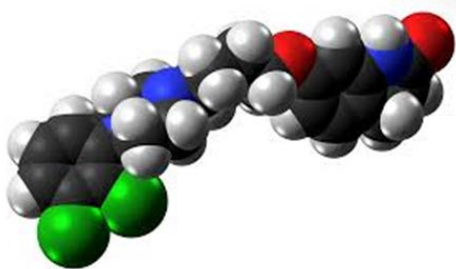
12:51 - 13 mar 2018

3 Retweet 1 Mi piace

<https://twitter.com/wvanwezenbeek/status/973527086685093893>

Gaucelm Faidit

I.
Ara nos sia guitz
lo vers dieus Iesu Cristz,
car de franca gen gaia
soi per Lui partitz,
on ai estat noiritz
et onratz e grazitz;
per so-l prec no-ill desplaia
s'ieu m'en vauc marritz.
A! gentils lemozis,
el vostr'onrat pais
lais de bella paria
seignors e vezis
e domnas ab pretz fis,
pros, de gran cortesia,
don plane e languis
e sospir nueg e dia.



Un po' di glossario

rdn! research data netherlands

ABOUT THE COURSE » START THE COURSE » LOGIN »

Essentials 4 Data Support » Start the course » I-Definitions » Research data

I-Definitions

- Research data
- Open data
- Research lifecycle

I DEFINITIONS

II PLANNING PHASE

III RESEARCH PHASE

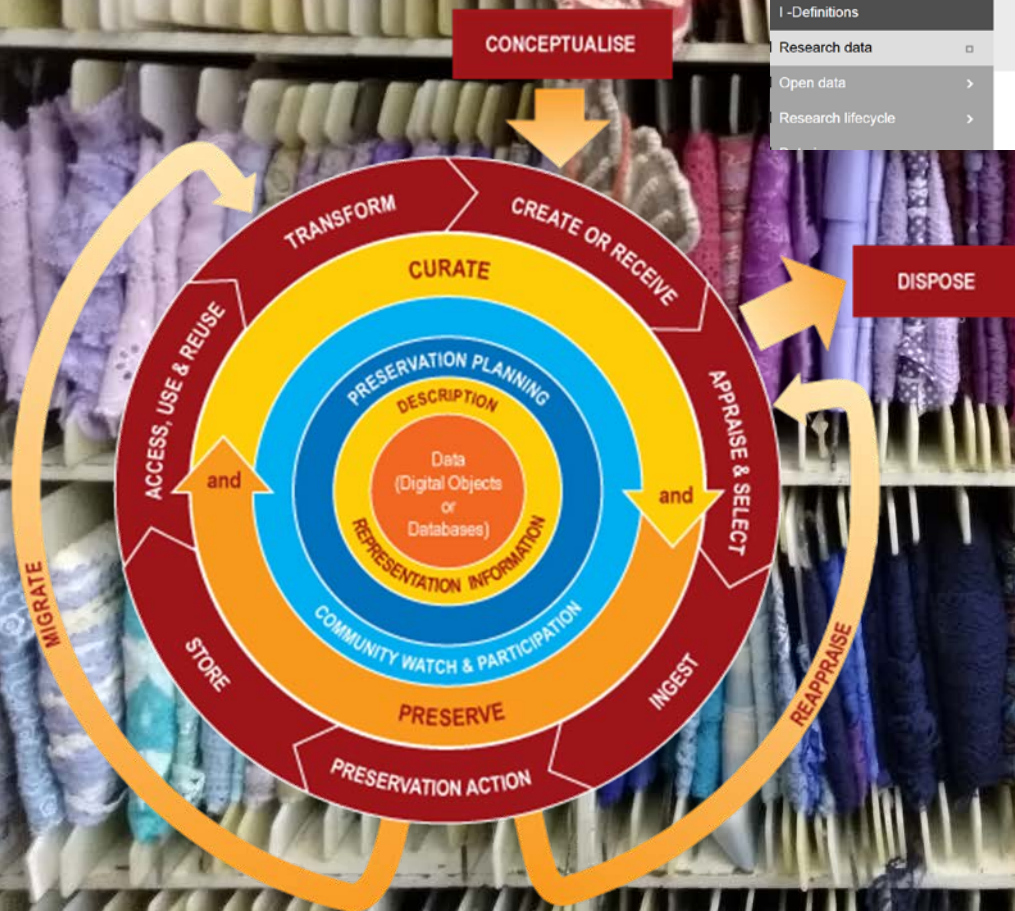
IV USER PHASE

V LEGISLATION & POLICY

VI DATA SUPPORT

Research data

Essentials4data



Un po' di glossario

5 modi per pensare i dati:

- come sono raccolti (esperimenti, simulazioni...)
- come si presentano (testi, questionari, video...)
- il loro formato elettronico (.txt, .mkv...)
- il loro volume (big data...)
- in che fase sono del ciclo (raw data...)

▣ The way the data is collected.

- ▣ By experimenting, simulations, observations, derived data, reference data.

▣ The data forms.

- ▣ For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, slides, artefacts, models, scripts, databases, metadata, etc.

▣ The formats for electronic storage of the research data.

▣ The size (volume) of the data files.

▣ The *research lifecycle* phase the data is in.



Due pilastri, anzi tre

 because good research needs good data

Home Digital curation About us News Events Resources Training Projects Com...

Home > Resources

In this section

- Briefing Papers
- How-to Guides & Checklists
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies
- Repository audit and assessment
- Standards
- Publications and presentations
- Roles
- Curation journals
- Informatics research
- External resources
- Online Store

Resources for digital curators

With just a few clicks, you can access any of the data curation resources collected by the DCC since we opened our doors in 2004.

Our comprehensive and easily accessible digital library is completely free to use and aims to provide you with everything you need to evaluate and implement those digital curation techniques most suited to your particular research project.

What's more, we regularly review and update our library, adding the latest resources developed for data curators.

Select from the links below and left to access high-level digital curation briefing papers; legal watch, standards watch and technology watch papers; case studies and interviews; and instalments from our detailed Curation Reference Manual.

International Journal of Digital Curation

Our highly regarded digital journal, published twice a year, features general articles and peer reviewed papers and serves as an invaluable channel for the

<http://www.dcc.ac.uk/resources>

 AUSTRALIAN NATIONAL DATA SERVICE

About ANDS | Events | Contact Us

<https://www.ands.org.au/>

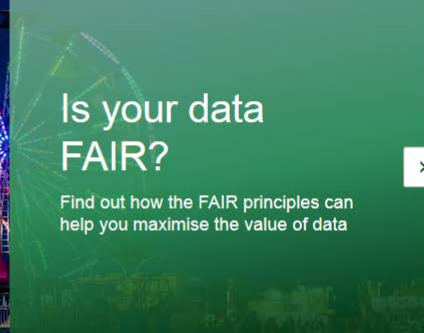
About us News and Events Partners and Communities Working with data Online Services Guides and resources

Findable Accessible Interoperable Reusable

Is your data FAIR?

Find out how the FAIR principles can help you maximise the value of data





 Data Archiving and Networked Services

Netherlands Contact Search this website

<https://dans.knaw.nl/en>

HOME DEPOSIT SEARCH TRAINING AND CONSULTANCY PROJECTS ABOUT DANS NEWS AND EVENTS

Welcome at DANS: the Netherlands institute for permanent access to digital research resources.

What can we do for you?



ARCHIVING

Deposit your datasets in DataverseNL or EASY or send research data for NARCIS.

DEPOSIT



REUSE

Find datasets, publications, researchers, projects and institutions via NARCIS and EASY.

SEARCH



TRAINING & CONSULTANCY

Let DANS advise you on data management and certification of digital archives.

ADVICE FROM DANS

SPOTLIGHT

DRYAD and DANS partner for long-term preservation research data

Dryad and DANS announce a new collaboration to ensure long-term preservation and accessibility to curated scientific data. Over 50,000 researchers who have already deposited research data with Dryad can count on continuous open access to their data packages with an extra layer of security and recoverability as a result of this partnership.



NEWS

CoreTrustSeal certification launched

The Data Seal of Approval (DSA) and ICSU World Data System (WDS) announce the launch of a new certification organization: CoreTrustSeal.

 DANS Retweeted

 Marta Terepek @martaterepek

Nice demo by @pkdoom @DANSKNW - tool to help historians decide which @re3data repository to use to archive their datasets: d4rs-dev.dans.knaw.nl/dcc18

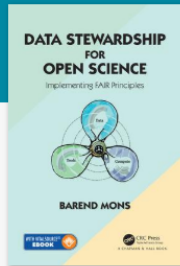
Feb 20, 2018

... e un maestro



Taylor & Francis Group
an informa business

<https://www.taylorfrancis.com/books/9781498753180>



Data Stewardship for Open Science Implementing FAIR Principles

the worst way imaginable to communicate the outcome of the scientific process. If science has become indeed data driven and *data is the oil of the 21st century*, we better put data centre stage and publish data as first-class research objects, obviously with supplementary narrative where needed, steward them throughout their life cycle, and make them available in easily reusable format.

Yet another recent study claimed that only about 12% of NIH funded data finds its way to a trusted and findable repository. Philip Bourne, when associate director for data science at the U.S.A. National Institutes of Health coined the term dark data for the 88% that is lost in amateur repositories or on laptops. When we combine the results of the general reproducibility related papers and the findability studies,

GET ACCESS

PREVIEW PDF



Monsense and more... @barendmons · 2 h

Finally! Tomorrow the book goes to the printer: Data Stewardship for Open Science: Implementing FAIR Principles

Traduci dalla lingua originale: inglese



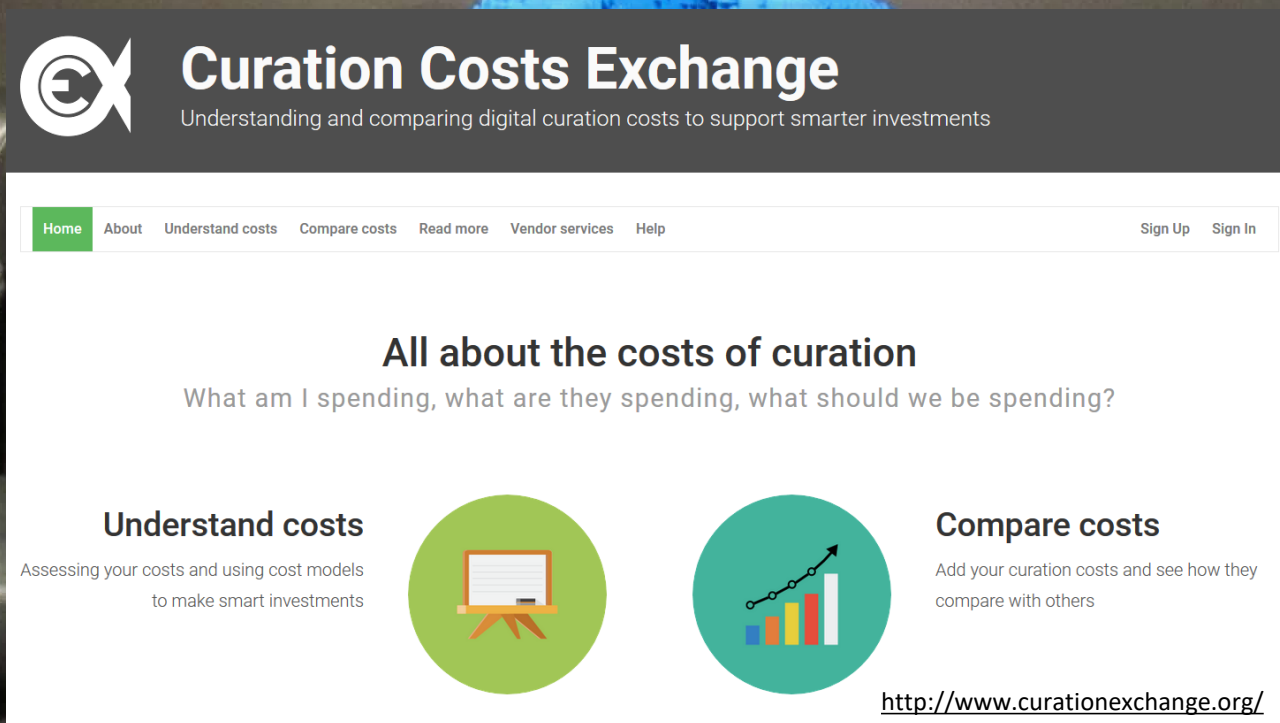
Data Stewardship for Open Science: Implementing ...
Data Stewardship for Open Science: Implementing FAIR Principles has been written with the intention of making scientists, funders, and innovators in all disciplines an...
crcpress.com



In conclusion to this paragraph, my statement in 2005: Text-mining? Why bury it first and then mine it again? [Mons, 2005] is still frighteningly relevant.

A good data steward publishes data with a supplementary article(Data(+)).

I costi



The screenshot shows the homepage of the Curation Costs Exchange. At the top, there is a dark header with the logo (a stylized 'C' with an 'X') and the title 'Curation Costs Exchange' in white. Below the title is the tagline 'Understanding and comparing digital curation costs to support smarter investments'. A navigation bar below the header contains links: 'Home' (highlighted in green), 'About', 'Understand costs', 'Compare costs', 'Read more', 'Vendor services', 'Help', 'Sign Up', and 'Sign In'. The main content area has a heading 'All about the costs of curation' followed by the question 'What am I spending, what are they spending, what should we be spending?'. Below this, there are two columns. The left column is titled 'Understand costs' and includes the text 'Assessing your costs and using cost models to make smart investments' next to a green circular icon of a whiteboard. The right column is titled 'Compare costs' and includes the text 'Add your curation costs and see how they compare with others' next to a teal circular icon of a bar chart with an upward arrow. At the bottom right of the screenshot is the URL <http://www.curationexchange.org/>.

Curation Costs Exchange
Understanding and comparing digital curation costs to support smarter investments

Home About Understand costs Compare costs Read more Vendor services Help Sign Up Sign In

All about the costs of curation

What am I spending, what are they spending, what should we be spending?

Understand costs

Assessing your costs and using cost models to make smart investments

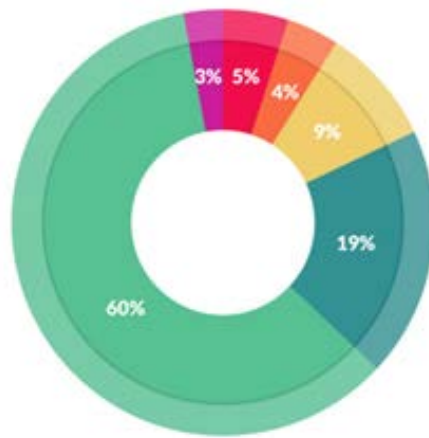
Compare costs

Add your curation costs and see how they compare with others

<http://www.curationexchange.org/>

**CI SONO COSTI PER CONSERVARE E GESTIRE I DATI...
MA PENSIAMO A QUANTO COSTEREBBE
NON CONSERVARLI E NON GESTIRLI**

Costi



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Data science report, 2016, cit. by Erik Schultes



Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

10,2 bn

16 bn

26,2 bn

Cost of not having FAIR data, 2018

Costi

CILIEGINO
CATI SICCA

€ 3,60



[DM costing tool](#)

[What's Zingtree?](#)

Data Management Costing Tool

Data Management costing tool



Welcome to the Data Management Costing Tool. This is for TU Delft researchers and staff to help determine costs and staffing requirements in project proposals. Let's start with some questions about your project which will help us estimate the data management needs of your project.



1. GESTIONE DEI DATI



because good research needs good data

... dall'inizio

Research Data Management:
Get it right from the beginning

May 2018



Good RDM = Higher quality, efficiency and value for your research

Add a "version management" tab to your spreadsheet.

Now, let me expand on this idea.

Start by adding an extra "version management" tab to a new spreadsheet. In this sheet, carefully write down a version name (name of the file, typically) in the first column, in the second column the date, and in a third column an explanation of all changes you made to the sheet. Carefully fill out this sheet every single time you move something around, or tinker with the sheet.

If you're a starting PhD student, start doing this the very next time you build a new sheet. Thank me later.

If you already have multiheaded monstrous sheets: start by managing them in this way, and take a few extra hours to redefine the logic behind what you did earlier. Your dissertation writing self will thank you.



Data Management expert guide



Data Management Expert Guide

- 1. Plan >
- 2. Organise & Document >
- 3. Process >
- 4. Store >
- 5. Protect >
- 6. Archive & Publish >
- 7. Discover >



Plan

In this introductory tour, you will become aware of what data management and a data management plan (DMP) are and why they are important. General concepts such as social science data and FAIR data will be explained. Based on our recommendations and good practice examples, you will be able to start writing your DMP.

Organise & Document

If you are looking for good practices in designing an appropriate data file structure, naming, documenting and organising your data files within suitable folder structures, this chapter is for you.

Process

Store

To be able to plan a storage and backup strategy, you will learn about different storage and backup solutions and their advantages and disadvantages. Also, measures to protect your data from unauthorised access with strong passwords and encryption will be explained.

Protect

This chapter highlights your legal and ethical obligations and shows how a combination of gaining consent, anonymising data, gaining clarity over who owns the copyright to your data and controlling access can enable the ethical and legal sharing of data.

Archive & Publish

When you arrive at this chapter you will have learnt to differentiate between currently available data publication services. You will also find a number of stepping stones on how to promote your data.

Discover

How can you discover and reuse existing or previously collected datasets?

Data management ABC – File naming



File naming conventions

The conventions comprise the following 13 rules. Follow the links for examples and explanations of the rules.

1. Keep file names short, but meaningful
2. Avoid unnecessary repetition and redundancy in file names and file paths.
3. Use capital letters to delimit words, not spaces or underscores
4. When including a number in a file name always give it as a two-digit number, i.e. 01-99, unless it is a year or another number with more than two digits.
5. If using a date in the file name always state the date 'back to front', and use four digit years, two digit months and two digit days: YYYYMMDD or YYYYMM or YYYY or YYYY-YYYY.
6. When including a personal name in a file name give the family name first followed by the initials.
7. Avoid using common words such as 'draft' or 'letter' at the start of file names, unless doing so will make it easier to retrieve the record.
8. Order the elements in a file name in the most appropriate way to retrieve the record.
9. The file names of records relating to recurring events should include the date and a description of the event, except where the inclusion of any of either of these elements would be incompatible with rule 2.
10. The file names of correspondence should include the name of the correspondent, an indication of the subject, the date of the correspondence and whether it is incoming or outgoing correspondence, except where the inclusion of any of these elements would be incompatible with rule 2.
11. The file name of an email attachment should include the name of the correspondent, an indication of the subject, the date of the correspondence, 'attach', and an indication of the number of attachments sent with the covering email, except where the inclusion of any of these elements would be incompatible with rule 2.
12. The version number of a record should be indicated in its file name by the inclusion of 'V' followed by the version number and, where applicable, 'Draft'.
13. Avoid using non-alphanumeric characters in file names.

Data management ABC – Versioning

University of Leicester

Good Practice and Guidance – Document Version Control Chart (Draft)

1. Create Document/File

- Save the document according to file naming guidance/good practice.

2. Document Identification

- Identify on the document e.g. in header or footer, the author, filename, page number and date the document is created/revised.

3. Version Control Table

- Versions and changes documented with Version Control Table where significant/formal/project based.

4. Version Number

- Current version number identified on the first page and where appropriate, incorporated into the header or footer of the document.
- Version number is included as part of the file name.

5. First Draft Version

- Named as version "0-1" (no full stops in electronic file names).
- Subsequent draft versions 0-2, 0-3, 0-4 ...

6. First Final/Approved Version

- When document is final/approved it becomes version 1-0.

7. Changes to Final Version

- Changed/revised final version becomes x-1.
- Subsequent drafts to Final version become e.g. 1-1, 1-2, 1-3 etc.

8. Further Final/Approved Documents

- Version number increased by "1-0" e.g. 1-0, 2-0, 3-0 etc.
- e.g. Amendments to Final 1-0 are 1-1, 1-2, 1-3 and as approved becomes 2-0.

Data Management



Home Articles About

About Pensoft

Books

E-Books

Blog

Jo

Project Report

Research Ideas and Outcomes 4: e26439
<https://doi.org/10.3897/rio.4.e26439>

May 9, 2018

Reviewable v1

Support Your Data: A Research Data Management Guide for Researchers

▼ John A Borghi, Stephen Abrams, Daniella Lowenberg, Stephanie Simms, John Chodacki

Abstract ▲

Researchers are faced with rapidly evolving expectations about how they should manage and share their data, code, and other research materials. To help them meet these expectations and generally manage and share their data more effectively, we are developing a suite of tools which we are currently referring to as "Support Your Data". These tools, which include a rubric designed to enable researchers to self-assess their current data management practices and a series of short guides which provide actionable information about how to advance practices as necessary or desired, are intended to be easily customizable to meet the needs of a researchers working in a variety of institutional and disciplinary contexts.

Suppl. material 5: Draft Guide - Preparing [doi](#)

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Getting your data ready for analysis" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

Filename: Draft Guide - Preparing.odt

[Download file](#) (59.52 kb)

Suppl. material 6: Draft Guide - Analyzing [doi](#)

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Analyzing your data and handling the outputs" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

Filename: Draft Guide - Analyzing.odt

[Download file](#) (51.82 kb)

Suppl. material 7: Draft Guide - Sharing [doi](#)

Authors: John Borghi

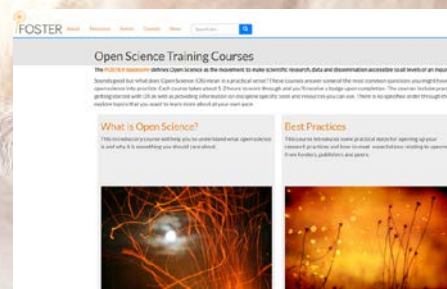
Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Sharing and publishing your data" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I have created plans for managing my data that are designed to stream its future use by myself and others.
Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it not necessarily standardized.	I organize my data so that others can navigate, understand, and use it without me being present.
Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed maximize opportunities for re-use by myself and others.
Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.
Analyzing your data and handling the outputs	I often have to redo my analyses or examine their products to determine what procedures or parameters were applied.	After I finish my analysis, I document the specific parameters, procedures, and protocols applied.	I regularly document the specifics of both my analysis workflow and decision making process while I am analyzing my data.	I have ensured that the specifics of my analysis workflow and decision making process can be understood and put into action by others.

<https://www.cessda.eu/>

Imparare a gestire



A large banner for the FOSTER 'Managing and Sharing Research Data' course. The banner has a background image of colorful, abstract shapes resembling bubbles or cells. The FOSTER logo is in the top left corner, followed by navigation links: About, Resources, Events, Courses, and News. A search bar is located on the right. The main title 'Managing and Sharing Research Data' is centered in a dark box. Below the banner, there is a section titled 'Data-driven research is becoming increasingly common...' followed by a list of learning objectives and a 'Start the Free Course' button.

Managing and Sharing Research Data

Data-driven research is becoming increasingly common in a wide range of academic disciplines, from Archaeology to Zoology, and spanning Arts and Science subject areas alike. To support good research, we need to ensure that researchers have access to good data. Upon completing this course, you will:

- understand which data you can make open and which need to be protected
- know how to go about writing a data management plan
- understand the FAIR principles
- be able to select which data to keep and find an appropriate repository for them
- learn tips on how to get maximum impact from your research data

[Start the Free Course](#)

<https://www.fosteropenscience.eu/node/2328>



Full details

Level of knowledge: Introductory: no previous knowledge is required

Topics



Preparare a proteggere



What are personal data?

Click the plus sign to expand the text box

- + What are personal data?
- + Protecting personal data
- + Legal requirements - EU General Data Protection Regulation (GDPR)
- + Legal requirements - GDPR research exemptions

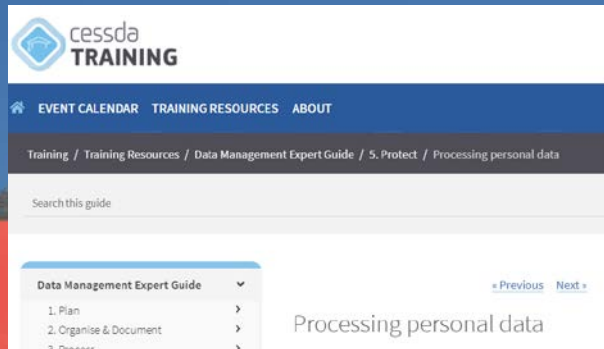
This course covers data protection in particular and ethics more generally. It will help you understand the basic principles of data protection and introduces techniques for implementing data protection in your research processes. Upon completing this course, you will know:

- what personal data are and how you can protect them
- what to consider when developing consent forms
- how to store your data securely
- how to anonymise your data

[Start the Free Course](#)



[dati personali]



I. Process lawfully, fair and transparent

The participant is informed of what will be done with the data and data processing should be done accordingly.

II. Keep to the original purpose

Data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

III. Minimise data size

Personal data that are collected should be adequate, relevant and limited to what is necessary.

IV. Uphold accuracy

Personal data should be accurate and, where necessary kept up to date. Every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay.

V. Remove data which are not used

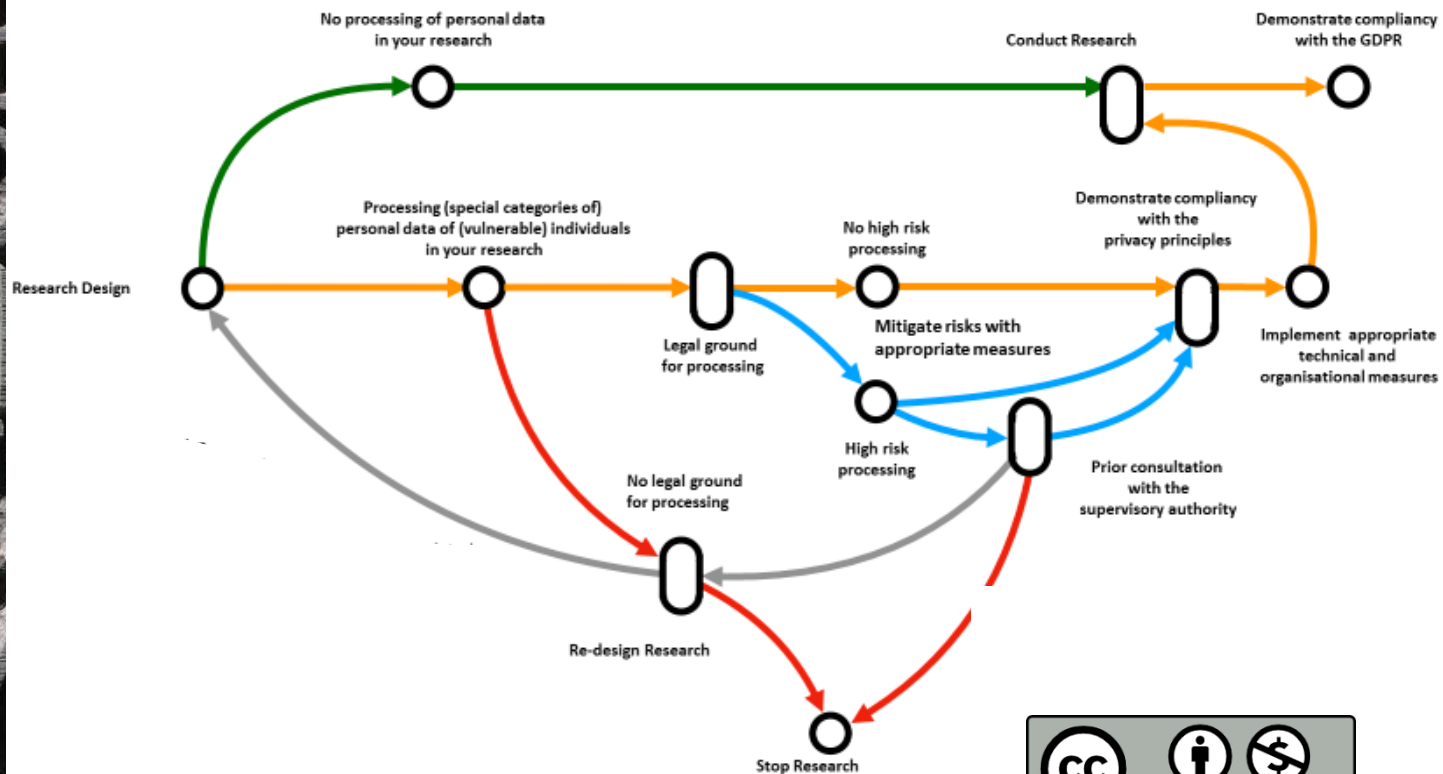
Personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

VI. Ensure data integrity and confidentiality

Personal data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

[Date GDPR]

The Privacy Impact Assessment (PIA) Route Planner for Academic Research Inspired by Harry Beck's London Metro Map



Erasmus University Rotterdam
marlon.domingus@eur.nl
February 2018

The Logic of a Privacy Impact Assessment (PIA) for Academic Research

Q1. Do you process (special categories of) personal data of (vulnerable) individuals in your research?



YES

NO
Proceed - no measures required for safeguarding privacy.

"Personal Data" (GDPR*, Article 4):

Any information relating to an identified or identifiable natural person: a name, an identification number, location data, an online identifier, one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

"Special Categories of Personal Data (Sensitive Data)" (GDPR, Article 9):

Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

Action

Records of processing activities (GDPR*, Article 30):

The university shall maintain a digital record of the processing activities in your research to demonstrate compliancy to the GDPR. This register contains:

1. The name and contact details of the researcher, the research partners and service providers;
2. The purposes of the processing;
3. A description of the categories of data subjects and of the categories of personal data;
4. The categories of recipients to whom the personal data have been or will be disclosed;

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Online available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>

** Article 29 Data Protection Working Party: Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Adopted on 4 April 2017. As last Revised and Adopted on 4 October 2017. Online available at: https://ec.europa.eu/newsroom/document.cfm?doc_id=47711

Q2. What is the legal ground for this processing?

Lawfulness of Processing (GDPR*, Article 6, 89):

1. The individuals participating in your research have freely given their explicit consent for one or more specific purposes.
2. Your research contributes to a legitimate interest, yet results in no high risks for the individuals participating in the research.
3. Your research has a scientific, historical or statistical purpose, yet results in no high risks for the individuals participating in the research.

Action

Data protection by design and by default (GDPR*, Article 25):

Implement appropriate technical and organisational measures:

1. **Individual participating in your research (data subject).** Is the participant well informed, aware of possible risks for her/him and aware of the purpose of the research?
2. **Data.** Is the data de-identified and encrypted?
3. **Access Management.** How is access managed and controlled for the PI / team (expanded) / public?
4. **Software / Platform.** Are the *Terms of Service* for used software / platform checked (where is the data and who has access and has which usage rights)?
5. **Devices.** Are devices used safe? Encrypted drive, encrypted communication, strong password / two factor authentication.
6. **Partners.** Are the research partners / service partners trusted and are appropriate legal agreements made, with regards to roles, rights and responsibilities?
7. **Safe and secure collaboration.** Is the ((cross border communication to, in and from the) collaboration platform end to end encrypted, are roles and permissions defined and implemented, is logging and monitoring implemented?

8. **Risk definition and mitigation.** Are risks defined and mitigated? Is a risk audit procedure started?

YES

NO
Stop research or redefine research.

Q3. Is this processing a high risk processing?

Criteria for high risk processing (WP29 - DPIA Guideline**):

1. Evaluation or scoring
2. Automated-decision making with legal or similar significant effect
3. Systematic monitoring
4. Sensitive data or data of a highly personal nature
5. Data processed on a large scale
6. Matching or combining datasets
7. Data concerning vulnerable data subjects
8. Innovative use or applying new technological or organisational solutions
9. When the processing itself prevents data subjects from exercising a right or using a service or a contract

YES

NO
Proceed - measures required for safe-guarding privacy.

Action

Prior consultation (GDPR*, Article 36):

1. The Data Protection Officer shall, on behalf of the researcher, consult the supervisory authority, prior to the processing (the research) when the processing would result in a high risk *in the absence of measures* to mitigate the risk.

Action

Principles relating to processing of personal data (GDPR*, Article 5):

Demonstrate compliancy with the principles: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality and accountability.

[anonimizzare i dati]

OpenAIRE

AMNESIA
Anonymize your datasets

AMNESIA allows end users to anonymize sensitive data in order to share them with a broad audience. The service allows the user to guide the anonymization process and decide on a flexible trade-off between privacy guaranty and data utility. The service is offered through a web interface that allows users to explore the anonymized data visually. Moreover, the service detects duplicate anonymized files when they are uploaded to Zenodo.

data anonymization research data management

Homepage Service

Usage

TECHNOLOGY READINESS LEVEL
8 - system complete and qualified

LIFECYCLE STATUS Beta

TARGET USERS

Research communities, Research Infrastructures, Universities, Research Centers, Hospitals. Any commercial provider that produces data and wants to share them or outsource them.

Service coverage

Countries serviced by AMNESIA

<http://catalogue.openaire.eu/service/openaire.amnesia>

Contractual Info

[Service level agreement →](#)
[Terms of use →](#)

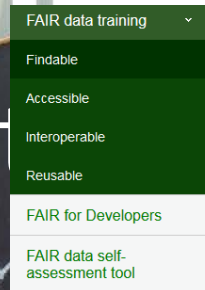
Support

[Helpdesk →](#)
[User manual →](#)
[Feedback →](#)
[Training information →](#)



2. RENDERE I DATI FAIR

FAIR in a nutshell



F1. (meta)data are assigned a globally unique and eternally

There are many resources created by the ARDC on the topic of [metadatas](#)

- [Metadata guide](#)
- [Data versioning](#)

The ARDC has information on persistent identifiers on three different levels

- [Persistent identifiers: awareness level](#)
- [Persistent identifiers: working level](#)
- [Persistent identifiers: expert level](#)

It is also a provider of services for minting persistent identifiers of many different types of data (e.g. of the data being identified):

- [Digital Object Identifier \(DOI\) System for research data](#)
- [Handle minting Service \(Identify My Data\)](#)
- [International Geo Sample Numbers \(IGSN\)](#)

Complementary to the assignment of persistent identifiers is their proper

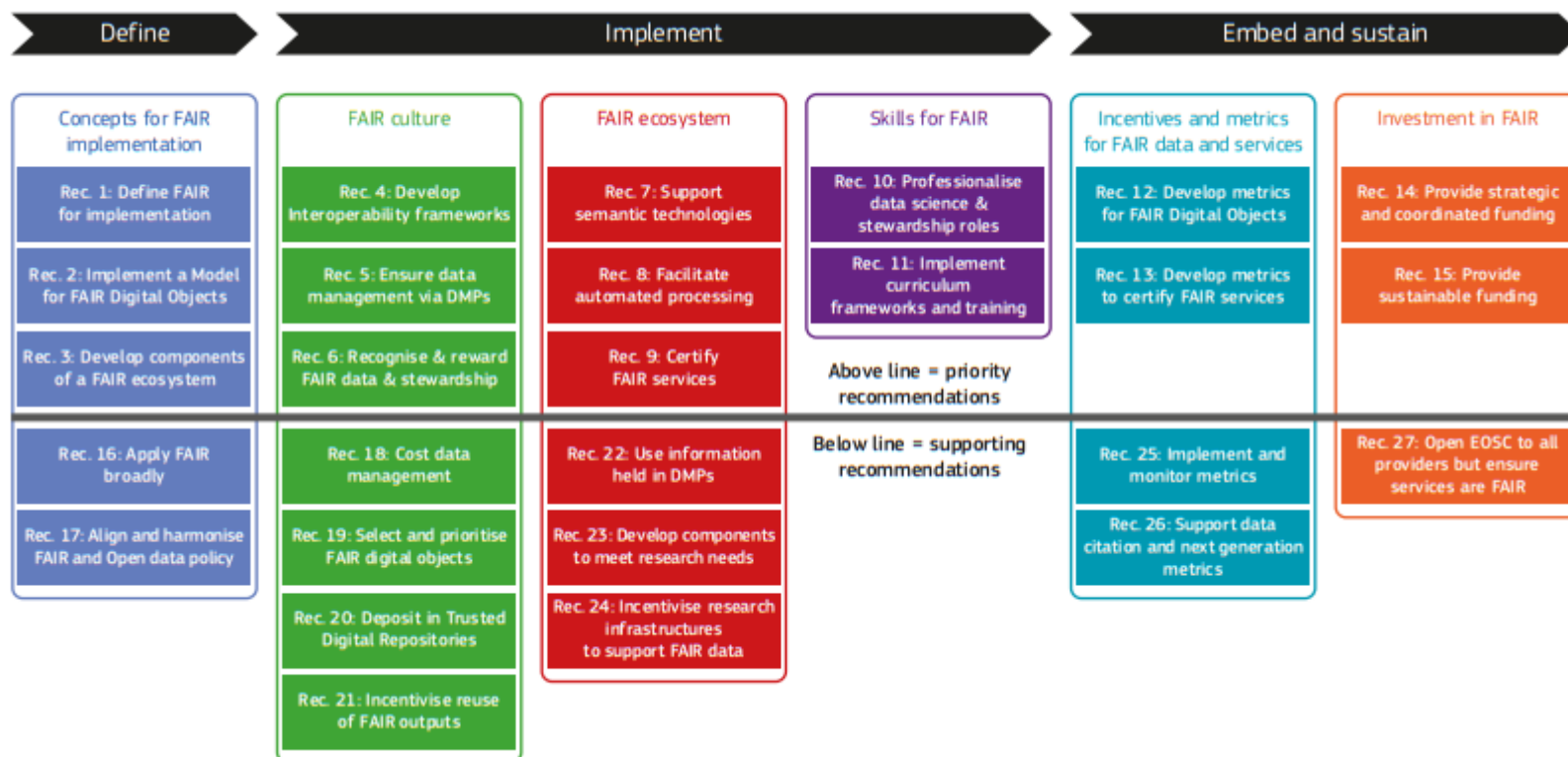




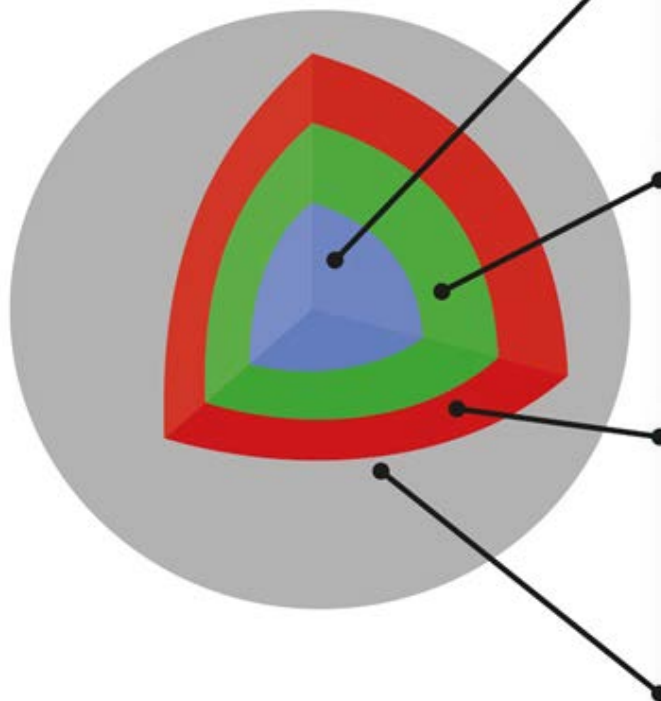
Nov. 20, 2018

Final Report and Action Plan
from the European
Commission Expert Group
on FAIR Data

TURNING
FAIR INTO



FAIR data - the ideal FAIR object



DIGITAL OBJECT

Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and support citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to process and analyse the data.

METADATA

Contextual documentation

In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

FAIR: technology VS domain

FAIR Principles

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Technology



Domain-relevant content

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;

... quanto siete FAIR?

Findable

Does the dataset have any identifiers assigned?

Is the dataset identifier included in all metadata records/files describing the data?

How is the data described with metadata?

What type of repository or registry is the metadata record in?

Accessible

How accessible is the data?

Is the data available online without requiring specialised protocols or tools once access has been approved?

Will the metadata record be available even if the data is no longer available?

ands Training

Search for Research Data Enter Keywords

Search the ANDS Site Enter Keywords

About us News and Events Partners and Communities Working with data Online Services Guides and resources

Working with data

The FAIR data principles

FAIR webinar series (Aug/Sep 2017)

FAIR data training

Findable

Accessible

Interoperable

Reusable

FAIR data training

ANDS | Working with data | The FAIR data principles

FAIR data training

If you intend to run workshops on FAIR data, or include FAIR in training that you are already running here are some ideas and resources.

- A basic checklist (or more comprehensive breakdown) as a tool for discussing the components of FAIR
- Use the FAIR data self-assessment tool in training or consultation
- Discussing the components via a process of transforming a dataset to be more FAIR
- Case studies of domain specific consideration of the principles

<https://www.ands-nectar-rds.org.au/fair-tool>

RDS **ands** **nectar**

home | news | events | programs | about

FAIR self-assessment tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

FAIR maturity evaluator

The FAIR Maturity Evaluation Service

Public Entry Points:

[Browse existing Maturity Evaluations](#)

[Browse MI Test Collections \(Begin New Maturity Evaluation\)](#)

[Register a New MI Test Collection](#)

[Browse Maturity Indicator Tests](#)

[Register a new Maturity Indicator Test](#)

[Search for MI Tests and Test Collections](#)

<https://linkeddata.systems:3000/>

<https://tinyurl.com/MeasuringFAIR>



(Feel free to add comments during/after the talk)

Examining FAIR
"Maturity"

Mark D Wilkinson
CBGP UPM-INIA, Madrid, Spain

presenting the work of:
Mark Wilkinson, Erik Schultes, Susanna Assunta
Sansone, Ricardo de Miranda Azevedo, Luiz
Olavo Bonino da Silva Santos, Philippe
Rocca-Serra, Peter McQuilton, Dominique
Dobson, Mark D Wilkinson



March 21, 2019

- OGGETTIVO
- MACHINE-READABLE ... COME I DATI FAIR

FAIR Data management wizard

Data Stewardship Wizard

Data integration

7

Data interpretation

3

Information and insight

14

ELIXIR Data Stewardship Wizard

Smart Data Management Plans for FAIR Open Science

For serious researchers and data stewards

Is there any pre-existing data?

Are there any data sets available?

Data Stewardship for C

- ☐ No
- ☐ Yes 

Will reference data be created?

Will any of the data that you create be used by others?

Data Stewardship for C

- ☐ No
- ☐ Yes 

Will you be storing samples?

Will you store licenses with the data?

What's up?

Always consider the use of your data beyond the original purpose. One of the issues with re-using other people's data is that they cannot be assumed to be reusable from an ethical or legal standpoint without explicit permission. Assuming that unlicensed data are 'free to use for whatever purpose' is intrinsically wrong, and in the case of for instance Pharmaceutical industry can lead to court cases later on. Therefore, whenever you publish a data set or any other kind of information or digital object, it is important to define a license for reuse. For software many licences exist, and for data, increasingly, standard licenses are available or under development. Please note that a given license is also a defined concept and therefore deserves a Persistent Identifier and a URI pointing to where the license can be studied (machine readable licenses are also under development in some areas). This means that in the metadata, the license under which the data or the workflow can be reused is 'just another PID in the right place'. Users can then specify in their search or workflow container that 'only data with the following licenses should be included'.

For instance, if you include some data in your analysis that cannot be used for commercial purposes, that decision may render your entire results not usable for commercial purposes (at least in the view of some lawyers). This means that not licensing your data at all, even if you don't care who uses them and for what purpose is very counterproductive and will severely undermine the actual reuse of your data by others and in particular by industry. It will also lower the attribution-rate (usually part of the license conditions) and thus the citation and the impact score of your data.

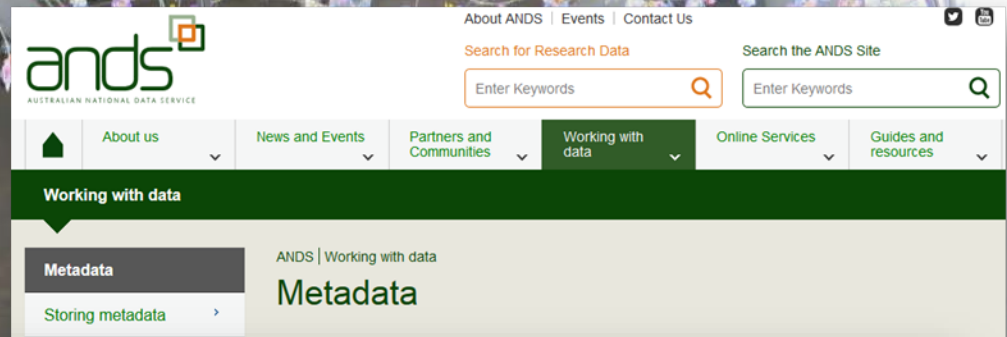
Do

- Always carefully choose a license to be attached to your data upon publication.
- Include and clearly mark the license PID as a concept + attributes in the metadata.
- Store and 'expose' the license as part of the metadata in Open Access environments where search engines can easily find the license, even if the data they describe are not (yet) FAIR or even highly restricted in access. The 'fact' that a data set with a specific license is 'out there' is a first step toward effective reuse of your data or information source.
- Make sure, especially when you restrict use of your data, that you are able to enforce the license you choose. Licenses that are not enforceable make no sense. (please note that the enforcement is usually not done by an individual research group but at institutional or repository level)

Don't

- Ever publish data without a license attached or choose a license lightly, without considerations of anticipated reuse of your data.
- Choose a license that is not transitive (i.e. can not be transferred with subsets of the data), but make sure its transitivity does not unduly restrict the reuse of your data.
- Choose an unnecessary complicated license with many clauses and wherever possible one that is already widely adopted in the research community for either software

F = Findable. Metadata



Types of metadata	Goal	Example
Descriptive metadata	<i>The minimal metadata, required to find a digital object.</i> <i>If there are additional contextual metadata, a user will have a better idea on how to use the data</i>	<i>Author, title, abstract, date</i> <i>Contextual metadata are for example location, time, data collection method (tools)</i>
Structural metadata	<i>These link the individual objects of a unity</i>	<i>Links to related digital objects, (e.g. the article written based on the linked research data)</i>
Technical metadata	<i>Information on the technical aspects of the data set</i>	<i>Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard</i>
Administrative metadata	<i>Metadata focusing on user rights and management of digital objects</i>	<i>License, possible reasons for an embargo, waivers</i>

F = Findable. Metadata standards

Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

Arts and Humanities [Edit](#)

- [Archaeology](#) [Edit](#)
- [Creative art and design](#) [Edit](#)
- [Heritage Studies](#) [Edit](#)
- [Historical and Philosophical Studies](#) [Edit](#)
- [History by Area](#) [Edit](#)
- [History](#) [Edit](#)
- [Law](#) [Edit](#)
- [Music](#) [Edit](#)

Engineering [Edit](#)

- [Architecture](#) [Edit](#)
- [Building Conservation](#) [Edit](#)

Life Sciences [Edit](#)

- [Agricultural Economics](#) [Edit](#)
- [Agricultural Science](#) [Edit](#)
- [Animal pathology](#) [Edit](#)
- [Animal physiology](#) [Edit](#)
- [Biochemistry](#) [Edit](#)
- [Biodiversity](#) [Edit](#)
- [Bioengineering](#) [Edit](#)
- [Biogeography](#) [Edit](#)
- [Bioinformatics](#) [Edit](#)

Physical Sciences & Mathematics [Edit](#)

- [Astronomy](#) [Edit](#)
- [Astrophysics](#) [Edit](#)
- [Cartography](#) [Edit](#)
- [Chemistry](#) [Edit](#)
- [Climatology](#) [Edit](#)
- [Crystallography](#) [Edit](#)
- [Environmental Science](#) [Edit](#)
- [Geology](#) [Edit](#)
- [Geoscience](#) [Edit](#)
- [Glaciology](#) [Edit](#)
- [Hydrogeology](#) [Edit](#)
- [Hydrography](#) [Edit](#)
- [Hydrology](#) [Edit](#)
- [Marine Science](#) [Edit](#)
- [Maritime Geography](#) [Edit](#)
- [Materials Science](#) [Edit](#)
- [Meteorology](#) [Edit](#)
- [Minerology](#) [Edit](#)
- [Nuclear and Particle Physics](#) [Edit](#)
- [Oceanography](#) [Edit](#)
- [Palaeontology](#) [Edit](#)
- [Physics](#) [Edit](#)
- [Planetary science](#) [Edit](#)
- [Remote Sensing](#) [Edit](#)
- [Soil Science](#) [Edit](#)
- [Solar physics](#) [Edit](#)

Social and Behavioral Sciences [Edit](#)

- [Anthropology](#) [Edit](#)
- [Demography](#) [Edit](#)
- [Economics](#) [Edit](#)
- [Geography](#) [Edit](#)
- [Health Policy](#) [Edit](#)
- [Human and Social Geography](#) [Edit](#)
- [Planning \(Urban, Rural and Regional\)](#) [Edit](#)
- [Politics](#) [Edit](#)
- [Sociology](#) [Edit](#)

General Research Data [Edit](#)

- [Multi-disciplinary](#) [Edit](#)

F = findable. Metadata tools

What CEDAR does

<https://metadatacenter.org/>

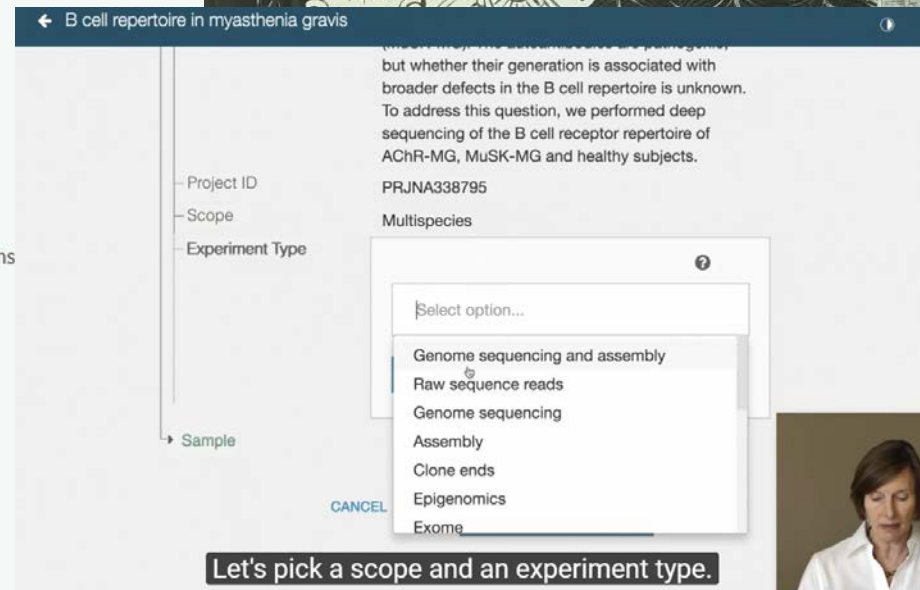
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools will create a metadata record to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools help users collect, store, and download the information that users have provided.

What can CEDAR do for me already?

As of its production release, in February 2017, CEDAR addresses these scenarios:

- create user-friendly, shareable forms for collecting metadata, with features like
 - nested and repeatable elements and fields
 - reusable elements
 - control over tool tips, field titles, and field descriptions
- share your forms and metadata
 - provide a link to your metadata editors, so they can enter metadata responses based on your forms
 - share your forms and other content with individuals or a group
 - create and manage groups to make permissions simpler
- associate your questions (fields) and possible answers (values) with controlled terms
 - select any term or collection of terms from the NCBO BioPortal semantic repository
 - combine different terms from different controlled vocabularies into a single set of options
 - create your own terms, or term lists ('value sets') that can be re-used
- view responses meeting your (simple) search criteria, in several forms
 - CEDAR Metadata Editor's metadata view
 - an in-line JSON-LD format, used by CEDAR for all its metadata instances
 - download of JSON-LD files via the [CEDAR REST API](#), for offline integration with your workflow
- use the Workbench Desktop interface to manage your content
 - use My Workspace to see your items, or Shared with Me to see other items you can access
 - select an item and control-click or use the 3-dot menu in the upper right to share it, copy it, delete it, or get info on it
- enable intelligent metadata suggestions in your template by using a field's Suggestions tab
 - CEDAR keeps track of metadata entered for that field
 - users will see a drop down list of the most popular metadata entries, and can select from them
- remotely access CEDAR content and capabilities using the [CEDAR REST API](#)

With these capabilities, you can capture simple or rich metadata for your project, build a repository of project metadata, or design particular needs. Advanced users can even submit metadata entries through CEDAR's REST API.



Let's pick a scope and an experiment type.



F = Findable. Persistent identifiers

The image is a collage of web pages related to digital identifiers. At the top, a banner for DataCite says 'WELCOME TO DATACITE' and 'the leading global provider of DOIs for research data.' Below this, the ORCID website is shown, featuring a navigation bar with 'FOR RESEARCHERS', 'FOR ORGANIZATIONS', 'ABOUT', 'HELP', and 'SIGN IN'. A search bar is at the top left. A green arrow points down to a section titled 'DISTINGUISH YOURSELF IN THREE EASY STEPS'. The steps are: 1. REGISTER (Get your unique ORCID identifier), 2. ADD YOUR INFO (Enhance your ORCID record), and 3. FIND YOUR DATA (Find datasets, software, images, and other research material). A link <https://orcid.org/> is highlighted. To the right, a quote bubble says 'Cite your research source with confidence, and receive praise when your work is re-used with DataCite!'. At the bottom, there are logos for 're3data.org' (Find an appropriate repository to access and deposit research data with re3data.org) and a citation tool (Generate your references automatically with our easy-to-use citation formatting tool). A link <https://www.datacite.org/> is at the bottom right.

USATE IL DOI

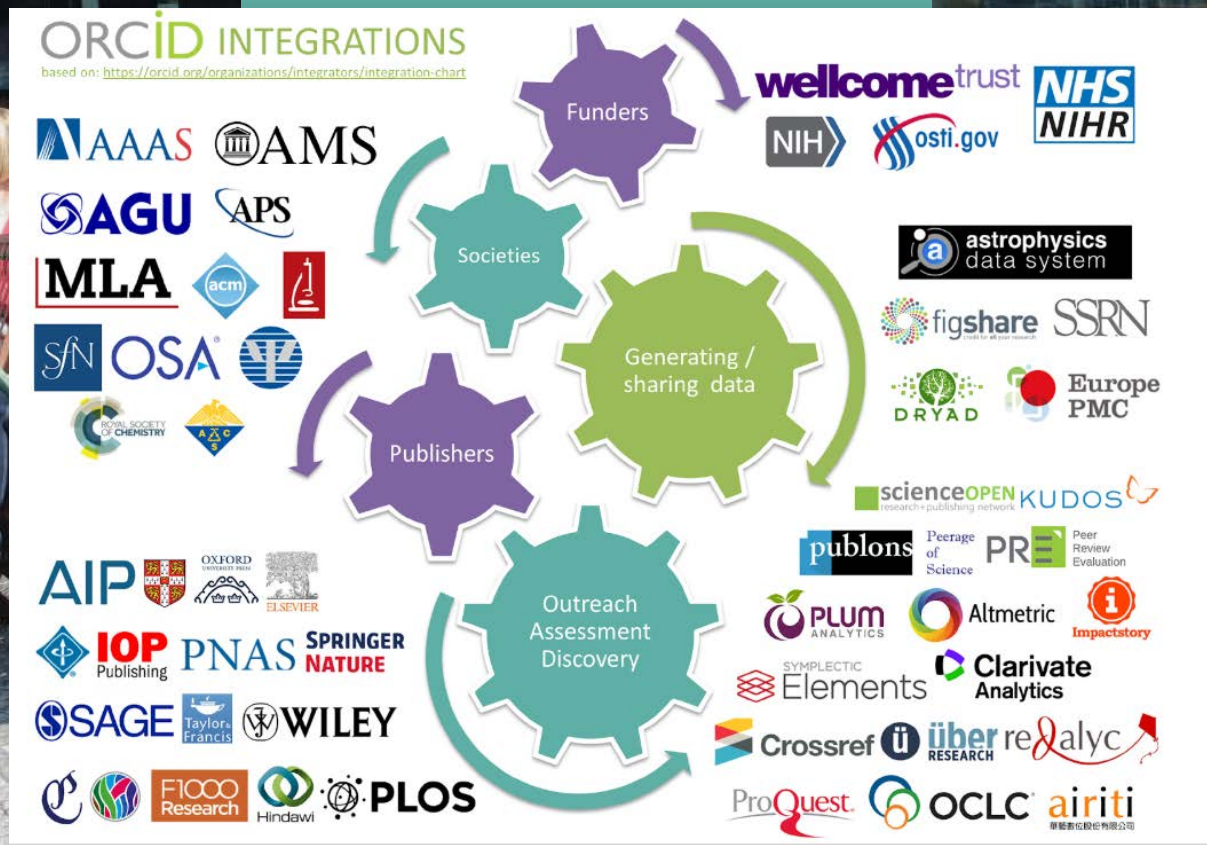
USATE ORCID

<https://orcid.org/>

<https://www.datacite.org/>

ORCID – ID, ognuno è unico

CREARE E MANTENERE UN
ORCID ID FA GIRARE TUTTE
QUESTE RUOTE



... [se non potete usare il DOI]



The Future of Research Communications and e-Scholarship

[ABOUT](#) [COMMUNITY](#) [GROUPS](#) [RESOURCES](#) [NEWS + BLOGS](#) [EVENTS](#) [PUBLICATIONS](#)

FORCE11 » [Community News](#) » [Introducing a new standard for the citation of research data](#)

INTRODUCING A NEW STANDARD FOR THE CITATION OF RESEARCH DATA

Posted by [Jennifer McLennan](#) | May 8, 2018 | [Sign In](#) or [Join Now!](#) to post comments

Rules, registry and recommendations

Compact Identifiers. A “compact identifier” is a string constructed by concatenating a namespace prefix, a separating colon, and a locally unique identifier (LUI), e.g. `pdb:2gc4`.

Provider Specification. To specify a specific provider, where multiple providers exist, prepend the provider code and a “/” to the compact identifier, e.g. `rcsb/pdb:2gc4`.

Provider Default. Where multiple providers exist, and the provider is not specified in the compact identifier, the resolver will determine where to resolve the request based on its own rules, e.g., taking into account uptime availability, regional preference, or other criteria.

Redirect Rule. A URL template associated with the provider code is maintained in the namespace registry, defining how to forward compact identifiers to any specific provider (see 4.2.3 below).

SCIENTIFIC DATA

May 8, 2018

Altmetric: 20

[More detail](#) »

Article | [OPEN](#)

Uniform resolution of compact identifiers for biomedical data

Sarala M. Wimalaratne, Nick Juty, John Kunze, Greg Janée, Julie A. McMurtry, Niall Beard, Rafael Jimenez, Jeffrey S. Grethe, Henning Hermjakob, Maryann E. Martone & Tim Clark

A reasonable solution to the identifiers problem is to assign Digital Object Identifiers (DOIs) to identify datasets. DOIs are already widely used in the publishing world as persistent identifiers for scholarly publications. They have been adopted by generalist data repositories such as Dryad, FigShare, Zenodo and Dataverse, as well as by domain data repositories outside of biomedicine. Handles¹⁴, which underlie the DOI system, may also be used directly. The DataCite consortium provides a robust central means for assigning DOIs to data.

However, DOIs are not commonly used for biomedical data, which is partitioned across over 600 autonomous repositories that are independently funded. Instead, in biomedicine there has been a

A = Accessible

ACCESSIBLE≠OPEN
«ACCESSO» SIGNIFICA ANCHE
RISTRETTO O SOTTO EMBARGO

- **Open access**

Data that can be accessed by any user whether they are registered or not. Data in this category shouldn't contain personal information (unless consent is given (see '[Informed consent](#)').

- **Access for registered users (safeguarded)**

Data that is accessible only to users who have registered with the archive. This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

- **Restricted access**

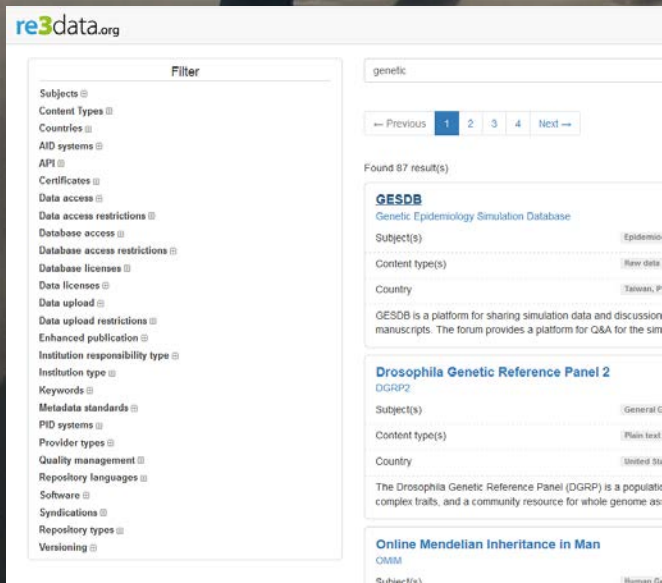
Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.

Restricted access requires long-term commitment of the researcher or person responsible for the data to handle the upcoming the permission requests.

- **Embargo**

Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after a certain period of time.

A = Accessible. Cercate un data repository?



2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management

By offering detailed information on more than 2,000 research data repositories, re3data has become the most comprehensive source of reference for research data infrastructures globally. Through the development and advocacy of a framework for discipline...

[Read more](#)

Three new DOI Fabrica features to simplify account management

Last month we launched DOI Fabrica, the modernized version of the DataCite Metadata Store (MDS) web frontend. It is the one place for DataCite providers and their clients to create, find, connect and track every single DOI from their organization...

[Read more](#)

One step closer towards instant DOI search results

Art Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to created it, did not exist. And the answer to what this illustrates – this is simply a distorted...

[Read more](#)

<https://www.re3data.org/>

A = Accessible. Data journals

Data Journals

Hier entsteht eine Liste von Data Journals, die vorwiegend Data Papers

- Atomic Data and Nuclear Data Tables [\(Elsevier\)](#)
- Biodiversity Data Journal [\(Pensoft Publishers\)](#)
- Biomedical Data Journal [\(Procon Ltd.\)](#)
- BMC Research Notes [\(Biomed Central\)](#)
- Chemical Data Collections [\(Elsevier\)](#)
- Data [\(MDPI\)](#)
- Data in Brief [\(Elsevier\)](#)
- Dataset Papers in Science [\(Hindawi Publishing Corporation\)](#)
- Earth System Science Data - ESSD [\(Copernicus Publications\)](#)
- Ecological Archives [\(Ecological Society of America - ESA\)](#)
- European Data Watch [\(European Data Watch\)](#)
- F1000Research [\(F1000 Research\)](#)
- Genomics Data [\(Elsevier\)](#)
- Geoscience Data Journal [\(Wiley\)](#)
- GigaScience [\(BioMed Central\)](#)
- Internet Archaeology [\(Internet Archaeology\)](#)
- Journal of Open Psychology Data (JOPD) [\(Ubiquity Press\)](#)
- Journal of Chemical & Engineering Data [\(ACS Publications\)](#)
- Journal of Physical and Chemical Data [\(AIP Publishing\)](#)
- Nuclear Data Sheets [\(Elsevier\)](#)
- Open Archaeology Data [\(Ubiquity Press\)](#)
- Open Data Journal for Agricultural Research [\(diverse\)](#)
- Open Health Data [\(Ubiquity Press\)](#)
- Open Journal of Bioresources [\(Ubiquity Press\)](#)
- Open Network Biology [\(BioMed Central\)](#)
- Research Data Journal for the Humanities and Social Sciences [\(Brill\)](#)
- Scientific Data [\(Nature Publishing Group\)](#)

Dataset Description

Object Name

- *walkers* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for records made by individual walkers during stage-one fieldwalking.
- *counts* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for potsherds counted during stage-one fieldwalking.
- *pottery* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main pottery database, assembled various artefact specialists.
- *petrography* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for those sherds sampled for thin section petrography.
- *lithics* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main lithics database.
- *other* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all non-ceramic and non-lithic finds.
- *structs* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all standing remains, except for terraces.
- *coast* – a vector polygon dataset (.shp and associated files) with the shape of Antikythera's coastline.
- *geology* – a vector polygon dataset (.shp and associated files) with the main bedrock units on Antikythera.
- *tracts* – a vector polygon dataset (.shp and associated files) with the main stage-one survey units.
- *grids* – a vector polygon dataset (.shp and associated files) with the main stage-two survey units.
- *terraces* – vector line dataset (.shp and associated files) with all observable agricultural terraces (i.e. the location

I don't need

- *other* – primarily Andrew Bevan (UCL), with further assistance from James Conolly (Trent)
- *geology* – a combination of fieldwork by Ruth Siddall (UCL) and remote sensing by Andrew Bevan (UCL)

Repository Location

UK Archaeology Data Service Collection 1115 (doi: 10.5284/1012484)

Publication Date

05/02/2012

Language

English (a Greek language summary of the project methods and results can be found at www.ucl.ac.uk/asp/ or www.tuarc.trentu.ca/asp/).

License

Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (<http://dx.doi.org/10.5284/1001000> to the fact that the ASP data locations, dates and structures and terraces

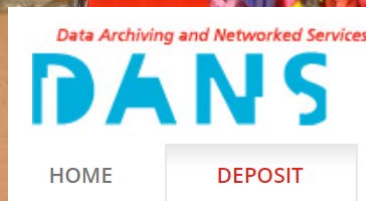
Data journals

Panayioti Polydoratou

Alexander Technological Educational Institute of Thessaloniki

European Commission Workshop
Alternative Open Access Publishing Models: Exploring New Territories in
Communication
Brussels, 12 October 2015

A = Accessible. Formati



Type	Preferred format(s)	Non-preferred format(s)
Text documents	PDF/A (.pdf)	- ODT (.odt) - MS Word (.doc, .docx) - RTF (.rtf) - PDF (.pdf)
Plain text	Unicode text (.txt)	Non-Unicode text (.txt)
Markup language	- XML (.xml) - HTML (.html) - Related files: .css, .xslt, .js, .es	SGML (.sgml)
Spreadsheets	- ODS (.ods) - CSV (.csv)	- MS Excel (.xls, .xlsx) - PDF/A (.pdf) - OOXML (.docx, .docm)
Databases	- SQL (.sql) - SIARD (.siard) - DB tables (.csv)	- MS Access (.mdb, .accdb) (v. 2000 or later) - dBase (.dbf) - HDF5 (.hdf5, .he5, .h5)
Statistical data	- SPSS Portable (.por) - SPSS (.sav) - STATA (.dta) - DDI (.xml) - data (.csv) + setup (.txt)	- SAS (.7dat; .sd2; .tpt) - R (* under examination)
Raster images	- JPEG (.jpg, .jpeg) - TIFF (.tif, .tiff) - PNG (.png) - JPEG 2000 (.jp2)	DICOM (.dcm) (by mutual agreement)

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tiff) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary file package

Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)

UK Data Service			
	About us	Get data	Use data
Home > Manage data > Format your data > Recommended formats			
Recommended formats			

Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHTML 1.0

A = Accessible. Conservazione

LONG TERM OR SHORT TERM?

Checksum Checker

Software for Digital Preservation

Download version 3.0.1, released 25 March 2014 AEST

Checksum Checker is free and open source software developed by the National Archives of Australia. Checksum Checker is a piece of software that is used to monitor the contents of a digital archive for data loss or corruption.

Checksum Checker is a component of the Digital Preservation Software Platform (DPSP).

Features

As part of the Digital Preservation Recorder (DPR) workflow, checksums are generated for each Archival Information Package (AIP). Checksum Checker generates a new checksum for each AIP and compares it against the stored checksum. If the checksums do not match, then the AIP is flagged as being corrupt.

Checksum Checker incorporates the following features:

- Checksum Checker functions as a service.
- Checksum Checker sends automated emails to a nominated administrator email address, coinciding with certain events (such as the start of a checking run or when an error is encountered).

Checksum Checker is released under the GPLv3, and is available for download. <http://checksumchecker.sourceforge.net/>

Search

 Home

 Download

 Docs

 F.A.Q

 Licensing

 External Links

 Contact Us

Storage Solutions	Advantages	Disadvantages	Suitable for
Personal Computer & Laptop	<i>Always available</i> <i>Portable</i>	<i>Drive may fail</i> <i>Laptop may be stolen</i>	<i>Temporary storage</i>
Networked drives File servers managed by your university, research group or facilities like a NAS-server	<i>Regularly backed up</i> <i>Stored securely in a single place</i>	<i>Costs</i>	<i>Master copy of your data</i> <i>(if enough storage space is provided ..)</i>
External storage devices USB flash drive, DVD/CD, external hard drive	<i>Low cost</i> <i>Portability</i>	<i>Easily damaged or lost</i>	<i>Temporary storage</i>
Cloud services	<i>Automatic synchronization between folders and files</i> <i>Easy to access and use</i>	<i>It's not sure whether data security is taken care of</i> <i>You don't have direct influence on how often backups take place and by whom</i>	<i>Data sharing</i>

1

2

3

4

5

6

Organize and document research data. Make digital versions of paper data documentation in a PDF/A format (suitable for long-term storage).

A = Accessible. Conservazione

Portable devices Cloud storage Local storage Networked drives



Laptops, tablets, external hard-drives, flash drives and Compact Discs

Advantages

- Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field.
- Low-cost solution.

Disadvantages/Risks

- Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk.
- Not robust for long-term storage or master copies of your data and files.
- Possible quality control issues due to version confusion.

Precautions for (sensitive) personal data

Use in cloud storage
encrypt data
password protect

Advantages

- Automatic backups.
- Often automatic version control.

Disadvantages/Risks

- Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens.
- Insufficient control over where the data is stored and how often it is backed up.
- Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes.
- Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.

Precautions for (sensitive) personal data

- Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see '[Security](#)' for more information on encryption; also see '[Protecting data](#)').

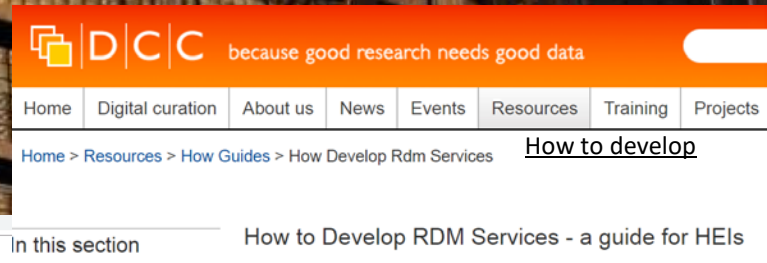
Recommendations

- Do: use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.
- Do: Read the terms of service. Especially focus on rights to use content given to the service provider.
- Do: Opt for European, national, or institutional cloud services which store data in Europe if possible.
 - [B2drop](#) (EUdat, n.d.) is an example of a European cloud storage solution.
 - [SWITCHdrive](#) (SWITCH, 2017) is a Swiss solution.
 - [DataVerseNL](#) (Data Archiving and Networked Services, 2017) is an example of a service for Dutch researchers that allows the storage and sharing of data both during and after the research period.
- Don't: make this your only storage and backup solution.
- Don't: use for unencrypted (sensitive) personal data.

CESSDA Guide

CI SONO ESIGENZE (E STRUMENTI) DIVERSI
DIPENDE DALLA FASE.

Cosa conservare?



The screenshot shows the top of the DCC (Digital Curation Centre) website. The header is orange with the DCC logo and the tagline "because good research needs good data". Below the header is a navigation menu with links: Home, Digital curation, About us, News, Events, Resources, Training, and Projects. A secondary navigation bar shows the breadcrumb "Home > Resources > How Guides > How Develop Rdm Services" and a link to "How to develop".

Establishing criteria for selection decisions

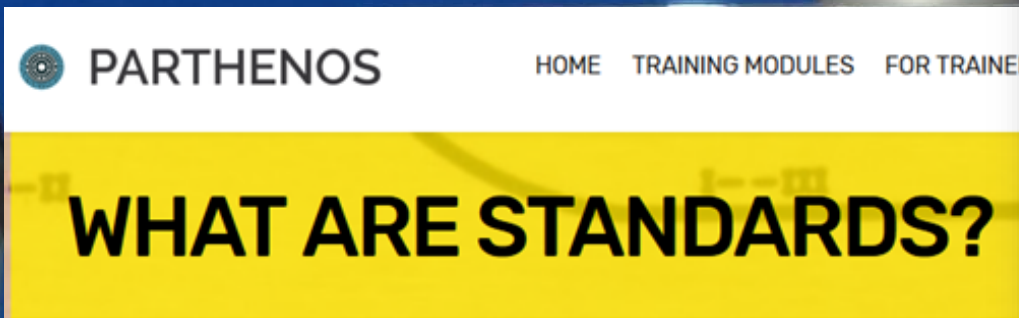
In this section

How to Develop RDM Services - a guide for HEIs

You should establish criteria to guide selection decisions. The DCC's How to Select and Appraise Research Data for Curation[56] proposes seven criteria as outlined below:

1. **Relevance to mission:** the resource content fits any priorities stated in the institution's mission, or funding body policy including any legal requirement to retain the data beyond its immediate use.
2. **Scientific or historical value:** is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
3. **Uniqueness:** the extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
4. **Potential for redistribution:** the reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
5. **Non-replicability:** it would not be feasible to replicate the data/resource or doing so would not be financially viable.
6. **Economic case:** costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
7. **Full documentation:** the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource's provenance and the context of its creation

I = Interoperable. Standards



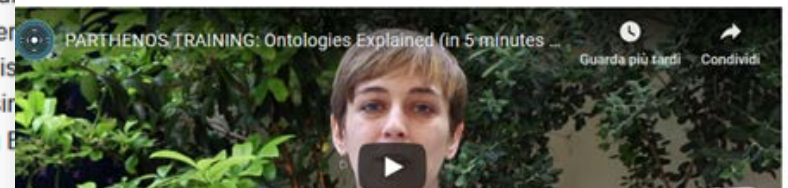
Even perfect metadata may not allow data to become interoperable if a different standard is used. A "standard" refers to a system that structures what types of information are captured about an item in a collection. In our .mp3 library system, a standard is expressed in the header categories such as 'name,' 'time,' 'artist,' and 'album' are listed, with every entry having this filled in. Standards are used to ensure that metadata is as useful as possible for organising a collection, ensuring that common questions (how many songs are there on the album "Big Bang") can be easily and accurately answered.

How Many Standards Are There and Who Decides Which One To Use?

Different standards have arisen in different kinds of cultural heritage institution: the most common standards in museums are different from those in archives, and those common in libraries are different again.



In addition to metadata and standardised metadata schemas, research infrastructures can also use other forms of "knowledge representation system" to enhance the researcher's experience of the interoperable data they present. When we talk about 'Knowledge Representation Systems' in research infrastructures, we usually mean a specific category of hierarchical systems of terms known more commonly as an 'ontology'. Before the digital age, philosophers referred to an ontology as "the study of the kinds of things that exist". Ontologies are similar to taxonomies, another knowledge organisation framework you probably remember from early lessons in biology.



What is Metadata?

What are Standards?

What Are Knowledge Representation Systems and 'Ontologies'?

Sustainability

Methods and Tools

Networks

I = Interoperable

To speed up discovery and uncover new insights, research data should be easily combined with other datasets by humans as well as computer systems.

INTEROPERABLE

10 Establish well documented machine-actionable APIs

Well documented and machine-actionable APIs - a set of subroutine definitions, protocols, and tools for building application software - allow for automatic indexing, retrieval and combining of (meta)data from different data repositories.



Document APIs well and make it possible to deliver the schema of the (meta)data model. Consider showing examples of how to successfully mine data from different endpoints and combine them into new data sets usable for new research.

11 Use open well-defined vocabularies

The description of metadata elements should follow community guidelines that use open, well defined and well known vocabularies. Such vocabularies describe the exact meaning of the concepts and qualities that the data represent.



Use vocabularies relevant to your field, and enrich and structure your research output accordingly from the start of your research project.



Give examples of vocabularies the research community may use, based on research domain specifics.

12 Document metadata models

Clearly documenting metadata models helps developers to compare and make mappings between metadata.



Publish the metadata models in use in your research infrastructure. Document technical specifications and define classes (groups of things that have common properties) and properties (elements that express the attributes of a metadata section as well as the relationships between different parts of the metadata). For metadata mapping purposes, list the mandatory and recommended properties.

13 Prescribe and use interoperable data standards

Using a data standard backed up by a strong community, increases the possibility to share, reuse and combine data collections.



Check with the repository where you want to deposit your data what data standards they use. Structure your data collection in this format from the start of your research project.



Clearly specify which data standard your institution uses, pool a community around them and maintain them especially with a perspective on interoperability. Good examples are CMDI (language studies) and the SIKB0102 Standard (archaeology).

14 Establish processes to enhance data quality

To boost (meta)data quality and, therefore, interoperability, establish (automatic) processes that clean up, derive and enrich (meta)data.



Establish procedures to minimise the risk of mistakes in collecting data. E.g. choose a date from a calendar instead of filling it in by hand.



Invest in tools to help clean up (meta)data and to convert data into standardised and interoperable data formats. Combine efforts to develop workflows and software solutions for such automatic processes, e.g. by using machine learning tools.

15 Prescribe and use future-proof file formats

All data files held in a data repository should be in an open, international, standardised file format to ensure long-term interoperability in terms of usability, accessibility and sustainability.

GUIDELINES
to FAIRify data
management
and make data
reusable

PARTHENOS

2019

[FAIRsharing / Interoperability]

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards

Databases

Policies

Collections

Add/Claim Content

Stats

Log In or Register

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

HOW CAN WE HELP?

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards

Databases

Policies

Collections

Add/Claim Content

Stats

Log In or Register

View as Table

View as Grid

Sort by

Name

Recommended Records

Recommended

Associated Publication?

No Publication

Has Publication

Claimed?

No Maintainer

Has Maintainer

Record Status

Uncertain

Deprecated

In development

Ready

Standard Type

Terminology Artifact 726

Model/Format 267

Reporting Guideline 154

Metric 14

Identifier Schema 11

Show More

Domains

Chemical Entity 128

Report 109

Natural Sciences

1120

Life Sciences

1087

Biology

454

Showing records 1 - 50 of 1294.

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	Is Collection/Recommendation	Status
ABA Adult Mouse Brain	ABA	Standard	None		None	None	None	None	None	None	R
Access to Biological Collection Data	ABCD	Standard	Biobanking Life Sciences		None	None	GBIF Atlas of Living Australia IPT - GBIF Australia Repository GBIF Spain IPT - GBIF Spain Repository Canada's IPT - GBIF Canada's Repository SIB Colombia IPT - GBIF Colombia Repository Plus 1 more...	ABCD EFG ABCDRIA	None	World Biobanking Information Network	R
Access to Biological Collection Data Extension for Geoscientists	ABCD EFG	Standard	Biobanking Life Sciences		None	None	GeoCase Data Portal	XML ABCD	None	None	R
Access to Biological Collection Data DNA extension	ABCDRIA	Standard	Biobanking Life Sciences		None	None	GenBank	ABCD	None	World Biobanking Information Network	Dev
ACE format	ACE format	Standard	Life Sciences		None	None	None	None	None	None	R
Adalab meta ontology	ADALAB.META	Standard	None		None	None	None	None	None	None	R
Adalab ontology	ADALAB	Standard	None		None	None	None	None	None	None	R
Achene Drug Reaction Markup Language	EU-ADR ML	Standard	None		None	None	Thesaurus Health Terminology Achene Markup Language Chemical Drug	None	None	None	U

<https://fairsharing.org>

I= Interoperable – a

FAIRDOM Platform

SEEK

Research datasets, models or simulations, processes and information about the people and organisations involved. The platform is based on the ISA-Tools format. When paired with our workflow, it simplifies upload and download of files.

data sharing within groups and consortia. In addition,

<https://fair-dom.org/platform/>



Rightfield is an open-source tool for adding ontology term selection to Excel spreadsheets. Rightfield is used by a 'Template Creator' to create semantically aware Excel spreadsheet templates. The Excel templates are then reused by Scientists to collect and annotate their data; without any need to understand, or even be aware of, Rightfield or the ontologies used. Rightfield embedded templates are used within the **Samples** framework of the **SEEK**.

MORE INFORMATION

<https://fair-dom.org/platform/rightfield/>

Metadata	Values (examples)		Notes
Asset Title			The name of the data file
Uploader			The person submitting the asset to SEEK
Uploader SEEK ID			If you add your own SEEK ID, this will help us link this asset with your profile
Project	Project		The project that the asset belongs to
ASSAY			
Assay SEEK ID			If referring to an existing Assay, you can link to it via the Assay SEEK ID.
Assay Title			The title of an existing assay
Assay_type	ExperimentalAssayType		The assay_type describes the type of experiment you are performing
Technology_type	amplification by-product_formation catabolic_response cell_growth_optimisation cell_size Comparative_genomic_hybridization comparative_genomics continuousEnzymatic		Describes the type of instruments and/or equipment used for the experiment
Description			A brief, human readable description.
Experimentalist			The names of the people who carried out the experiments. These can either be SEEK members or external scientists
Date			The start date for the experiment if different from the upload date
SOP			Links to SOPs and protocols used to carry out the experiment. If they are already in SEEK, you can refer to them by their SEEK ID
Publication (optional)			If this data appears in a publication, you can link it directly, or via the assay or study. If it is already registered in SEEK, you can use the PubMed ID or DOI as a reference.
Experimental_conditions			
Item	ExperimentalConditions	ExperimentalConditions	The name of the experimental condition you are fixing in your experiment (e.g. temperature, concentration, pH etc). If there is more than 1, please list them in columns across the spreadsheet
Compound (if concentration)			The compound name is only required if the item is concentration.
Unit			The SI units of the experimental conditions measurements.
Start_value (optional)			This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
End_value (optional)			This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
Comments			Additional information that would be useful for people reading this data file

R = Riusable. Documentazione



research
data
netherlands

Essentials 4 Data Support

Data documentation is describing the characteristics of a dataset, occurring at various levels, such as:

- A description of the **process** a researcher uses to collect data. Documentation takes place in, for instance a codebook, lab journal, log or diary.
- A description of the **data itself** (how much, what data format, what software to use to read the data).
- A description of the **changes of the dataset in time**. This is used to create a historical report of all uses and edits of the research data over a period of time. In data jargon this is called **data provenance**. In order to make a historical report, a description of the data collection process and of the data itself is also essential.

Proper data documentation ensures that research data are traceable and unambiguously understood and used by current and future users (including the researcher).

Due to the great diversity of datasets, the choices for documenting the data are not always obvious.

<http://datasupport.researchdata.nl/en/start-the-course/iii-the-research-phase/data-documentation/>

PER ESSERE RIUSABILI I DATI DEVONO AVERE

1. DOCUMENTAZIONE

2. LICENZE

R = Reusable. Documentazione



Project-level documentation



Project-level documentation explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc. In the accordion the questions which your project-level documentation should answer are stated in more detail:

detail:

- ⊕ 1. For what purpose was data created
- ⊕ 2. What does the dataset contain
- ⊕ 3. How was data collected
- ⊕ 4. Who collected the data and when
- ⊕ 5. How was the data processed
- ⊕ 6. What possible manipulations were done to the data
- ⊕ 7. What were the quality assurance procedures
- ⊕ 8. How can data be accessed

Data-level documentation

Data-level or object-level documentation provides information at the level of individual objects such as pictures or interview transcripts or variables in a database. You can embed data-level information in data files. For example, in interviews, it is best to write down the contextual and descriptive information about each interview at the beginning of each file. And for quantitative data variable and value names can be embedded within the data file itself.



⊖ Quantitative data

Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.



Data-level documentation for quantitative data

For quantitative data document the following:

- **Information about the data file**
Data type, file type and format, size, data processing scripts.
- **Information about the variables in the file**
The names, labels and descriptions of variables, their values, a description of derived variables, if applicable, for variables that are not in the data file.

R = Reusable. Documentation

≡ ✓ protocols.io

Make your science more reproducible
protocols.io is the #1 open access repository for science methods

Editing Fixation of yeast cells for RNA-FISH

DESCRIPTION
GUIDELINES & WARNINGS
MATERIALS
STEPS

INOCULATION AND GROWTH

1. Around 10am, start a cell culture in a 50ml tube...
2. Grow for 8-10 hours in a shaker at 30 °C.
3. Measure OD in the evening and dilute into 250ml...

4. Transfer to 50ml falcon tubes.
5. Add 5ml of Formaldehyde, invert a few times, set...

4 Fixation
Transfer to 50ml falcon tubes.

new step paste from text paste from buffer

5 NEW SECTION (optional)

Fixation

Add 5ml of Formaldehyde, invert a few times, set...

Optional: Transfer to gentle rocking over recommended per Anne Dodson, Marc...

Replay

Delete

ROBERT BLANK DELETE

RStudio

<https://www.rstudio.com/>

RStudio

Open source and enterprise-ready professional software for R

What is an Open Notebook?

Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.

Notebook web application

The notebook web application enables users to:

- Edit code in the browser, with automatic syntax highlighting, indentation, and tab completion/introspection.
- Run code from the browser, with the results of computations attached to the code which generated them.
- See the results of computations with rich media representations, such as HTML, LaTeX, SVG, PDF, etc.
- Create and use interactive JavaScript widgets, which bind interactive user interface controls and visualizations to reactive kernel side computations.



<http://jupyter.org/index.html>

The Jupyter Notebook

The Jupyter Notebook is an open-source web application that lets you create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data transformation, numerical simulation, statistical modeling, machine learning, and much more.

Try it in your browser Install the Notebook



...WHY NOT?

- PROTOCOLS.IO
- OPEN LAB NOTEBOOK

PER TENERE TRACCIA DI TUTTO L'ESPERIMENTO
[DIFFICILE ALL'INIZIO, POI TUTTO FATTO...]

R= Reusable. Licenze

Copyright: protects the STRUCTURE, selection or arrangement of their contents" (Art. 3) NOT THE DATA

Sui generis database right: protects the «substantial effort» in OBTAINING data [NOT «CREATING»]... the right owner often is the institution

Database=a collection of independent works, data or other materials arranged in a systematic or methodical way (Art.1)

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 11 March 1996

on the legal protection of databases

COUNCIL OF THE EUROPEAN UNION,
European Community, and in particular Article 37 (2), 66 and 100a thereof,

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?

RICORDA: NESSUN
COPYRIGHT SUI DATI
(NON CREATIVI)

semplici dati e
informazioni

nessuna tutela

database
non creativo

solo diritto
sui generis

database
creativo

diritto sui generis
+ diritto d'autore

livello diritto
d'autore

livello diritto
sui generis

[webinar]



The screenshot shows the OpenAIRE website with a navigation bar at the top containing 'SERVICES', 'SUPPORT', 'OPEN SCIENCE IN EUROPE', and 'ABOUT'. The main content area has a header with the OpenAIRE logo and a link to 'More Information about the 2019 webinar series.' Below this is a breadcrumb trail: 'data management plan | OA to research data | open science'. The central focus is a large green and blue graphic with the title 'Aspetti legali nella gestione dei dati della ricerca' (Legal aspects in the management of research data) and the speaker's name 'Thomas Margoni, University of Glasgow - CREATE OpenAIRE project'. To the left of this graphic are the logos for RDA (Research Data Alliance Europe) and the Italian National Node. To the right is a dark blue sidebar with a 'Support' section containing links for 'RESOURCES' (Open Science Primers, Guides, Factsheets, Use cases), 'HELPDESK' (FAQs, Ask a Question), and 'TRAINING' (Webinars, Workshops, Community of Practice).

OpenAIRE

SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

More Information about the 2019 webinar series.

data management plan | OA to research data | open science

RDA
RESEARCH DATA ALLIANCE
EUROPE

IT NATIONAL
NODE

OpenAIRE

Aspetti legali nella gestione dei dati della ricerca

Thomas Margoni
University of Glasgow - CREATE
OpenAIRE project

Support

RESOURCES

- Open Science Primers
- Guides
- Factsheets
- Use cases

HELPDESK

- FAQs
- Ask a Question

TRAINING

- Webinars
- Workshops
- Community of Practice

[non suoniamo tutti la stessa musica]

Obstacles to the trans-European archiving and sharing of research data

Making research data as openly available as possible is a widely recognised goal. For researchers working on an interdisciplinary project involving several countries, it can be difficult to fully comprehend in which ways open access to research data can be legally obtained. European national laws still diverge.

[A report from Knowledge Exchange](#) (Knowledge Exchange, 2011) concludes that it will remain difficult to predict when particular files of research data are protected because of:

- **Diversity in copyright protection**

Even though most research data will fail to meet the criteria for copyright protection because they are not likely to be considered as “works” (they mainly concern facts), the lack of harmonisation of the criteria for copyright protection in Europe is tricky. E.g., whereas Germany, Denmark and the Netherlands have a relatively similar (higher) originality standard, the UK has a very low standard (skill, judgment and labour) making it

- **Diversity in copyright owner**

If protection applies, the right holder's consent is required for sharing the data. However, the designation of the copyright owner is also different in different jurisdictions. Although in many cases the maker of the work will be considered to be the author and therefore the right holder, only Dutch and UK law designate the employer as the right holder if the work was made in the course of employment.

[CESSDA guide](#)

⊕ Czech Republic

⊕ Finland

⊕ Sweden

⊕ Switzerland

⊕ UK

[regole chiare dall'inizio]

Don't even
think of park-
ing here! :)



Safe to be open 2016

Safe to be open

Study on the protection
of research data and
recommendations
for access and usage

Edited by Lucie Guibault
and Andreas Wiebe

... FISSATE REGOLE CHIARE DA SUBITO

- CHI HA I DIRITTI SUI DATI (SE ESISTE)
- CHI DEVE PROVVEDERE A CONSERVARE
 - CHI PUÒ SFRUTTARE

R = Reusable. License

How to License Research Data

This guide will help you decide how to apply a licence to your research data, and which licence would be most suitable. It should provide you with an awareness of why licensing data is important, the impact licences have on future research, and the potential pitfalls to avoid. It concentrates on the UK context, though some aspects apply internationally; it does not, however, provide legal advice. The guide should interest both the principal investigators and researchers responsible for the data, and those who provide access to them through a data centre, repository or archive.

Creative Commons at a glance

Good for

- very simple, factual datasets
- data to be used automatically

Watch out for

- versions: use v. 4 or later
- attribution stacking
- the NC condition: only use with dual licensing
- the SA condition as it reduces interoperability
- the ND condition as it severely restricts reuse

ODC-By at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking

ODC-ODbL at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking
- the copyleft condition as it reduces interoperability
- the DRM clause as it may put off some reusers

Public domain at a glance

Good for

- most databases and datasets
- data to be used by anyone or any tool
- data to be used for any purpose

Watch out for

- lack of control over how database is reused
- lack of protection against unfair competition

Creative Commons



FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE V0.1

This information guide contains questions and responses to common concerns surrounding open science and the implications of licensing data under Creative Commons licences. It is intended to aid researchers, teachers, librarians, administrators and many others using and encountering Creative Commons licences in their work.

<https://doi.org/10.5281/zenodo.840651>

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the [CC0 Public Domain Dedication](#), which is first and foremost a waiver, but [can act as a licence](#) when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is not protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the [CC0 Public Domain Dedication](#), which is first and foremost a waiver, but [can act as a licence](#) when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is **not** protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

But I would like attribution when others use my dataset. In that case, shouldn't I use a CC BY licence?

We recommend that you avoid using a CC BY licence. Here's why:

While attribution is a genuine, recognisable concern, not only might using a CC BY licence be legally unenforceable when no underlying copyright or SGDR protects the work, but it may also communicate the wrong message to the world. A better solution is to use CC0 and [simply ask for credit](#) (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease. Such requests are consistent with scholarly norms for citing source materials.

Legally speaking, datasets that are **not** subject to copyright or related rights (and are thus in the public domain) cannot be the object of a copyright licence. Despite this, agreements based in contract law may be enforceable. Creative Commons licences, however, are copyright licences. Therefore, where the conditions for a copyright or related right are not triggered, copyright licences, such as the CC BY licence, **are unenforceable**.

In some cases, however, rights may exist (like the *sui generis* database right previously mentioned), and permission for others to use your dataset may be legally required. These rights are meant to protect the maker's investment, rather than originality. As such, database rights do not include the moral right of attribution. So by using a CC BY licence, you signal to users that you restrict access to your dataset beyond the protections provided by the law. We are not saying that this cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

Commons Op

USARE CC0

- CHIEDERE CHE VENGA DATO CREDITO ALL'AUTORE
- PROPORRE GIÀ LA CITAZIONE-TIPO (non citare la fonte è scorretto scientificamente)

It sounds like you're really pushing for the use of CC0 for open science datasets.

Exactly. Data is only open if anyone is free to use, reuse, and distribute it. This means it must be made available for both commercial and non-commercial purposes under non-discriminatory conditions that allow for it to be modified.

When data is made available for all reuse, others can create new knowledge from combining it. This leads to the enrichment of open datasets and further dissemination of knowledge. Accordingly, CC0 is ideal for open science as it both protects and promotes the unrestricted circulation of data.

And remember, it's bad science not to cite the source of data you use. To help others cite your data [include a citation](#) that users can copy and paste to give you credit for your hard work.

cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

I'm uncomfortable with others using my research for commercial purposes. Should I use a non-commercial licence for my dataset?

We recommend you avoid using a non-commercial licence. Here's why:

For legal purposes, drawing a line between what is and is not 'commercial' can be tricky; it's not as black and white as you might think. For example, if you release a dataset under a non-commercial licence, it would clearly prohibit an organisation

I'm uncomfortable per use of my research for all purposes. Should I use Derivatives' (ND) licence?

We recommend you avoid using a Derivatives' licence. Here's why:

Similar to how a non-commercial licence can have the same effect on someone from recombining and data for new research. For data Access, it must permit these im reuse.

«RINUNCIA» AI DIRITTI FERMA RESTANDO LA NORMATIVA DEL PAESE [ITALIA: DIRITTI MORALI INALIENABILI, QUINDI RESTANO]

UN MODO
STRUTTURATO DI
PENSARE AI DATI

REGOLE CHIARE=MENO
ERRORI DA SUBITO

È UN «LIVING DOCUMENT»,
CRESCE COL PROGETTO

Top tip - keep it short and specific!

FOSTER toolkit

This very short extract from a presentation by Peter Dukes, Medical Research Council (MRC) back in 2012, provides really useful advice on writing a DMP from the funding body perspective. While it is an example from the Life Sciences, the advice applies to all disciplines. The quality of the video isn't great, unfortunately, but the advice provided definitely is!



Advice on writing Data Management Plans

Research Data Improved Data Management Plans

4. Keep it simple

- Informative: two audiences
- Specific: e.g. name standards
- Concise: < 1/4 to 3 pages
- Don't forget: your achievements & innovation

DOVE METTERE TUTTE QUESTE INFORMAZIONI?
NEL DATA MANAGEMENT PLAN

Vantaggi di un DMP



⊖ Benefit 3. Clarifies needed budget

Data management is not free. You do not want to find yourself running out of funding before the end of the project because you have ignored or underestimated the cost of structured, detailed, and safe data management. Therefore, an important aspect of a DMP is its use in calculating how much money will be required for managing your research data during your research project.

A DMP can be useful in the process of applying for funding. Grant applications should not only include time and resources for collecting, analysing, and publishing on data in their budget, time and resources for careful documentation as well as server space, backup solutions, and documentation software need to be included as well. A DMP is also useful once funding is granted to plan and manage your expenses. Many research funders require a DMP as part of the application and decision-making process. The arguments for making data available are several, the most popular being that the data produced by public funds should be used to the greatest extent possible and available to the public. Unless there are legal, ethical or commercial barriers, data should also be openly available so that research results can be verified, replicated and reused.

Examples of Data Management cost assessments are given by the [University of Utrecht](#) (n.d.) and the Dutch Landelijk Coördinatiepunt Research Data Management ([LCRDM](#), n.d.) inspired by the '[Data management costing tool](#)' by UK Data Service, 2013.

CESSDA Guide

È FONDAMENTALE PER
STIMARE I COSTI DI GESTIONE

DMP Core Requirements

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS



When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

- ☐ **1. Data description and collection or re-use of existing data**
 - a. How will new data be collected or produced and/or how will existing data be re-used?
 - b. What data (for example the kinds, formats, and volumes) will be collected or produced?

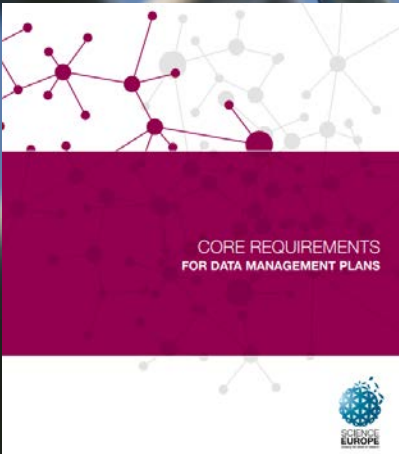
- ☐ **2. Documentation and data quality**
 - a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
 - b. What data quality control measures will be used?

- ☐ **3. Storage and backup during the research process**
 - a. How will data and metadata be stored and backed up during the research process?
 - b. How will data security and protection of sensitive data be taken care of during the research?

- ☐ **4. Legal and ethical requirements, codes of conduct**
 - a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
 - b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
 - c. How will possible ethical issues be taken into account, and codes of conduct followed?

- ☐ **5. Data sharing and long-term preservation**
 - a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
 - b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
 - c. What methods or software tools will be needed to access and use the data?
 - d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

- ☐ **6. Data management responsibilities and resources**
 - a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
 - b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?



DMP Core Requirements

Translating the Core Requirements into a DMP template

The following example of a data management plan template is based on the core requirements for DMPs.⁶ These core requirements should be considered as a minimum standard, leaving the flexibility to formulate additional guidelines according to the needs of specific domains or to national or local legislation.

The template presented below refers to the 15 questions covering six core requirements for good data management. Additional guidance and explanations are provided to help researchers fill out such a template and to assure that all relevant aspects of research data management are covered. The below table is an example of how the core requirements can be transformed into a DMP template. It will be up to the individual organisations and disciplines to develop templates that fit their needs.

GENERAL INFORMATION

- Administrative information**
- Provide information such as name of applicant, project number, funding programme, version of DMP.

1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA

1 a

How will new data be collected or produced and/or how will existing data be re-used?

- Explain which methodologies or software will be used if new data are collected or produced.
- State any constraints on re-use of existing data if there are any.
- Explain how data provenance will be documented.
- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

2 DOCUMENTATION AND DATA QUALITY

2 a

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data.
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
- Use community metadata standards where these are in place.
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.
- Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.

2 b

What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.



DMP Core Requirements

Compatibility with the FAIR Data Principles

THE FAIR DATA PRINCIPLES		Core Requirements for DMPs (CR)	Criteria for the Selection of Trustworthy Repositories
To be Findable			
F1	(meta)data are assigned a globally unique and eternally persistent identifier	CR 5d	Criterion 1
F2	data are described with rich metadata	CR 2a	Criterion 2
F3	metadata clearly and explicitly include the identifier of the data they describe	CR 5d	Criterion 1, Criterion 2
F4	(meta)data are registered or indexed in a searchable resource		Criterion 2
To be Accessible			
A1	(meta)data are retrievable by their identifier using a standardised communications protocol	CR 5c	Criterion 1, Criterion 2
A1.1	the protocol is open, free, and universally implementable	CR 5c	Criterion 2
A1.2	the protocol allows for an authentication and authorisation procedure, where necessary	CR 4b, CR 5a, CR 5c	Criterion 3
A2	metadata are accessible, even when the data are no longer available	CR 4c, CR 5a, CR 5d	Criterion 2c

THE FAIR DATA PRINCIPLES		Core Requirements for DMPs (CR)	Criteria for the Selection of Trustworthy Repositories
To be Interoperable			
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	CR 1b, CR 2a	Criterion 2d
I2	(meta)data use vocabularies that follow FAIR principles	CR 2a, CR 2b	Criterion 2
I3	(meta)data include qualified references to other (meta)data	CR 2a, CR 5a, CR 5c	Criterion 2b
To be Re-usable			
R1	meta(data) are richly described with a plurality of accurate and relevant attributes	CR 2a, CR 2b	Criterion 2
R1.1	(meta)data are released with a clear and accessible data usage license	CR 4b, CR 5a	Criterion 3d
R1.2	(meta)data are associated with detailed provenance	CR 1a, CR 1b, CR 2b	Criterion 1c, Criterion 2, Criterion 3b, Criterion 4a
R1.3	(meta)data meet domain-relevant community standards	CR 1b, CR 2a	Criterion 2d



DMP questions



Overview

Title of the project

Date of this plan

Description of the project

- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data

- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers

- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)

- What are their contact details and their roles in the project?

Funder (if applicable)

- If funding is granted, what is the reference number of the funding granted?

Data producer

- Which organisation has the administrative responsibility for the data?

Project data contact

- Who can be contacted about the project after it has finished?

Data owner(s)

- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles

- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs

- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses for (preparing the data for) archiving covered?

Organising and documenting your data

Data collection

- How will the data be collected?
- Is specific software or hardware or staff required?
- Who will be responsible for the data collection?
- During which period will the data be collected?
- Where will the data be collected?

Data organisation

- How will you organise your data?
- Will the data be organised in simple files or more complex databases?
- How will the data quality during the project be ensured?
- If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

Data type and size

- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- After the project: What is the total amount of data collected (in MB/GB)?

File format

- In what format will your data be?
- Does the format change from the original to the processed/final data?
- Will your (final) data be available in an open format?

Folder structure and names

- How will you structure and name your folders?

File structure and names

- How will you structure and name your files?

Documentation

- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

Metadata

- What metadata will be provided with the collected/ generated/ reused data?
- How will metadata for each object be created?
- Is there any program that can be used to document the data?
- Can metadata be added directly into the files or will the metadata be produced in another program or document?

Metadata standard (if applicable)

- What metadata standard(s) will you use?

Basic Information.

- State the purpose of the data collection/generation.
- Explain the relation to the objectives of the project
- Consider what data will be collected or created as part of the study (RAW data).
- Consider what data will be produced by processing the RAW data (Secondary, processed data).
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- Specify the types and formats you plan to use for the data generated/collected (raw, processed, published).
- Consider what data will be published as the result of your study (Published data).

Volume and Life Cycle of the Data.

If you are using FAIRDOM, we will look after data that will be retained and potentially exchanged by your projects. It will help with local storage for temporarily-held local data prior to processing.

For RAW data, please consider the following:

- How much RAW data you think will be produced (Estimates, per month, year, full project duration)?
- Will all of the RAW data be kept for the duration of the study or will the RAW data be deleted once it is processed?
- For large scale RAW data (images, sequence) have you planned the local storage capacity necessary for processing?
- Do you require help to organise a suitable local management system for RAW data?
- Do you have policies that govern the management and usage of RAW data?
- How long will RAW data be kept? Will there be a long-term archive?

For Secondary and Published data, please consider the following:

- What data processing is foreseen in the project?
- How much processed data will be produced, and stored (can you make estimates per month, year, full project)?
- How much of this data will be published? (Estimates per month, year, full project)?
- Does your institution, or the project funders, have policies governing the access and usage of processed data?

Additional for personally sensitive data (e.g medical data)

- When looking at the data flow through the project, define what data is:
 - aggregated (typically safe to share, if names cannot be recovered)
 - anonymized (name cannot be recovered from the data)
 - pseudonymized (name can be recovered by some)
 - non-anonymized (name linked to data)
- Determine which organisational boundaries have to be traversed by which data.
- Make sure with your "local" data protection officer and ethics commission that the data can be shared with your partners along the flow described with the anonymisation levels as described. Why local? Some laws change across surprising boundaries. E.g. in Germany Universities and other public organisations are subject to another data protection law than enterprises. Why seek advice? In some cases you may be required to be able to recover the name-data-relation, e.g. to enable study participants to "leave" a study.

Data Management Checklist

<https://fair-dom.org/knowledgehub/data-management-checklist/>

Data Management Checklist

Making data findable (documentation and metadata management)

- What documentation and metadata will accompany the data (assist its discoverability)? (Details on methodology, definitions, procedures, SOPs, vocabularies, units, dependencies, etc)
- What information is needed for the data to be read and interpreted in the future?
- What naming conventions will be used?
- How will you approach versioning your data?
- How will you capture / create this documentation and metadata?
- How do you ensure the completeness of the captured data?

Making Data Accessible

- Specify which data will be made openly available taking into consideration
 - What ethics and legal compliance issues do you have if any? Do you need consent for data preservation and sharing? Do you have to protect certain data? Is any data sensitive?
 - Do you think you might have Intellectual Property Rights issues? Have you considered ownership of the data, licensing, restrictions on use?
 - Do you think you will need to embargo any data?
- How will you make the data available? (consider the platforms you will use: databases, repositories, etc)
- What methods or software tools are needed to access the data? You should list where the software can be obtained. You should also document how to use the software to access the data. The documentation should be as complete as possible, including examples. If you distribute your system, include the access software and its documentation as part of any distribution.
- If there are any restrictions on accessibility, how will you provide access?

Making Data Interoperable

- What standards (metadata vocabularies, formats, checklists) or methodologies will you use?
- How do you address data and model quality? What validation steps do you foresee?
- Will you use standardised vocabulary for all data types to allow inter-disciplinary interoperability?
- Where you can not use standardised vocabulary for all types of data, can you map to more commonly used ontologies?

Making data Re-usable

- How will you licence your data to permit the widest re-use possible?
- When will the data be made available for re-use? Does this include an embargo period? (if yes, please detail why)
- Which data will be available for re-use during/after the project? For data that is not re-usable, please detail why
- What are your data quality assurance processes?
- How long do you expect your data to remain re-usable?

PERSONALIZZABILE

DMP online

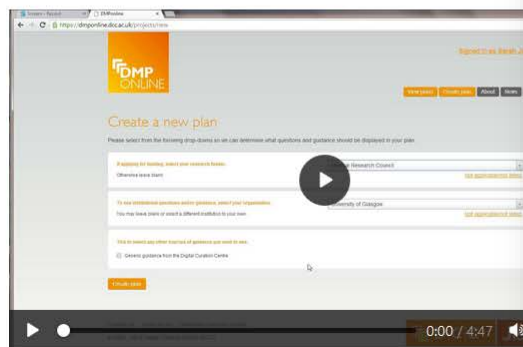


Home About Future plans Help Change language

Welcome.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curatorial Studies (UC3).

Screencast on how to use DMPonline



<https://dmponline.dcc.ac.uk/>

Sign in

Veteran tapes

Project Details Plan overview Write Plan Share Download

expand all | collapse all

13/13 answered

Data Collection (2 / 2)

What data will you collect or create?

B I [List Icon] [Link Icon] [Grid Icon]

The "Veteran tape " project will collect and generate different types of datasets:

Type of data	Volume	Format	Storage format
Video recordings	600 x 1Gb	.mkv	.mkv
Transcriptions	600 x 1500Kb	MS Word	.txt
Structured interview text	1 x 500Kb	MS word	.txt

For the video recordings the selected format is .mkv; the same .mkv format will be used for the long-term preservation .

Transcriptions will be written in MS Word and then stored as .txt files.

We checked the format compatibility against EASY File format
<https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats>

As the total volume of data is greater than 50Gb, DANS requires a fee for the storage. We are currently in touch with EASY to determine the costs of archiving.

Save

Guidance

Comments (1)

DCC

DCC guidance

Guidance

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.

2. Data Management Plan – general definition

Data Management Plans (DMPs) are a *key element* of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of 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 access and
- how data will be curated and preserved (including after the end of the project).

A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt submit a DMP on a voluntary basis.

IN HORIZON 2020
DMP ENTRO M6
COME DELIVERABLE

4. Research data management plans during the project life cycle

Once a project has had its funding approved and has started, you **must submit a first version of your DMP** (as a deliverable) within the first 6 months of the project. The Commission provides a DMP template in annex, the use of which is recommended but voluntary.

The DMP needs to be **updated** over the course of the project whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP should be updated as a minimum in time with the periodic evaluation/assessment of the project. If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest. Furthermore, the consortium can define a timetable for review in the DMP itself.

Periodic reporting

For general information on periodic reporting please check the following sections of the online manual

- How to fill in [reporting tables for publications, deliverables](#)
- [Process for continuous reporting](#) in the grant management system.



H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

Guide 2016

Version 3.0
26 July 2016

... buon lavoro!

ONE DAY OR
DAY ONE
you decide.

...grazie!
elena.giglia@unito.it