OPEN SCIENCE DALLA A ALLA Z
5-DATA MANAGEMENT

Elena Giglia

Elena Giglia
Perugia, 27 febbraio 2020
elena.giglia@unito.it
@egiglia

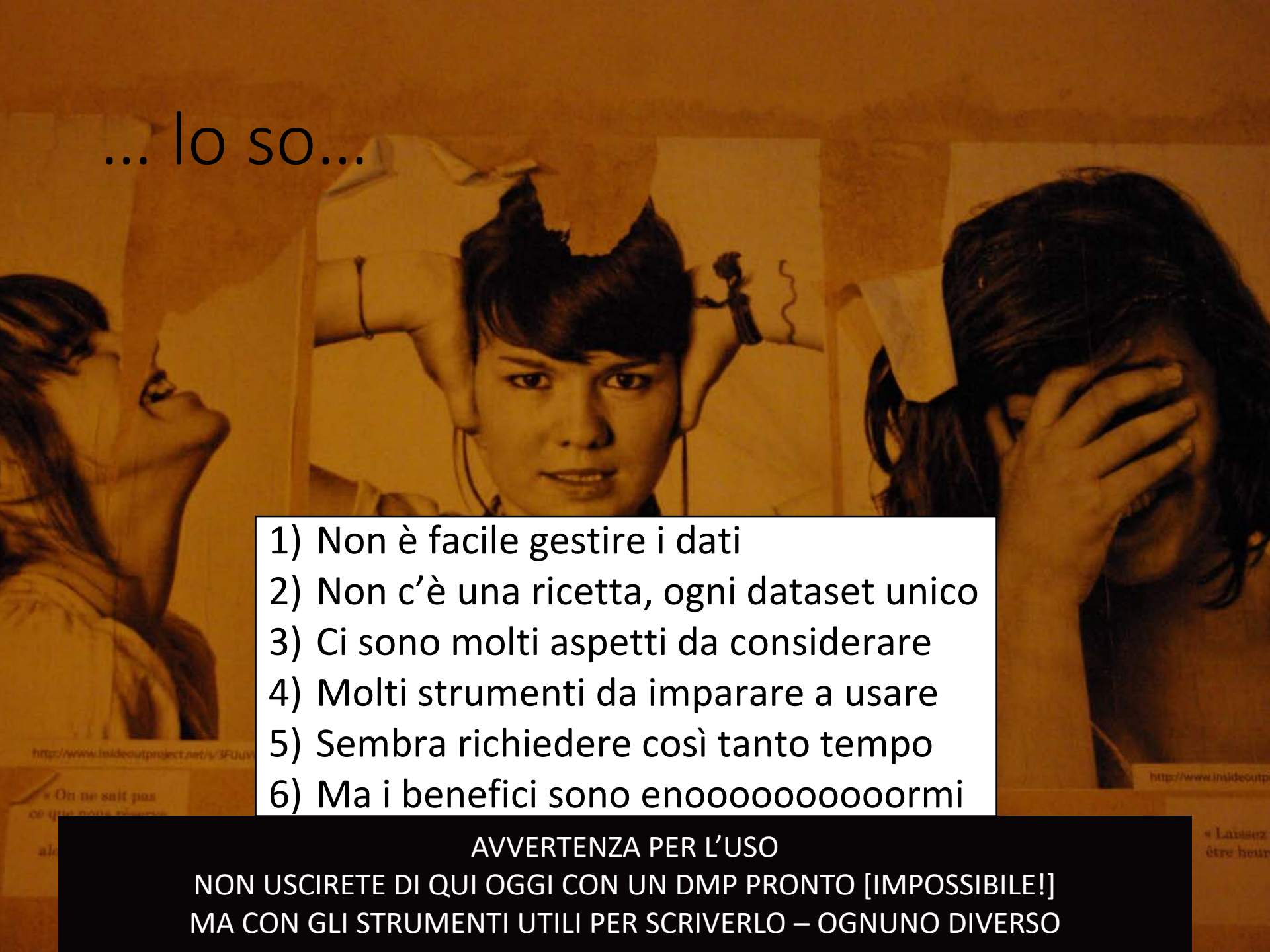
In questo modulo impareremo:

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 se i dati sono gestiti bene dall'inizio possono poi essere resi FAIR e se possibile aperti

… lo so…



1) Non è facile gestire i dati
2) Non c'è una ricetta, ogni dataset unico
3) Ci sono molti aspetti da considerare
4) Molti strumenti da imparare a usare
5) Sembra richiedere così tanto tempo
6) Ma i benefici sono enooooooooormi

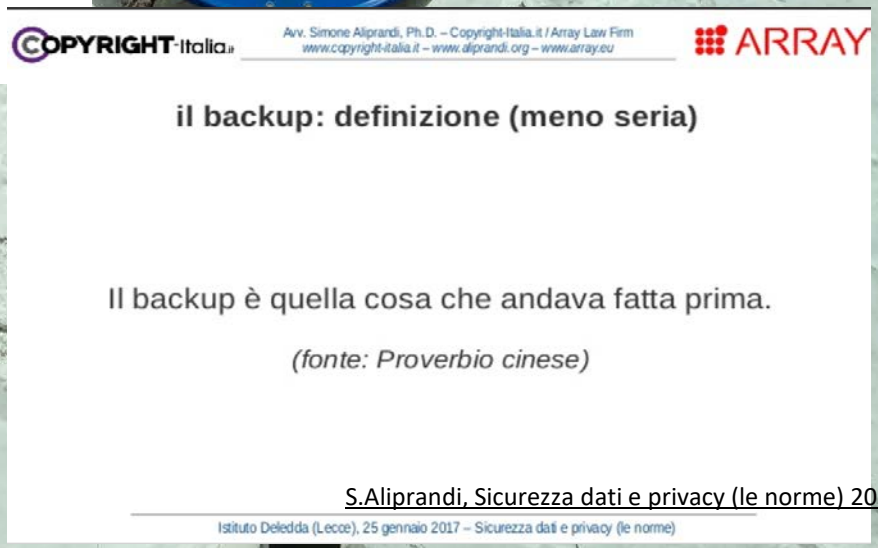
AVVERTENZA PER L'USO
NON USCIRETE DI QUI OGGI CON UN DMP PRONTO [IMPOSSIBILE!]
MA CON GLI STRUMENTI UTILI PER SCRIVERLO – OGNUNO DIVERSO

…parliamo di dati

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

QUANTI DI VOI AVREBBERO PROBLEMI SE PERDERSTE LA CHIAVE USB?

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



S.Aliprandi, Sicurezza dati e privacy (le norme) 2017

I dati. Perché vanno curati?

https://www.youtube.com/watch?v=N2zK3sAtr-4&ecver=2

… 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

… 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

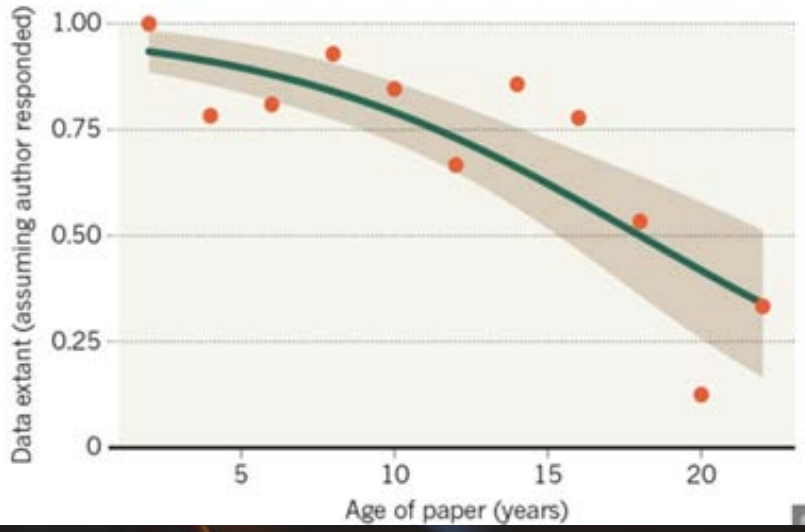
19 December 2013

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


UniTO cosa offre per lo storage e la condivisione dei dati?
…DOVE conservate i dati?

Perché occuparsi 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)

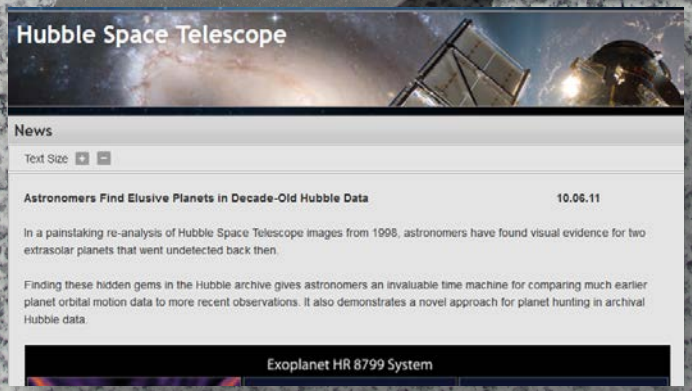
PERMETTERE VALIDAZIONE E CONTROLLI

(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]

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.

Perché occuparci di dati?

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

2013

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

DATA AB INITIO

AGING RESEARCH DATA RIGHT, FROM THE START

K.Birney, 2015

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.**

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, selective exclusion of available data, and unconventional weighting of summary statistics lead to serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period. Our finding is that when properly calculated, the average real GDP growth rate for countries carrying a public-debt-to-GDP ratio of over 90 percent is actually 2.2 percent, not −0.1 percent as published in Reinhart and Rogoff. That is, contrary to RR, average GDP growth at public debt/GDP ratios over 90 percent is not dramatically different than when debt/GDP ratios are lower.

We also show how the relationship between public debt and GDP growth varies significantly by time period and country. Overall, the evidence we review contradicts Reinhart and Rogoff's claim to have identified an important stylized fact, that public 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

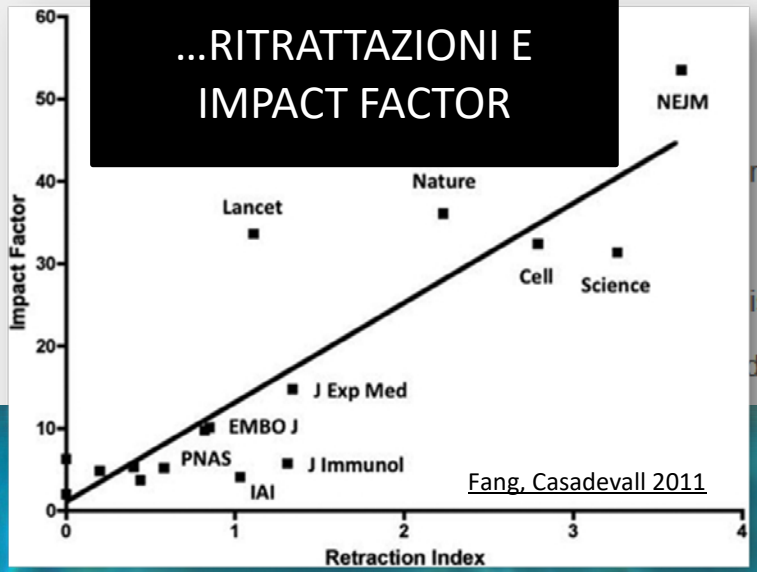
https://memegenerator.net/instance/64979477/case-closed-judge-judy-data-or-it-didnt-happen

Perché occuparci dei dati

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

…RITRATTAZIONI E IMPACT FACTOR



Fang, Casadevall 2011

PLOS | MEDICINE

OPEN ACCESS

ESSAY

How to Mak

John P. A. Ioannidis

Published: October 21,

PLOS | MEDICINE 2005 BROWSE PUBL

OPEN ACCESS

ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

Tony Ross-Hellauer ha ritwittato

Gr@h@m Steel @McDawg · 8 set

If you missed Prof John Ioannidis's Plenary at #osfair2017

Traduci dalla lingua originale: inglese

OSFair2017

OSFair2017

youtube.com

2017

OSFPREPRINTS

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

ries or "successes"

issemination of research

ds and statistical literacy

nature human behaviour

Altmetric: 2520 Citations: 63 More detail »

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

http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001747

Perché occuparsi dei dati?

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

Perché occuparci dei dati?

The Vienna Declaration on the European Open Science Cloud

Vienna, 23 November 2018

e u 2 0 1 8 · a t

Vienna, Nov.23, 2018

PERCHÉ ORA ABBIAMO EOSC

We, Ministers, delegates and other participants attending the launch event of the European Open Science Cloud (EOSC):

1. Recall the challenges of data driven research in pursuing excellent science as stated in the "EOSC Declaration" signed in Brussels on 10 July 2017.

2. Reaffirm the potential of the European Open Science Cloud to transform the research landscape in Europe. Confirm that the vision of the European Open Science Cloud is that of a research data commons, inclusive of all disciplines and Member States, sustainable in the long-term.

3. Recognise that the implementation of the European Open Science Cloud is a process, not a project, by its nature iterative and based on constant learning and mutual alignment. Highlight the need for continuous dialogue to build trust and consensus among scientists, researchers, funders, users and service providers.

4. Highlight that Europe is well placed to take a global leadership position in the development and application of cloud services for Science. Rea and open to the world, reaching out over time to

SEAMLESS ACCESS TO OPEN BY DEFAULT
FAIR DATA

5. Recall that the Council roadmap and the federated

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.

[è la professione del futuro]

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.

e-IRG e-Infrastructure Reflection Group

e-Infrastructure Reflection Group

Open Working

An Experiment in Open Working from 4TU.Centre for Research Data & TU Delft Research Data Services (Note! This is a tes

HOME | ABOUT OPEN WORKING AT TU DELFT | DRAFT DATA MANAGEMENT PLAN CATALOGUE | DATA STEWARDSHIP | CONTACT

FEBRUARY 23, 2018

We are hiring (again!) – Data Steward position at TU Delft

We need 500.000 respected data stewards to operate the European Open Science Cloud

04/05/16 09:08

At the e-IRG workshop in Amsterdam, we had the opportunity to talk to Barend Mons who is chairing the High Level Expert Group on the European Open Science Cloud, an advisory group to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it, Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose

[profilo del data steward]

National Coordination of Data Steward Education in Denmark

Final report to the National Forum for Research Data Management (DM Forum)

Results and recommendations
January 2020

Jan 31, 2020

THE ADMINISTRATOR

- Establish good practices in compliance and data privacy
- Fast learner with a structured and analytical mindset
- Focus on execution and seek challenges in strategic development
- Implement solutions and educating end-users about them
- Passion for policy and IT security
- Positive attitude on cloud solutions
- Risk assessments while having disciplinary knowledge
- Team player with can-do attitude towards processes and operations

THE ANALYST

- Ensure data quality
- Enthusiasm in cloud solutions
- Fast learner and innovative on building custom software and databases
- Good at multitasking
- Programming skills in statistical and data analysis
- Seek challenges, have positive attitude towards reporting

THE DEVELOPER

- FAIR principles advisor and good at data planning and governance
- Focus on collaboration and knowledge sharing to raise business awareness
- Innovative thinker who develops procedures and guidelines
- Innovative thinking concerning master data management
- Passionate about process optimization via good project management
- Working in a team with compliance and data privacy experts trying to establish good practices

THE AGENT OF CHANGE

- Agile mindset and enthusiasm
- Client and customer oriented, understanding both users and processes and operations
- Developing user friendly procedures and guidelines
- Educate users on ethics and the responsible conduct of research
- Focus on execution of policy and strategy awareness
- Passionate to implement solutions via project and change management

Students with Bachelor degree

Directed corporate employment

- *One-year master programme*
  For students who fulfill requirements to basic programming skills, study skills, subject knowledge and academic language level.
- *Two year master programme*
  For students who do not meet the requirement for programming and study skills follow a *pre-master's year before the master's* and then continue with the one-year master.
- *Two-year candidatus*
  Traditional university candidatus combining theory, methods and internships. A dissertation in the area of DS is required.

Students with a PhD or equivalent

Directed corporate or academic employment

For PhDs from any field at university faculties or part of research teams in industry. Educational programmes are a collaborative endeavor between faculties, library or knowledge centers, Centers for Information Security, Data Steward Community, Research coordinators, Project PI's, System Developers, Communication and Teaching teams, the Graduate School (for PhD training) & Human Research Ethics Committees.

- *Requirements*: Short courses, workshops, mentorships, online modules, summer school programmes & MOOCs, that immerse the student in the DS Community.

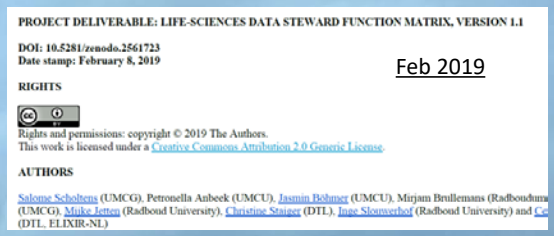
Continuing and professional education

Directed corporate or academic employment

Flexible master programme

- For students who have professional experience and wish to improve their DS skills but keep working full-time, or already have a master's degree in any field.
- The programme is a part-time flexible master continuing education program with a prescribed period of study, that the student individually plans, and a total of 60 ECTS-points.
- The student combines elements from established programmes, thus it is a requirement that DS courses, such as the one year master and two year candidatus, are available through the Open University.

[competenze]



PROJECT DELIVERABLE: LIFE-SCIENCES DATA STEWARD FUNCTION MATRIX, VERSION 1.1

DOI: 10.5281/zenodo.2561723
Date stamp: February 8, 2019

Feb 2019

RIGHTS

AUTHORS

Salome Scholtens (UMCG), Petronella Anbeek (UMCU), Jasmin Böhmer (UMCU), Mirjam Brullemans (Radboudumc (UMCG), Mijke Jetten (Radboud University), Christine Staiger (DTL), Inge Slouwerhof (Radboud University) and Co (DTL, ELIXIR-NL)

Mappings to existing data stewardship principles and approaches:
(*: indicates that this topic involves potentially all areas, with emphasis on the indicated areas)

Research data lifecycle: (https://www.ukdataservice.ac.uk/manage-data/lifecycle)

① creating data; ② processing data; ③ analysing data; ④ preserving data; ⑤ giving access to data; ⑥ reusing data.

FAIR principles: (http://www.nature.com/articles/sdata201618)

Ⓕ findable; Ⓐ accessible; Ⓘ interoperable; Ⓡ reusable data

Purdue competence areas (https://docs.lib.purdue.edu/lib_fsdocs/136)

① databases and data formats; ② discovery and acquisition of data; ③ data management and organization; ④ quality assurance; ⑤ data conversion and interoperability; ⑥ metadata; ⑦ curation and re-use; ⑧ cultures of practice; ⑨ data preservation; ⑩ data analysis; ⑪ data visualization; and ⑫ ethics, including citation of data.

DAMA knowledge areas (https://dama.org/content/body-knowledge)

① data governance; ② data architecture; ③ data modelling and design; ④ data storage and operations; ⑤ data security; ⑥ data integration and interoperability; ⑦ documents and content; ⑧ reference and master data; ⑨ data warehousing and business intelligence; ⑩ metadata; and ⑪ data quality.

[Data stewardship competence center]

GO FAIR

FAIR Principles | Implementation Networks | News | Events | Resources | About GO FAIR

Data Stewardship Competence Centers (DSCC)

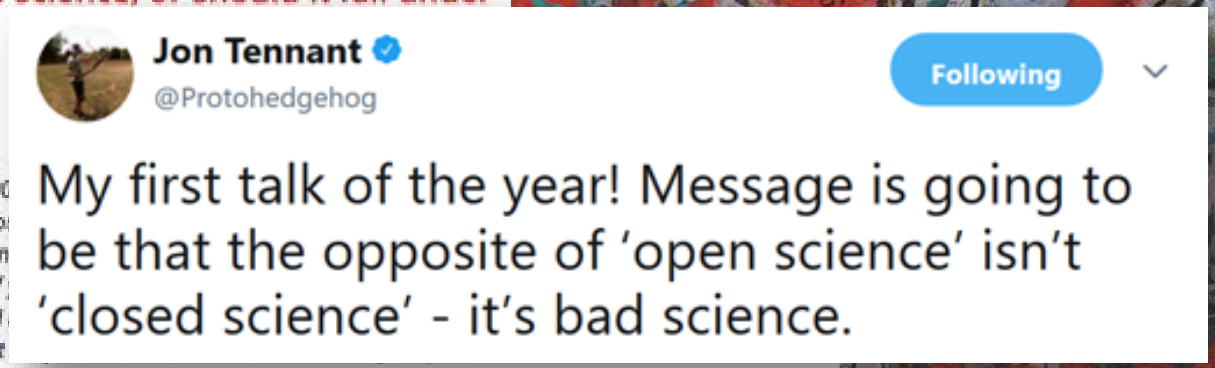
https://www.go-fair.org/implementation-networks/overview/dscc/

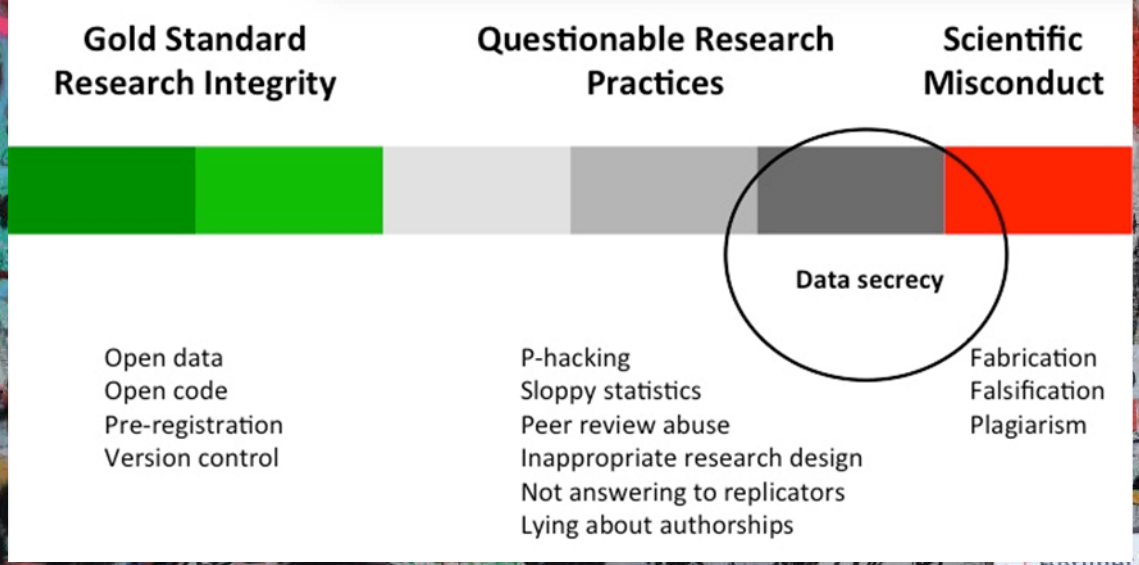
This mutual support and knowledge transfer might contain, for example:

- Co-develop and share awareness building methods and good practices in FAIR research data stewardship.
- Jointly evaluate optimal technologies, software, open source code, repositories, etc to build and maintain (an IN connectable) national IT infrastructure also usable within the new generation research workspaces (cloud, FAIR, VRE).
- Develop and share FAIR metadata templates (generic or discipline-specific), and coordinate this over various disciplines and domains, in close collaboration with international organisations.
- Jointly develop strategies to deal with data privacy, licensing, security and copyright.
- Test and feedback to funders on recommendable tooling for FAIR compliant RDS.
- Jointly develop requirements and skill sets needed and training/education requirements for data stewards.
- Share and where needed participate in the development of e-learning tools, courses and curricula for professional data stewards.
- Share and co-develop approved incentives for research data stewardship.
- Share/compare strategies to make DSCC sustainable in the long-term.

No data?

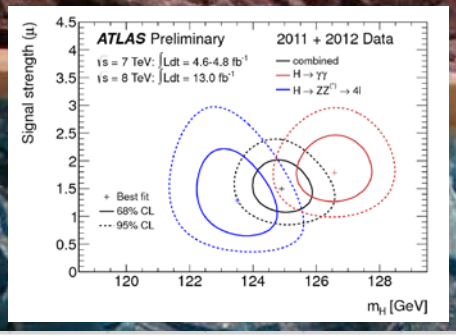
https://twitter.com/alastairdunning/status/968453078218395648





Parliamo di dati

«pezzi»
di conoscenza osservabili



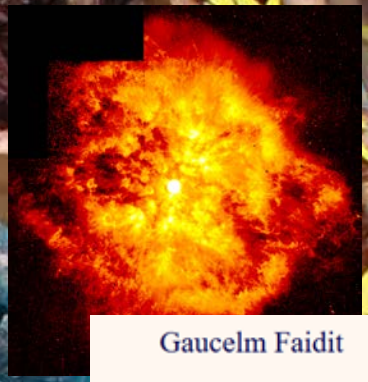


Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample from the cohort.

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	

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.

Wilma van Wezenbeek
@wvanwezenbeek

Following

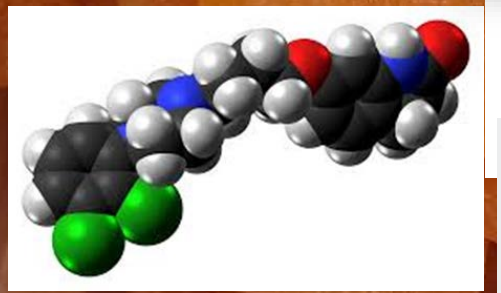
#osc2018 Wolfram Horstmann wants us to talk about datadiversity, like we do with biodiversity #openscience

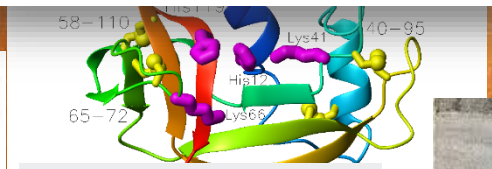
Traduci il Tweet

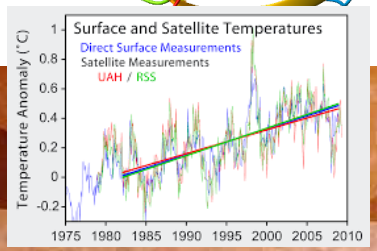
12:51 - 13 mar 2018

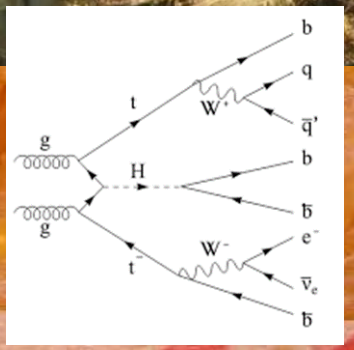
3 Retweet **1** Mi piace

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

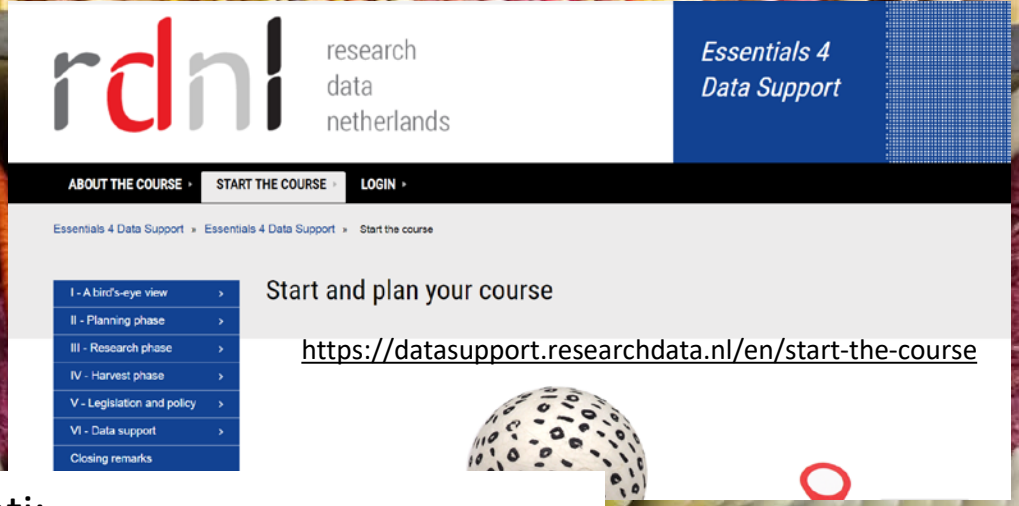








Le basi



5 modi per pensare i dati:

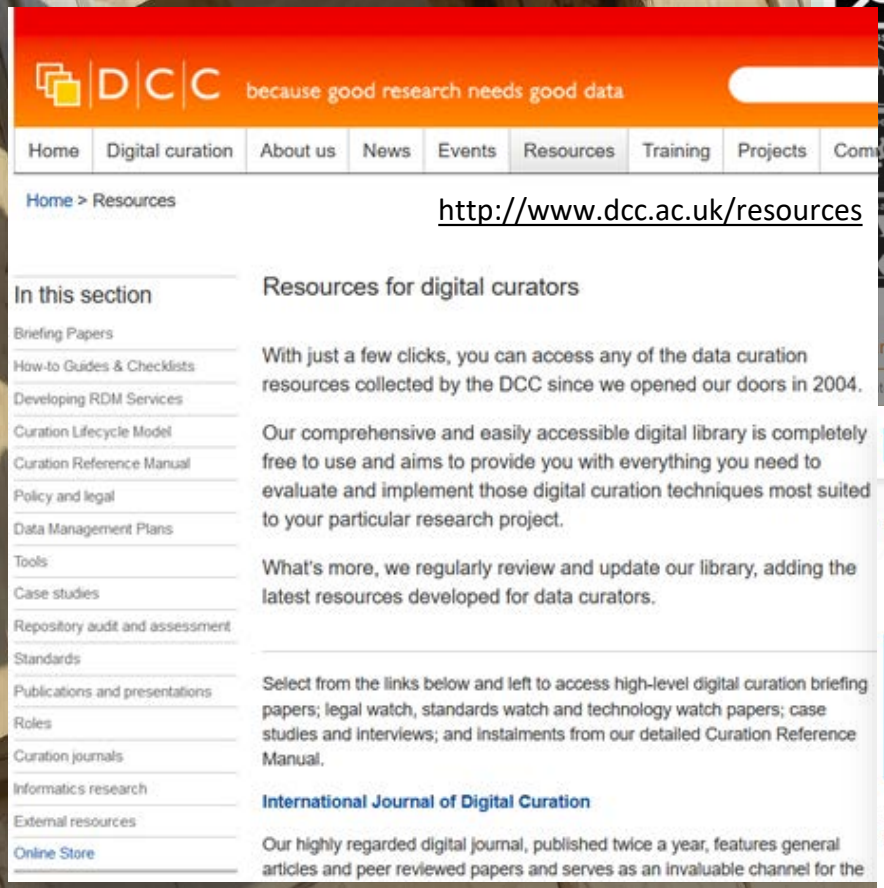
- come sono raccolti (esperimenti, simulazioni…)
- come si presentano (testi, questionari, video…)
- il loro formato elettronico (.txt, .mkv…)
- il loro volume (big data…)
- in che fase sono del ciclo (raw data…)

RICHIEDONO STRUMENTI E TRATTAMENTI DIVERSI

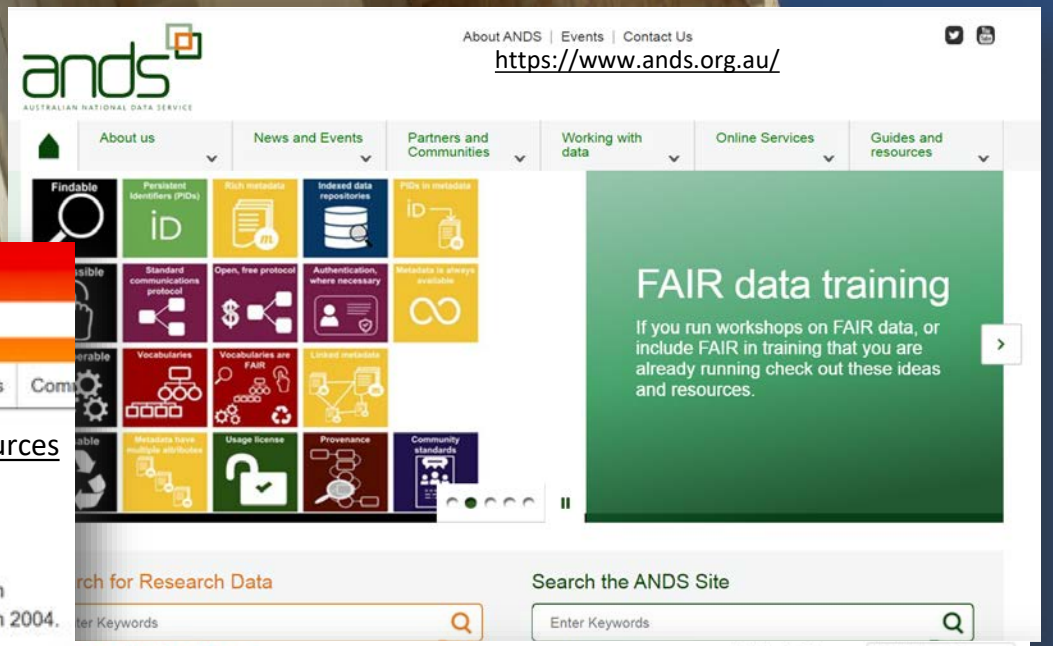
- **The way the data is collected**.
  - By experimenting, simulations, observations, derived data, reference data.
- **The data forms**.
  - For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples,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

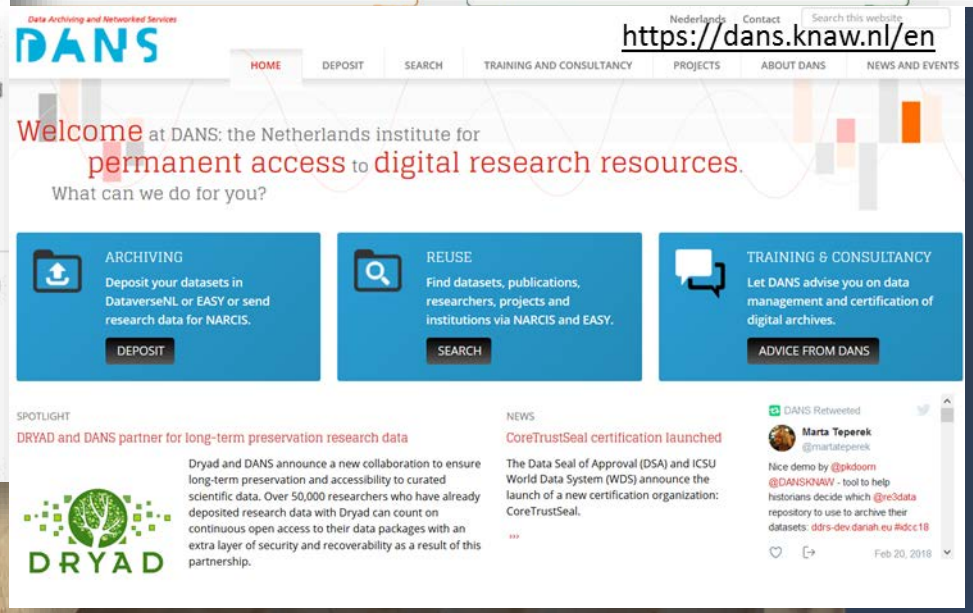
3 pilastri



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



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

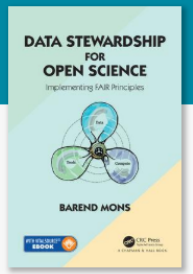


https://dans.knaw.nl/en

… e un maestro

Taylor & Francis Group
an informa business

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



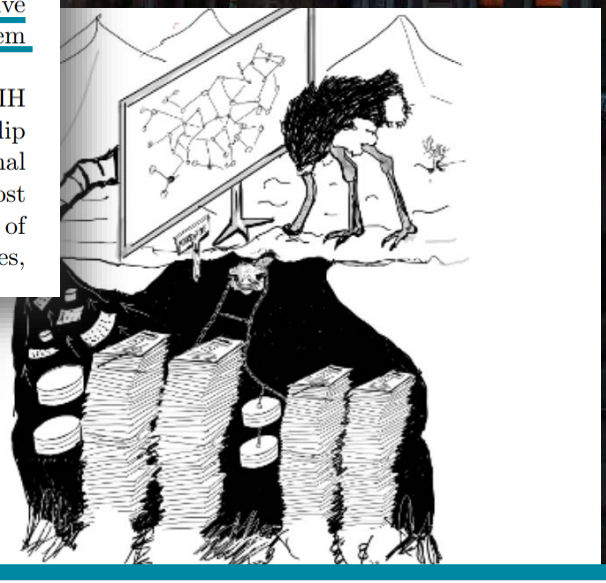
Data Stewardship for Open Science

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

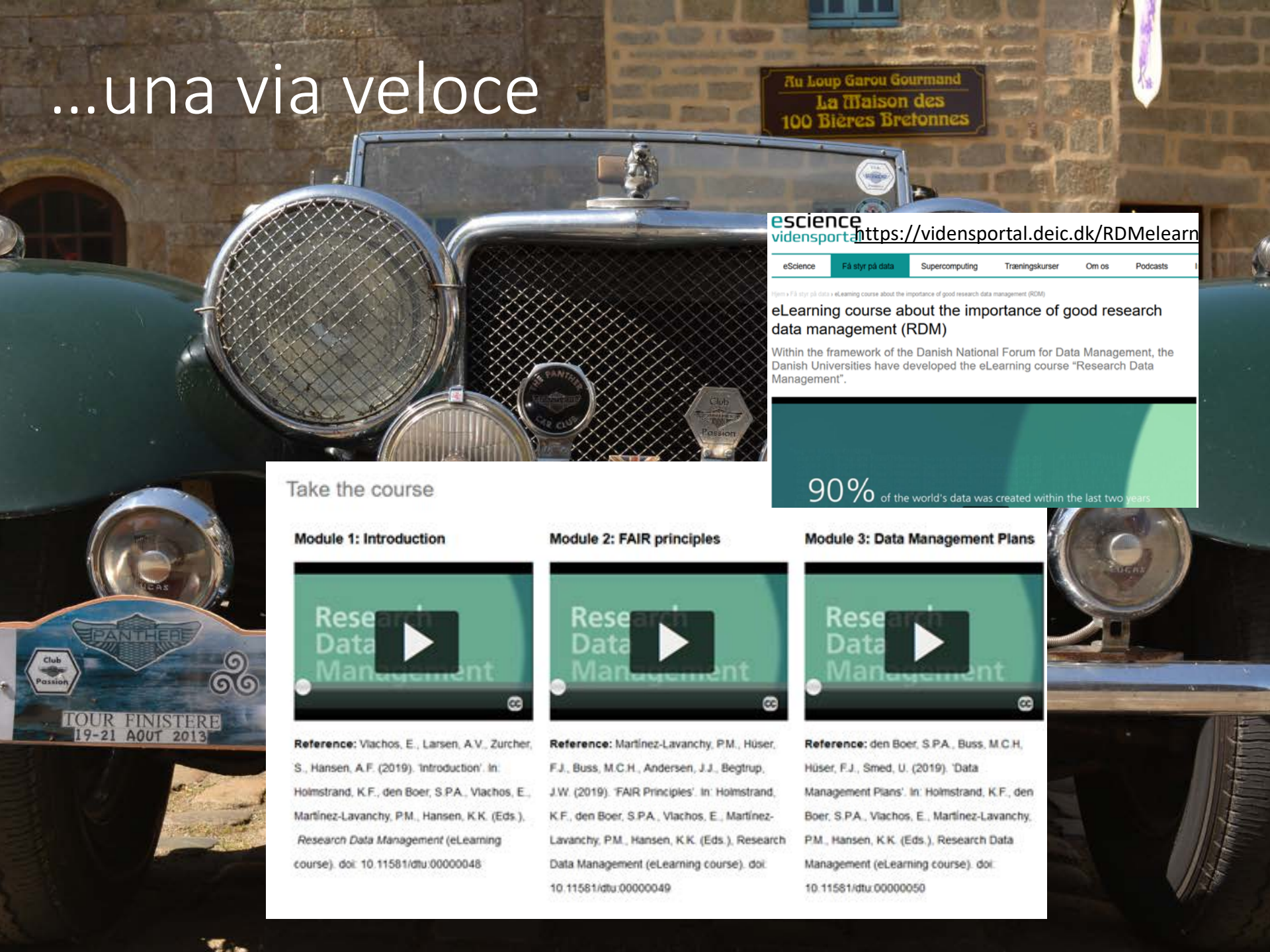
Traduci dalla lingua originale: inglese

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



escience
vidensportal https://vidensportal.deic.dk/RDMelearn

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

Hjem › Få styr på data › eLearning course about the importance of good research data management (RDM)

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

Costi

DM costing tool

What's Zingtree?

Zingtree

Data Management Costing Tool

Data Management costing tool

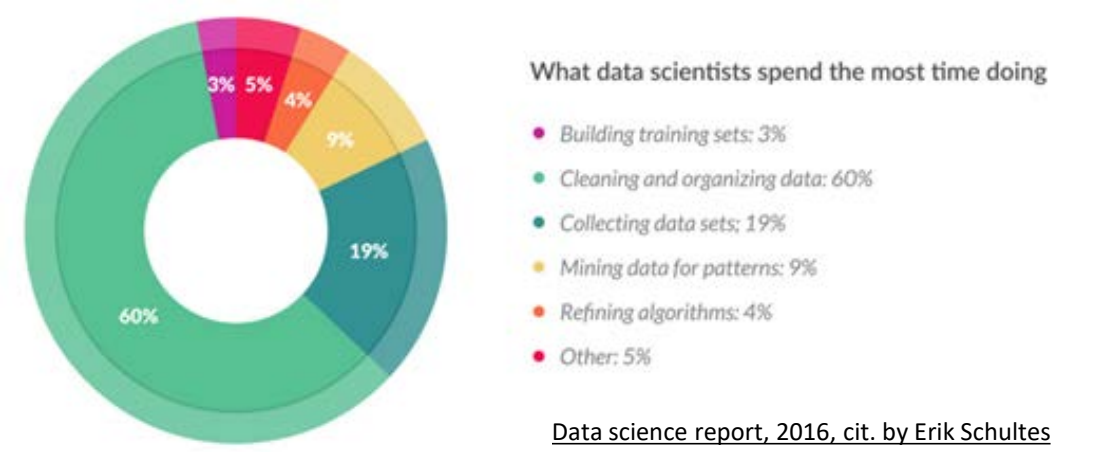
4TU. CENTRE FOR RESEARCH DATA

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

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


…un passo indietro…

[il fondamento]

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

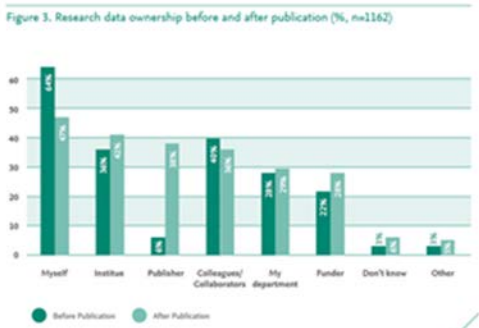
linkedin.com

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.

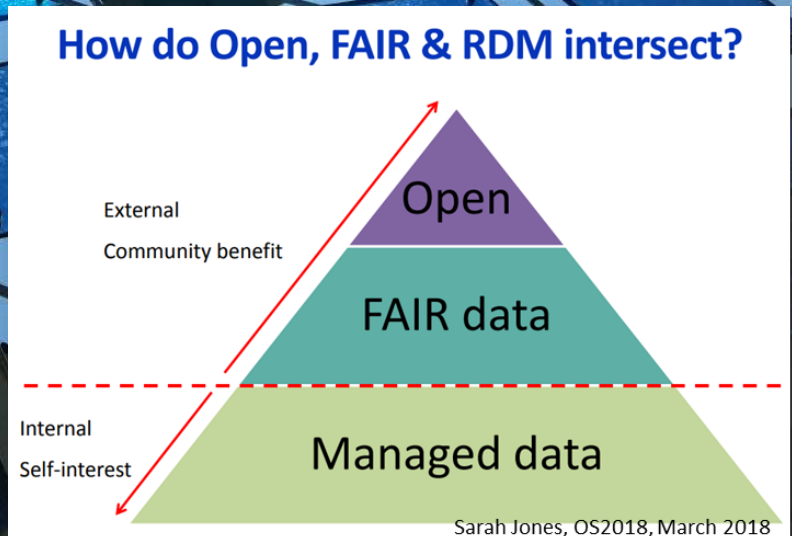
Of course, the study reports

[i tre passi fondamnetali]

Open

FAIR

Dati gestiti



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.

ORGANIZZAZIONE (file naming, versioning…)

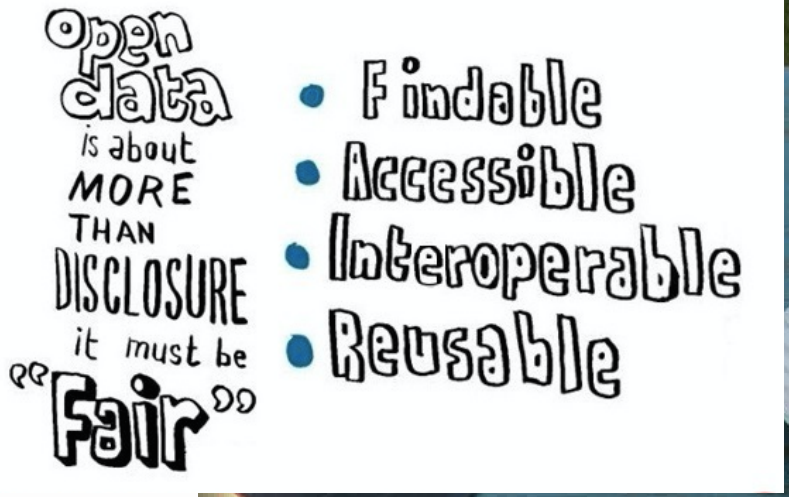
METADATI

BACKUP E STORAGE

CONSERVAZIONE SUL LUNGO PERIODO

ASPETTI LEGALI

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
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

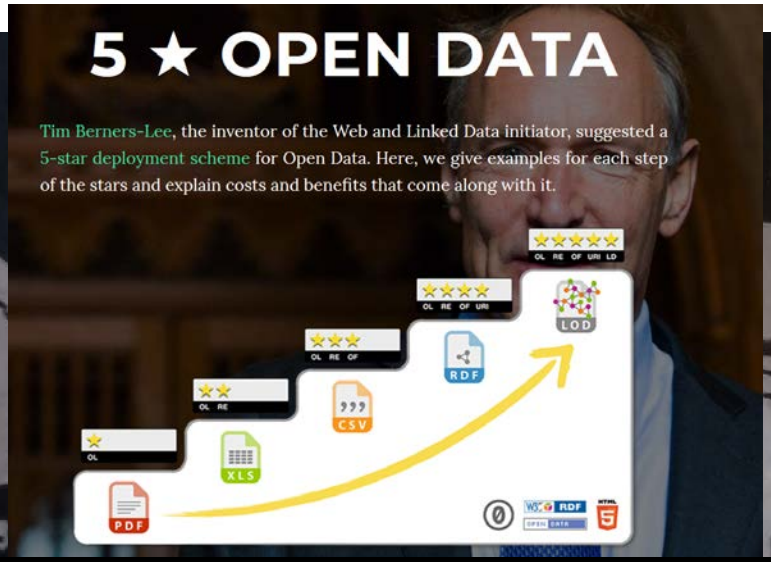
TO BE RE-USABLE:

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.

https://www.force11.org/group/fairgroup/fairprinciples

«ACCESIBLE» NON SIGNIFICA «OPEN». SIGNIFICA SOLO DICHIARARE LE CONDIZIONI SECONDO CUI I DATI SONO ACCESSIBILI

3. I DATI POSSONO ESSERE OPEN



★ make your stuff available on the Web (whatever format) under an open license[1]

★★ make it available as structured data (e.g., Excel instead of image scan of a table)[2]

★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)[3]

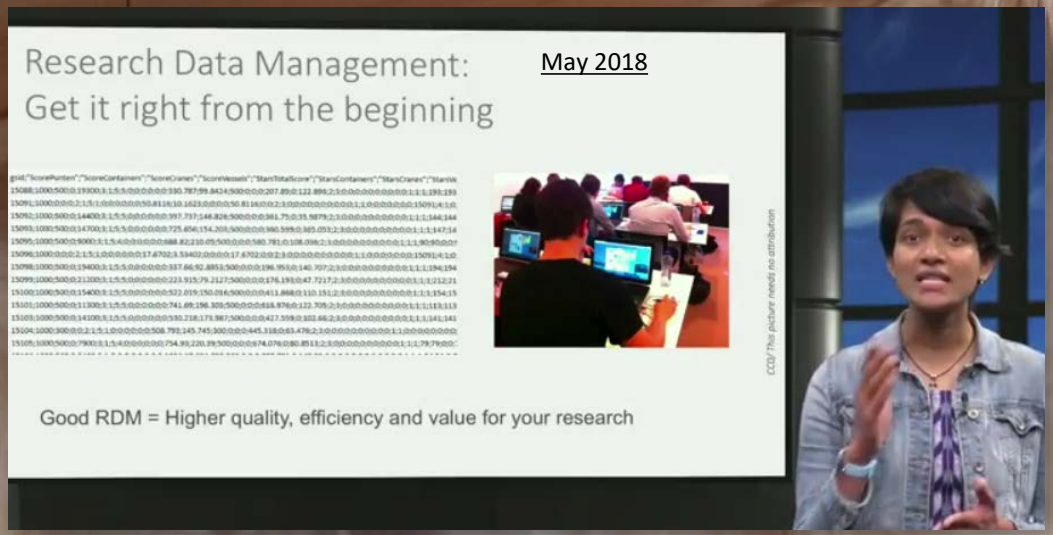
★★★★ use URIs to denote things, so that people can point at your stuff[4]

★★★★★ link your data to other data to provide context[5]

GESTIRE I DATI

DCC because good research needs good data

… dal primo giorno



May 2018

Add a "version management" tab to your spreadsheet.

Now, let me expand on this idea.

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

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

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

2013

Serve formazione

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

AT THE END OF EACH STEP, THERE IS A SECTION «ADAPT YOUR DMP» ACCORDING TO WHAT YOU HAVE JUST LEARNT

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:

cessda TRAINING

⊕ 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:

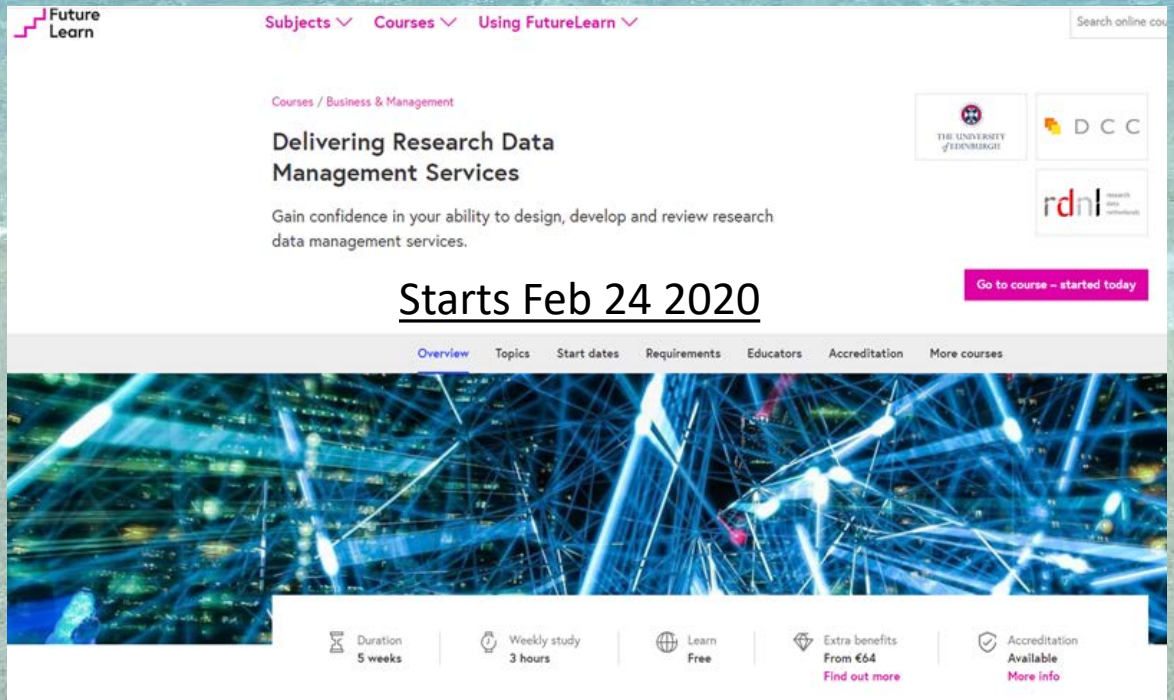
- Will established *terminologies/ontologies* (i.e. structured controlled vocabularies) be used in the project? If not, how does yours relate to established ones?
- Which *coding* is used (if any)? How does this relate to other research?

⊖ 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?

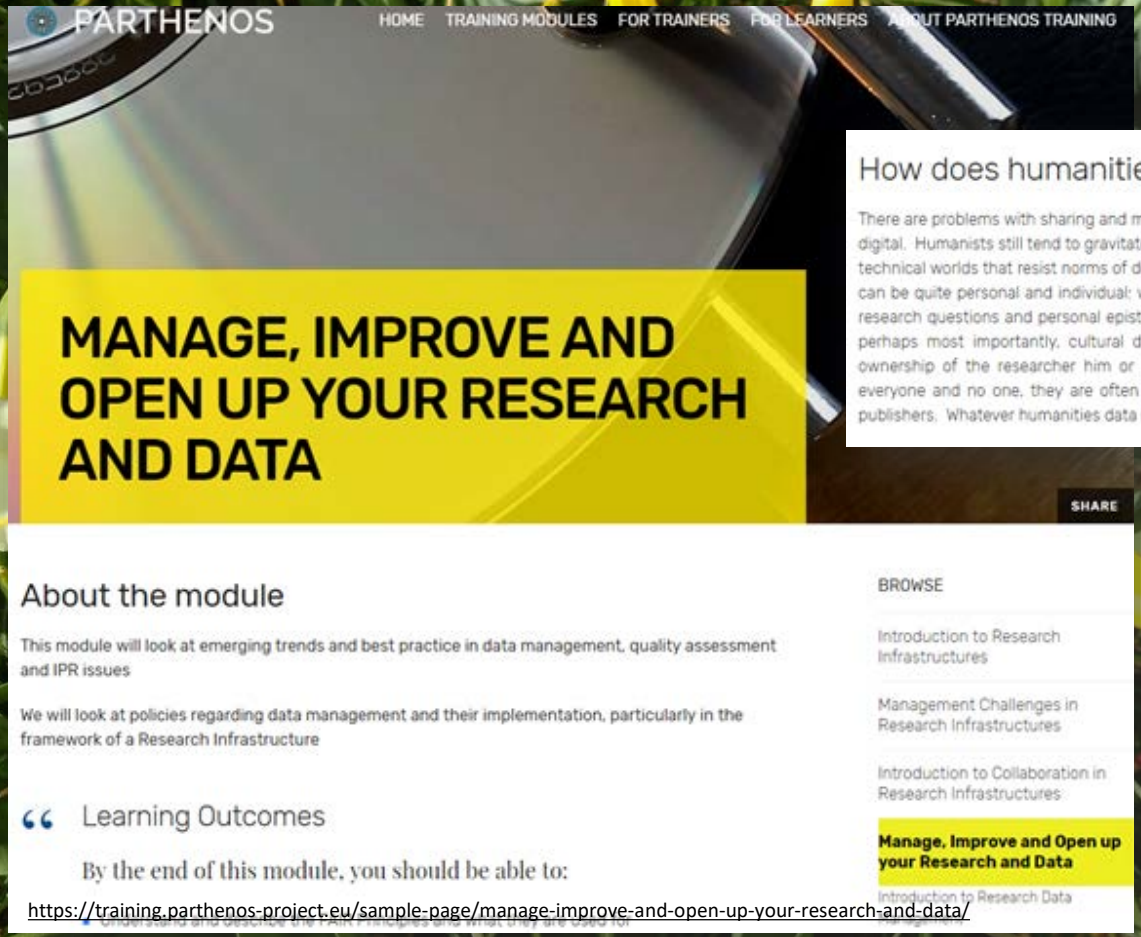
CESSDA Guide

[un corso a distanza]



Starts Feb 24 2020

…e le scienze umane?

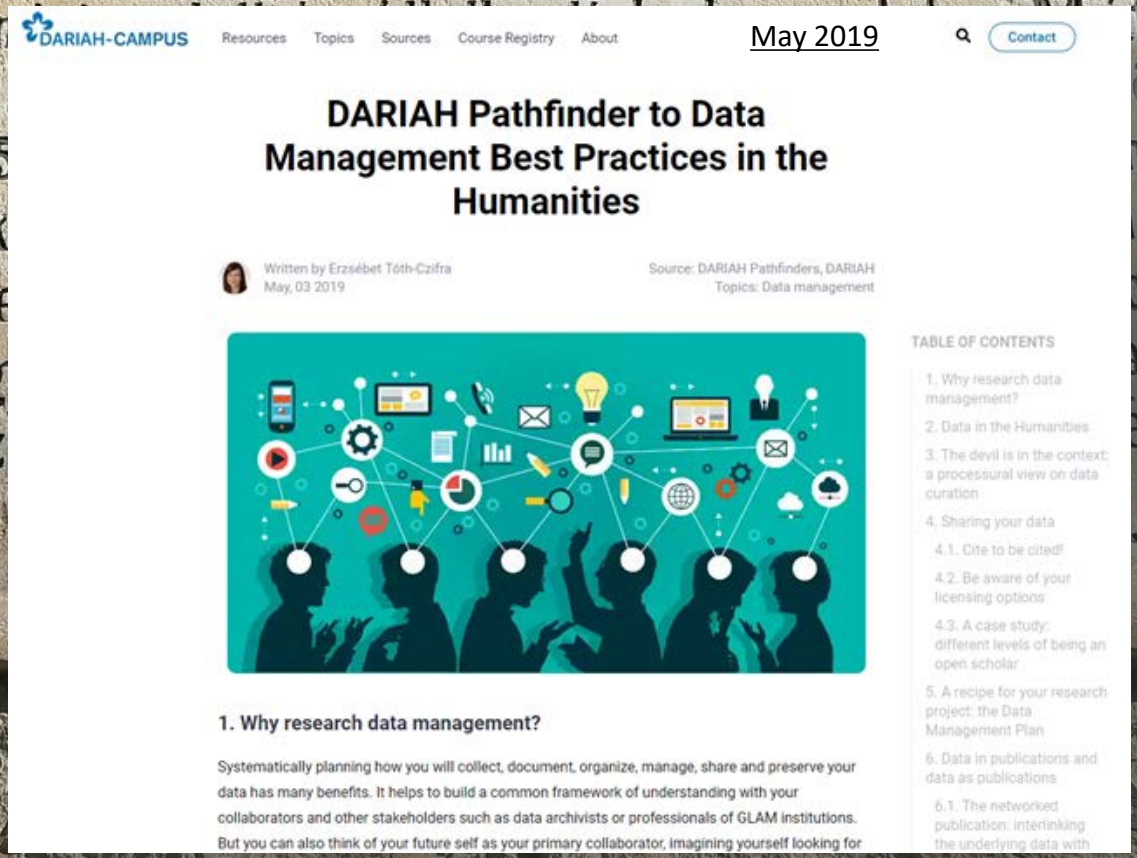


https://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/

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?



DARIAH-CAMPUS Resources Topics Sources Course Registry About Contact

May 2019

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?
2. Data in the Humanities
3. The devil is in the context: a processural view on data curation
4. Sharing your data
   4.1. Cite to be cited!
   4.2. Be aware of your licensing options
   4.3. A case study: different levels of being an open scholar
5. A recipe for your research project: the Data Management Plan
6. Data in publications and data as publications
   6.1. The networked publication: interlinking the underlying data with

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

Data management ABC – File naming

File naming conventions

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

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

Data management ABC – File naming / 2

Data versioning

ands
AUSTRALIAN NATIONAL DATA SERVICE

About us

Working with data

Data versioning

+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 trustworthines
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

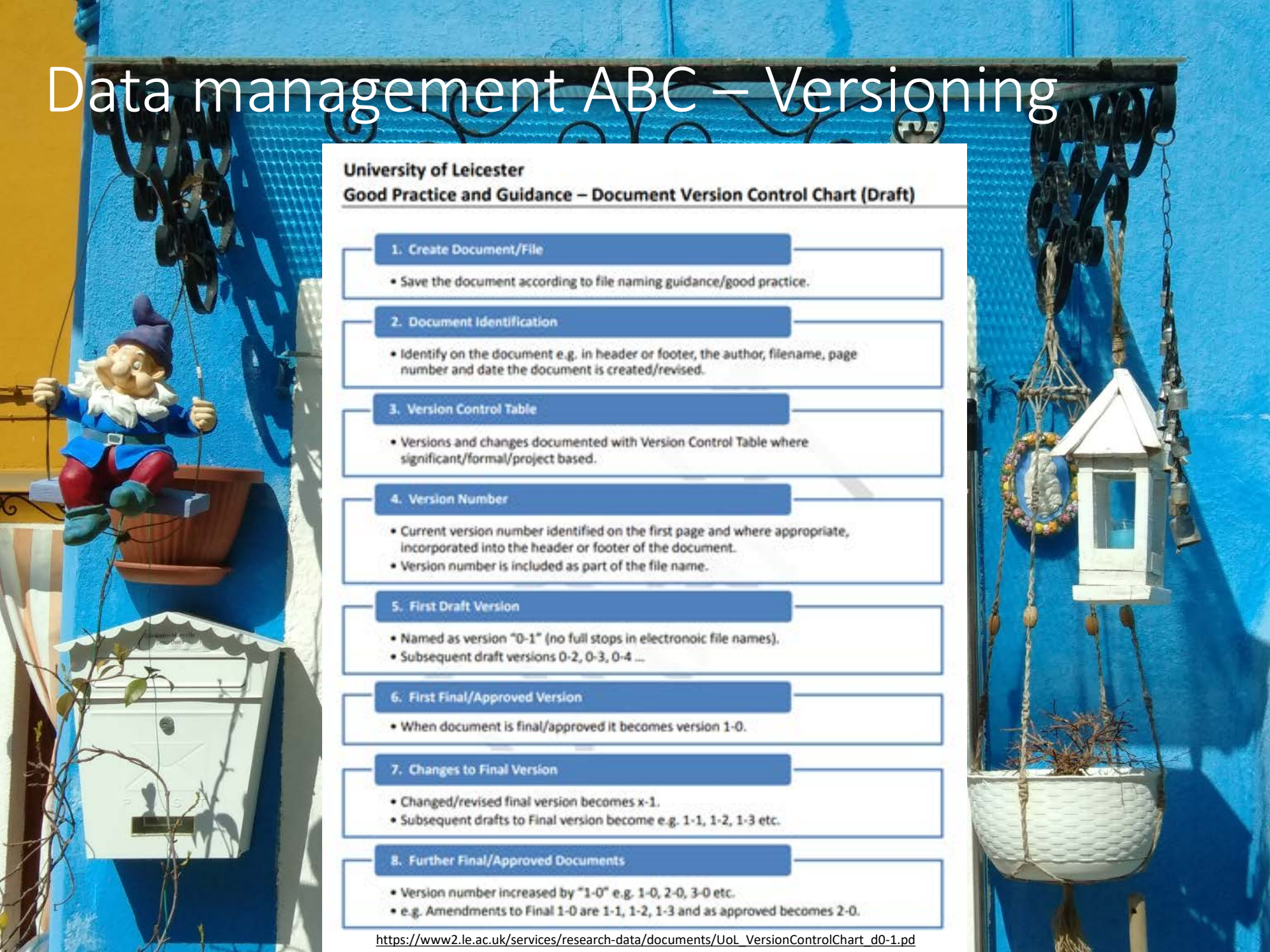
Data versioning follows a similar path to software versioning, usually applying a two-part numbering rule: Major.Minor (e.g. V2.1). Major data revision indicates a change in the formation and/or content of the dataset that may bring changes in scope, context or intended use. For example, a major revision 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

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.

https://www2.le.ac.uk/services/research-data/documents/UoL_VersionControlChart_d0-1.pd

Data management ABC – Versioning

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 Manageme

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

RIO Home Articles About | About Pensoft Books E-Books Blog

Project Report

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

May 9, 2018

Reviewable v1

Support Your Data: A Research Data Management Guide for Researchers

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

Abstract

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

Suppl. material 5: Draft Guide - Preparing doi

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

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

Filename: Draft Guide - Preparing.odt
Download file (59.52 kb)

Suppl. material 6: Draft Guide - Analyzing doi

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

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

Filename: Draft Guide - Analyzing.odt
Download file (51.82 kb)

Suppl. material 7: Draft Guide - Sharing doi

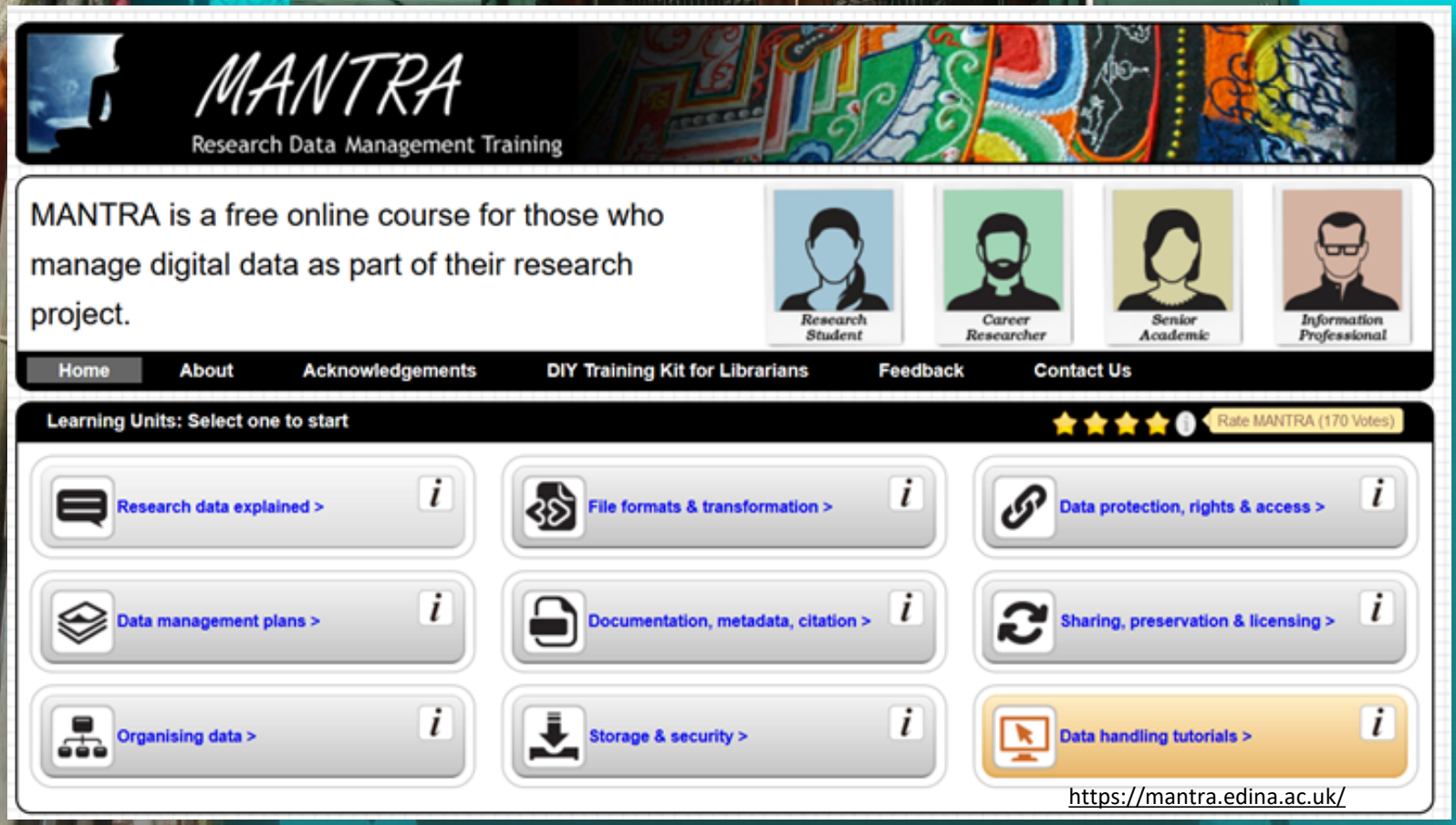
Authors: John Borghi

Data type: OpenDocument Text (.odt) file

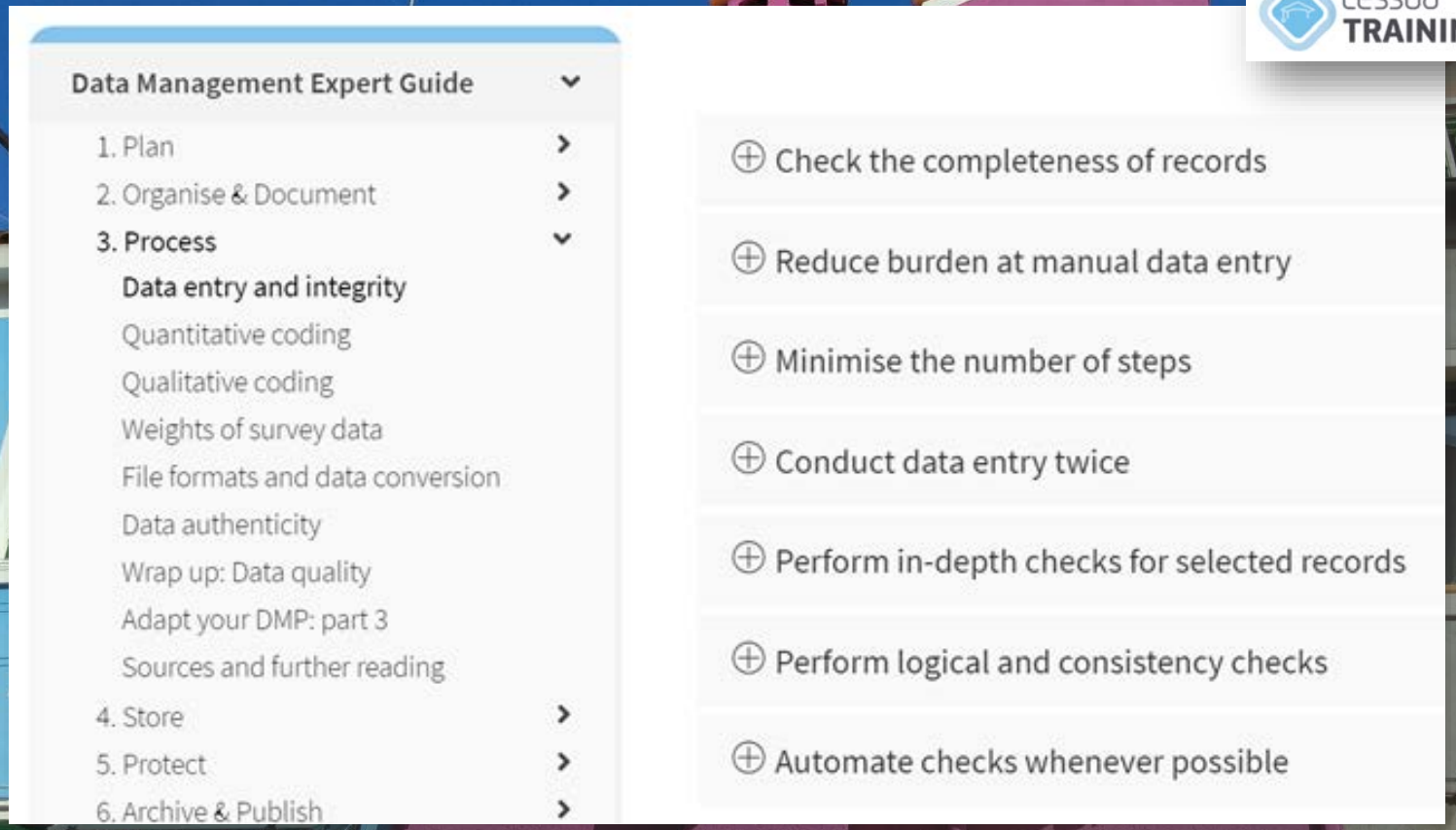
Brief description: A draft guide that corresponds with the "Sharing and publishing your data" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red

https://www.cessda.eu/

Formazione

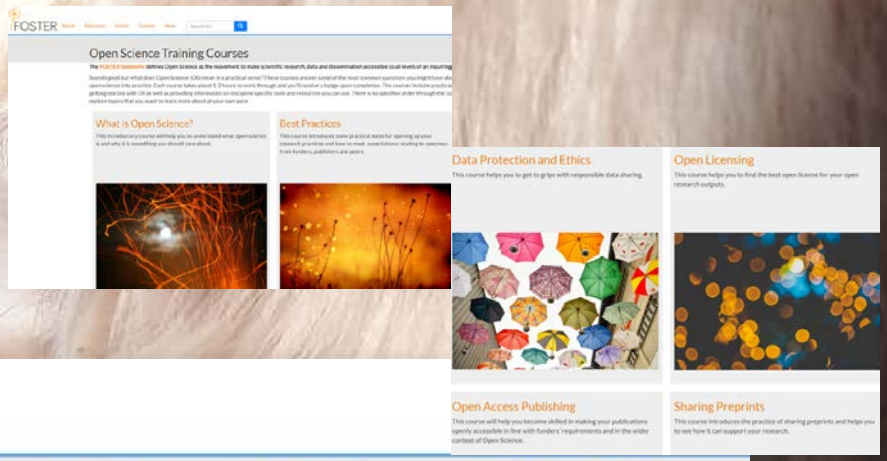

MANTRA
Research Data Management Training
MANTRA is a free online course for those who manage digital data as part of their research project.
Research Student
Career Researcher
Senior Academic
Information Professional
Home
About
Acknowledgements
DIY Training Kit for Librarians
Feedback
Contact Us
Learning Units: Select one to start
Rate MANTRA (170 Votes)
Research data explained >
File formats & transformation >
Data protection, rights & access >
Data management plans >
Documentation, metadata, citation >
Sharing, preservation & licensing >
Organising data >
Storage & security >
Data handling tutorials >
https://mantra.edina.ac.uk/

Data management ABC – Data entry



CESSDA Guide

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

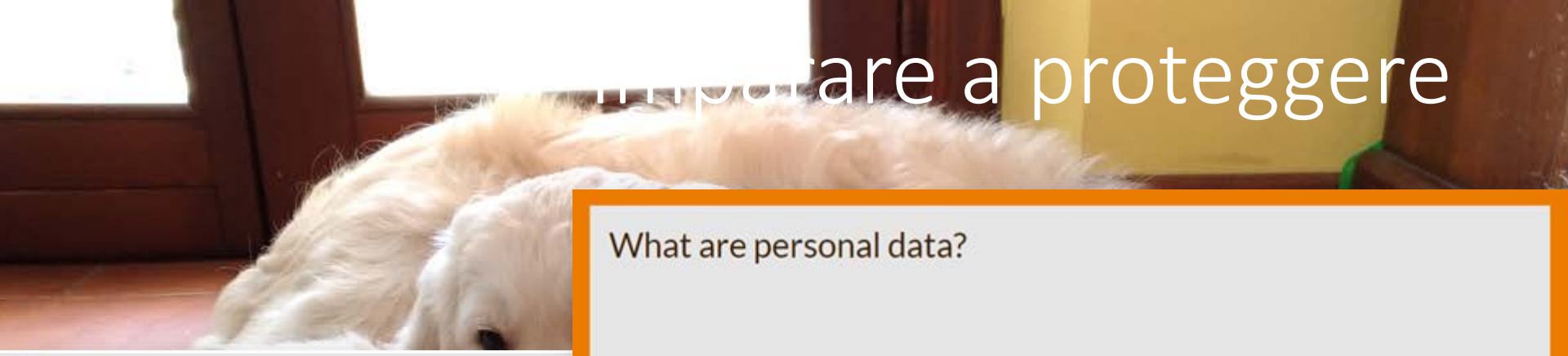
Full details

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

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

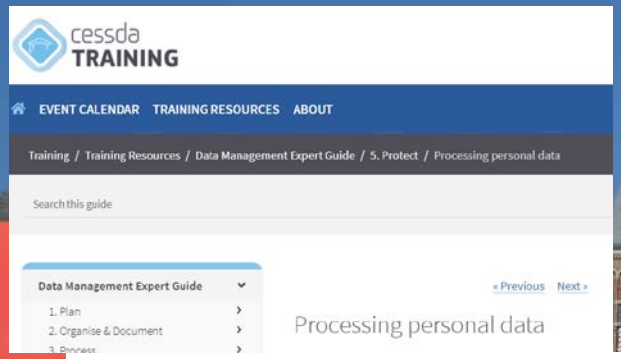
Start the Free Course

Full details

Level of knowledge: Introductory: no previous knowle
is required

Topics

[dati personali]



I. Process lawfully, fair and transparent

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

II. Keep to the original purpose

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

III. Minimise data size

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

IV. Uphold accuracy

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

V. Remove data which are not used

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

VI. Ensure data integrity and confidentiality

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

[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

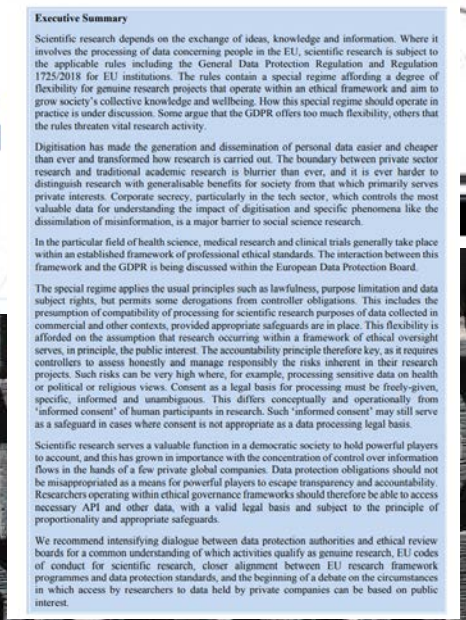
Menu

Webinar Video: GDPR & What It Means For Researchers

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.

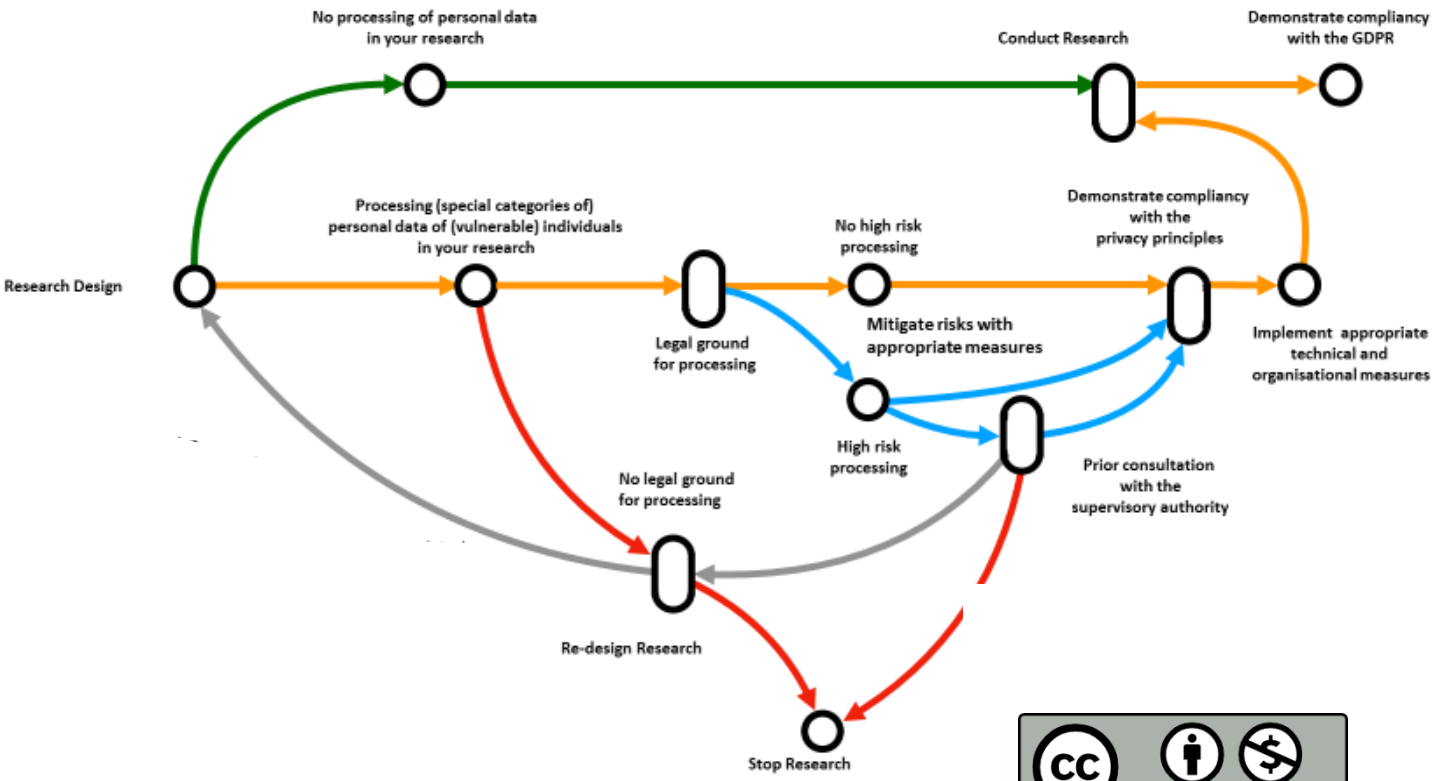
The special regime applies the usual principles such as lawfulness, purpose limitation and data subject rights, but permits some derogations from controller obligations. This includes the presumption of compatibility of processing for scientific research purposes of data collected in commercial and other 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 safeguardingp rivacy.

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.

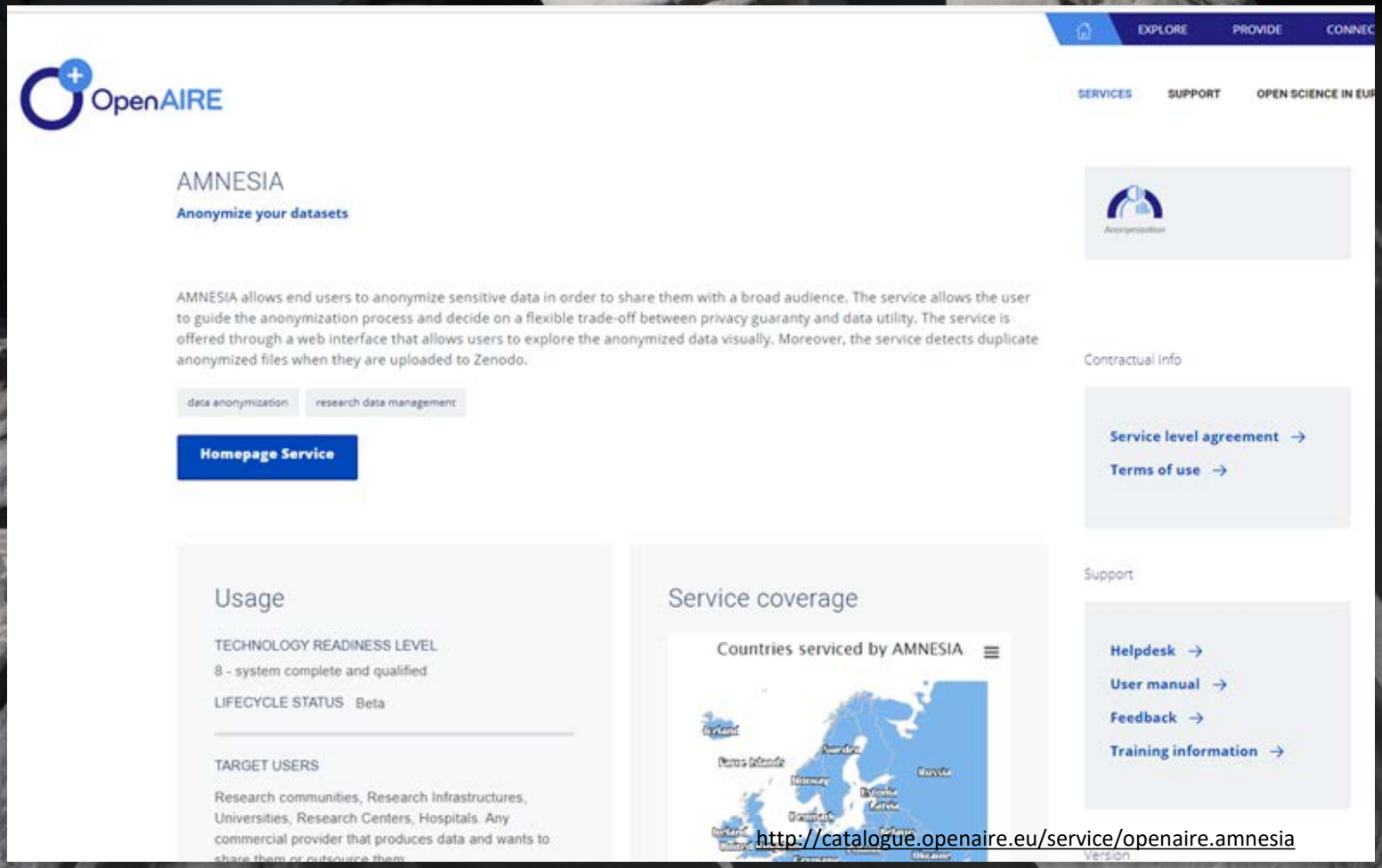
Action

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 data

Findable

discoverable with machine readable metadata, identifiable and locatable by means of a standard identification mechanism

Accessible

available and obtainable to both human and machine

Interoperable

both syntactically parseable and semantically understandable, allowing data exchange and reuse among scientific disciplines, researchers, institutions, organisations and countries

Reusable

sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible across scientific disciplines and borders, and the least cumbersome integration with other data sources

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

FAIR principles for dummies

GO FAIR | GO FAIR Initiative | Implementation Networks | FAIR Principles

FAIR Principles

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an esse

F1. (Meta)data are assigned a globall

F2. Data are described with rich meta

F3. Metadata clearly and explicitly in

F4. (Meta)data are registered or inde

Accessible

Once the user finds the required data, s
including authentication and authorisat

A1. (Meta)data are retrievable by the protocol

A1.1 The protocol is open, free, a

A1.2 The protocol allows for an a necessary

What does this mean?

Principle F1 is arguably the most important because it will be hard to achieve other aspects of FAIR without globally unique and persistent identifiers. Hence, compliance with F1 will already take you a long way towards publishing FAIR data (see **10 ways identifiers can help with data integration**).

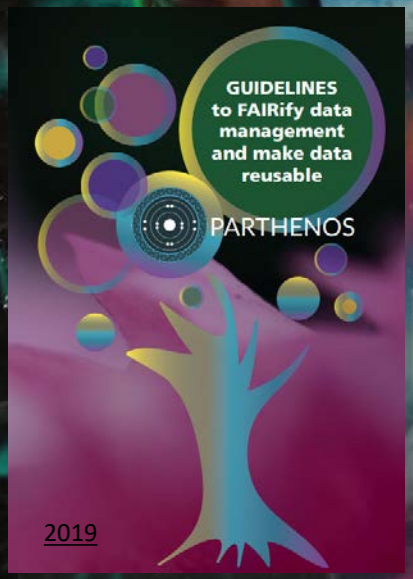
Globally unique and persistent identifiers remove ambiguity in the meaning of your published data by assigning a unique identifier to every element of metadata and every concept/measurement in your dataset. In this context, identifiers consist of an internet link (e.g., a URL that resolves to a web page that defines the concept such as a particular human protein: **http://www.uniprot.org/uniprot/P98161**). Many data repositories will automatically generate globally unique and persistent identifiers to deposited datasets. Identifiers can help other people understand exactly what you mean, and they allow computers to interpret your data in a meaningful way (i.e., computers that are searching for your data or trying to automatically integrate them). Identifiers are essential to the human-machine interoperation that is key to the vision of **Open Science**. In addition, identifiers will help others to properly cite your work when reusing your data.

Of course, identifiers are one thing, but their meaning is another (see principles I1-I3). F1 stipulates two conditions for your identifier:

1. It must be globally unique (i.e., someone else could not reuse/reassign the same identifier without referring to your data). You can obtain globally unique identifiers from a registry service that uses algorithms guaranteeing the uniqueness of newly minted identifiers.
2. It must be persistent. It takes time and money to keep web links active, so links tend to

https://www.go-fair.org/fair-principles/

…FAIR data



Introduction

Once upon a time in the beautiful kingdom of Datamania lived a prince named Prince Fairhair. Though he was gentle as few, and good looking too, his father would not let him choose the love of his life on his own. No, he was destined to marry a woman from the neighbouring kingdom. He did not even know her name, only that she was referred to as My Fair Lady. Before the father of My Fair Lady could accept the marriage, he had a quest for Prince Fairhair. Only by fulfilling the quest, would he be able to marry the princess. His quest was to find out how to turn water into gold. A quest that would require gathering loads of data chests and look for clues that could lead to the recipe.

Luckily, Prince Fairhair was not alone in his quest. One of the castle wings housed a number of wiz-
ards who could help him decrypt and investigate the data chests. However, it was impossible for the data
wizards to go and hunt for data themselves. Thus to assist them, a huge number of elves were trained to

Findable #1:

(Meta)data are assigned globally unique and persistent identifiers

The elves returned one by one to the castle, and some of them were really frustrated. They had been following paths to data chests that had been meticulously described, but somehow the data chests had been removed, just leaving holes in the ground. Fimble was one of these elves, who came back quite puzzled about some strange codes he had found. He could not decipher them and therefore did not know where to go.

"Look" said Fimble to the data wizard, "*I have this strange code 10.1234/abbb, and I don't know what it means?*"

"*Oh, these are very useful indeed*" said the data wizard. "*We can look up the codes in these huge books. Now let me see. 10 is the great country of Datavalley, and we should look in the house number 1234.*" He showed a map to Fimble in the book. "*This is where you should go*".

"*Are you sure it's still there?*" said Fimble, not wanting to waste a single more step on hunting down data chests he could not find.

"*Absolutely. These books are magic. If someone moves the data chest to a new location, the book will know.*"

"*Great*" said Fimble, and took off in a sprint. He soon returned happy carrying a data chest.

FAIR data

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]

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 data

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 following the publication of the FAIR Principles states the authors' position clearly: "The recognition that computers must be capable of accessing a data publication autonomously, unaided by their human operators, is core to the FAIR Principles. Computers are now an inseparable companion in every research endeavour". In recent surveys, the time reportedly spent by PhD students and other researchers in projects dealing with discovering and reusing multiple data sources – so called 'data munging' – has been pegged at 80% [19]. Were these colleagues and their machine-assistants only having to deal with FAIR data and services, this wasted time would be reduced to a fraction of what it is today. The avoidance of time-wasting would be a first return on investment in good data stewardship. To serve this potentially enormous cost reduction, FAIR compliant (meta)data and services should be actionable by machines without human supervision whenever and 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

FAIR per le scienze umane

Feb. 2020

IDENTIFY

Research Data in the Humanities

Introduction

In the humanities, we all use research data, although we may not be aware of it. It is like in the case of Monsieur Jourdain, the title character of Molière's *Le Bourgeois gentilhomme*, who learnt, to his great satisfaction, that unwittingly he had been speaking prose all his life. With research data in the humanities it is exactly the same: you are using it, even if you don't know it, and once you realise it, it will affect your research workflow forever.

Although the term 'data' intuitively seems to be more at place in natural or social sciences (e.g. survey data, experimental data), currently, due to

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.

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

FAIR / Open

Turning FAIR into reality, 2018

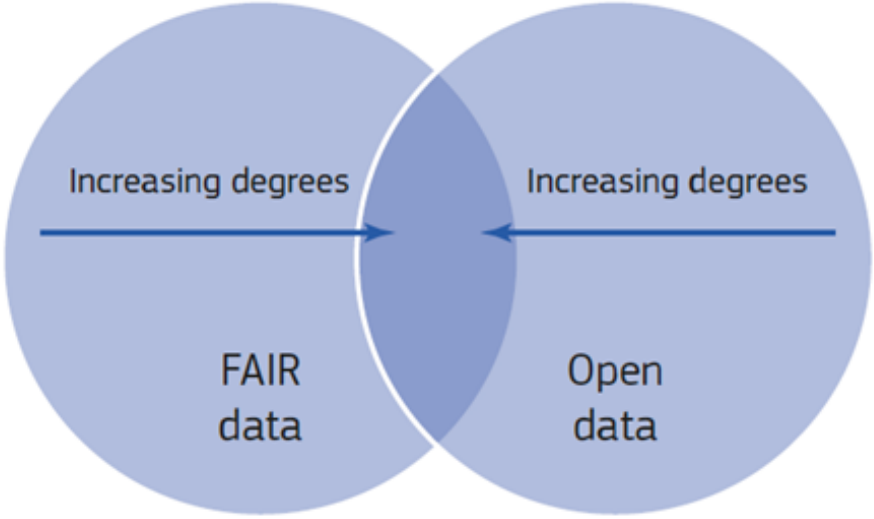


Figure 4. The relationship between FAIR and Open

Data can be FAIR or Open, both or neither. The greatest benefits come when data are both FAIR and Open, as the lack of restrictions supports the widest possible reuse, and reuse at scale. To maximise the benefits of making FAIR data a reality, and in the context of Open Science initiatives, the FAIR principles should be implemented in combination with a policy requirement that research data should be Open by default - that is, Open unless there is a good reason for restricting access or reuse. In recent European Commission formulations, the maxim 'as open as possible, as closed as necessary' has been introduced, which is a helpful articulation of the principles

Rec. 17: Align and harmonise FAIR and Open data policy

Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim 'as Open as possible, as closed as necessary' should be applied proportionately with genuine best efforts to share.

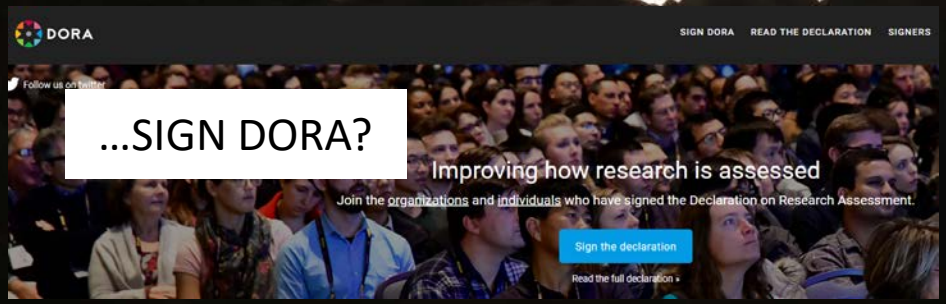
FAIR data – incentivi

The system of incentives and rewards must also be addressed in a fundamental way. From the perspective of measuring and rewarding research contributions, the full diversity of outputs should be taken into account including FAIR data, code, workflows, models, and other digital research objects as well as their curation and maintenance. In the 21st century, traditional publications and journal articles are far from being the only significant contributions to the advancement of knowledge.

…STOP WORSHIPPING IMPACT FACTOR

…AND JOURNAL ARTICLES

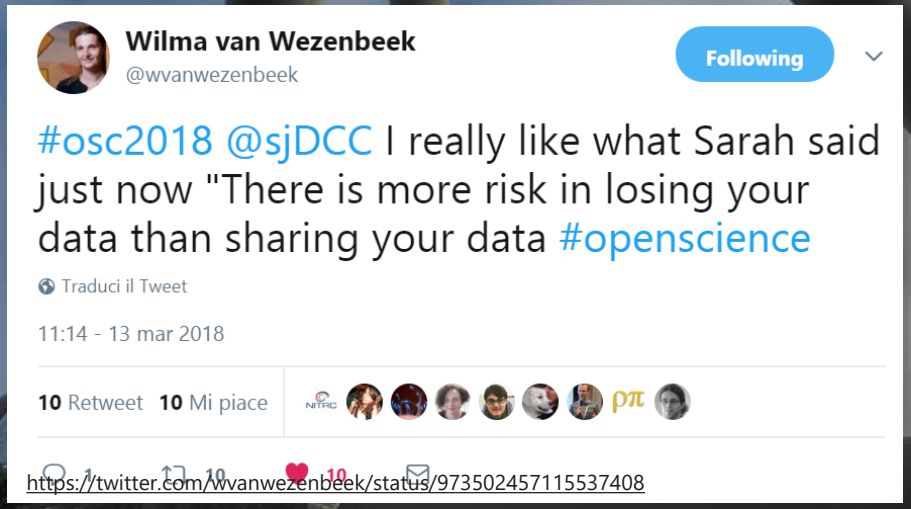
…SIGN DORA?



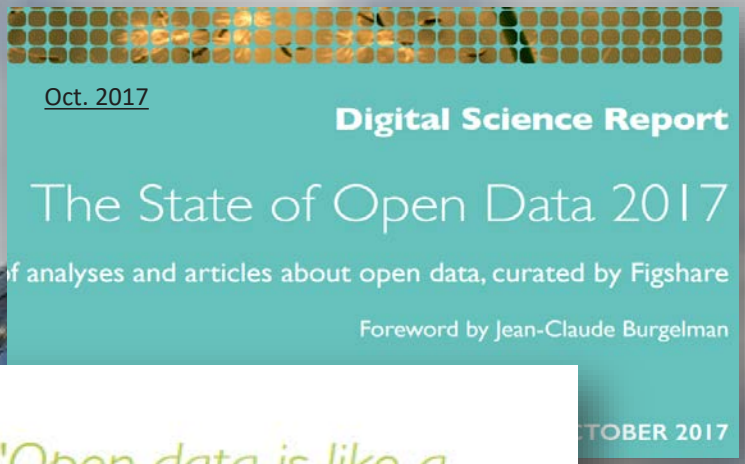
Turning FAIR into reality, 2018

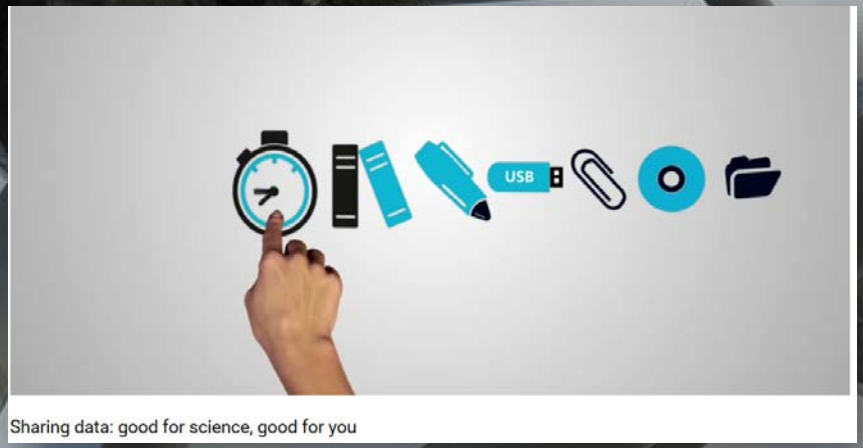
DATI APERTI

Perché i dati aperti?



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





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

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."

People will contact me to ask about stuff

YES, I KNOW. FRANKENSTEIN WAS THE DOCTOR, NOT THE MONSTER. FROM FLICKR BY CHOP SHOP GARAGE.

Christopher and Alex (C&A) say: "This is usually an objection of people who feel overworked and that [data sharing] isn't part of their job…" I would add to this that science is all about learning from each other – if a researcher is opposed to the idea of discussing their datasets, collaborating with others, and generally being a good science citizen, then they should be outed by their community as a poor participant.

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 misinterpret it by accident will be grateful for the help." From the UK Data Archive: "Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data."

It's worth mentioning, however, a second point C&A make: "Publishing may actually be useful to counter willful misrepresentation (e.g. of data acquired through Freedom of Information legislation), as one can quickly point to the real data on the web to refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how interesting or useful it is — even niche datasets hav people that care about them." I'd also add that it's impossible to decide wheth dataset has value to future research. Consider the many datasets collected bef "climate change" was a research topic which have now become invaluable to documenting and understanding the phenomenon. From the UK Data Archive:

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 burner due to other obligations? That said, this is a real concern.

C&A suggest the embargo route: "One option is to have an automatic or optional embargo; require people to archive their data at the time of creation but it becomes public after X months. You could even give the option to renew the embargo so only things that are no longer cared about become published, but nothing is lost and eventually everything can become open." Researchers like to have a say in the use 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 complicated, it could harm your professional [standing]." I would add that if it's too complicated to share, then it's too complicated to reproduce, which means it's arguably not real scientific progress. This can be solved by more documentation.

My data is embarrassingly bad

C&A: "Many eyes will help you improve your data (e.g. spot inaccuracies)… people will accept your data for what it is." I agree. All researchers have been on the back end of making the sausage. We know it's not pretty most of the time, and we can accept that. Plus it helps you strive will be at managing and organizing data during your next collection phase.

It's not a priority and I'm busy

Good news! Funders are *making* it your priority! New sharing mandates in the OSTP memorandum state that any research conducted with federal funds must be accessible. You can expect these sharing mandates to drift down to you, the researcher, in the very near future (6-12 months).

CARLY STRASSER

Open Science + Data Sharing Advocate

Home

Blog Posts

Previous Research

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

Closed Data… Excuses, Excuses

… «as open as possible»…

L 172/56 EN Official Journal of the European Union 26.6.2019

DIRECTIVE (EU) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 20 June 2019

on open data and the re-use of public sector information

(recast)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Carlos Moedas
@Moedas

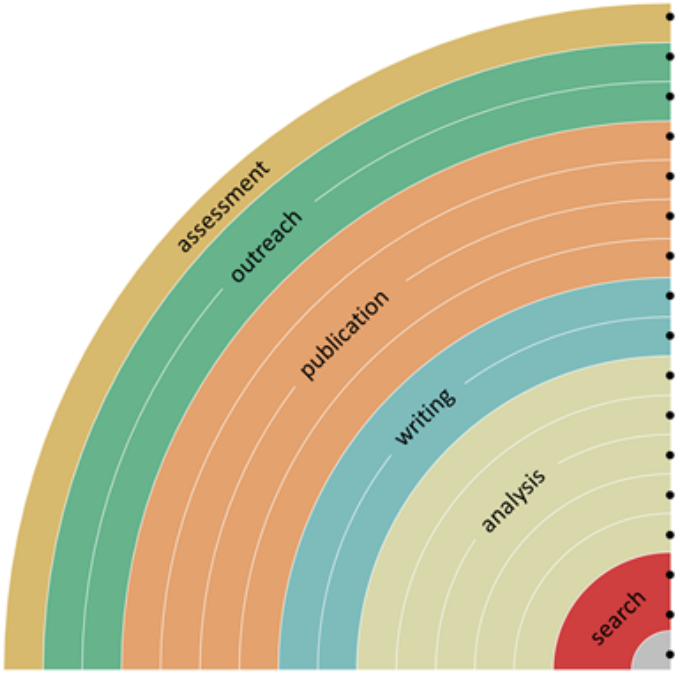
Segui

2/4 "Open as possible, as closed as necessary" is the new principle for all #data from publicly funded #research in Europe #openaccess

RETWEET 76 MI PIACE 32

Open Science: perché solo i dati?

You can make your workflow more open by …



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

DOI: 10.5281/zenodo.1147025

Pausa?