INFORMAZIONI DI SERVIZIO

PER GLI «ESTERNI» NON UNITO
CHI AVESSE BISOGNO DEL CERTIFICATO DI PARTECIPAZIONE
DEVE ISCRIVERSI UTILIZZANDO IL MODULO
CHE TROVATE ALLA PAGINA SU OA.UNITO/CORSI

PER FAVORE SPEGNETE MICROFONI E WEBCAM!

ISCRIVETEVI ALLA FINE DEL CORSO…
DOVRETE DICHIARARE DI AVERLO SEGUITO!

LE SLIDES SONO DISPONIBILI SU ZENODO
https://doi.org/10.5281/zenodo.3900981

POI SARANNO SULLA PAGINA
https://www.oa.unito.it/new/open-science-open-access-fair-data-eosc-webinar-in-collaborazione-con-openaire/
INSIEME ALLA REGISTRAZIONE

REGISTRAZIONE INUTILIZZABILE SE TENETE MICROFONI APERTI

OPEN SCIENCE DALLA A ALLA Z
5 - GESTIONE DEI DATI, FAIR, OPEN

Elena Giglia
25 giugno 2020
elena.giglia@unito.it
@egiglia

La strada davanti a noi

1

- Perché occuparsi dei dati
- EOSC è qui fra noi [e Raccomandazioni, Direttive…]
- Dati gestiti correttamente, dati FAIR, dati Open

2

- FAIR in pratica
- Data Management Plan, questo sconosciuto

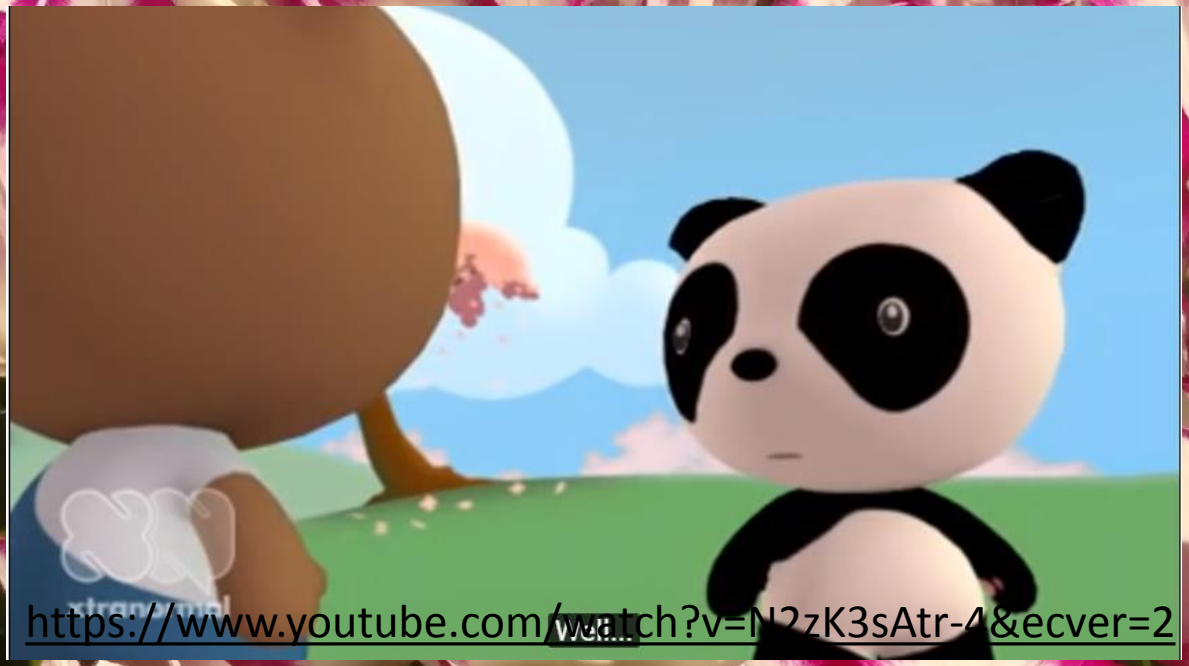
Open Science, Open Innovation, EOSC, FAIR data: esserci!

Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century

There is value and risk of being a first mover, but there is higher risk of being a follower.

Perché occuparci dei dati?

[Video]



… è l'incubo del data steward:

- nessun backup
- nessun software di accompagnamento
- nessuna legenda dati

… e in più:

- dati prodotti con fondi pubblici
- pubblicati su Science che li richiede
- utili a una ricercatrice di area diversa

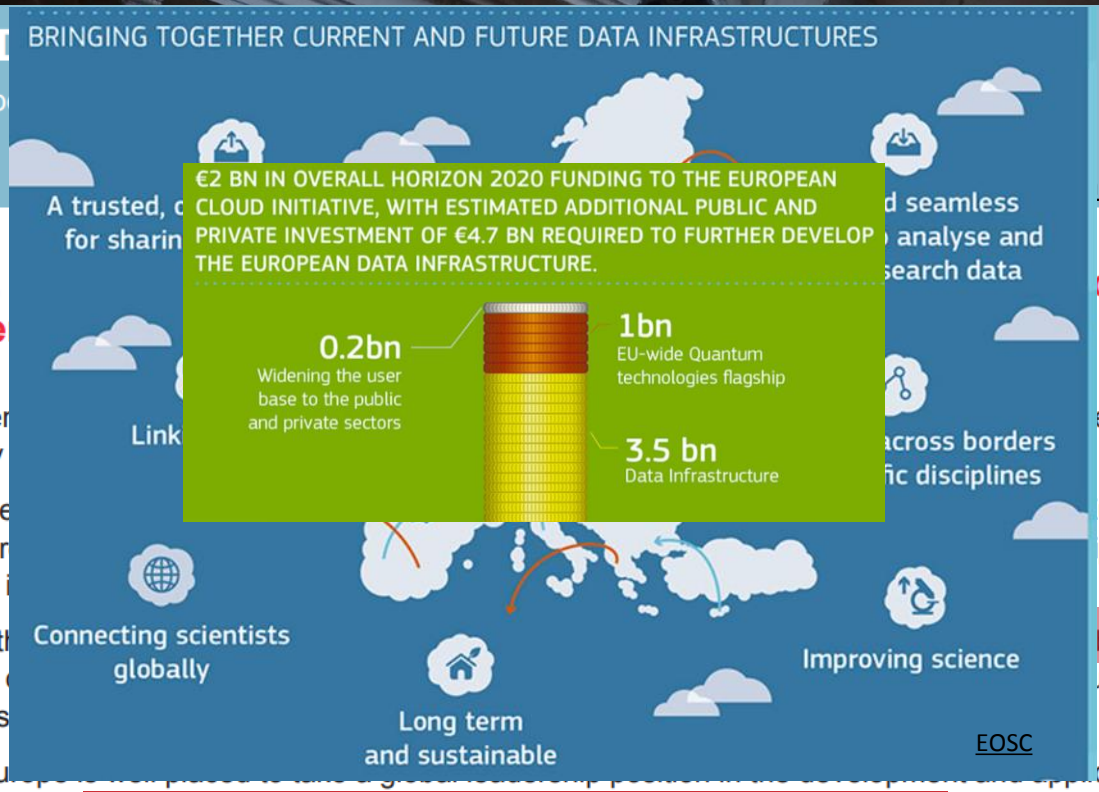
[THE EOSC!]

The Vienna D

Vienna, 23 Novemb

u 1 8
· a t

na, Nov.23, 2018



We, Ministers, European Ope

1. Recall the challe Brussels on 10 July

2. Reaffirm the pote the vision of the Eur States, sustainable

3. Recognise that t iterative and based consensus among s

4. Highlight that Eu services for Science. Rea reaching out over time to

5. Recall that the Council

SEAMLESS ACCESS TO OPEN BY DEFAULT
FAIR DATA

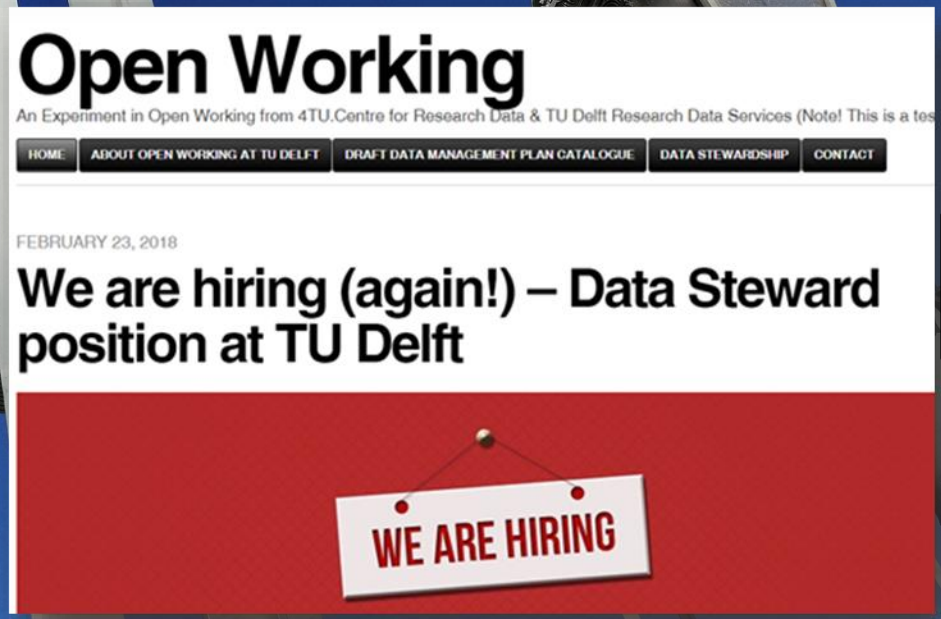
9. Call for the European Open Science Cloud to provide all researchers in Europe with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing and processing research data supported by FAIR data principles.

Science Cloud a reality, hinting at the need to further strengthen the ongoing dialogue across institutions and with stake-holders, for a new governance framework to be launched in Vienna, on 23 November 2018.

[EOSC e I data stewards]

Report, 2016

The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation[9]. The success of the EOSC depends upon it.



≠ DATA SCIENTIST

≠ DATA ANALYST

[competenze]

KØBENHAVNS UNIVERSITET 06/04/202 0

Education core content

This 1-year degree should build upon students' educational/job background through domain specific data knowledge and leverage with theoretical and practical competences.
The education can be viewed as a Data Steward specialisation within the domain of their previous degree/jobs. The education contains **60 ECTS** and is expected to finish with a 15 ECTS project.

Preliminary Content

The 60 ECTS should be distributed among the following main areas:

- 22,5-30 ECTS: IT competences – including computational thinking, data modelling, data management, data harvesting, cleaning, and storing, infra-structure (storage & compute). An introduction to data science, machine learning, and their derived data needs.
- 7,5-15 ECTS: Legal and ethical competences – including GDPR, FAIR, data security, and data & AI ethics.
- 7,5-15 ECTS: Domain specific data competences – including knowledge about data, infrastructure, and practice within the students primary domain, e.g., health, life-science, finance/fintech, or the public sector.
- 15 ECTS: Graduate project (possibly in collaboration with academia, industry, or the public sector)

Competences such as project management, communication skills, and change management should be incorporated as well

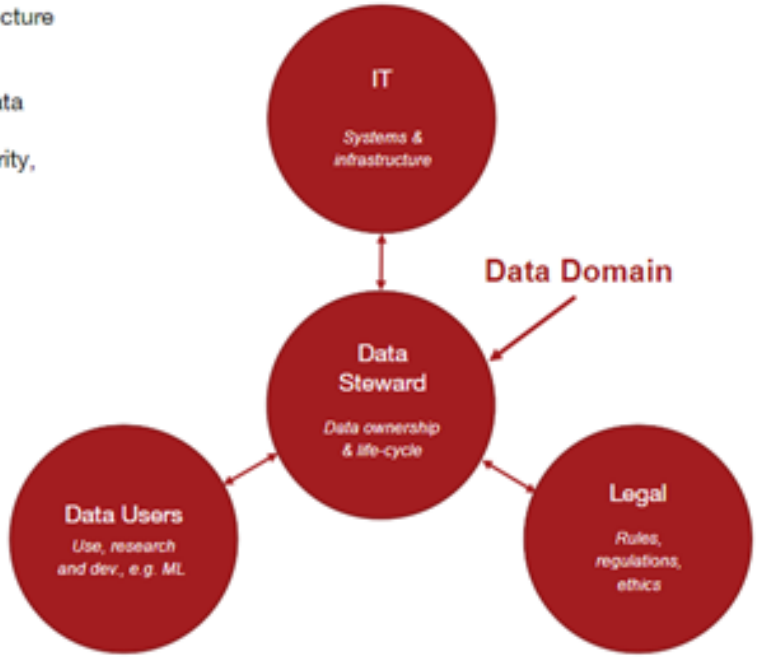
Competence Profile

A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

The future Data Steward must assume ownership and responsibility for data, data quality, and the data life-cycle as their primary function. They should ensure collaboration and coherence between IT competences, quality assurance, security, rules & regulations, and facilitate the application and use of data internally and externally in the organisation.

Competence profile examples

- Domain-specific data understanding
- Ability to ensure that structured and unstructured data and meta data is modelled, harvested, stored, and maintained in a documented, and regulated fashion with focus and findability, accessibility, interoperability, and reusability.
- Competences to facilitate HPC (High Performance Computing) during development and research through handling of large-scale data in public and private enterprises.
- Understanding of and competences within legal, ethical and security aspects of data handling, data sharing, e.g., integrity and GDPR.



Copenhagen Univ. June 17 2020

…e non dimentichiamo H2020

DMP VA FATTO SEMPRE, NON SOLO SE CI SONO DATI OPEN.
OGNI PROGETTO GESTISCE DATI (E DEVE FARLO CORRETTAMENTE).
CHE POI LI APRA O NO SI VEDE APPUNTO NEL DMP…

https://goo.gl/sryNTg

3. Open access to research data (Extended Open Research Data Pilot)

What?

Beneficiaries of actions that participate in the Open Research Data Pilot must give **open, free-of-charge access** to the end-user to **digital research data** generated during the action (⚠ new in Horizon 2020).

⚠ As of the Work Programme 2017, the Open Research Data pilot has been extended to all thematic areas of Horizon 2020 (except ERC PoC actions, SME instrument Ph1 actions, ERA-NET Cofund actions that do not produce data, EJP Cofund actions, and prizes).

Participation is therefore now in principle **the default**. However, actions may **opt out** at any stage — both before signing the GA and afterwards (through an amendment; *see Article 55*) —, if:

- participation is incompatible with the obligation to protect results *(see Article 27)*
- participation is incompatible with the security obligations *(see Article 37)*
- participation is incompatible with rules on protection of personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate/collect any research data or
- there are other legitimate reasons not to take part.

GRANT AGREEMENT ART. 29.3

- PILOT PROJECT EXTENDED TO ALL DISCIPLINES
- OPT OUT CLAUSES

PRINCIPLE: **«AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»**

...e non dimentichiamo H2020

https://goo.gl/sryNTg

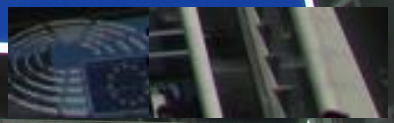
Actions participating in the pilot must draw up a **data management plan (DMP)** within the first 6 months of the project implementation.

The data management plan must support the management life-cycle for all data that will be collected, processed or generated by the action. It must cover how to make data findable, accessible, interoperable and re-usable (FAIR), including:

- the handling of data during and after the project
- what data will be collected, processed or generated
- what methodology and standards will be applied
- whether data will be shared / made open access (and how) and, if any, what data will not be shared / made open access (and why)
- how data will be curated and preserved.

The data management plan should be updated (and become more precise) as the project evolves. New versions should be created whenever important changes to the project occur *(e.g. new data sets, changes in consortium policies, etc.),* at least as part of the mid-term review (if any) and at the end of the project.

DATA MANAGEMENT PLAN (DMP)
- DELIVERABLE
- WITHIN M6
- TO BE UPDATED



In questo modulo impareremo:

1. come gestire i propri dati correttamente e rendere la ricerca più efficace
2. la differenza fra dati FAIR e Open

MESSAGGI CHIAVE

- gestire bene i dati è nell'interesse di chi fa ricerca
- solo dati gestiti bene possono essere resi FAIR e se possibile aperti
- NON ci sarà una ricetta per DMP, ma strumenti utili (da imparare)

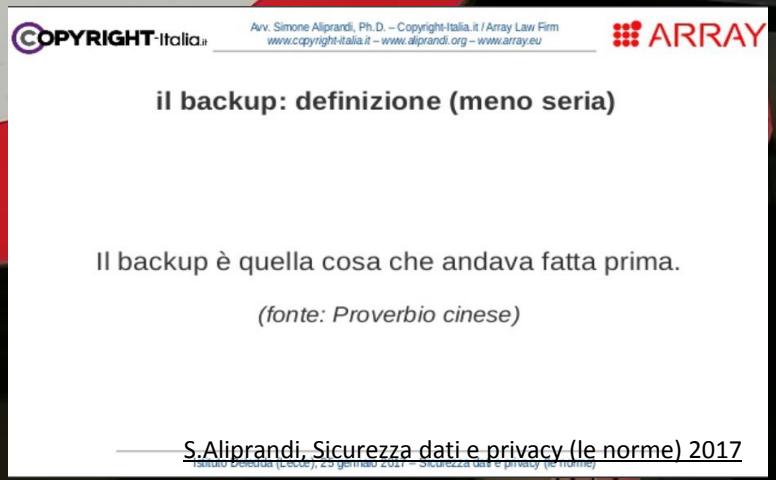
[DMP]

...parliamo di dati

QUANTI DI VOI AVREBBERO PROBLEMI SE ORA IL VOSTRO PC SI ROMPESSE?

QUANTI DI VOI AVREBBERO PROBLEMI SE PERDESTE LA CHIAVE USB?

QUANTI DI VOI AVREBBERO PROBLEMI SE I FILES SU DRIVE SCOMPARISSERO?



… i dati sono fragili

Scientists losing data at a rapid rate

Decline can mean 80% of data are unavailable after 20 years.

Elizabeth Gibney & Richard Van Noorden

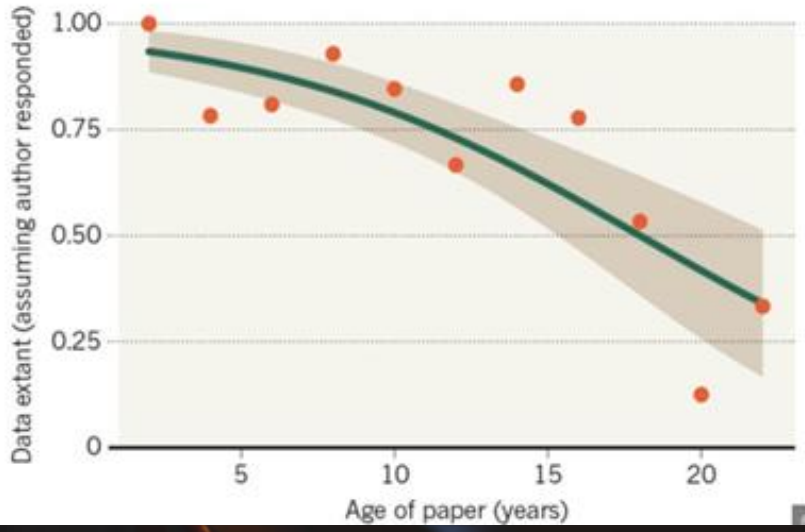
19 December 2013

Rights & Permissions

80% saranno persi in 20 anni

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416

…ECCO A COSA SERVE IL DATA MANAGEMENT PLAN.
NON È SOLO L'ENNESIMA NOIA BUROCRATICA

Level	Data Classification and Examples (abridged version)
5	***Information that would cause severe harm to individuals or the University if disclosed.***
	• Research information classified as Level 5 by an IRB or otherwise required to be stored or processed in a high security environment and on a computer not connected to the Harvard data networks • Certain individually identifiable medical records and genetic information, categorized as extremely sensitive
4	***Information that would likely cause serious harm to individuals or the University if disclosed.***
	• High Risk Confidential Information (HRCI) and research information classified as Level 4 by an IRB • Personally identifiable financial or medical information • Information commonly used to establish identity that is protected by state, federal, or foreign privacy laws and regulations • Individually identifiable genetic information that is not Level 5 • National security information (subject to specific government requirements) • Passwords and Harvard PINs that can be used to access confidential information
3	***Information that could cause risk of material harm to individuals or the University if disclosed.***
	• Research information classified as Level 3 by an IRB • Information protected by the Family Educational Rights and Privacy Act (FERPA) to the extent it is not covered under Level 4 including non-directory student information and directory information about students who have requested a FERPA block • HUIDs associated with names or any other information that could identify individuals
1	***Public information.***
	• Research data that has been de-identified in accordance with applicable rules • Published research • Published information about the University • Course catalogs • Directory information about students who have not requested a FERPA block • Faculty and staff directory information

Harvard security

S
e

…PERCHÉ IL PROBLEMA NON È SOLO DATI APERTI/CHIUSI A FINE RICERCA…
MA, BEN PIÙ IMPORTANTE,
DOVE LI CONSERVO MENTRE CI LAVORO?
CHI HA ACCESSO?
CHE SISTEMA DI SICUREZZA È PREVISTO?

TU Delft

May 10, 2020

Data Management Concept Note

Main Points

Data acquisition

- Check the type, source of the data and how to gather/collect it
  - Data types (to help define sensitivity of collected and processed data)
  - Data format (to help define the tools and infrastructure needed)
  - Data size (to help define storage and infrastructure needs)
- Check the ownership of the collected and processed data
  - Check with the data source about the ownership, attribution and use conditions (e.g. licence)
  - Check the need to make a data processing agreement with the data source on the ownership / access control
  - Are there (own) institutional policies that apply to the data?
  - Can the data be shared with other parties (if so, under which circumstances?)
- Confidentiality of the data (if applicable):
  - Register crucial information regarding data access / sharing
  - Ensure security of confidential data (personal data, commercially sensitive data, or data that would harm society when publicly disclosed)
  - Ensure compliance with General Data Protection Regulation / Algemene verordening gegevensbescherming when processing personal data
  - Ensure there are procedures in place to avoid data breaches, with the support of a privacy advisor/data protection officer

Data collection

- Establish a workflow for data collection
  - How will the data be collected?
  - Who has access to which data in short / long term?
  - What resources are needed for data analysis?
  - How will the data be exchanged / transferred among relevant stakeholders?
- Storage arrangement
  - Check available storage capacity and backup strategy

Data storing / backup

- Create a clear folder structure and consistent file naming convention
- Make a backup strategy where data is stored at least two different physical locations and preferably automatically backed up
- Access control to confidential data
- Apply encryption at disk or folder level if needed
- Create a consistent and standard versioning of the data files
- Determine the minimal documentation of the data that is required to find it, understand it and use it

Data sharing

- Create proper data sharing procedures
  - Consider agreements established in the Data acquisition phase, and evaluate/assess data sharing with other parties
  - Be aware of the permission and consequence of sharing confidential data
- Copyright / Licensing
  - How should others use the data
  - Who should be attributed for creating/gathering the data

Organizational Implications

In addition to the above mentioned actions, there are also a few things to consider to make data management a standard practice in daily operations.

I dati?

HARVARD UNIVERSITY

Research Data Management @Harvard

Vision

Research data are an important asset to our University and our researchers. Thus, we want to support robust data management and documentation practices that ensure long term-access and re-use of Harvard research data.

learn

Model Policy for Research Data Management (RDM) at Research Institutions/Institutes

1. PREAMBLE

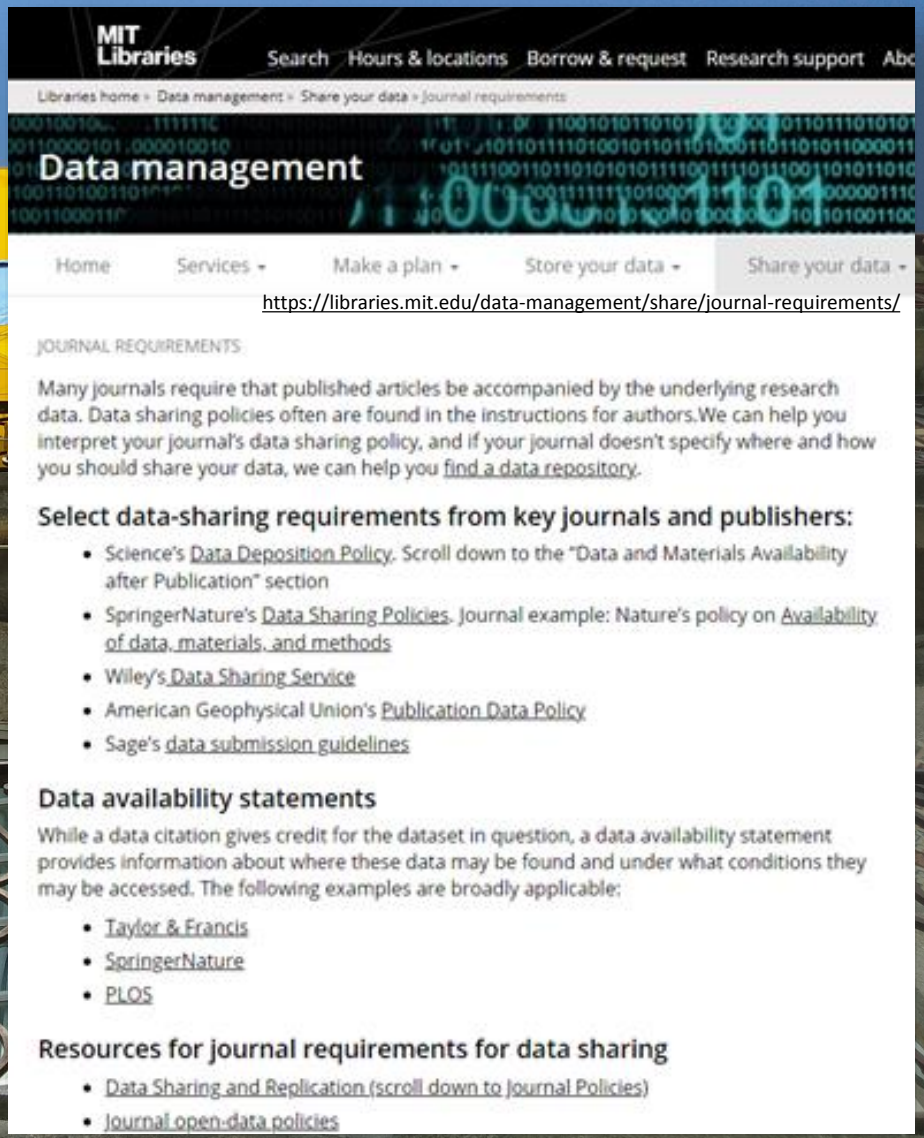
The [*name of research institution*] recognizes the fundamental importance of research data[1] and the management of related administrative records in maintaining quality research and scientific integrity, and is committed to pursuing the highest standards. The [*name of research institution*] acknowledges that correct and easily retrievable research data are the foundation of and integral to every research project. They are necessary for the verification and defence of research processes and results. RDM policies are highly valuable to current and future researchers. Research data have a long-term value for research and academia, with the potential for widespread use in society.

PERCHÉ I DATI SONO IL FONDAMENTO DELLA SCIENZA E DI UNA RICERCA RESPONSABILE

DCC because good research needs good data

Perché occuparci dei dati?

PER INTERESSE…
PERCHÉ SE I DATI NON SONO «ESPOSTI» LE RIVISTE NON ACCETTANO IL LAVORO…



MIT Libraries — Search | Hours & locations | Borrow & request | Research support | Abc

Libraries home › Data management › Share your data › Journal requirements

Data management

Home | Services | Make a plan | Store your data | Share your data

https://libraries.mit.edu/data-management/share/journal-requirements/

JOURNAL REQUIREMENTS

Many journals require that published articles be accompanied by the underlying research data. Data sharing policies often are found in the instructions for authors.We can help you interpret your journal's data sharing policy, and if your journal doesn't specify where and how you should share your data, we can help you find a data repository.

Select data-sharing requirements from key journals and publishers:

- Science's Data Deposition Policy. Scroll down to the "Data and Materials Availability after Publication" section
- SpringerNature's Data Sharing Policies. Journal example: Nature's policy on Availability of data, materials, and methods
- Wiley's Data Sharing Service
- American Geophysical Union's Publication Data Policy
- Sage's data submission guidelines

Data availability statements

While a data citation gives credit for the dataset in question, a data availability statement provides information about where these data may be found and under what conditions they may be accessed. The following examples are broadly applicable:

- Taylor & Francis
- SpringerNature
- PLOS

Resources for journal requirements for data sharing

- Data Sharing and Replication (scroll down to Journal Policies)
- Journal open-data policies

Perché occuparci dei dati?

Health · Second Opinion

'We're opening everything': Scientists share coronavirus data in unprecedented way to contain, treat disease

Feb.1, 2020

The Washington Post
Democracy Dies in Darkness

Jan. 24 2020

Science

Scientists are unraveling the Chinese coronavirus with unprecedented speed and openness

Just 10 days after a pneumonia-like illness was first reported among people who visited a seafood market in Wuhan, China, scientists released the genetic sequence of the coronavirus that sickened them. That precious bit of data, freely available to any researcher who wanted to study it, unleashed a massive collaborative effort to understand the …idly spreading in China and beyond.

…n an open-access repository for genetic …y Mesecar, a professor of cancer structural …d his laboratory to start analyzing the DNA …e to that of severe acute respiratory syndrome …ed more than 8,000 people and killed nearly 800. Scientists at the National Institutes of Health's Rocky Mountain Laboratories in Montana asked a company to turn the information from a string of letters on a computer screen into actual DNA they could study in lab dishes.

PERCHÉ NELLE CRISI SI CAPISCE LA LORO IMPORTANZA

At unprecedented speed, scientists are starting experiments, sharing data and revealing the secrets of the pathogen — a race that is made possible by new scientific tools and cultural norms in the face of a public health emergency.

AAAS Become a Member

Science Contents News Careers Journals

Two papers relying on hospital records of COVID-19 patients have been retracted because the company that purportedly analyzed the raw data won't allow their validity to be independently validated. AP PHOTO/MANU FERNANDEZ

Two elite medical journals retract coronavirus papers over data integrity questions

By Charles Piller, Kelly Servick | Jun. 4, 2020 , 5:30 PM

June 4, 2020

https://www.gisaid.org/

GISAID 2008 - 2020

About us Database Features Events Collabor

In Focus

30,000 viral genomic sequences of hCoV-19 shared with unprecedented speed via GISAID

Since the start of the COV… pandemic virus, laborator… genome sequence data w… progress in the understan… and development of candi… are essential to design an…

CORD-19

COVID-19 Open Research Dataset

The Semantic Scholar team at the Allen Institute for AI has partnered with leading research groups to provide CORD-19, a free resource of more than 63,000 scholarly articles about the novel coronavirus for use by the global research community.

https://www.semanticscholar.org/cord19

[OCDE]

OECD Innovation
@OECDinnovation

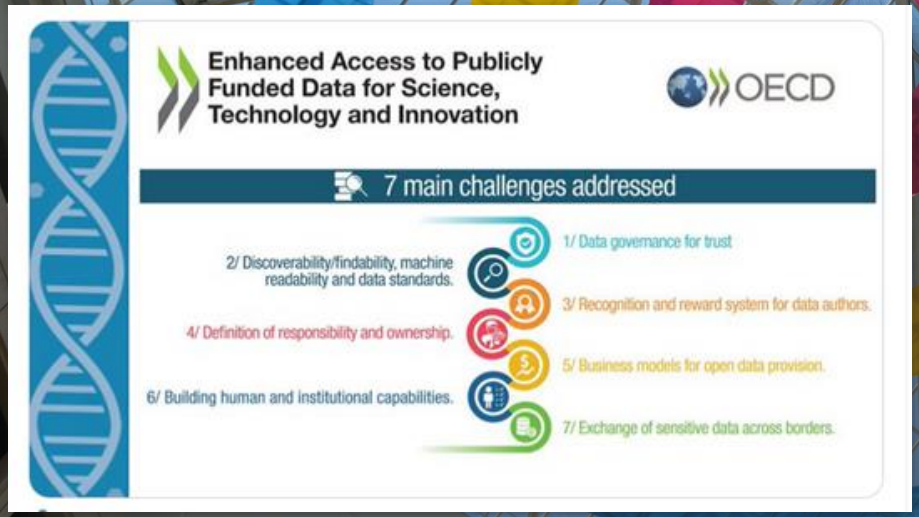
May 20, 2020

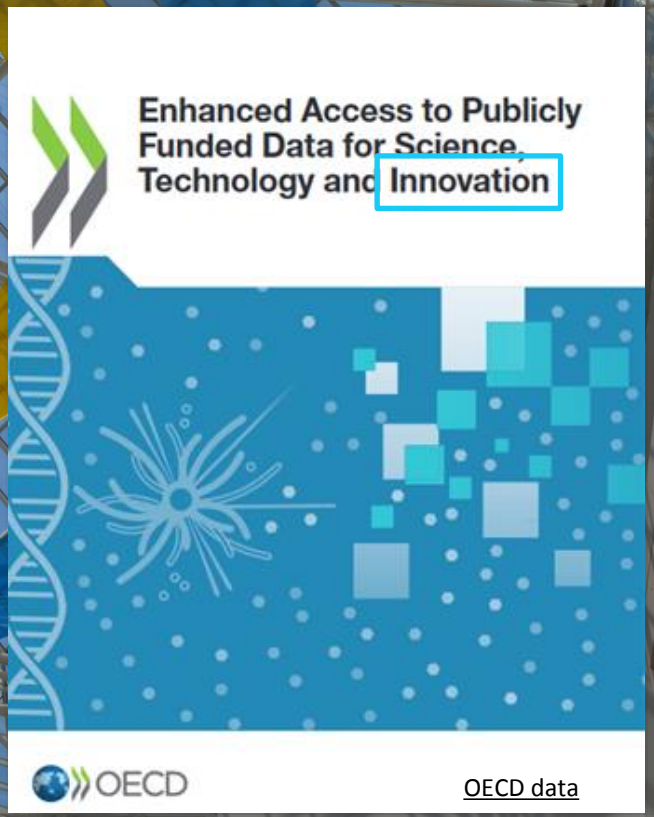
Access to publicly funded data has become more important than ever during the COVID-19 crisis.

We look at what countries can do to encourage #DataAccess in our report 👉 oe.cd/2ZO

#researchdata #opendata

Traduci il Tweet





OECD data

Perché occuparci dei dati?

Jisc

How and why you should manage your research data: a guide for researchers

An introduction to engaging with research data management processes.

JISC Guide

EVITARE DI PERDERLI

ORGANIZZARLI PER RENDERE PIÙ EFFICACE LA RICERCA

(SE APERTI) ESSERE PIÙ VISIBILI

ALCUNI SONO UNICI E IRRIPETIBILI (meteorologia)

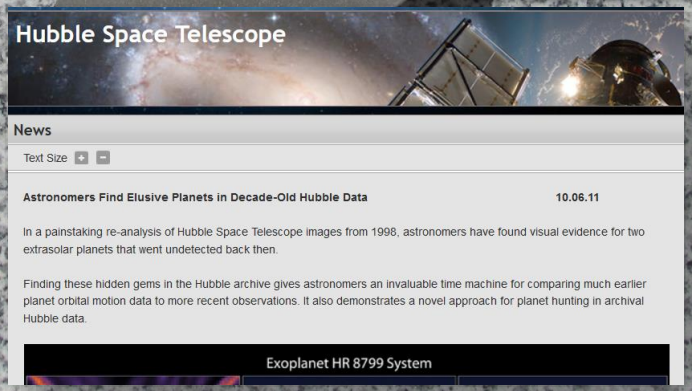
PERMETTERE VALIDAZIONE E CONTROLLI

(SE APERTI) FAVORIRE COLLABORAZIONI

MIGLIORARE INTEGRITÀ DELLA RICERCA

(SE APERTI) FAVORIRE RIUSO INEDITO

ESSERE RIPRODUCIBILI



«the coolest thing to do with your data will be thought of by someone else» [R.Pollock]

Perché occuparci di dati?

Il debito pubblico deprime la crescita? Il clamoroso errore di Carmen Reinhart e Kenneth Rogoff

Pubblicato da keynesblog il 18 aprile 2013 in consigliati, Economia, ibt, Teoria economica

2013

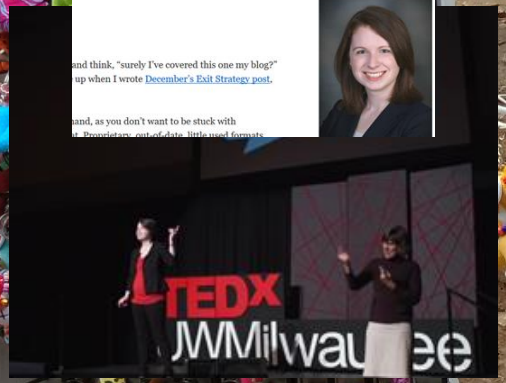
DATA AB INITIO

AGING RESEARCH DATA RIGHT, FROM THE START

K.Birney, 2015

and think, "surely I've covered this one my blog?"
up when I wrote December's Exit Strategy post.

hand, as you don't want to be stuck with
t, Proprietary, out-of-date, little used formats



Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Thomas Herndon* Michael Ash Robert Pollin

April 15, 2013

Herndon, 2013

JEL CODES: E60, E62, E65

Abstract

We replicate Reinhart and Rogoff (2010a and 2010b) and find that coding errors,

1. **l'esclusione selettiva di alcune osservazioni nei dati;**
2. **uno schema di bilanciamento dei dati non convenzionale;**
3. **un errore di codice nel foglio di calcolo originale utilizzato per selezionare i dati.**

debt loads greater than 90 percent of GDP consistently reduce GDP growth.

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

2015

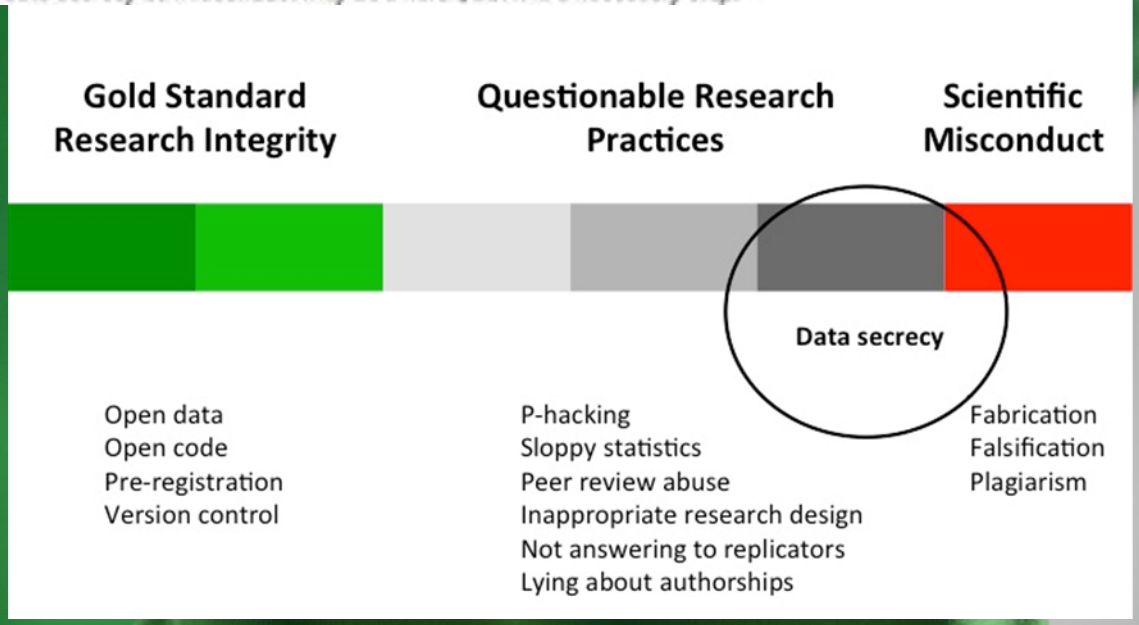
https://memegenerator.net/instance/64979477/case-closed-judge-judy-data-or-it-didnt-happen

No data?

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

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

A recent study sent data requests to 200 authors of economics articles where it was stated 'data available upon request'. Most of the authors refused. What does the scientific community think about those withholding their data? Are they guilty of scientific misconduct? Nicole Janz argues that if you don't share your data, you are breaking professional standards in research, and are thus committing scientific misconduct. Classifying data secrecy as misconduct may be a harsh, but it is a necessary step.



I dati creano ponti

Data creates a bridge between traditional disciplines, spawning discovery and innovation from the humanities to the hard sciences. Data dissolves barriers, opening up new channels of communication, lines of research, and commercial opportunities. Data will be the engine, the spark to create a better world for all.

World Economic Forum 2012, http://goo.gl/ExaGW

I dati creano ponti fra le discipline…
…e non è indifferente nel contesto Mission-oriented di Horizon Europe…

Missions in Horizon Europe

Mission areas

5 mission areas have been identified, each with a dedicated mission board an help specify, design and implement specific missions in Horizon Europe.

Mission area: Adaptation to climate change including societal transformation

Mission area: Cancer

Mission area: Climate-neutral and smart cities

Mission area: Healthy oceans, seas, coastal and inland waters

Mission area: Soil health and food

Horizon Europe

I dati rendono la ricerca più trasparente

Box 1. Some Research Practices that May Help Increase the Proportion of True Research Findings

- Large-scale collaborative research
- Adoption of replication culture
- Registration (of studies, protocols, analysis codes, datasets, raw data, and results)
- Sharing (of data, protocols, materials, software, and other tools)
- Reproducibility practices
- Containment of conflicted sponsors and authors
- More appropriate statistical methods
- Standardization of definitions and analyses
- More stringent thresholds for claiming discoveries or "successes"
- Improvement of study design standards
- Improvements in peer review, reporting, and dissemination of research
- Better training of scientific workforce in methods and statistical literacy

PLOS | MEDICINE 2014

OPEN ACCESS

ESSAY

How to Make More Published Research True

John P. A. Ioannidis

Published: October 21, 20

PLOS | MEDICINE 2005 BROWSE

OPEN ACCESS

ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://

OSF PREPRINTS https://dx.doi.org/10.17605/OSF.IO/2DXU5

The Preregistration Revolution

Brian Nosek, Charles Ebersole, Alexander DeHaven, David Mellor

Created on: June 16, 2017 | Last edited: March 12, 2018

The European Code of Conduct for Research Integrity

REVISED EDITION

Research Integrity

nature human behaviour

Altmetric: 2520 Citations: 63

Perspective | OPEN

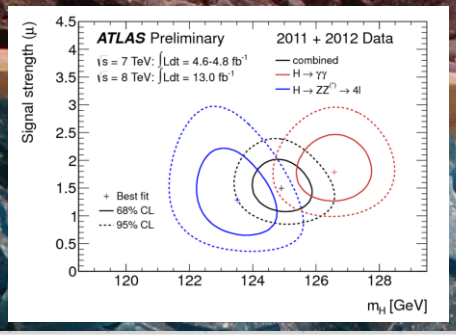
A manifesto for reproducible science

Marcus R. Munafò, Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware & John P. A. Ioannidis

Manifesto, Jan 2017

Parliamo di dati

«pezzi»
di conoscenza osservabili



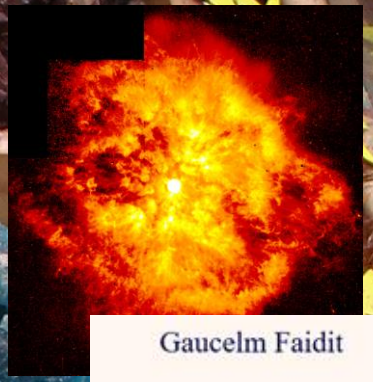


Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample from the cohort.

OTU	A01_TP1	A01_TP2	A01_TP3	A03_TP1	A03_TP3	A04_TP1	A04_TP2	A04_TP3	A05_T
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	
							19	85	
							646	214	
							292	37	
							18	170	
							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

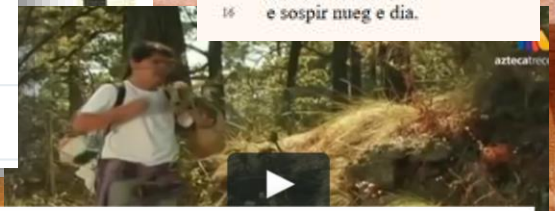
3 Retweet **1** Mi piace

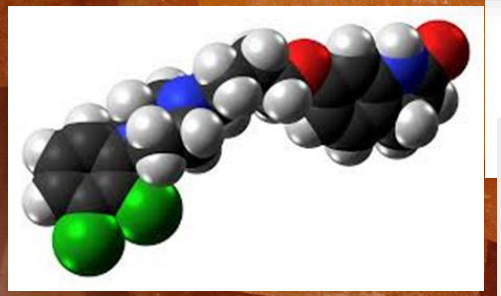
https://twitter.com/wvanwezenbeek/status/973527086685093893

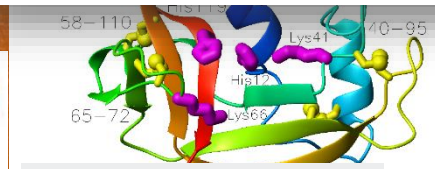
Gaucelm Faidit

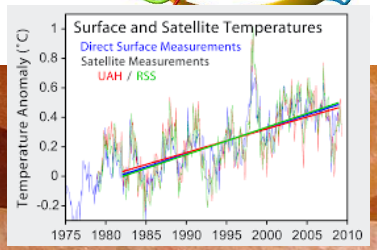
I.

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 planc e languis
e sospir nueg e dia.









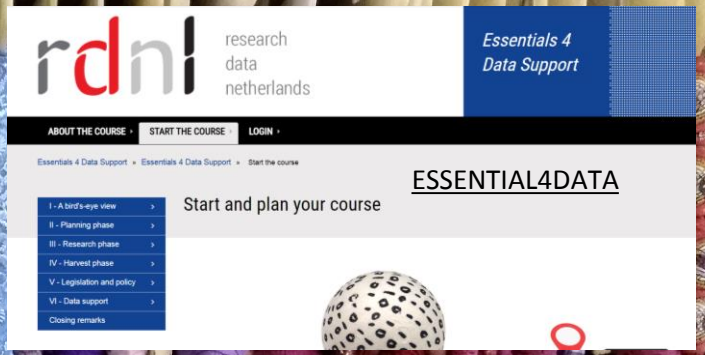
We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets.

[una premessa]

- …da qui in poi: panoramica sugli strumenti
- vanno «assaggiati» e adattati al proprio contesto…
- IMPARARE A USARLI PER SUGGERIRLI (impensabile che un ricercatore scenda così nel dettaglio)
- fondamentale il supporto
- **FONDAMENTALE UNA POLITICA ISTITUZIONALE CHE CHIARISCA RUOLI E RESPONSABILITÀ E DEFINISCA IL LIVELLO DEI SERVIZI**
- CREARE UNA **RETE DI DATA STEWARDS** [COMPETENZE DI DOMINIO + TECNICHE]

Le basi

[DMP]



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…)

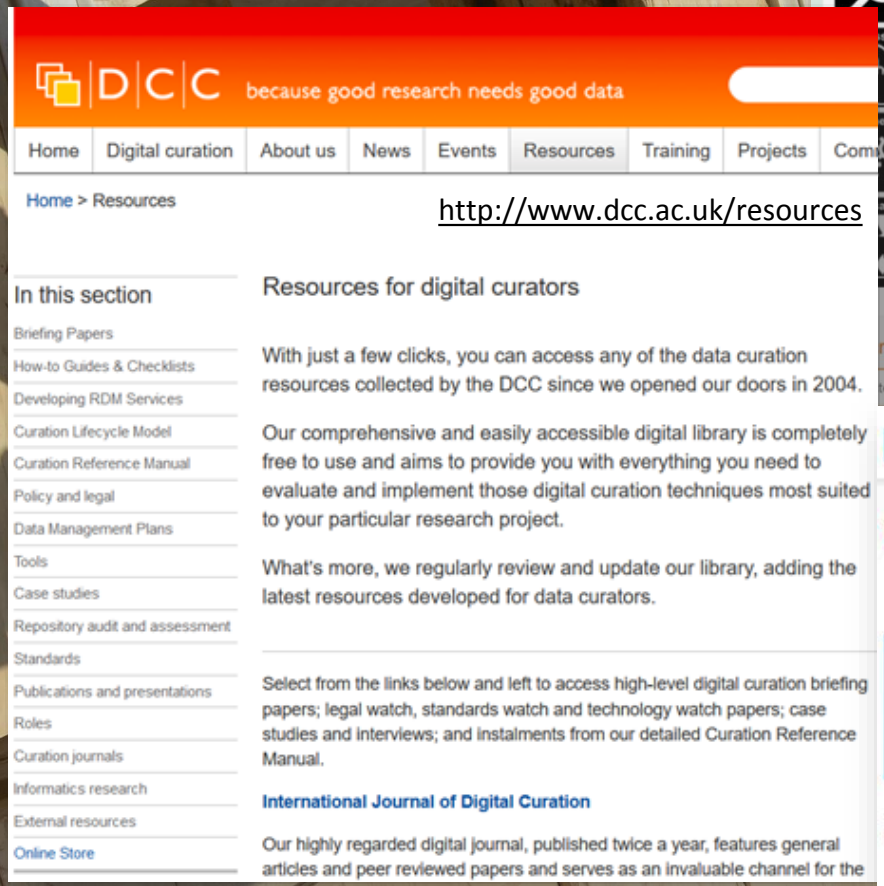
RICHIEDONO STRUMENTI E TRATTAMENTI DIVERSI

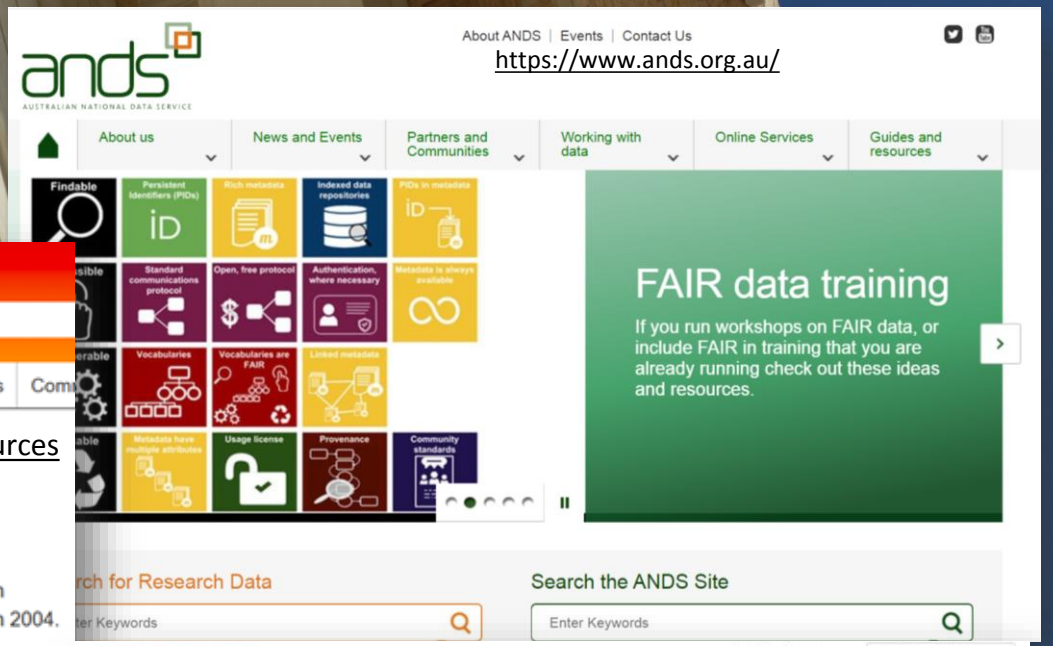
- **The way the data is collected**.
  - By experimenting, simulations, observations, derived data, reference data.
- **The data forms**.
  - For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, 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.**

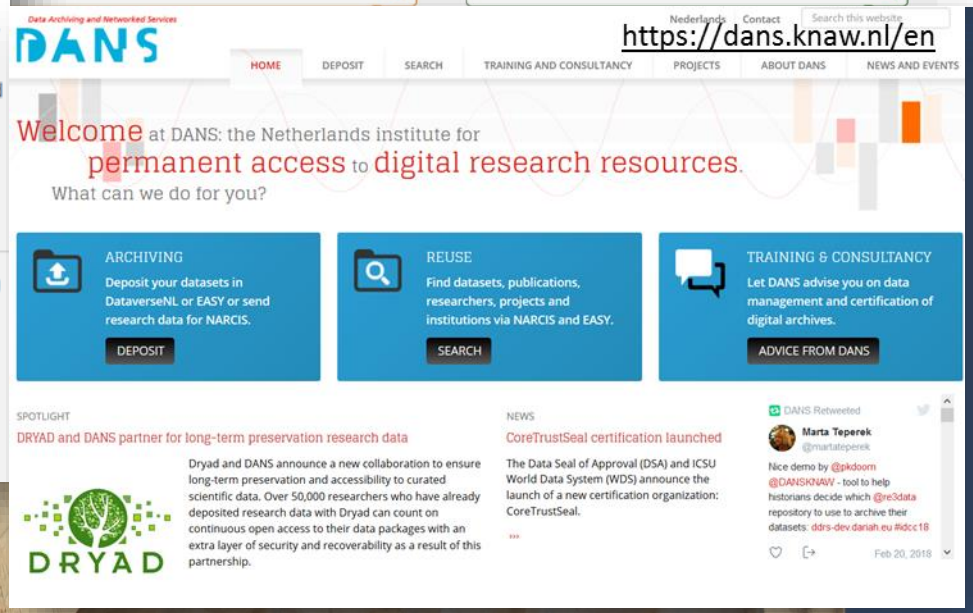
https://eprints.soton.ac.uk/403440/1/introducing_research_data.pdf

[DMP]

3 pilastri



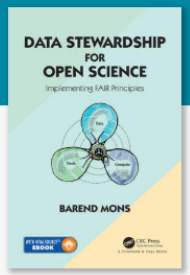




… e un maestro

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

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(+)).

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

…una via veloce

[DMP]

RDM video

escience
vidensportal

eScience | Få styr på data | Supercomputing | Træningskurser | Om os | Podcasts

eLearning course about the importance of good research data management (RDM)

Within the framework of the Danish National Forum for Data Management, the Danish Universities have developed the eLearning course "Research Data Management".

90% of the world's data was created within the last two years

Take the course

Module 1: Introduction

Reference: Vlachos, E., Larsen, A.V., Zurcher, S., Hansen, A.F. (2019). 'Introduction'. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), *Research Data Management* (eLearning course). doi: 10.11581/dtu:00000048

Module 2: FAIR principles

Reference: Martínez-Lavanchy, P.M., Hüser, F.J., Buss, M.C.H., Andersen, J.J., Begtrup, J.W. (2019). 'FAIR Principles'. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), Research Data Management (eLearning course). doi: 10.11581/dtu:00000049

Module 3: Data Management Plans

Reference: den Boer, S.P.A., Buss, M.C.H., Hüser, F.J., Smed, U. (2019). 'Data Management Plans'. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), Research Data Management (eLearning course). doi: 10.11581/dtu:00000050

[DMP]

Costi

DM costing tool

What's Zingtree?

Data Management Costing Tool

Data Management costing tool

4TU. CENTRE FOR RESEARCH DATA TU Delft

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.

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

OpenAIRE

What will it cost to manage and share my data?

What to cost in?

Infrastructure costs
- Digitisation
- Storage
- Licensing and Security
- Sharing and Re-use
- Archiving

...and

Skills costs
- Data wrangling
- Description and Documentation
- Metadata generation
- Formatting and Cleaning
- Consent and Anonymisation

A Data Management Plan (DMP) can help to identify activities and potential costs at the outset ouf your project. Identifying RDM costs before you begin the project ensures that you will be able to request adequate funds to support good data management and enable data sharing.

Things to consider...

- *Eligible costs:* When applying for funding, remember that there are typically two types of eligible costs; 'Direct costs', usually referring to staff time, travel, equipment, etc., and 'Indirect costs', generally covering things like administrative and financial management.
- *Avoid 'double dipping':* Most funders will cover justifiable costs related to RDM. However, if something is covered by indirect costs (e.g. institutional storage) you can't also claim it as a direct cost. Check with your institution on how best to include these in grant proposals.

Useful costing guides:

- OpenAIRE: How to identify and assess Research Data Management (RDM) costs
- LCRDM: Guide Research Data Management and Costs
- Horizon 2020 Costing Guide
- UK Data Service: Data management costing tool and checklist

How much could management & deposit cost?

Some factors that affect RDM costs...

Security of potenially sensitive data

Dataset size

Length of preservation required

Remember:
Different repositories apply different charging models. Some apply a fixed-fee per data package plus an amount over a certain volume, while others only apply variable fees depending on the data volume. Some may not charge at all.

Based on these examples, we have performed some comparative calculations. The cheapest repository changes at different points so shop around!

20GB of data could cost you between €0 - €109 to deposit

75GB of data could cost you between €245 - €340 to deposit

200GB of data could cost you between €790 - €906 to deposit

Developed for:

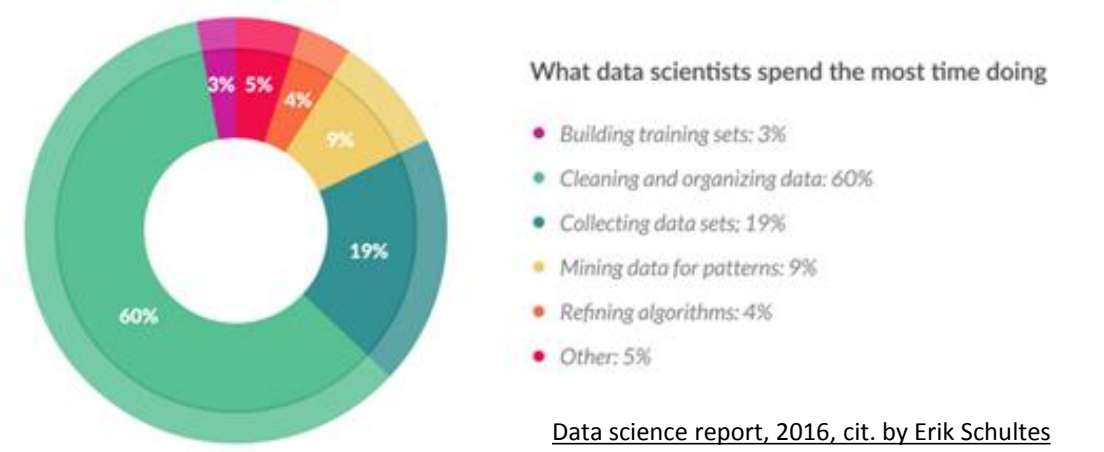
OpenAIRE by DCC

OpenAIRE-Advance receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777541.

More OpenAIRE Advance resources available here:

May 15, 2020

Costi



Data science report, 2016, cit. by Erik Schultes

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

cost of not having FAIR data, 2018

Il 5%

nature Feb. 25, 2020 Subscribe

WORLD VIEW · 25 FEBRUARY 2020

Invest 5% of research funds in ensuring data are reusable

It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons

I tell research institutions that, on average, 5% of overall research costs should go towards data stewardship. With €300 billion (US$325 billion) of public money spent on research in the European Union, we should expect to spend €15 billion on data stewardship. Scientists, especially more experienced ones, are often upset when I say this. They see it as 5% less funding for research.

Bunk. First, taking care of data is an ethical duty, and should be part of good research practice. Second, if data are treated properly, researchers will have significantly more time to do research. Consider the losses incurred under the current system. Students in PhD programmes spend up to 80% of their time on 'data munging', fixing formatting and minor mistakes to make data suitable for analysis – wasting time and talent. With 400 such students, that would amount to a monetary waste equivalent to the salaries of 200 full-time employees, at minimum. So, hiring 20 professional data stewards to cut time lost to data wrangling would boost effective research capacity. Many top universities are starting to see that the costs of not sharing data are significant and greater than the associated risks. Data stewardship offers excellent returns on investment.

- PRENDERSI CURA DEI DATI È ETICO
- ASSUMERE DATA STEWARDS FA RISPARMIARE TEMPO
- FAIR=FULLY ARTIFICIAL INTELLIGENCE READY

Funders hold the stick: they should disburse no further funding without a properly reviewed and budgeted data-stewardship plan. The carrot is that FAIR data allow much more effective artificial intelligence (FAIR can also mean 'fully AI ready'), which will open up unprecedented research opportunities and increase reproducibility.


Pausa?

Parte II

GESTIRE I DATI, RENDERLI FAIR, SE POSSIBILE OPEN


…un passo indietro…

[il fondamento] [DMP]

Information Guide: Introduction to Ownership of Rights in Research Data. CREATe, University of Glasgow, 2018

Burrow, S., Margoni, T. and McCutcheon, V. (2018) Information Guide: Introduction to Ownership of Rights in Research Data. CREATe, University of Glasgow, 2018. Documentation. University of Glasgow.

http://eprints.gla.ac.uk/171314/

OpenAIRE

Guides for Researchers

How do I know if my research data is protected?

Learn more about what is research data and their protection by intellectual property rights

OpenAIRE

I DATI NON SONO «MIEI»
NON ESISTE COPYRIGHT
PERCHÉ NON SONO CREATIVI

Wainer Lusoli
@w_lusoli

Following

repeat with me: #researchdata is NOT mine. I was paid to get it, I'll get a #nobel 4 it, but it's NOT mine linkedin.com/pulse/repeat-m … #opendata

Traduci dalla lingua originale: inglese

Repeat with me: research data is not mine

Seldom do I see something that truly shakes me at work. You know, work is work, I am no neurosurgeon, no médecin sans frontières nor am I a social

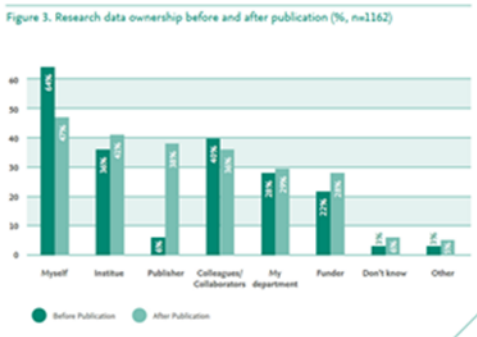
linkedin.com

11:18 - 12 apr 2017

14 Retweet **18** Mi piace

Lusoli, Apr.2017

This time though it happened. What it was: **64% of researchers believe they own the data they generated for their research.**



The result comes from a solid piece of academic research based on equally solid (open) data. The study and the report 'Open Data - the Researcher Perspective' were done by CWTS / Leiden and Elsevier. Credit giving, check.

Of course, the study reports

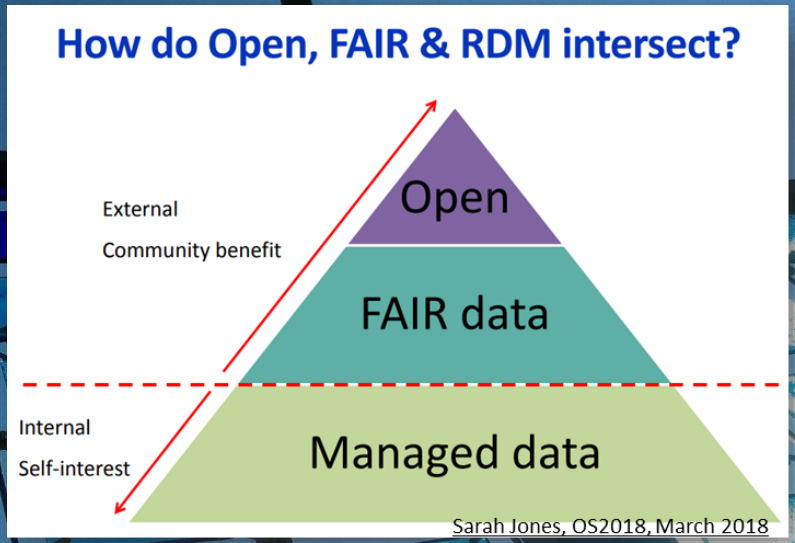
[DMP]

[webinar]

- POSSONO ESSERCI ALTRE FORME DI PROTEZIONE DEI DATI (ES. CONTRATTI)
- PER DATI CHE RICADONO SOTTO GDPR VA SEMPRE ESPLICITATA LA BASE LEGALE SULLA QUALE SI CONDUCE LA RICERCA

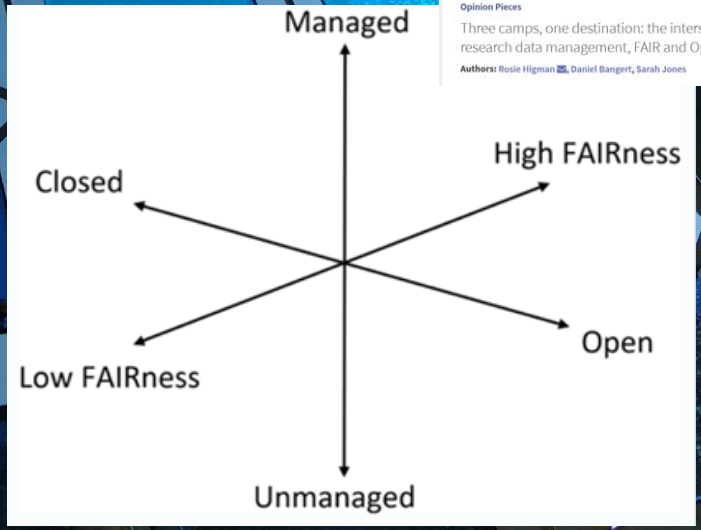
[i tre passi fondamentali]

[i tre passi fondamentali]



Sarah Jones, OS2018, March 2018

May 27, 2019



1. I dati vanno gestiti

Data management is an active process by which digital resources remain discoverable, accessible and intelligible over the longer term, a process that invests data and datasets with the potential to accrue value as assets enjoying far wider use than their creators may have anticipated. In the world of research, such a value-adding process is a significant contributor to the much desired achievement of impact.

2. I dati DEVONO essere FAIR

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE 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.

TO BE INTEROPERABLE:

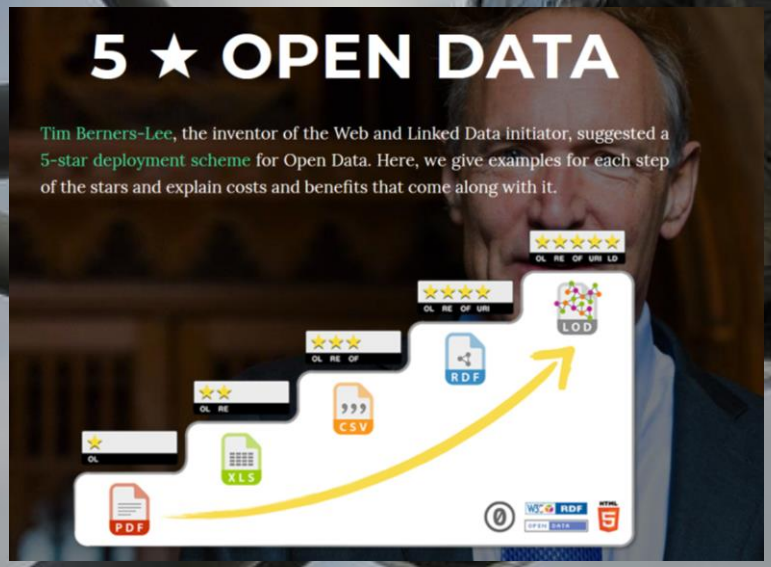
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have 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 their provenance.
R1.3. (meta)data meet domain-relevant community standards.

«ACCESSIBLE»
≠«OPEN»
= DOVE E A QUALI CONDIZIONI I DATI SONO ACCESSIBILI

3. I dati POSSONO essere Open



1. GESTIRE I DATI

DCC because good research needs good data

… dal primo giorno

Research Data Management:
Get it right from the beginning

May 2018

CC0/ this picture needs no attribution

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

PERCHÉ RISPARMIATE
ENERGIA E LA RICERCA
SCORRE FLUIDA


PERCHÉ I PRIMI RI-UTILIZZATORI DEI VOSTRI DATI SIETE VOI…
IL DMP NON È UNO SFORZO INUTILE!!!

Serve formazione?

[DMP]

cessda TRAINING

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?

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

Con un supporto pratico

[DMP]

ALLA FINE D OGNI MODULO TROVATE «ADAPT YOUR DMP» PER APPLICARE I CONCETTI CHE AVETE APPENA IMPARATO

Adapt your DMP: part 6

This is the sixth 'Adapt your DMP' section in this tour guide. To adapt your DMP, consider the following elements and corresponding questions:

⊕ Versioning

⊖ Interoperability

In order to be able to link your work to other research, it might be useful to build on established terminologies as well as commonly uses coding and soft- and hardware wherever this is possible.

- Which *software and hardware* will you use? How does this relate to other research?

If applicable:

- Will established *terminologies/ontologies* (i.e. structured controlled vocabularies) be used in the project? If not, how does yours relate to established ones?
- Which *coding* is used (if any)? How does this relate to other research?

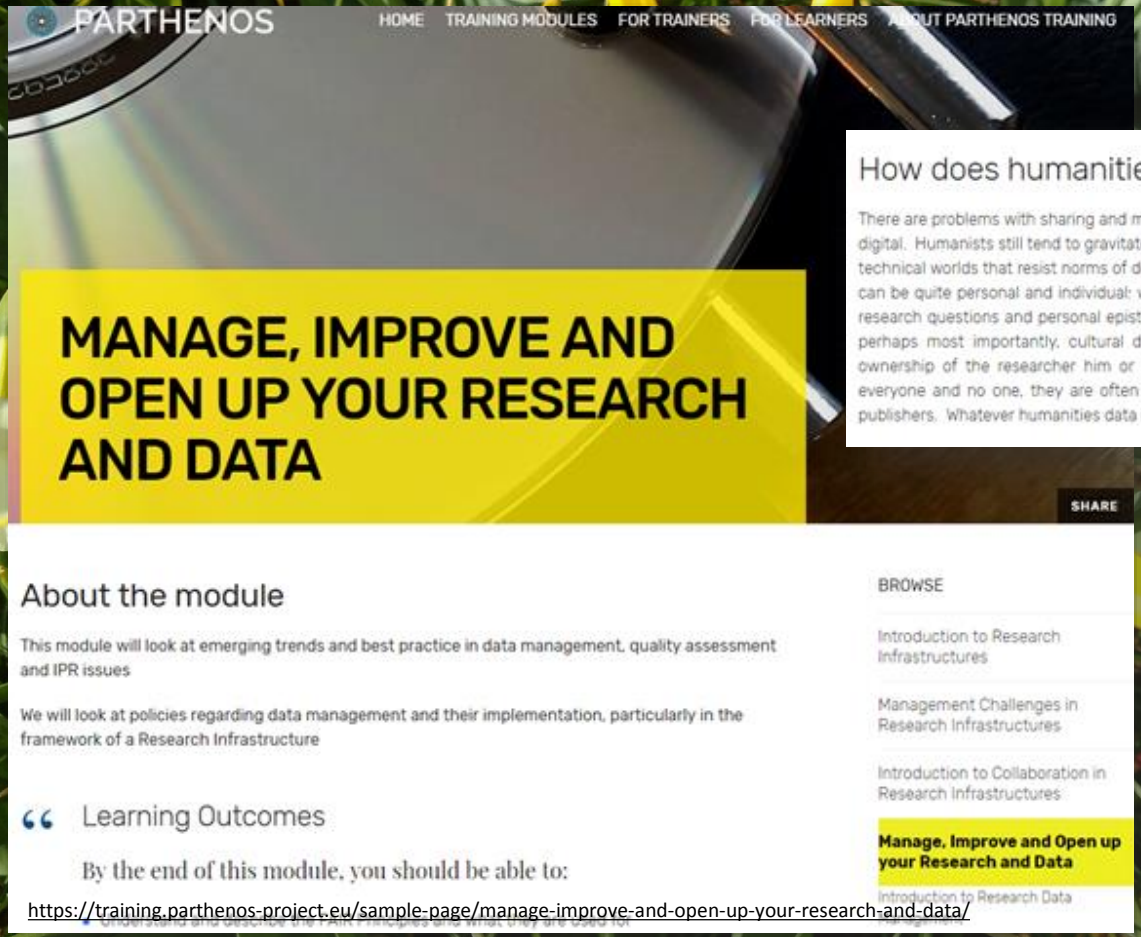
⊖ Deposit your data

- Will the data you produce and/or used in the project be useable by third parties, in particular after the end of the project?
- Which data and associated metadata, documentation and code will be deposited?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- What data quality assurance processes will you apply?

CESSDA Guide

[DMP]

…e le scienze umane?



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

How does humanities data tend to be different?

There are problems with sharing and managing the humanistic data, however. First of all, much of it is not digital. Humanists still tend to gravitate toward multimodal knowledge creation systems, hybrid digital and technical worlds that resist norms of deposit and reuse. Second, the semiotic systems of humanities data can be quite personal and individual: we prepare our sources to be useful for us, and what works for our research questions and personal epistemic instruments may not work at all for anyone else. Finally, and perhaps most importantly, cultural data is seldom if ever 'raw,' and seldom, if ever, under the sole ownership of the researcher him or herself. The records of human activity and creativity belong to everyone and no one, they are often preserved and curated by dedicated public institutions or private publishers. Whatever humanities data is, it is not simple!

E le scienza umane?

May 2019

DARIAH-CAMPUS Resources Topics Sources Course Registry About Contact

DARIAH Pathfinder to Data Management Best Practices in the Humanities

Written by Erzsébet Tóth-Czifra
May, 03 2019

Source: DARIAH Pathfinders, DARIAH
Topics: Data management

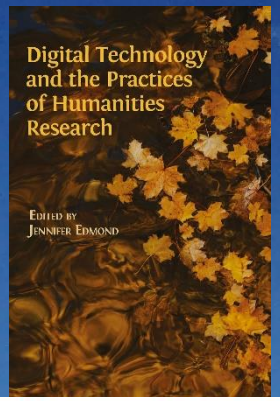
TABLE OF CONTENTS

1. Why research data management?

Systematically planning how you will collect, document, organize, manage, share and preserve your data has many benefits. It helps to build a common framework of understanding with your collaborators and other stakeholders such as data archivists or professionals of GLAM institutions. But you can also think of your future self as your primary collaborator, imagining yourself looking for

E le scienze umane?

Edmond, 2020



10. THE RISK OF LOSING THE *THICK DESCRIPTION*: DATA MANAGEMENT CHALLENGES FACED BY THE ARTS AND HUMANITIES IN THE EVOLVING FAIR DATA ECOSYSTEM 235

Erzsébet Tóth-Czifra

Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | **February 2020**

RECOMMENDATIONS

» Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.

» Familiarise yourself with the FAIR Data Principles before you start collecting data and building corpora e.g. FORCE11: the FAIR Data Principles, GO-FAIR: FAIR Data Principles and discuss with colleagues and experts to build a better understanding.

» Digitally document all your research and data collection work -- at the beginning of a project it is difficult to judge which information of the research process will be important and valuable later on.

» Use well-established tools to facilitate your research work, as many of them allow data sharing e.g. MIT Libraries Digital Humanities: Tools and Resource Recommendations.

» Browse humanities datasets and consider whether your own assets could be published in a similar fashion (e.g. Humanities Commons, UK Data Archive, ARCHE re3data.org filtered for humanities).

» When you start producing data, keep this maxim of Open Science in mind: data should be 'as open as possible and as closed as necessary'.

Data management ABC – File naming

[DMP]

SCEGLIERNE UNA… ED ESSERE CONSISTENTI!

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, 'attch', 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.

https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions

[DMP]

Data management ABC – File naming

CESSDA training

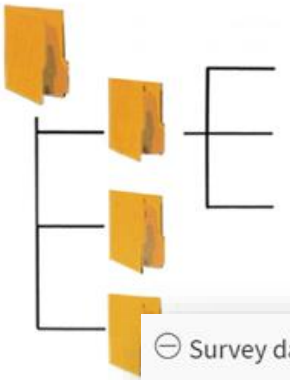
Folder structure

Structuring your data files in folders is important for making it easier to locate and organise files and versions. A proper folder structure is especially needed when collaborating with others.

The decision on how to organise your data files depends on the plan and organisation of the study. All material relevant to the data should be entered into the data folders, including detailed information on the data collection and data processing procedures.

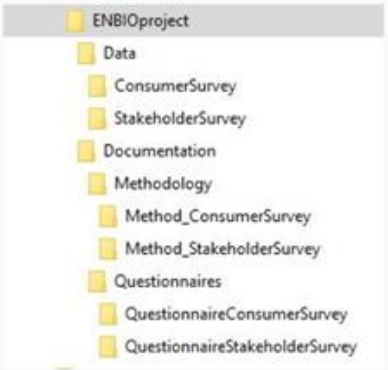
Consider the best hierarchy of your files and decide whether a deep or shallow hierarchy is preferable. If you have several independent data collections, it is advisable to create a separate data folder for each

inspiration, have a look at the examples in the accordion below.



Survey data

For this survey, data and documentation files are held in separate folders. Data files are to data type and then according to research activity. Documentation files are organised documentation file and research activity. It helps to restrict the level of folders to three more than ten items on each list.



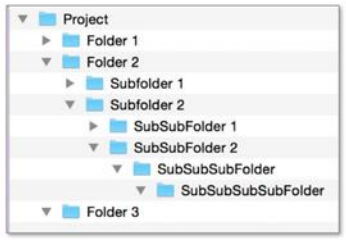
Data management ABC – Readme file

[DMP]

Sample_README_fileOrg.docx

Folder structure:

Sketch out here or insert a screenshot of your folder structure. Note, if including a screenshot, expand all folders to show the full hierarchy.



File naming schema:

File type: *Microscope image*
Filename schema: *[date]_[microscope]_[imageNumber]*
Schema key: *date: date of image capture in YYYYMMDD format*
microscope: name/model of microscope used
imageNumber: written in sequential formatting 00X - XXX
Example filename: *20180118_mic53_001.jpg*

MIT Libraries

README: File & Folder Schema (Example)

*This document is for recording your file-naming schemas and folder structures developed in the **Naming and organizing your files and folders worksheet**. This example README includes descriptions and examples for your guidance. See the **README: File & Folder Schema (Template)** for a blank version.*

For guidance on creating readmes to document information on datasets, see: Guide to writing "readme" style metadata. Cornell Research Data Management Service Group.
https://data.research.cornell.edu/content/readme

Overview:

Project/Lab Name: *Name the project for which this file organization documentation refers. If it documents the organization schema for a research/lab group, include that here.*
Ex: Our Lab, Project 123

Creator: *Who created the file organization schema? This is important information as a user may need to get clarification, suggest a revision of the schema, etc. Include the institution/address/email for contacting this person.*

MIT Libraries

README: File & Folder Schema (Example)

File type	Filename schema	Schema key	Example filename
Microscope image	*[Date]_[microscope]_[image Number]*	*Date: Date of image capture in YYYYMMDD format* *microscope: name of microscope used* *imageNumber: written in sequential formatting 00X*	*20180118_mic53_001.jpg*

Filename abbreviations

Use this section to document any abbreviations used in the file-naming schemes described above.

Filename descriptor	Abbreviations key
Ex: Location	*ATL: Atlanta* *BOS: Boston*
Ex: Microscope (name)	*mic53: microscope 53, located in room 1...*

MIT data management

Data management A
Readme file

[DMP]

1. *Introductory information*
   - **Title of the dataset**
   - **For each file or group of similar files, a short description of what data it contains**
   - Explain the file naming convention, if applicable
   - Format of the file if not obvious from the file name
   - If the data set includes multiple files that relate to each other, the relationship between the files or a description of the file structure that holds them
   - Contact information; in case users have questions regarding the data files

2. *Methodological information*
   - **Method description for collecting or generating the data, as well as the methods for processing data, if data other than raw data are being contributed**
   - Any instrument-specific information needed to understand or interpret the data
   - Software (including version number) used to produce, prepare, rende compress, analyze and/or needed to read the dataset, if applicable
   - Standards and calibration information, if appropriate

Guidelines for README

4TU. CENTRE FOR RESEARCH DATA

Guidelines for creating a README file

A readme file provides information about a dataset and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data.
A readme file must be submitted along with the dataset file(s).
The outline below should be completed with information relevant to the submitted dataset.

Best practices

- **Create one readme file for each dataset**
- **Name the file README;** not readme, read_me, ABOUT, etc.
- **Write your readme document as a plain text file;** save as README.txt or README.md when writing in Markdown. Or use README.pdf when text formatting is important for your file.

3. *Data specific information*
   - **Full names and definitions (spell out abbreviated words) of column headings for tabular data**
   - **Units of measurement**
   - **Definitions for codes or symbols used to record missing data**
   - **Specialized formats or abbreviations used**

4. *Sharing and Access information*
   - Licenses or restrictions placed on the data;
     Licenses allow you to specify the 'terms-of-use' for your data. The archive provides a license that is explained in its terms of use and applies this license as default selection. You can use this licensing wizard to help you to pick a more appropriate license for the use of your data. This license will then be displayed in the metadata.

Data management ABC – Versioning

[DMP]

Data versioning

ands AUSTRALIAN NATIONAL DATA SERVICE

About us

Working with data

SHARE

What do we mean by the term ‘data versioning’?

A version is “a particular form of something differing in certain respects from an earlier form or other forms of the same type of thing □”. In the research environment, we often think of versions as they pertain to resources such as manuscripts, software or data. We may regard a new version to be created when there is a change in the structure, contents, or condition of the resource.

In the case of research data, a new version of a dataset may be created when an existing dataset is reprocessed, corrected or appended with additional data. Versioning is one means by which to track changes associated with ‘dynamic’ data that is not static over time.

Why is data versioning important?

Increasingly, researchers are required to cite and iden
to support research reproducibility and trustworthiness
accurately indicate exactly which version of a dataset
particularly challenging where the data to be cited are
accessed via a web service.

Unlike the software domain, the data community doesn't yet have a standard numbering system. Three representative data version numbering patterns in use include:

- Numbering system 1
- Numbering system 2
- Numbering system 3

What tools are available for data versioning?

There is no one-size-fit-all solution for data versioning and tracking changes. Data come in different forms and are managed by different tools and methods. In principle, data managers should take advantage of data management tools that support versioning and track changes.

Example approaches include:

Git (and Github) for Data □ (with size <10Mb or 100k rows) which allows:

- effective distributed collaboration – you can take my dataset, make changes, and share those back with me (and different people can do this at once)
- provenance tracking (i.e. what changes came from where)
- sharing of updates and synchronizing datasets in a simple, effective, way.

Data versioning at ArcGIS □

- Users of ArcGIS can create a geodatabase version, derived from an existing version. When you create a version, you specify its name, an optional description, and the level of access other users have to the version. As the owner of the version, you can change these properties or delete a version at any time

Numbering system 1

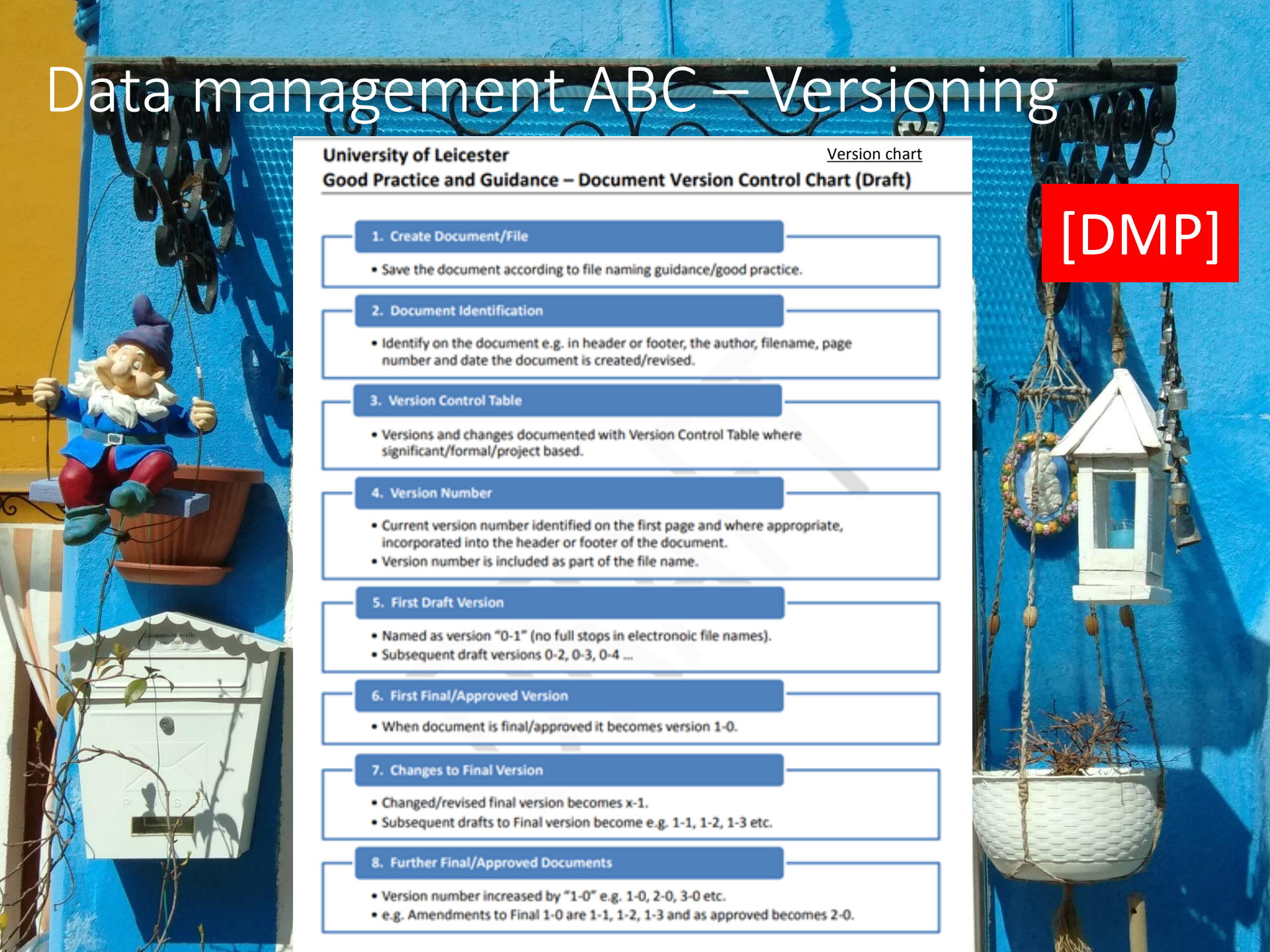
Data versioning follows a similar path to software versioning, usually applying a two-part numbering rule: Major.Minor (e.g. V2.1). Major data revision indicates a change in the formation and/or content of the dataset that may bring changes in scope, context or intended use. For example, a major revision may increase or decrease the statistical power of a collection, require change of data access interfaces, or enable or disable answering of more or less research questions. A Major revision may incorporate:

- substantial new data items added to /deleted from a collection
- data values changed because temporal and/or spatial baseline changes
- additional data attributes introduced
- changes in a data generation model
- format of data items a changed
- major changes in upstream datasets.

Minor revisions often involve quality improvement over existing data items. These changes may not affect the scope or intended use of initial collection. A Minor revision may include:

- renaming of data attribute
- correction of errors in existing data
- re-running a data generation model with adjustment of some parameters
- minor changes in upstream datasets.

Data management ABC – Versioning



University of Leicester Version chart

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 electronoic 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.

[DMP]

Data management ABC – Versioning

[DMP]

cessda TRAINING

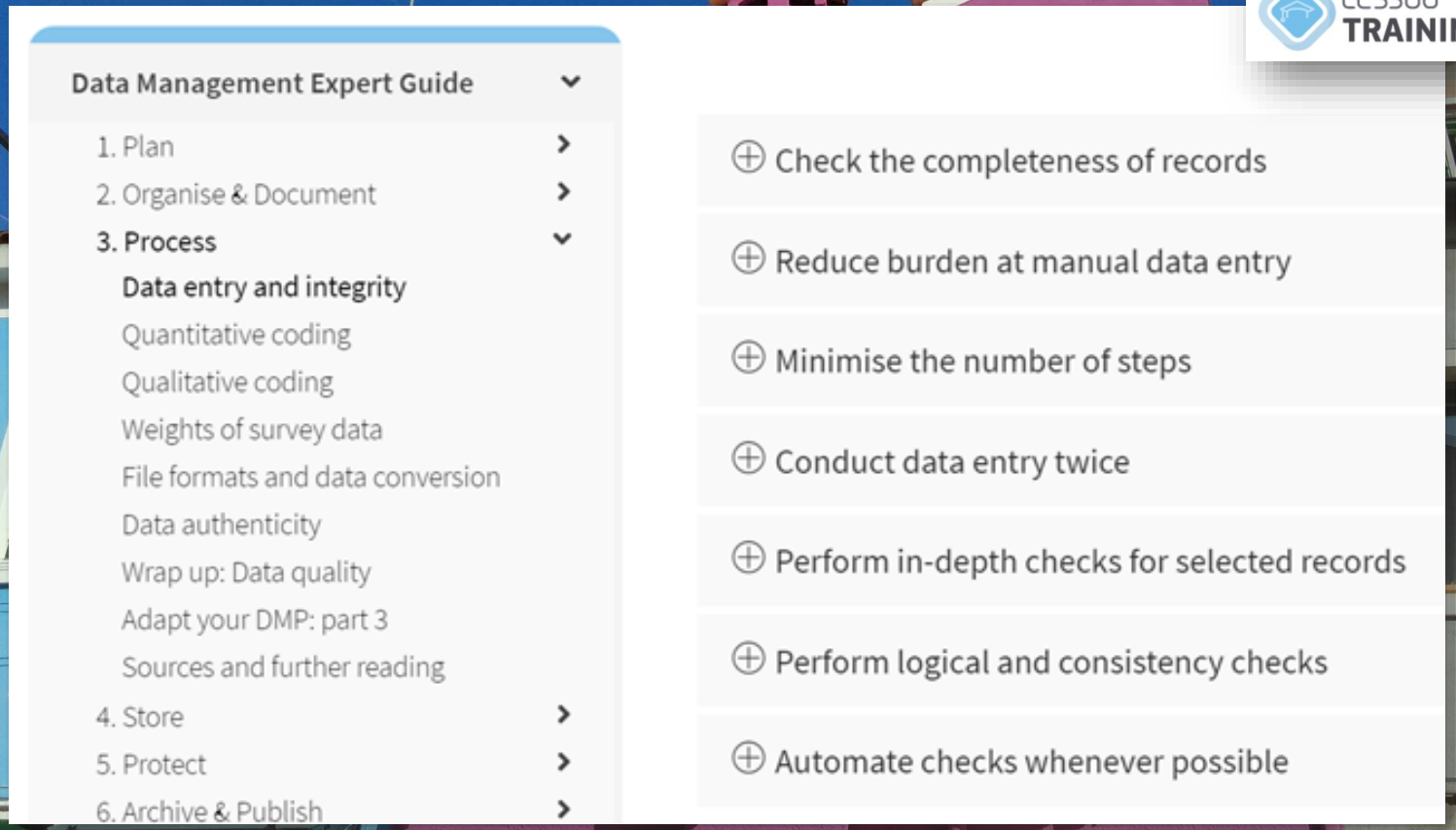
Version control

Version control can be done through:

- Uniquely identifying different versions of files using a systematic naming convention, such as using version numbers or dates (date format should be YYYY-MM-DD, see 'File naming');
  - Record the date within the file, for example, 20010911_Video_Twintowers;
  - Process the version numbering into the file name, for example, HealthTest-00-02 or HealthTest_v2;
  - **Don't** use ambiguous descriptions for the version you are working on. Who will know whether MyThesisFinal.doc, MyThesisLastOne.doc or another file is really the final version?
- Using version control facilities within the software you use;
- Using versioning software like Subversion (2017);
- Using file-sharing services with incorporated version control (but remember that using commercial cloud services as the Google cloud platform, Dropbox or iCloud comes with specific rules set by the provider of these services. Private companies have their own terms of use which applies for example to copyrights);
- Designing and using a version control table. In all cases, a file history table should be included within a file. In this file, you can keep track of versions and details of the changes which were made. Click on the tab to have a look at an example which was taken from the UK Data Service (2017c).

CESSDA training

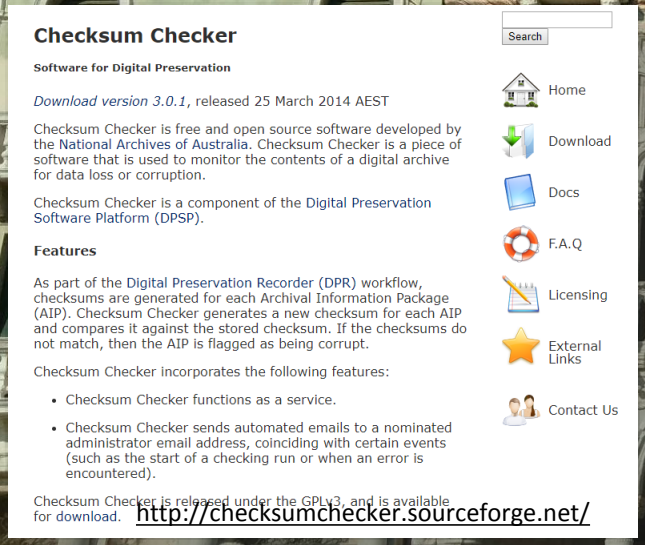
Data management ABC – Data entry



CESSDA Guide

Data Management ABC - conservazione

LUNGO O BREVE TERMINE?



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.

Search

Home

Download

Docs

F.A.Q

Licensing

External Links

Contact Us

http://checksumchecker.sourceforge.net/

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

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).

Data Management ABC- backup and storage

Portable devices | Cloud storage | Local storage | Networked driv

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

Advantages	Disadvantages/Risks	Pre (ser data
• 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.	• 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.	Use enc pass

Advantages	Disadvantages/Risks	Precautions for (sensitive) personal data
• Automatic backups. • Often automatic version control.	• 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.	• 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 strumenti diversi per esigenze diverse (durante/al termine). Durante, dovete anche poterci lavorare con il team

Cosa conservare?

DCC because good research needs good data

Home | Digital curation | About us | News | Events | Resources | Training | Projects

Home > Resources > How Guides > How Develop Rdm Services

How to develop

n this section

How to Develop RDM Services - a guide for HEIs

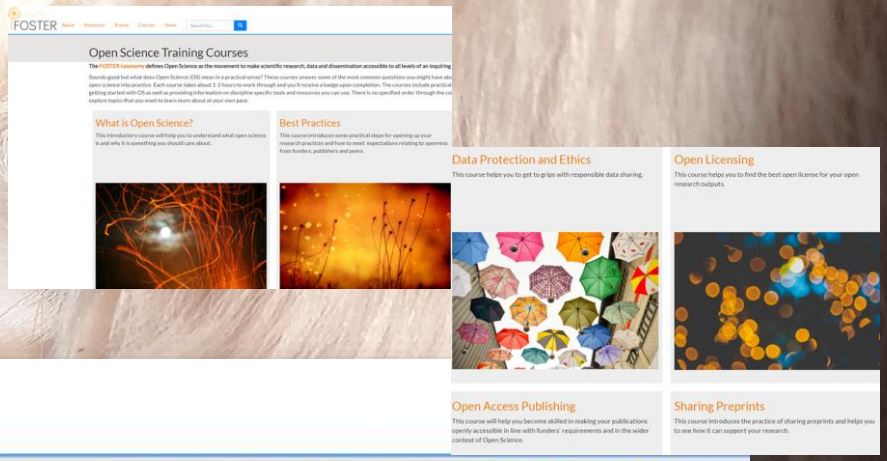
Establishing criteria for selection decisions

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

- Rilevanti per la missione dell'ente
- Valore storico
- Unicità
- Potenziale di riuso
- Non replicabili
- Costo/benefici
- Documentazione completa

Imparare a gestire



FOSTER About Resources Events Courses News Search for...

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

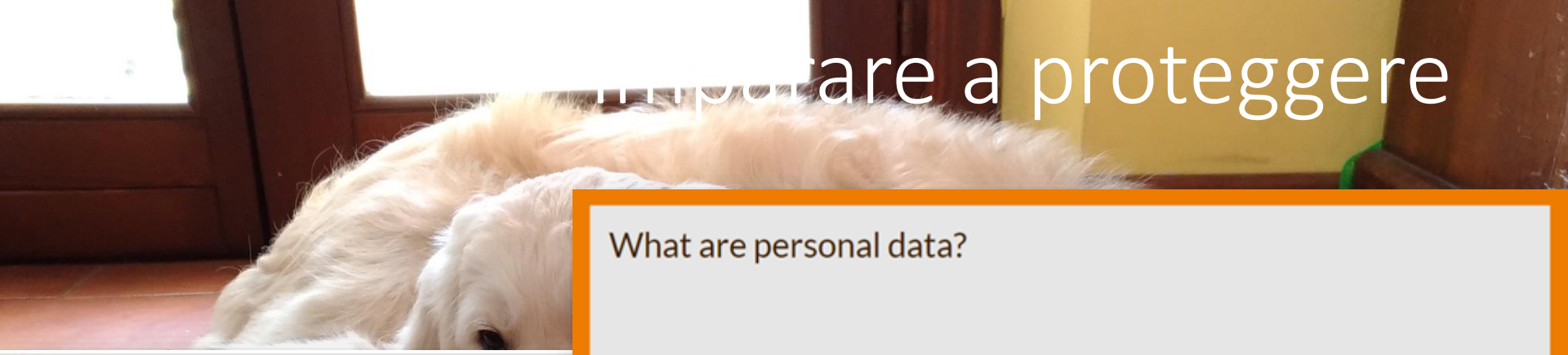
Full details

Level of knowledge: Introductory: no previous knowledge is required

Topics

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

Imparare 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

FOSTER About Resources Events Courses News

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:

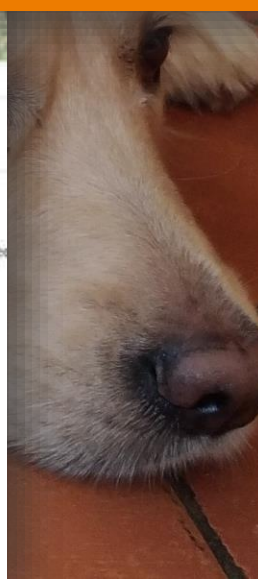
- what personal data are and how you can protect them
- what to consider when developing consent forms
- how to store your data securely
- how to anonymise your data

Start the Free Course

Full details

Level of knowledge: Introductory: no previous knowle
is required

Topics



[dati personali]

Legal Basis

Personal data can only be processed when there is a valid legal basis to do so. The GDPR recognises six bases (grounds):

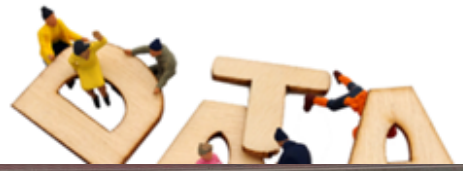
- consent of the data subject
- necessary for the performance of a contract
- legal obligation placed upon the data controller
- necessary to protect the vital interests of the data subject
- carried out in the public interest or in the exercise of official authority (public task)
- legitimate interest pursued by the data controller

The research exemption

The GDPR contains an exemption which entails that some of the principles above are slightly different when you collect and process personal data for research purposes. This is called the 'research exemption'.

> *Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subjected to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner | General Data Protection Regulation, Article 89.*

In practice, this means that Principle II. and V. are less strict. Further processing of personal data for the purposes of archiving, scientific or historical research purposes and statistical purposes is not



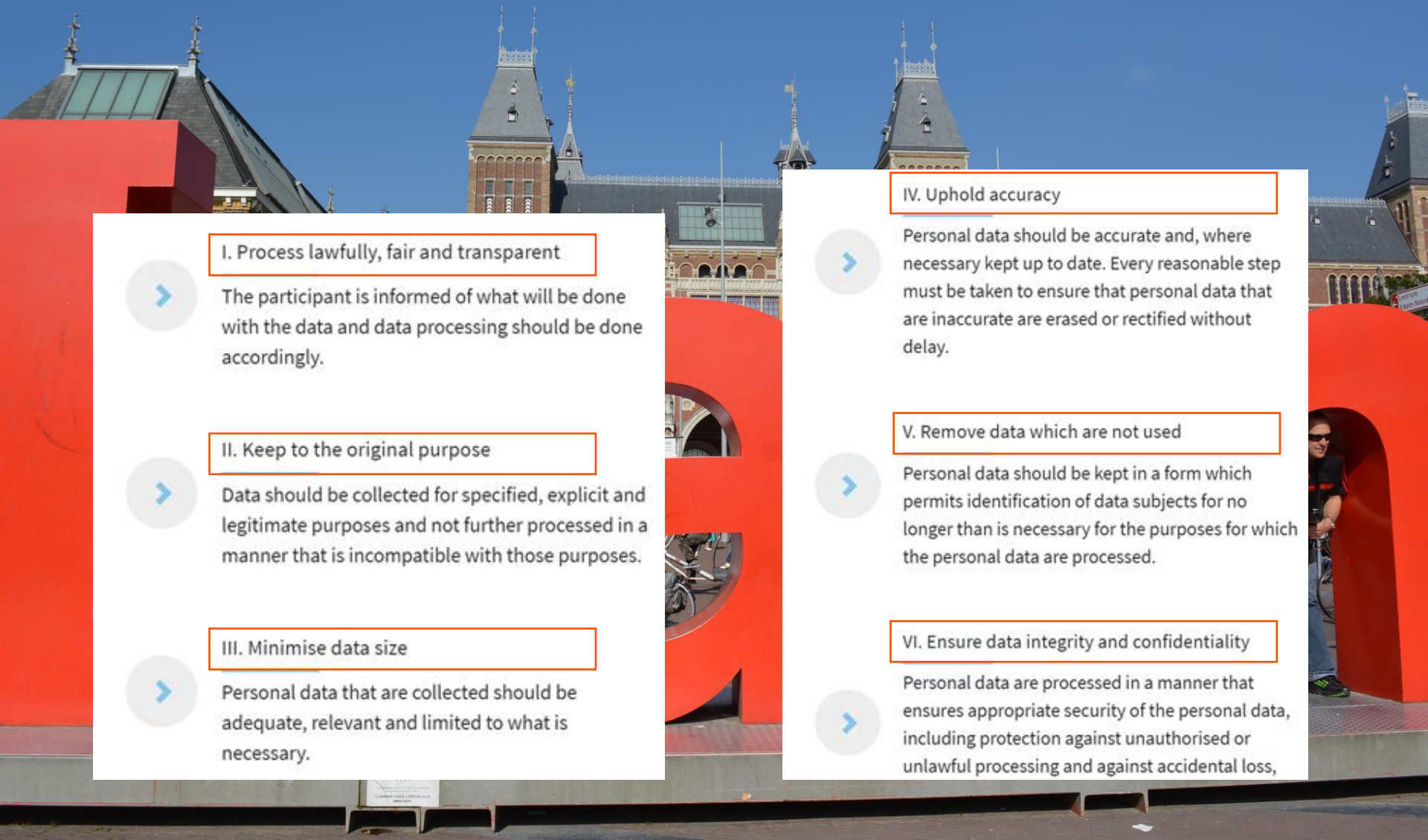
ART. 89 ECCEZIONI PER LA RICERCA MA SEMPRE SU UNA BASE LEGALE

cessda TRAINING

CESSDA guide
Data Management Expert Guide

[personal data]

CESSDA guide
Data Management Expert Guide



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,

[leggi applicabili]

Science Europe 2018

Privacy

- **Personal Data Protection Acts** are present in all European countries and concern general laws regulating the protection of personal data. They are based on European Directive 95/46/EC.[9] This Directive will be replaced in the near future by the General Data Protection Regulation (GDPR),[10] which all EU Member States will have to implement in their national legislation by May 2018.
- **Obligations to Report Data Leakage Acts** are additions to the Personal Data Protection Acts. They deal with the publication of personal data and contain sanctions in the form of penalties.
- **Medical Treatment Agreement Acts** regulate the use and preservation of personal (patient) data in and for medical research.
- **Scientific Medical Research with Humans Acts** regulate scientific research in the medical field, in particular how to handle personal health-related data. These make ethical reviews compulsory for all medical research projects.

Intellectual Property Rights

- **Copyright Acts** regulate the rights of the creator of a work. One distinguishes between exploitation rights and personal intellectual rights ('moral rights').
- The **Database Rights Act** recognises the investments made in creating and/or compiling a database. It is based on European Directive 96/9/EC.[11]
- **Related Rights Acts** or **Neighbouring Rights Acts** mostly refer to the rights of performers, phonogram producers, and broadcasting organisations.
- **Patent Acts** are for the protection of patents. Publication of research results (including data) is restricted during the application stage of a patent.

Public data

- **Public Records Acts** (Public Archives Acts) oblige all public administration offices and services to preserve their documents and transfer these, after appraisal and selection, to public archives.
- **Public Sector Information Acts** (concerning re-usability of public data) are based on European Directive 2013/37/EU[12] that focuses on the economic aspects of the re-use of public information. It encourages Member States to make as much of this information as possible available for re-use. This also covers content held by museums, libraries, and archives, but does not apply

- **Freedom of information Acts** regulate and enable citizen access to documents held by public authorities or companies carrying out work for a public authority. They do not specifically deal with access to research data.
- **Heritage Acts** are relevant for archaeological research data in so far as that they regulate ownership of documentation (data) from archaeological excavations.
- **Statistical Information Acts** regulate the competencies of the statistics authorities in data gathering as well in access to data.
- **Land Registry Acts** (cadastral information) regulate the competencies of the national land registries and access to their data, with special provisions concerning personal data contained in their various databases.

Codes of Conduct/Ethical Issues

- **Codes of Conduct**, where these exist on a national level or in an institution, should be taken into account in DMPs. They contain the general principles of good academic teaching and research.
- **Codes of Practice** for the use of personal data in scientific and scholarly research are based on the Personal Data Protection Acts[13] and prescribe how to handle personal data in research practice.
- **Codes of Conduct** for Medical Research regulate how researchers should handle medical personal data. They may be based on Medical Treatment Agreement Acts.

[Dati e GDPR]

Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

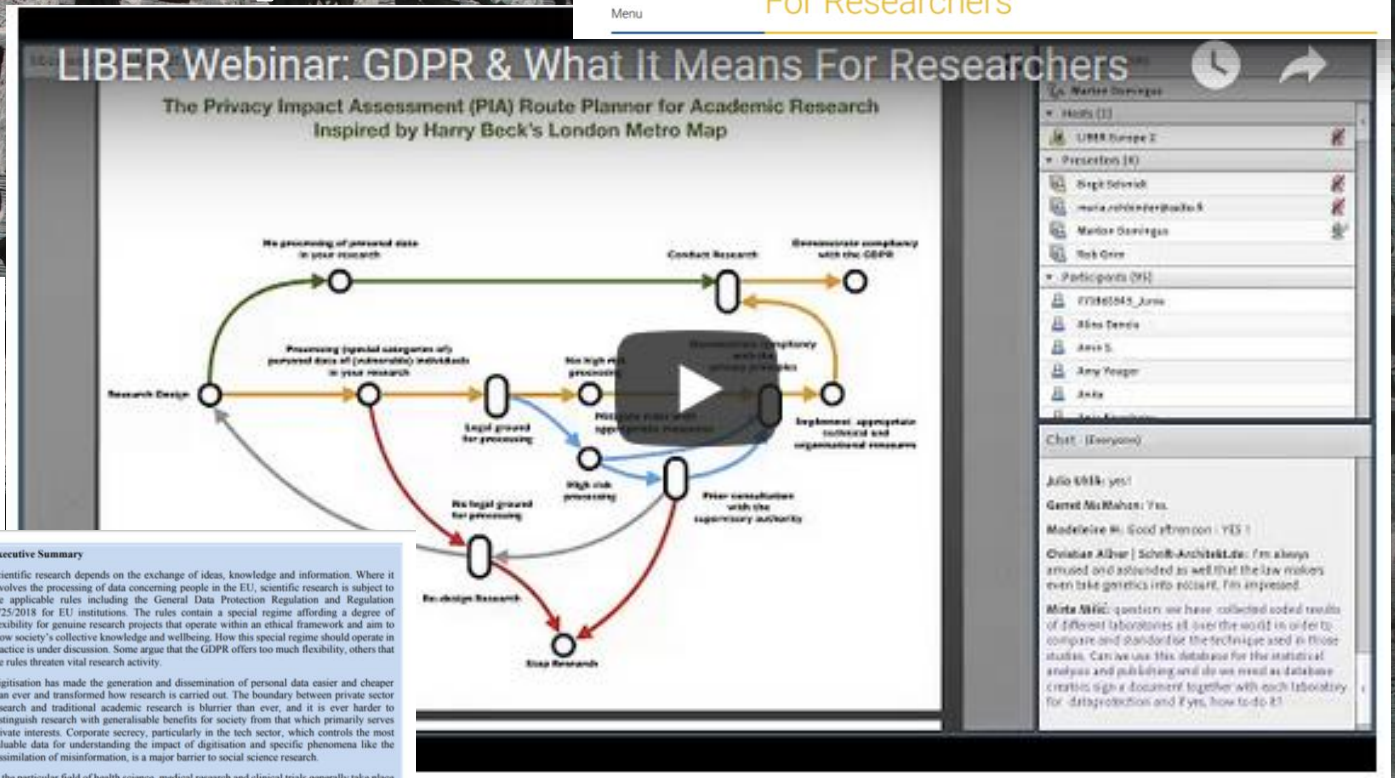
About Us | Network | Strategy | Login

Scholarly Communication | Digital Skills & Services | Research Infrastructure | Events

May 22 2018

Webinar Video: GDPR & What It Means For Researchers

Menu



EUROPEAN DATA PROTECTION SUPERVISOR

A Preliminary Opinion on data protection and scientific research

Jan 6, 2020

Executive Summary

Scientific research depends on the exchange of ideas, knowledge and information. Where it involves the processing of data concerning people in the EU, scientific research is subject to the applicable rules including the General Data Protection Regulation and Regulation 1725/2018 for EU institutions. The rules contain a special regime affording a degree of flexibility for genuine research projects that operate within an ethical framework and aim to grow society's collective knowledge and wellbeing. How this special regime should operate in practice is under discussion. Some argue that the GDPR offers too much flexibility, others that the rules threaten vital research activity.

Digitisation has made the generation and dissemination of personal data easier and cheaper than ever and transformed how research is carried out. The boundary between private sector research and traditional academic research is blurrier than ever, and it is ever harder to distinguish research with generalisable benefits for society from that which primarily serves private interests. Corporate secrecy, particularly in the tech sector, which controls the most valuable data for understanding the impact of digitisation and specific phenomena like the dissimilation of misinformation, is a major barrier to social science research.

In the particular field of health science, medical research and clinical trials generally take place within an established framework of professional ethical standards. The interaction between this framework and the GDPR is being discussed within the European Data Protection Board.

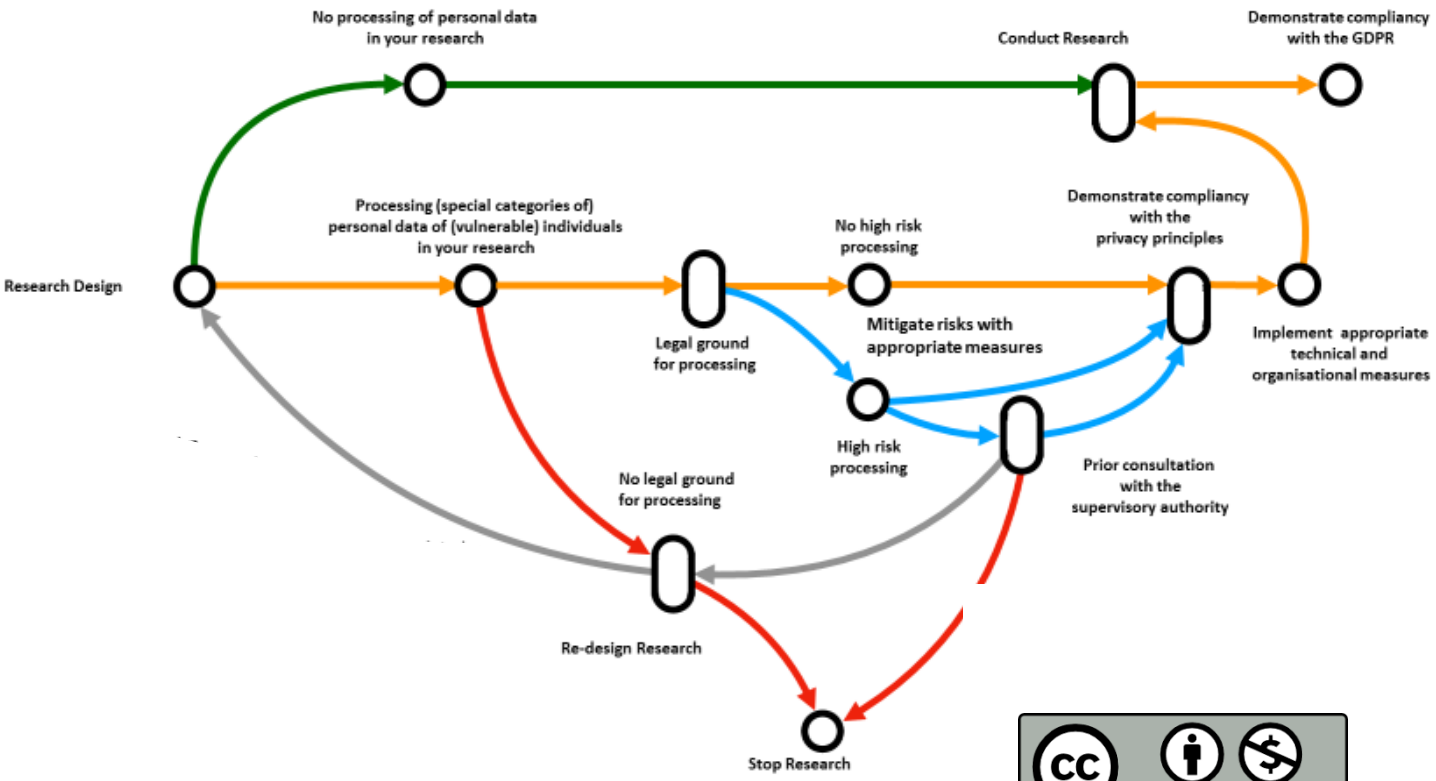
The special regime applies the usual principles such as lawfulness, purpose limitation and data subject rights, but permits some derogations from controller obligations. This includes the presumption of compatibility of processing for scientific research purposes of data collected in commercial and other contexts, provided appropriate safeguards are in place. This flexibility is afforded on the assumption that research occurring within a framework of ethical oversight serves, in principle, the public interest. The accountability principle therefore key, as it requires controllers to assess honestly and manage responsibly the risks inherent in their research projects. Such risks can be very high where, for example, processing sensitive data on health or political or religious views. Consent as a legal basis for processing must be freely-given, specific, informed and unambiguous. This differs conceptually and operationally from 'informed consent' of human participants in research. Such 'informed consent' may still serve as a safeguard in cases where consent is not appropriate as a data processing legal basis.

Scientific research serves a valuable function in a democratic society to hold powerful players to account, and this has grown in importance with the concentration of control over information flows in the hands of a few private global companies. Data protection obligations should not be misappropriated as a means for powerful players to escape transparency and accountability. Researchers operating within ethical governance frameworks should therefore be able to access necessary API and other data, with a valid legal basis and subject to the principle of proportionality and appropriate safeguards.

We recommend intensifying dialogue between data protection authorities and ethical review boards for a common understanding of which activities qualify as genuine research, EU codes of conduct for scientific research, closer alignment between EU research framework programmes and data protection standards, and the beginning of a debate on the circumstances in which access by researchers to data held by private companies can be based on public interest.

[Data and 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

https://surfdrive.surf.nl/files/index.php/s/BPRxchnZ44NZAgW

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 → **Q2. What is the legal ground for this processing?**

NO
Proceed - no measures required for safeguardingprivacy.

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.

YES → **Q3. Is this processing a high risk processing?**

NO
Stop research or redefine research.

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.

"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.

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?

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.

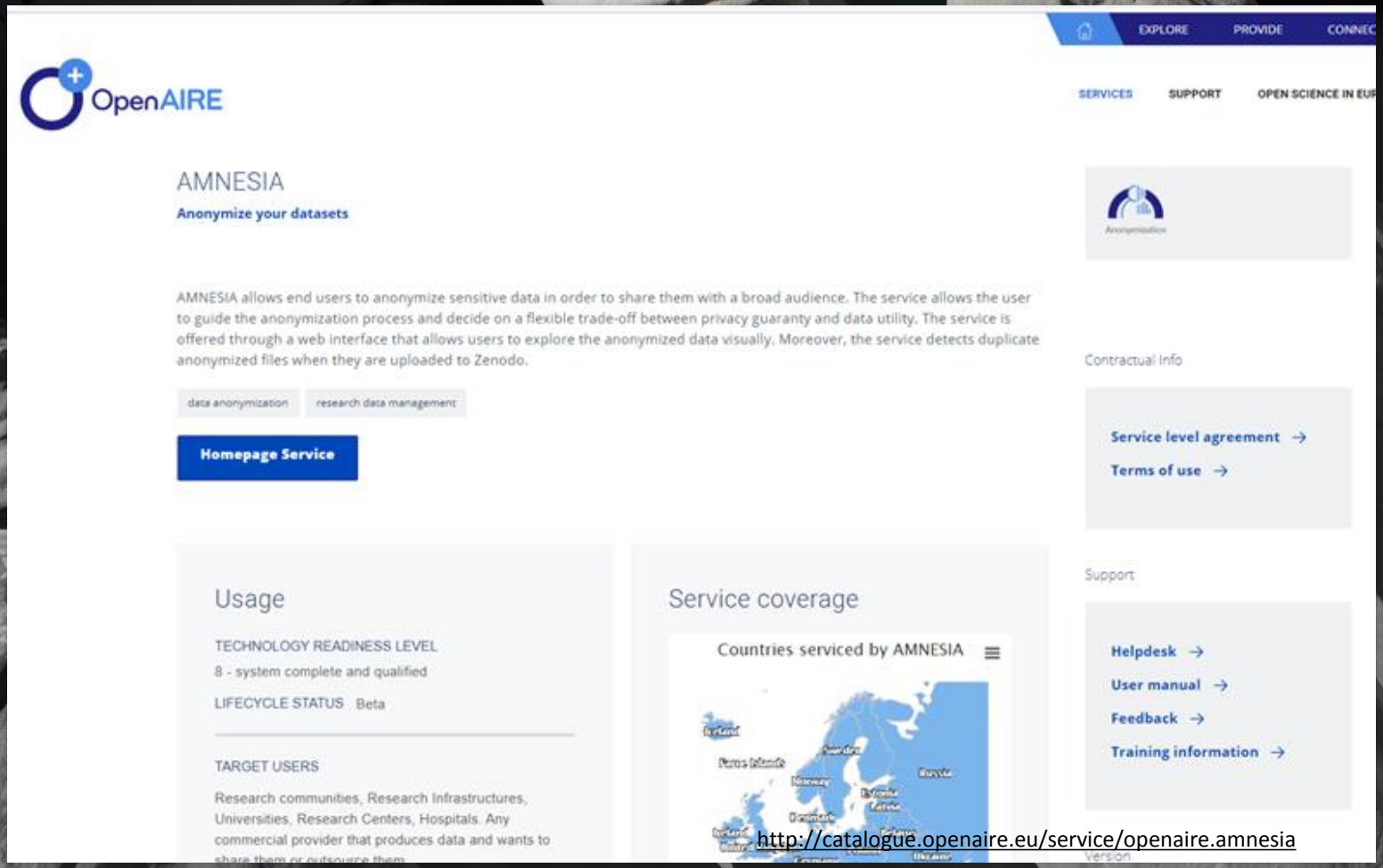
Action

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.

* 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

[anonimizzare]



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

DATI FAIR

…FAIR SIGNIFICA
[anche e soprattutto per le macchine]


SCIENTIFIC DATA

We'd like to understand how you use our websites in order to imp

Open Access | Published

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

FAIR guide, Nature, March 2016

FINDABLE

- IDENTIFICATIVI
- METADATI

ACCESSIBLE

- DOVE SONO CONSERVATI E A QUALI CONDIZIONI DI ACCESSO
- **NON SIGNIFICA «OPEN»**
- FORMATI APERTI

INTEROPERABLE

- STANDARDS
- ONTOLOGIE

REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE
- LEGGIBILI DALLE MACCHINE

FAIR in a nuts

FAIR data training
- Findable
- Accessible
- Interoperable
- Reusable

FAIR for Developers

FAIR data self-assessment tool

+SHARE

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

There are many resources created by the ARDC on the topic of metadata

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different leve

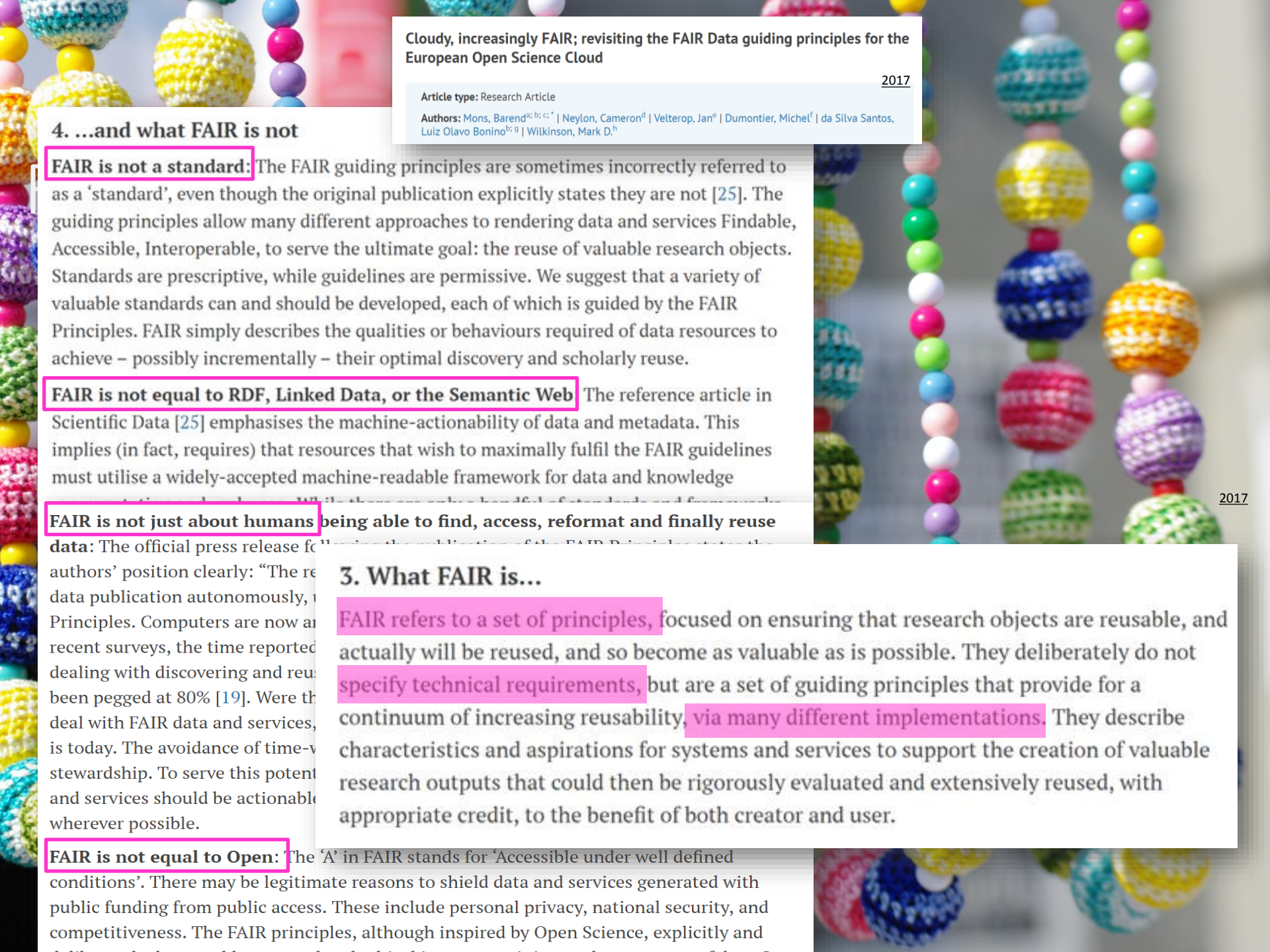
- Persistent identifiers: awareness level
- Persistent identifiers: working level
- Persistent identifiers: expert level

It is also a provider of services for minting persistent identifiers of many d 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

https://www.ands.org.au/working-with-data/fairdata/training



Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud

2017

Article type: Research Article

Authors: Mons, Barend[a; b; c; *] | Neylon, Cameron[d] | Velterop, Jan[e] | Dumontier, Michel[f] | da Silva Santos, Luiz Olavo Bonino[b; g] | Wilkinson, Mark D.[h]

4. ...and what FAIR is not

FAIR is not a standard: The FAIR guiding principles are sometimes incorrectly referred to as a 'standard', even though the original publication explicitly states they are not [25]. The guiding principles allow many different approaches to rendering data and services Findable, Accessible, Interoperable, to serve the ultimate goal: the reuse of valuable research objects. Standards are prescriptive, while guidelines are permissive. We suggest that a variety of valuable standards can and should be developed, each of which is guided by the FAIR Principles. FAIR simply describes the qualities or behaviours required of data resources to achieve – possibly incrementally – their optimal discovery and scholarly reuse.

FAIR is not equal to RDF, Linked Data, or the Semantic Web The reference article in Scientific Data [25] emphasises the machine-actionability of data and metadata. This implies (in fact, requires) that resources that wish to maximally fulfil the FAIR guidelines must utilise a widely-accepted machine-readable framework for data and knowledge

FAIR is not just about humans being able to find, access, reformat and finally reuse data: The official press release fo
authors' position clearly: "The re
data publication autonomously,
Principles. Computers are now a
recent surveys, the time reported
dealing with discovering and reu
been pegged at 80% [19]. Were th
deal with FAIR data and services,
is today. The avoidance of time-
stewardship. To serve this potent
and services should be actionable
wherever possible.

FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired by Open Science, explicitly and

2017

3. What FAIR is...

FAIR refers to a set of principles, focused on ensuring that research objects are reusable, and actually will be reused, and so become as valuable as is possible. They deliberately do not specify technical requirements, but are a set of guiding principles that provide for a continuum of increasing reusability, via many different implementations. They describe characteristics and aspirations for systems and services to support the creation of valuable research outputs that could then be rigorously evaluated and extensively reused, with appropriate credit, to the benefit of both creator and user.

FAIR/Open

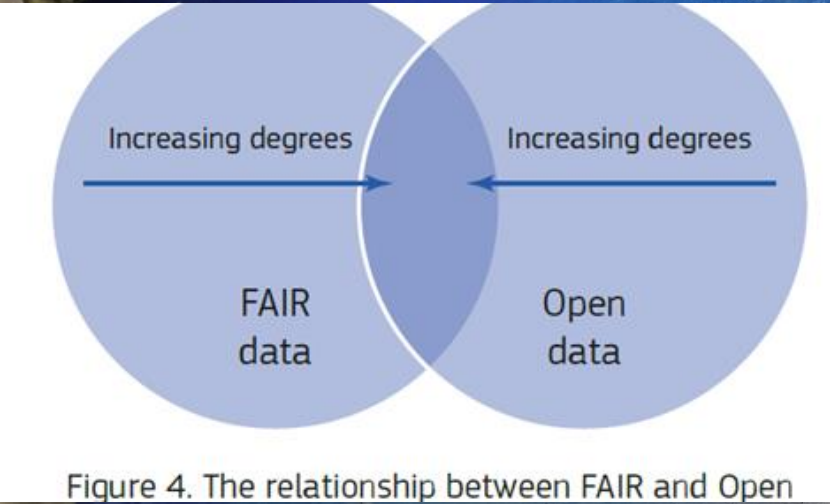


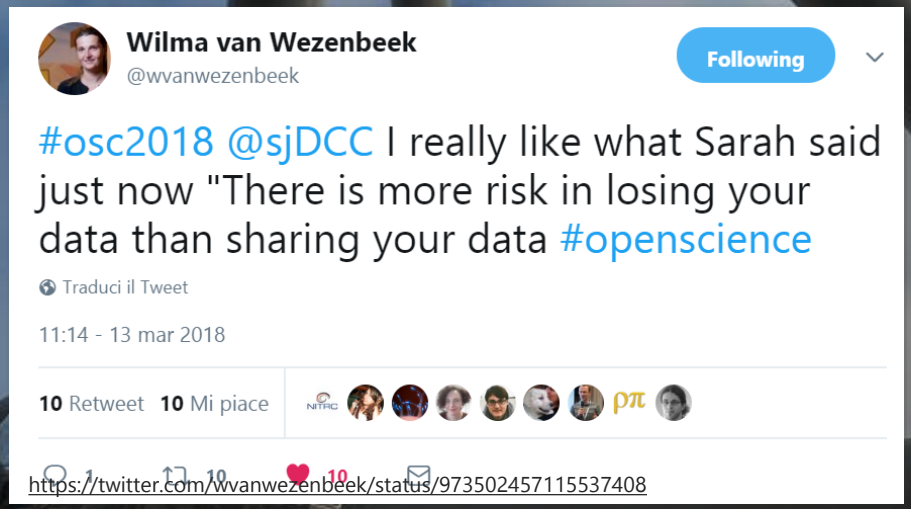
Figure 4. The relationship between FAIR and Open

Turning FAIR into reality, 2018

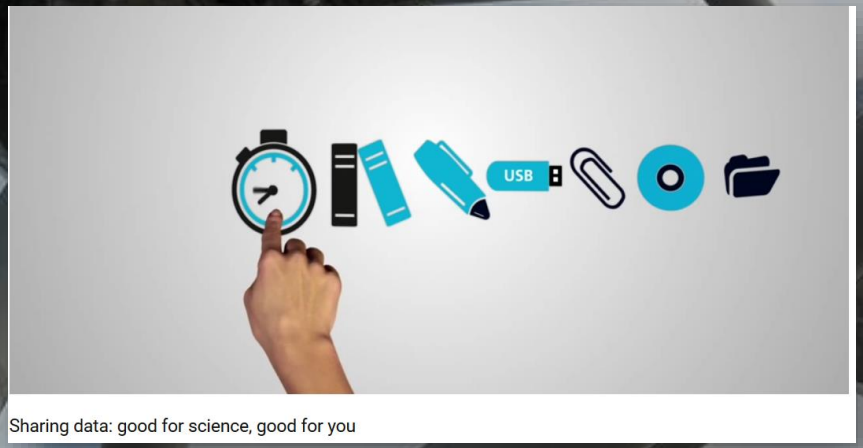
A TENDERE, I DUE INSIEMI SARANNO SEMPRE PIÙ SOVRAPPOSTI. MA EISTERANNO SEMPRE DATI PERFETTAMENTE FAIR CHE NON POSSONO ESSERE OPEN

DATI APERTI

Perché i dati aperti?

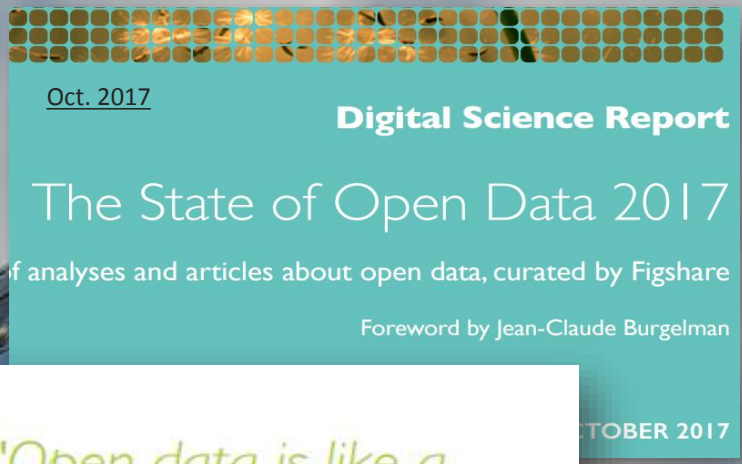


https://twitter.com/wvanwezenbeek/status/973502457115537408



Sharing data: good for science, good for you

https://www.youtube.com/watch?v=[illegible]&feature=youtu.be



People will contact me to ask about stuff

Christopher and Alex (C&A) say: "This is usually
objection of people who feel overworked and th
[data sharing] isn't part of their job…" I would a
learning from each other – if a researcher is opp
datasets, collaborating with others, and genera
they should be outed by their community as a p

COMUNICAZIONE SCIENTIFICA= GRANDE CONVERSAZIONE

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that mis r the help." From the UK Data Archive: "P ding contextual information for your res chers to correctly use and understand your

IMPOSSIBILE, SE NELLA «R» DI FAIR AVETE DOCUMENTATO

It's worth mentioning, however, a s actually be useful to counter willful misrepr Freedom of Information legislation), as one c refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how intere hav people that care about them." I'd a neth dataset has value to future researc be "climate change" was a research t to documenting and understanding ives

EHM… ALLORA PERCHÉ USA FONDI PUBBLICI?

I might want to use it in a research paper

Anyone who's discussed data sharing with a researcher is familiar with this excuse. The operative word here is *might*. How many papers have we all considered writing, only to have them shift to the back bur s is a real concern.

EMBARGO PERFETTAMENTE COMPATIBILE CON FAIR

C&A suggest the embargo route: "One embargo; require people to archive the es public after X months. You could even g ly things that are no longer cared about b eventually everything can become ope e of their datasets, but I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository.

I would also add that, as the original collector of the data, you are at a huge advantage compared to others that might want to use your dataset. You have knowledge about your system, the conditions during collection, the nuances of your methods, et cetera that could never be fully described in the best metadata.

I'm not sure I own the data

My data is too complicated.

C&A: "Don't be too smug. If it turns out it's not that compl professional [standing]." I would add that if it's too compl complicated to reproduce, which means it's arguably not can be solved by more documentation.

IMPOSSIBILE, SE NELLA «R» DI FAIR AVETE DOCUMENTATO

My data is embarrassingly bad

C&A: "Many eyes will help you imp people will accept your data for what it is." I a ck end of making the sausage. We know it's accept that. Plus it helps you strive will be at m r next collection phase.

MA COME SI FA A FARE RICERCA CON DATI «BRUTTI»???

It's not a priority and I'm busy

CARLY STRASSER
Open Science + Data Sharing Advocate
Home
Blog Posts
Previous Research

http://carlystrasser.net/closed-data-excuses-excuses/

Closed Data… Excuses, Excuses

SEMPRE MAGGIOR NUMERO DI FINANZIATORI E DI RIVISTE LO CHIEDE…
STA DIVENTANDO UNA PRIORITÀ

… «as open as possible»…


Grazie!