OPEN SCIENCE DALLA A ALLA Z
5 – GESTIONE DEI DATI

https://doi.org/10.5281/zenodo.4483434

Elena Giglia
Università di Torino
elena.giglia@unito.it
@egiglia

INRIM,
febbraio 2021

In questo modulo impareremo:

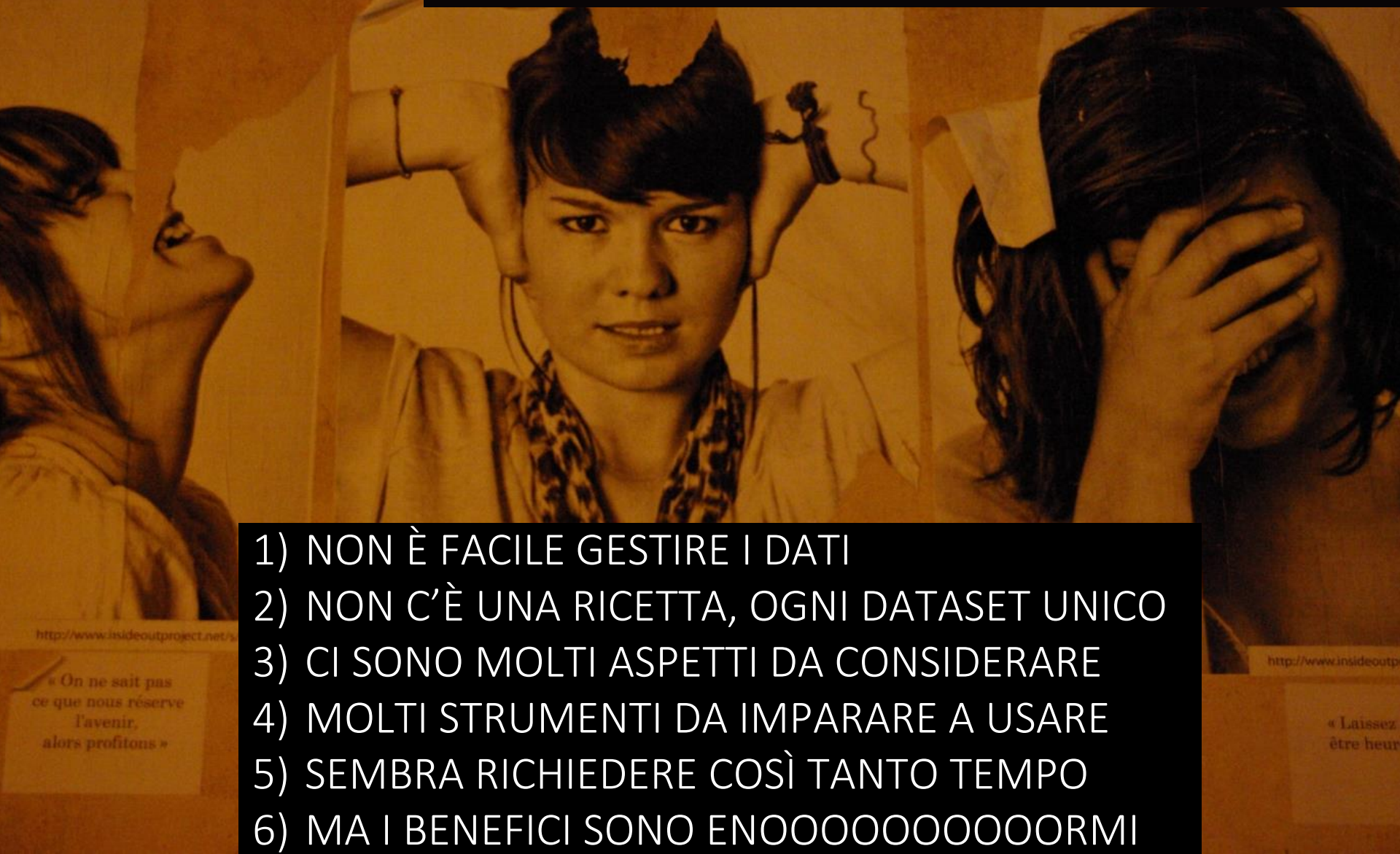
1. come gestire i propri dati correttamente e rendere la ricerca più efficace
2. la differenza fra dati FAIR e Open

MESSAGGI CHIAVE

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

… lo so…

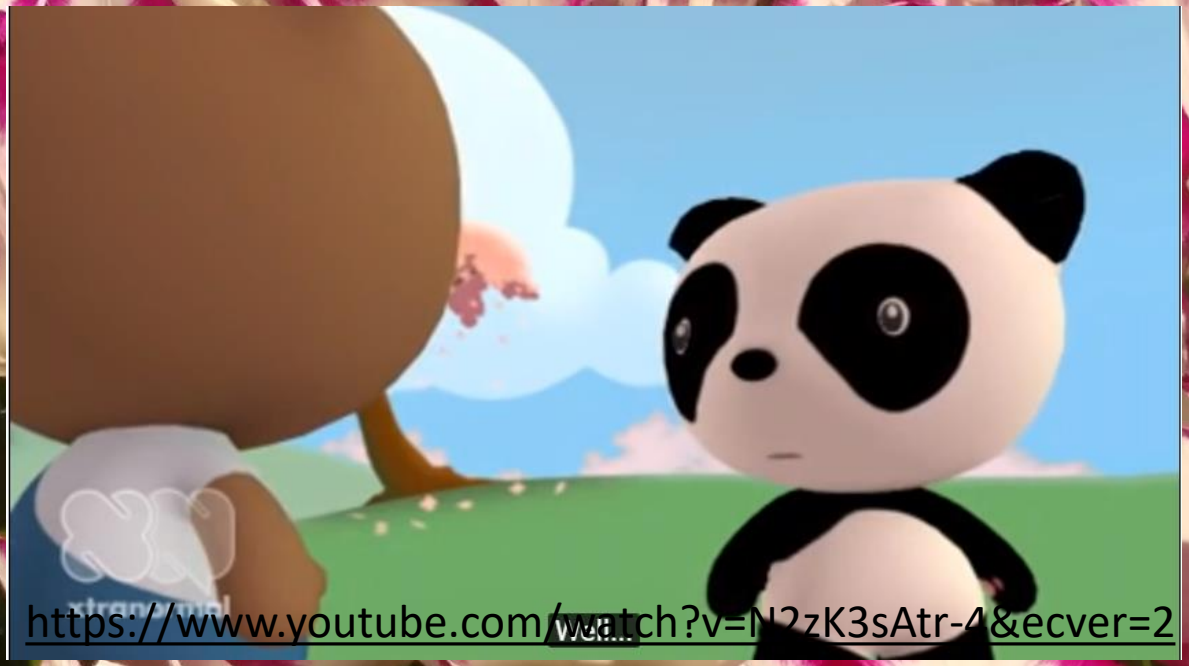
NON USCIRETE DI QUI OGGI CON UN DMP PRONTO
MA CON GLI STRUMENTI PER SCRIVERLO – OGNUNO DIVERSO



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 ENOOOOOOOOOORMI

Perché occuparci dei dati?

[Video]



… È L'INCUBO DEL DATA STEWARD:

- NESSUN BACKUP
- NESSUN SOFTWARE DI ACCOMPAGNAMENTO
- NESSUNA LEGENDA DATI

… E IN PIÙ:

- DATI PRODOTTI CON FONDI PUBBLICI
- PUBBLICATI SU SCIENCE CHE LI RICHIEDE
- UTILI A UNA RICERCATRICE DI AREA DIVERSA

Perché occuparci dei dati?

1.1 PERCHÉ DOBBIAMO. DIRETTIVA OPEN DATA

Open data directive

26.6.2019 IT Gazzetta ufficiale dell'Unione europea L 172/56

DIRETTIVA (UE) 2019/1024 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO

del 20 giugno 2019

relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico

(rifusione)

- Stimulate the publishing of dynamic data and the uptake of Application Programme Interfaces (APIs).
- Limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data.
- Enlarge the scope of the Directive to:
  - data held by public undertakings, under a specific set of rules. In principle, the Directive will only apply to data which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination;
  - research data resulting from public funding – Member States will be asked to develop policies for open access to publicly funded research data. New rules will also facilitate the re-usability of research data that is already contained in open repositories.
- Strengthen the transparency requirements for public–private

I DATI DELLA RICERCA SONO INCLUSI NELLA DIRETTIVA «OPEN DATA».
DEVONO ESSERE APERTI

…perché occuparci dei dati?

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

https://goo.gl/sryNTg

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

What?

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

…he Open Resear… …C actions, SME i… …tions, and prizes…

…le **the default**. However, actions may **opt out** at any stage — …ds (through an amendment; *see Article 55*) —, if:

…th the obligation to protect results *(see Article 27)*

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

DATA MANAGEMENT PLAN (DMP)
- DELIVERABLE
ENTRO M6
DA AGGIORNARE (MID TERM/END)

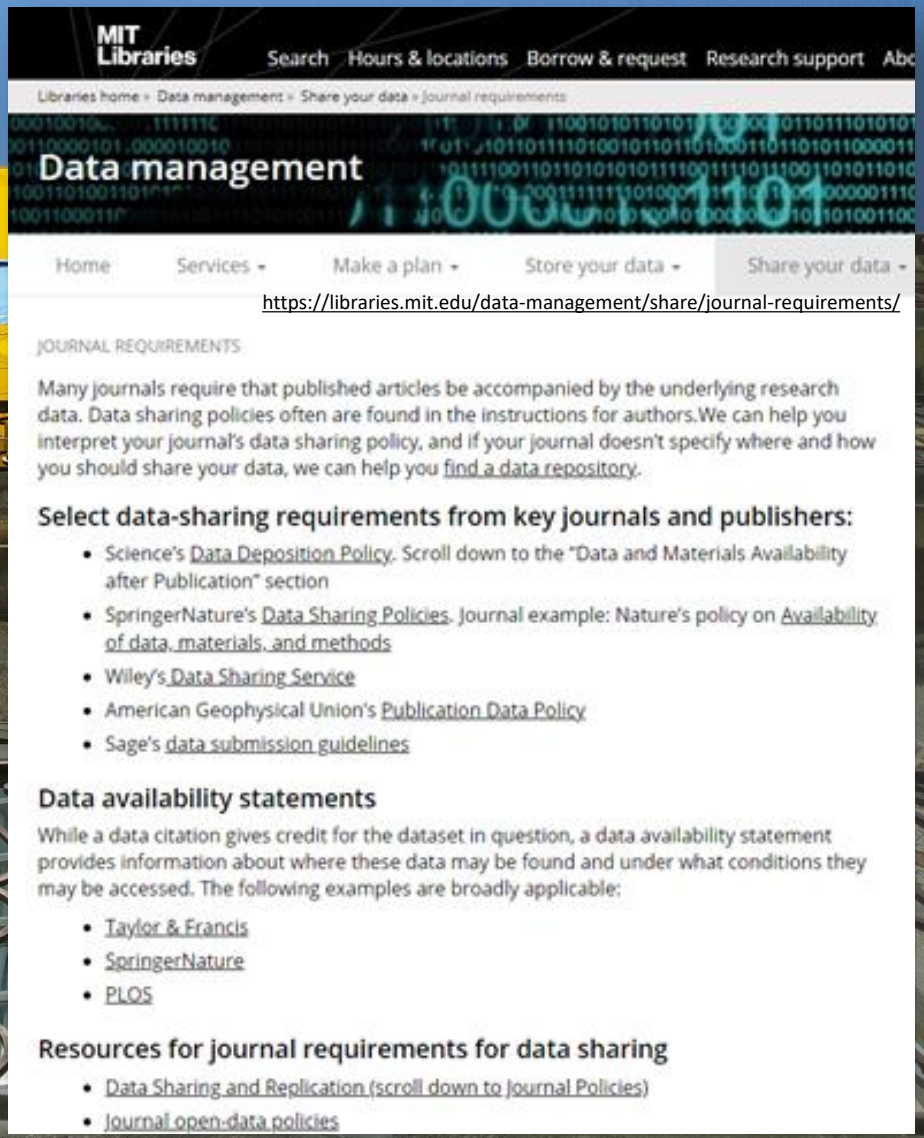
1.2 PERCHÉ DOBBIAMO.
HORIZON 2020 E HORIZONEUROPE

GRANT AGREEMENT ART. 29.3

PROGETTO PILOTA ESTESO A TUTTE LE AREE
CLAUSOLE OPT OUT
PRINCIPIO: «AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»

Perché occuparci dei dati?

1.3. PERCHÉ DOBBIAMO. UN NUMERO CRESCENTE DI RIVISTE CHIEDE DI PUBBLICARLI INSIEME ALL'ARTICOLO



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

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

Data management

Home Services Make a plan Store your data Share your data

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

JOURNAL REQUIREMENTS

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

Select data-sharing requirements from key journals and publishers:

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

Data availability statements

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

- Taylor & Francis
- SpringerNature
- PLOS

Resources for journal requirements for data sharing

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

Perché occuparci dei dati?

2. PERCHÉ CI CONVIENE.
NELLE CRISI SI CAPISCE LA LORO IMPORTANZA

The Value of RDA for COVID-19

Home » Get involved » The Value of RDA for... » The Value of RDA for COVID-19

13 July 2020 | 16426 reads | Facebook | Twitter

Under public health emergencies, and particularly the COVID19 pandemic, it is fundamental that data is shared in both a timely and an accurate manner. This coupled with the harmonisation of the many diverse data infrastructures is, now more than ever, imperative to share preliminary data and results early and often. It is clear that open research data is a key component to pandemic preparedness and response.

RDA

RESEARCH DATA ALLIANCE

https://www.gisaid.org/

GISAID 2008 - 2020

About us | Database Features | Events | Collabor

In Focus

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

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

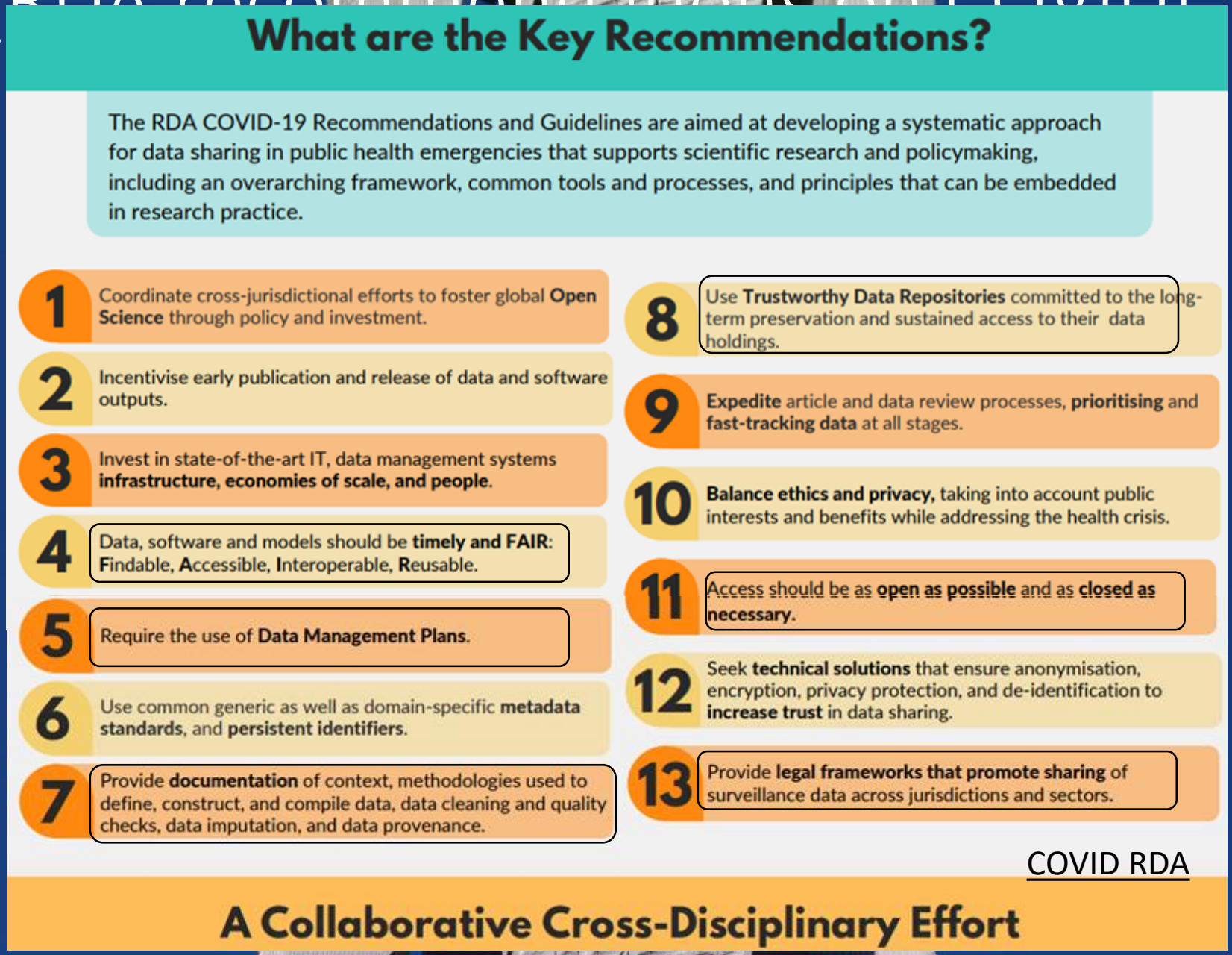
CORD-19

COVID-19 Open Research Dataset

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

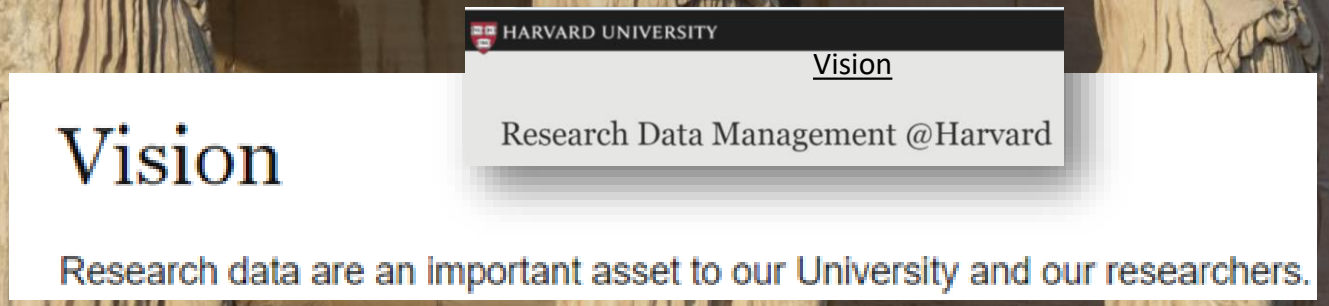
https://www.semanticscholar.org/cord19

[RDA recommendations on COVID]



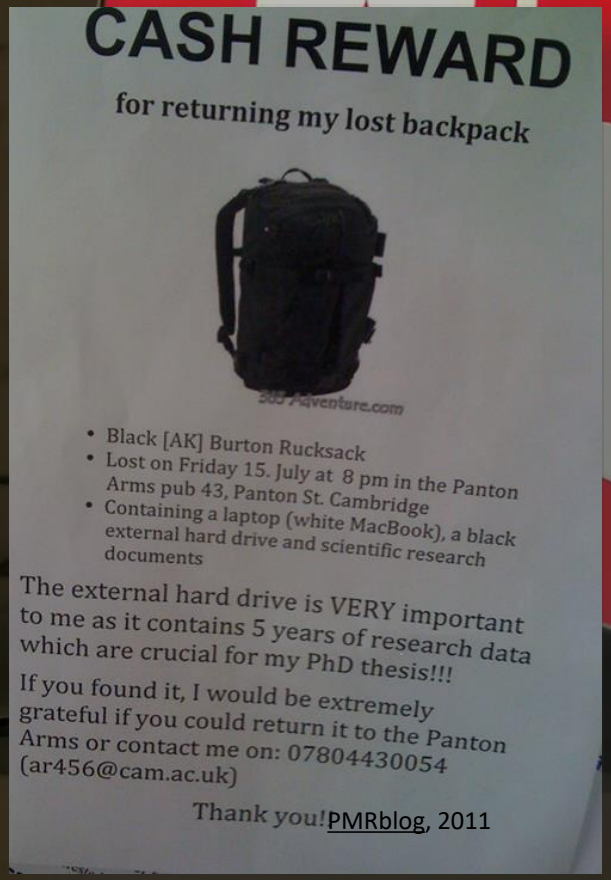
Perché occuparci dei dati?

3. PERCHÉ SONO IL FONDAMENTO DI UNA SCIENZA SOLIDA E DI UNA RICERCA RESPONSABILE



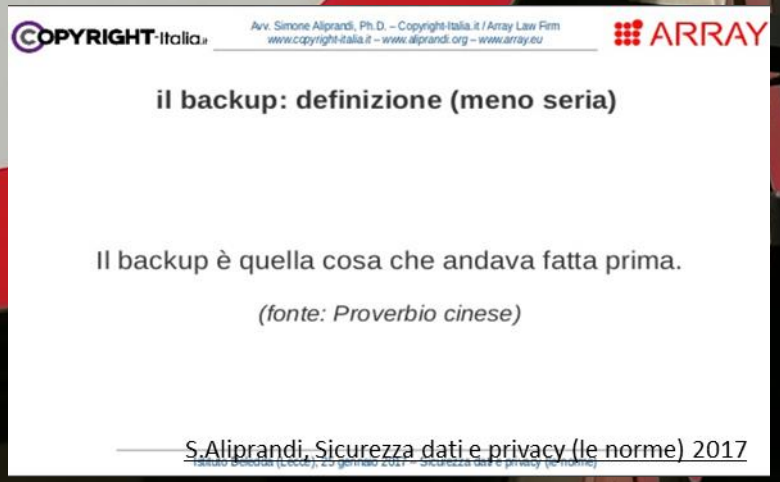
DCC because good research needs good data

…perché occuparci dei dati?



PMRblog, 2011

4. PER INTERESSE. PERCHÉ I DATI SI PERDONO



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

Perché occuparci dei dati?

Scientists losing data at a rapid rate

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

Elizabeth Gibney & Richard Van Noorden

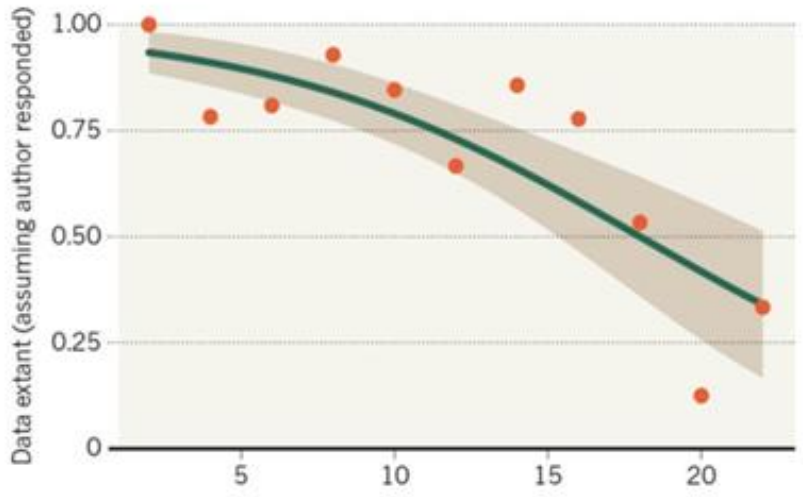
19 December 2013

Rights & Permissions

80% PERSI
ENTRO 20 ANNI

MISSING DATA

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



2013

5. PERCHÉ I DATI SONO FRAGILI… E DOVERLI PRODURRE DI NUOVO COSTA

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

Perché occuparci dei dati?

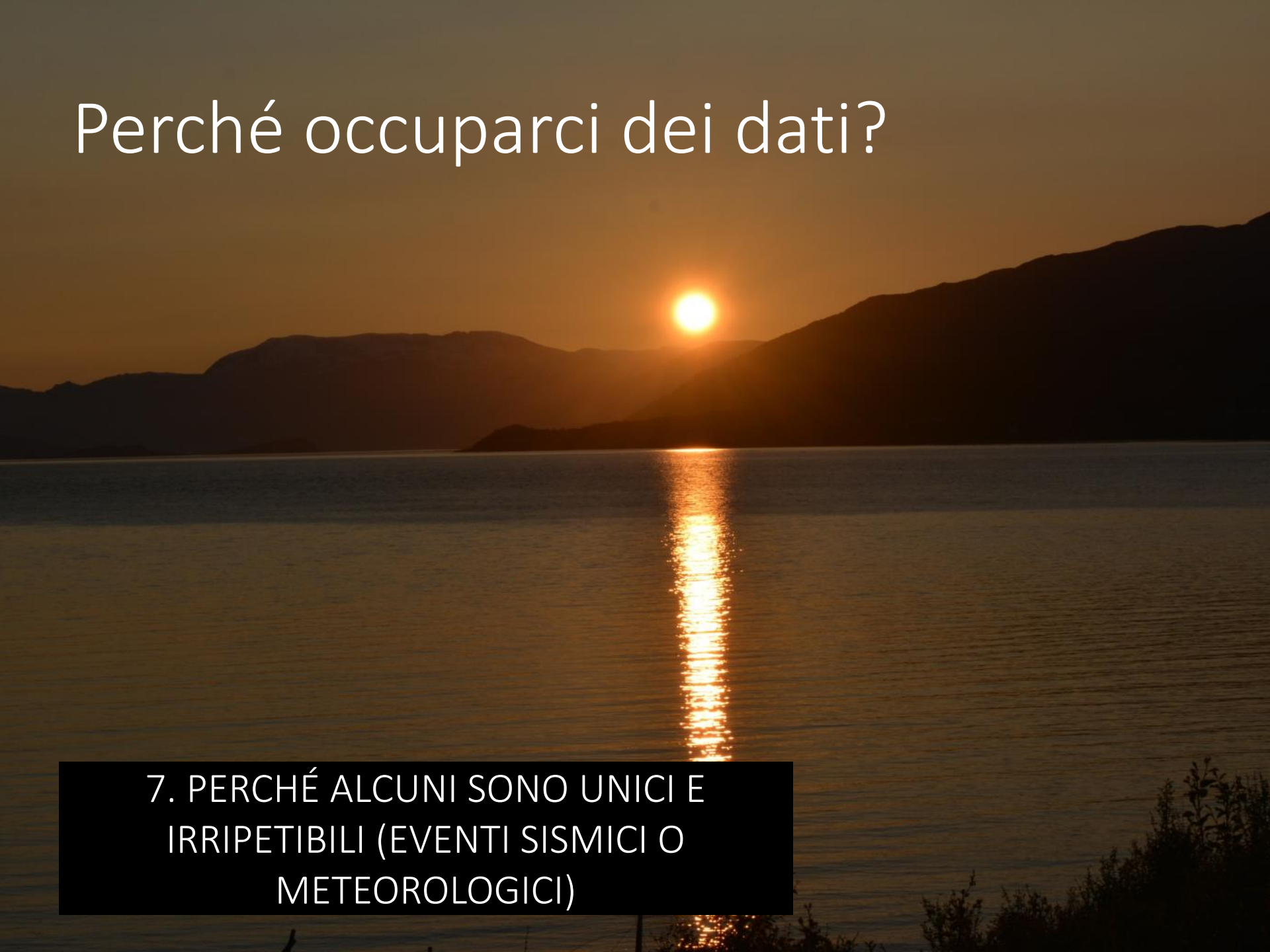
Nikolai Ivanovich Yezhov was head of the People's Commissariat for Internal Affairs until fell from Stalin's favor and power. Among art historians, he also has the nickname "The Vanishing Commissar" because after his ecution, his likeness was retouched out of an official ess photo; he is among the best-known examples of e Soviet press making someone who had fallen out of vor "disappear".

e Newseum (1 September 1999). " "The Commissar Vanishes" in The Vanishing Commissar". hived from the original on 8 February 2007.

6. PER GARANTIRE INTEGRITÀ. I DATI POSSONO ESSERE MANIPOLATI…È VOSTRO INTERESSE PRIMARIO EVITARLO

Lazzeri-Di Donato RDM, 2020

Perché occuparci dei dati?



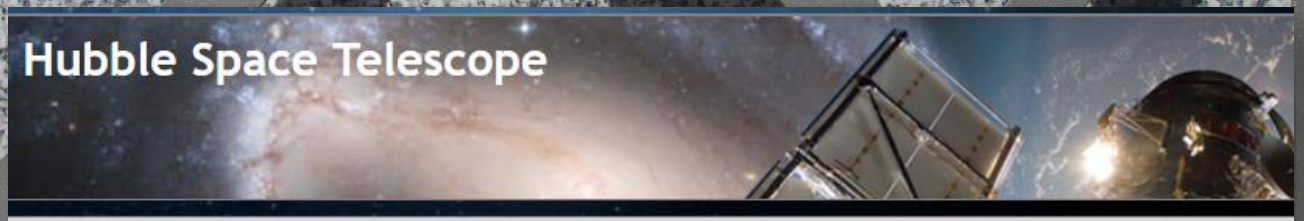
7. PERCHÉ ALCUNI SONO UNICI E IRRIPETIBILI (EVENTI SISMICI O METEOROLOGICI)

Perché occuparci dei dati?

8. PERCHÉ POSSONO ESSERE RIUTILIZZATI

… SPESSO IN MODO INEDITO

«THE COOLEST THING TO DO WITH YOUR DATA WILL BE THOUGHT OF BY SOMEONE ELSE» [R.POLLOCK]



News

Text Size

Astronomers Find Elusive Planets in Decade-Old Hubble Data **10.06.11**

In a painstaking re-analysis of Hubble Space Telescope images from 1998, astronomers have found visual evidence for two extrasolar planets that went undetected back then.

Finding these hidden gems in the Hubble archive gives astronomers an invaluable time machine for comparing much earlier planet orbital motion data to more recent observations. It also demonstrates a novel approach for planet hunting in archival Hubble data.

Exoplanet HR 8799 System

Perché occuparci dei dati?

9. PERCHÉ L'ACCESSO AI DATI FAVORISCE L'INNOVAZIONE

OECD data

Perché occuparci 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

Missions in Horizon Europe

Mission areas

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

Mission area: Adaptation to climate change including societal transformation

Mission area: Cancer

Mission area: Climate-neutral and smart cities

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

Mission area: Soil health and food

Horizon Europe

10. I DATI CREANO PONTI FRA LE DISCIPLINE…
…E NON È INDIFFERENTE PER LE MISSIONS DI HORIZON EUROPE…

Perché occuparci dei dati?

11. PER ESSERE RIPRODUCIBILI

12. PER L'INTEGRITÀ DELLA RICERCA

Perché occuparci di dati?

2015

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

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

2013

13. PER PERMETTERE VALIDAZIONI E CONTROLLI (E SCOPRIRE ERRORI)

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

- ESCLUSIONE SELETTIVA DI DATI
- SCHEMA NON CONVENZIONALE DI PESATURA DEI DATI
- ERRORE NEL FOGLIO DI CALCOLO PER SELEZIONARLI

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

scienzainrete

Jan 19, 2021

Sostieni Scien

Curiosamente, la posizione espressa dal professor Doshi è stata criticata non su basi scientifiche (che so: errori di metodo, ambiguità negli obiettivi da raggiungere, utilizzo di software e strumentazioni inadatte) ma dell'opportunità politica e del principio di autorità. Cioè su presupposti **diametralmente opposti** a quelli di una

Le ragioni di Peter Doshi sui vaccini: "Fidati, ma verifica"

COVID-19/Filosofia

di Andrea Monti

Chi pratica il metodo scientifico ha la testarda abitudine (incomprensibile ai più) di trarre conclusioni dall'analisi di dati secondo i criteri di un'ipotesi di ricerca e applicando un metodo che consente la verificabilità intersoggettiva dei risultati. Questo atteggiamento mentale è diametralmente opposto a chi basa le proprie opinioni e—peggio— decisioni sulla "fiducia" (spesso tramutata in "fede") e dunque sull'autorità di eminenze varie. Non discuto questo atteggiamento nell'ambito religioso; ma in quello laico, quello della scienza, sì. Se un dogma esiste, nella pratica del metodo scientifico, è quello della **metodicità del dubbio**, insieme a quello dell'**assenza di certezze**. Un teoria sp
quando vale. È successo, tanto per fare un es
di fuori dei laboratori, con la gravitazione di

thebmjopinion Jan. 4, 2021 Latest Authors Topic

Peter Doshi: Pfizer and Moderna's "95% effective" vaccines—we need more details and the raw data

January 4, 2021

Five weeks ago, when I raised questions about the results of Pfizer's and Moderna's covid-19 vaccine trials, all that was in the public domain were the study protocols and a few press releases. Today, two journal publications and around 400 pages of summary data are available in the form of multiple reports presented by and to the FDA prior to the agency's emergency authorization of each company's mRNA vaccine. While some of the additional details are reassuring, some are not. Here I outline new concerns about the trustworthiness and meaningfulness of the reported efficacy results.

Perché occuparci dei dati?

14. PER UNA SCIENZA SOLIDA

UN ARTICOLO SENZA I DATI È SOLO LA «PUBBLICITÀ» DELLA RICERCA

1995

WaveLab and Reproducible Research

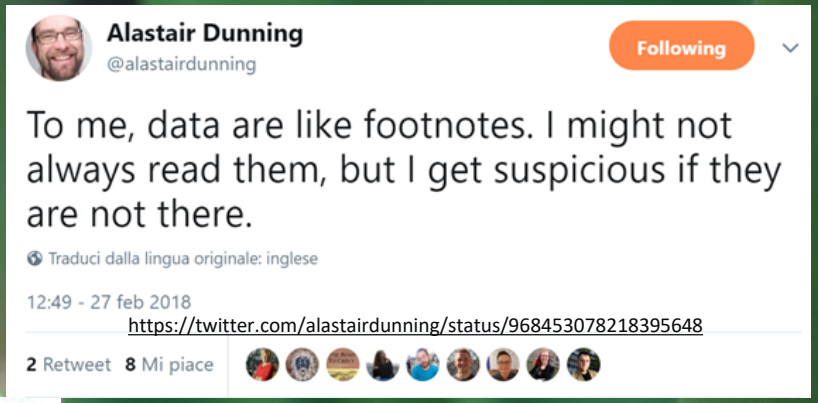
Jonathan B. Buckheit and David L. Donoho

Stanford University, Stanford CA 94305, USA

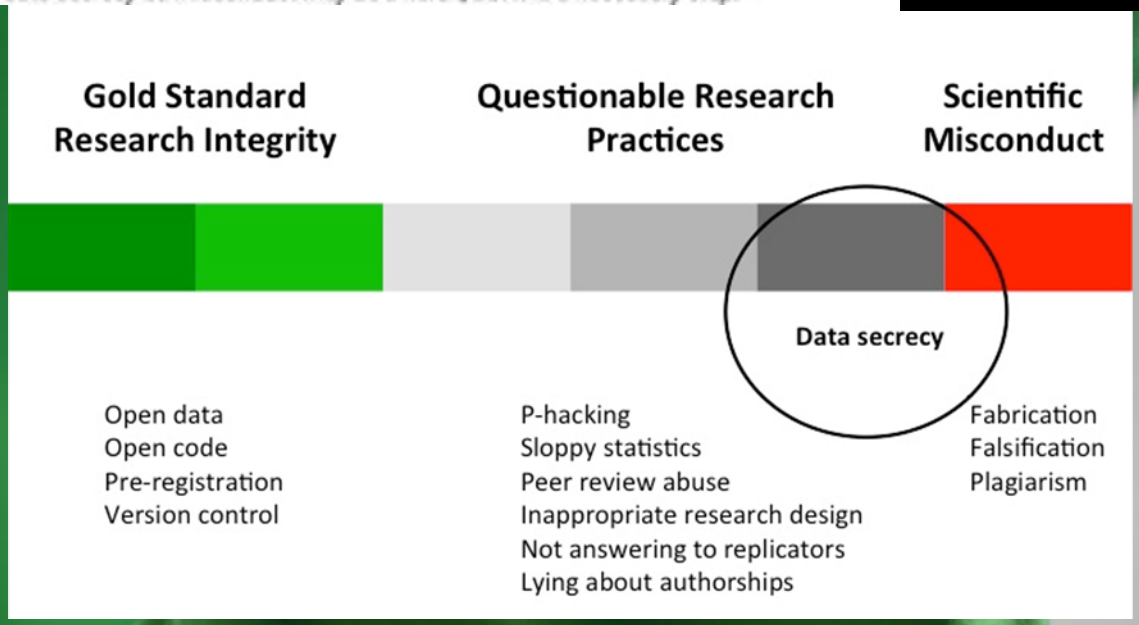
An article about computational science in a scientific publication is **not** *the scholarship itself, it is merely* **advertising** *of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

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

Nessun dato?

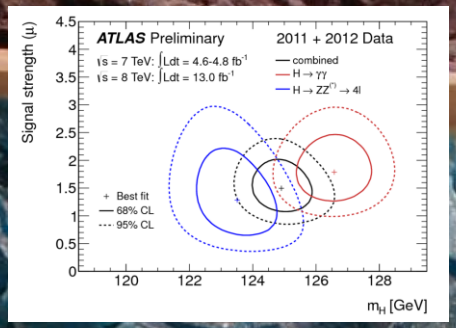


NESSUN DATO? PIGRIZIA O FRODE? I DATI COME NOTE A PIE' PAGINA: POSSO NON LEGGERLE, MA DIVENTO SOSPETTOSO SE NON CI SONO



Parliamo di dati

> *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.*



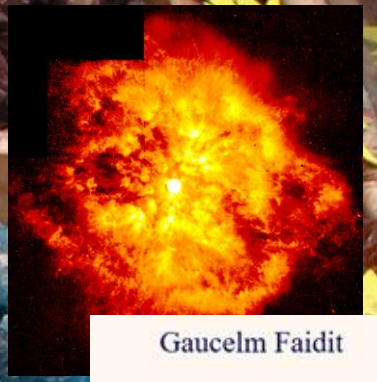
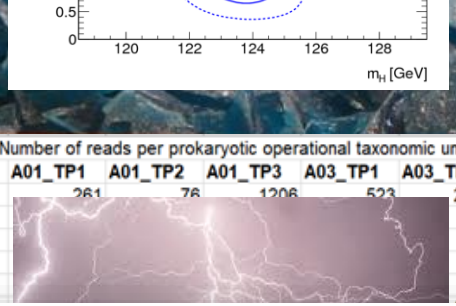


Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample fro							
OTU	A01_TP1	A01_TP2	A01_TP3	A03_TP1	A03_TP3	A04_TP1	A04_T
OTU_1	261	76	1206	523	2131	25707	6
OTU_2					206	119	
OTU_9					151	73	2
OTU_6					8	382	
OTU_4					17	15	



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

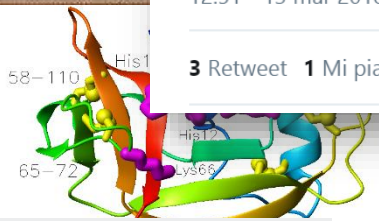
#osc2018 Wolfram Horstmann wants us to talk about datadiversity, like we do with biodiversity #openscience

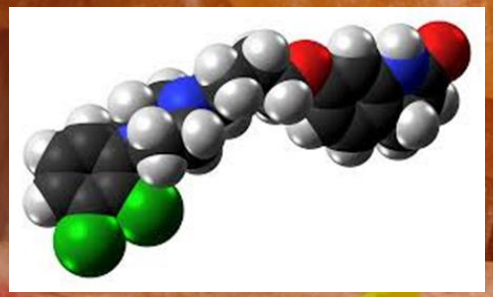
Traduci il Tweet

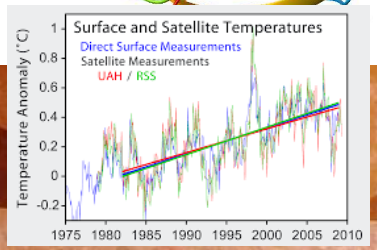
12:51 - 13 mar 2018

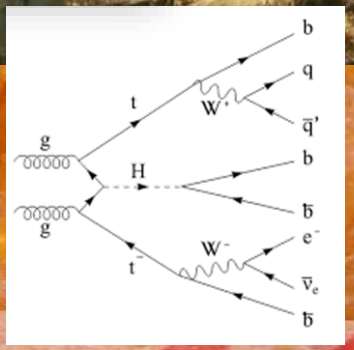
3 Retweet **1** Mi piace

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

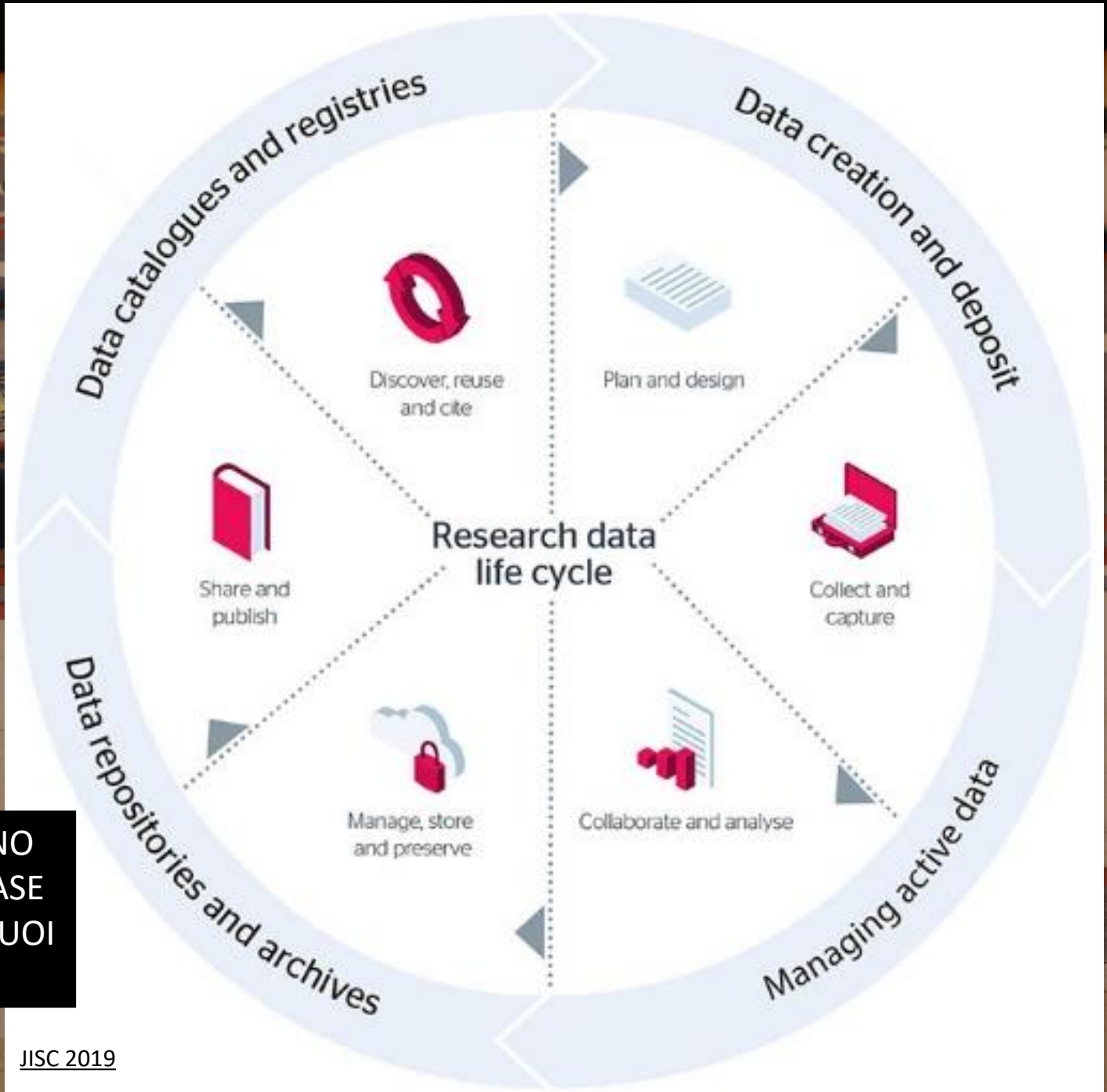








Il cicl

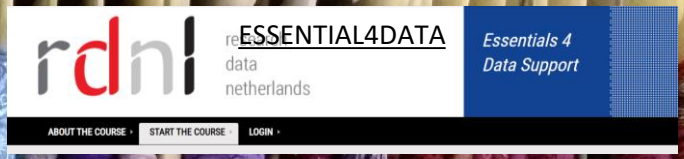


I DATI NON SONO STATICI. OGNI FASE DEL CICLO HA I SUOI STRUMENTI

JISC 2019

Le basi

[DMP]



ESSENTIAL4DATA

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


…un passo indietro…

[il fondamento] [DMP]

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

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

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

OpenAIRE

Guides for Researchers

How do I know if my research data is protected?

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

OpenAIRE

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

Wainer Lusoli
@w_lusoli

Following

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

Traduci dalla lingua originale: inglese

Repeat with me: research data is not mine

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

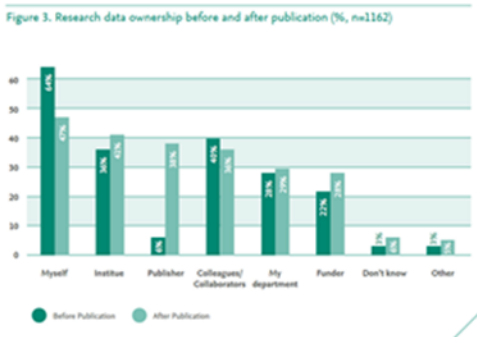
linkedin.com

11:18 - 12 apr 2017

14 Retweet **18** Mi piace

Lusoli, Apr.2017

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



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

Of course, the study reports

[DMP]

[webinar]

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

[i tre passi fondamentali]

Dati Open

Dati FAIR

Dati gestiti correttamente

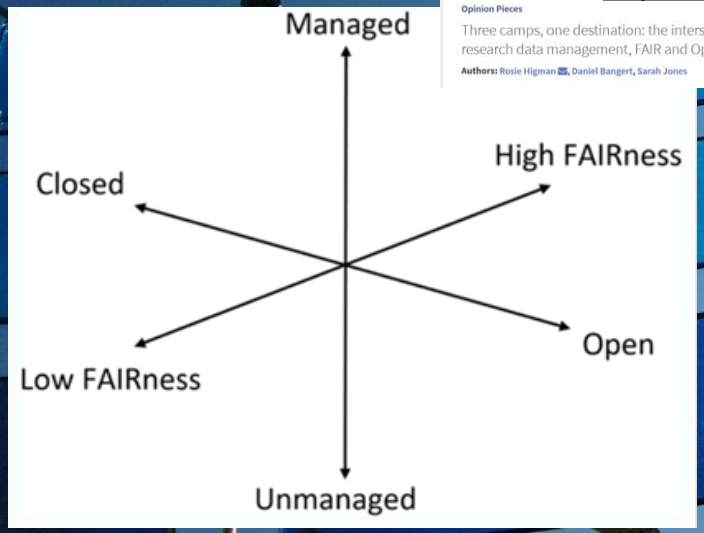
1. I DATI DEVONO ESSERE «AS OPEN AS POSSIBLE»

2. MA SE I DATI NON SONO «FAIR», APRIRLI COMPORTA RISCHI (USO SCORRETTO, CATTIVE INTERPETAZIONI, …)

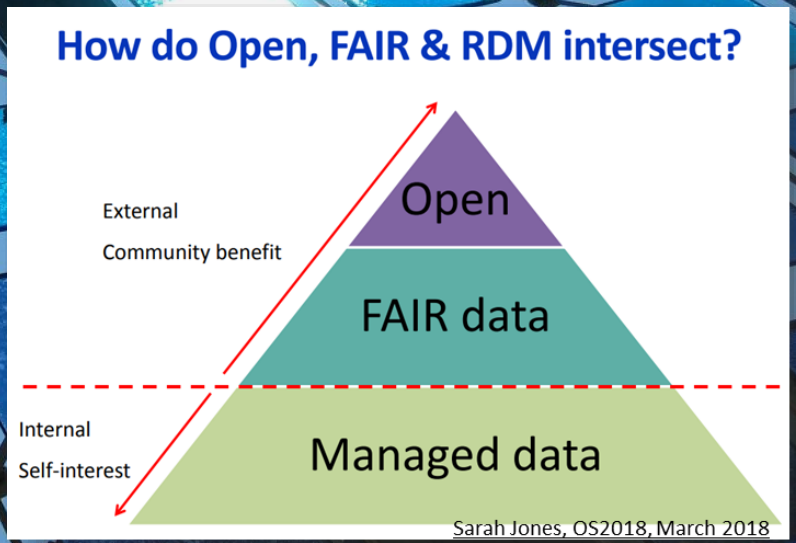
3. MA SE I DATI NON SONO CORRETTAMENTE GESTITI, RENDERLI «FAIR» COSTA TROPPO TEMPO E DENARO

E GESTIRE I DATI CORRETTAMENTE È NELL'INTERESSE PRIMARIO DEL RICERCATORE, PERCHÉ L'INTERA RICERCA SCORRE PIÙ FLUIDA

[i tre passi fondamentali]

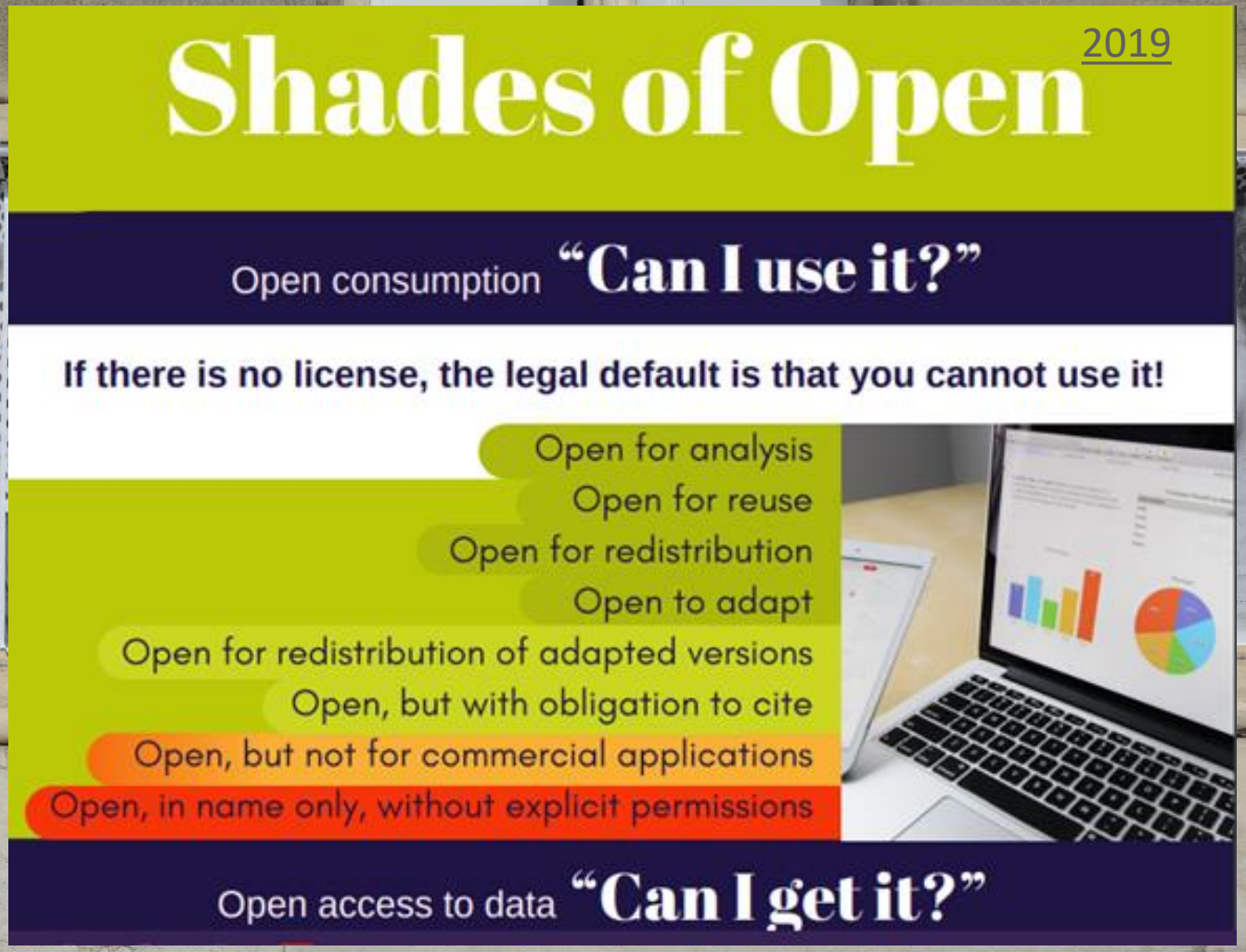


May 27, 2019



Sarah Jones, OS2018, March 2018

Open non FAIR è Open???



1. I dati vanno gestiti

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

2. I dati DEVONO essere FAIR

To be Findable:

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

TO BE ACCESSIBLE:

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

TO BE INTEROPERABLE:

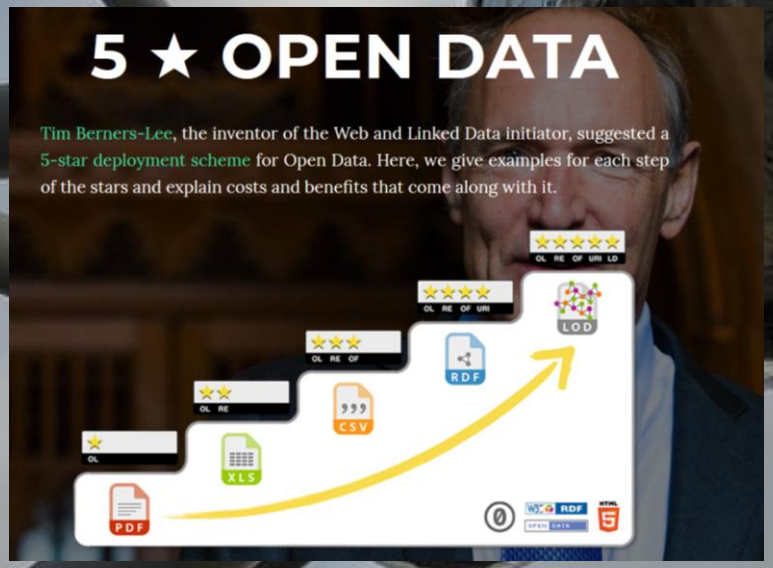
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

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

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

3. I dati POSSONO essere Open



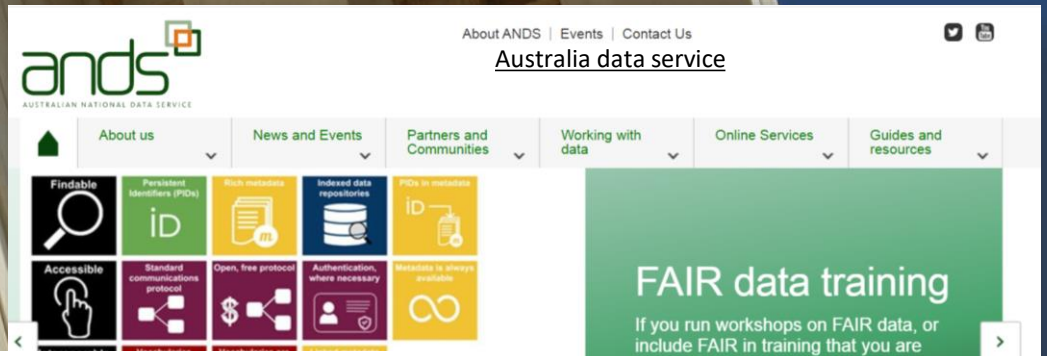
Pausa?

[una premessa]

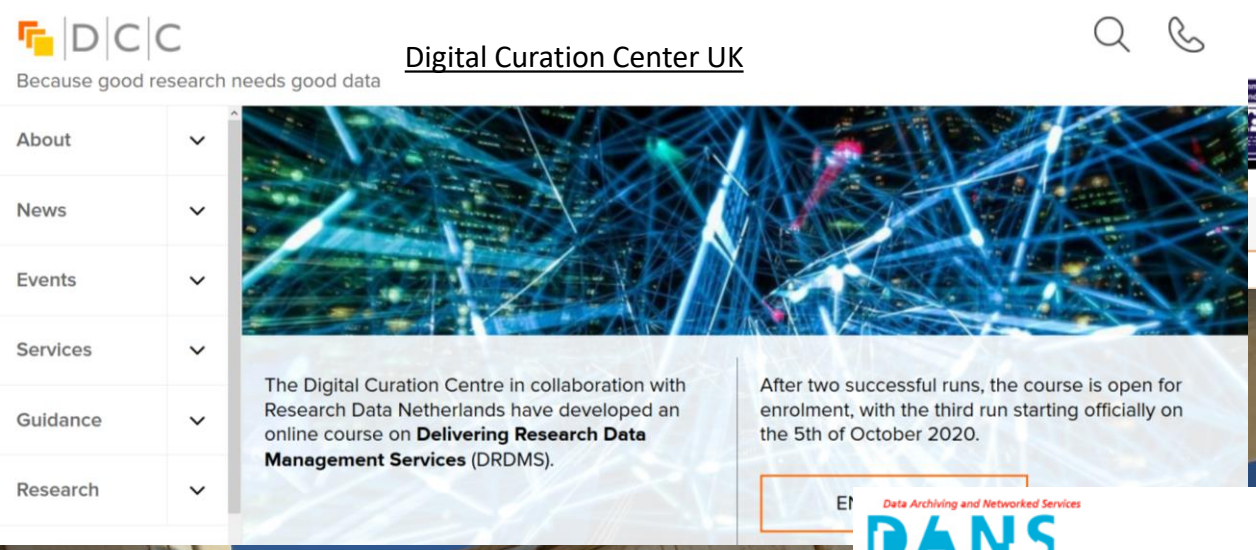
- …DA QUI IN POI: PANORAMICA SUGLI STRUMENTI
- VANNO «ASSAGGIATI» E ADATTATI AL PROPRIO CONTESTO…
- IMPARARE A USARLI PER SUGGERIRLI (IMPENSABILE CHE UN RICERCATORE SCENDA COSÌ NEL DETTAGLIO)
- FONDAMENTALE IL SUPPORTO
- **FONDAMENTALE UNA POLITICA ISTITUZIONALE CHE CHIARISCA RUOLI E RESPONSABILITÀ E DEFINISCA IL LIVELLO DEI SERVIZI**
- CREARE UNA **RETE DI DATA STEWARDS** [COMPETENZE DI DOMINIO + TECNICHE]

[DMP]

4 pilastri

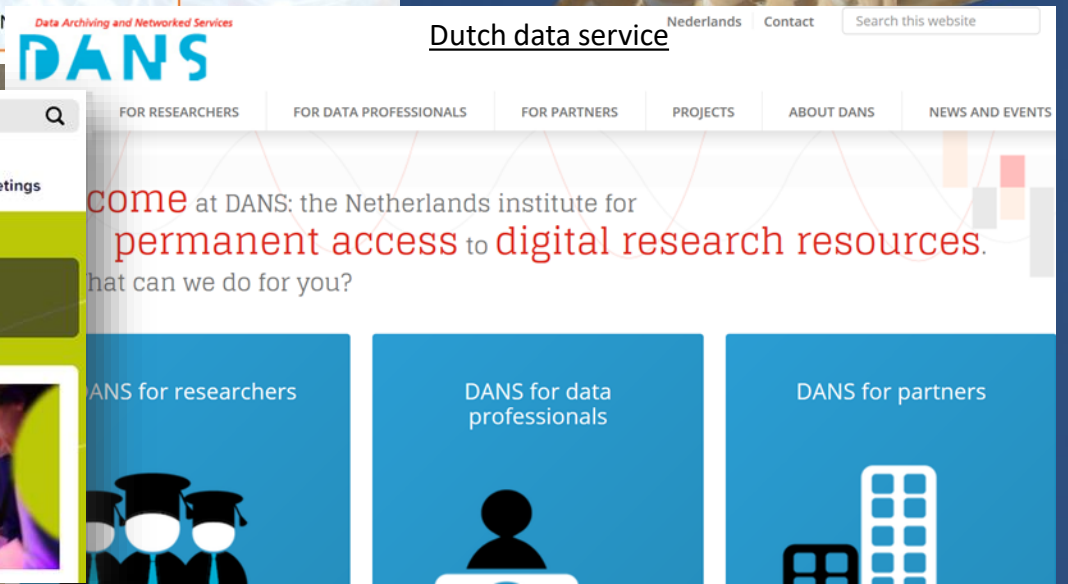


Australia data service



Digital Curation Center UK

Dutch consortium

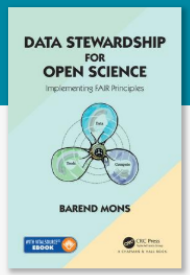


Dutch data service

… 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

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

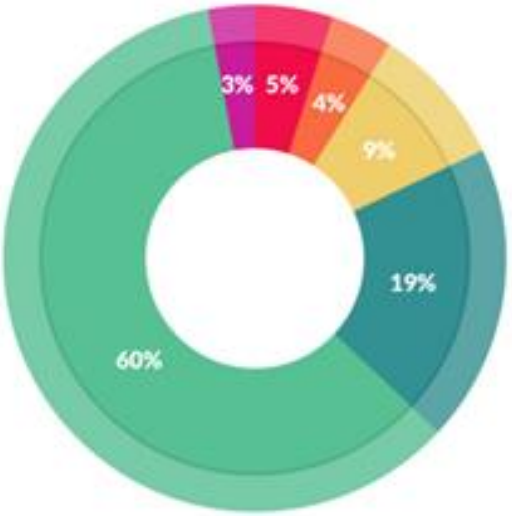
I tell research institutions that, on average, 5% of overall research costs should go towards data stewardship. With €300 billion (US$325 billion) of public money spent on research in the European Union, we should expect to spend €15 billion on data stewardship. Scientists, especially more experienced 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.

Costi



What data scientists spend the most time doing

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

Data science report, 2016, cit. by Erik Schultes

CI SONO COSTI PER GESTIRE, RENDERE FAIR E CONSERVARE I DATI…

MA PENSIAMO

- A QUANTO COSTEREBBE **NON CONSERVARLI E NON GESTIRLI**
- **A QUANTO TEMPO PERDETE PER «PULIRLI» PRIMA DI POTERLI USARE**

Costi

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

OpenAIRE

What will it cost to manage and share my data?

What to cost in?

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

...and

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

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

Things to consider...

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

Useful costing guides:

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

May 15, 2020

How much could management & deposit cost?

Some factors that affect RDM costs...

Security of potenially sensitive data

Dataset size

Length of preservation required

Remember:

Different repositories apply different charging models. Some apply a per data package plus an amount over a certain volume, while oth apply variable fees depending on the data volume. Some may not char

[DMP]

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

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

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

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

Developed for:

OpenAIRE by DCC

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

More OpenAIRE Advance resources available here:

[DMP]

Cost

DM costing tool

What's Zingtree?

Zingtree

Data Management Costing Tool

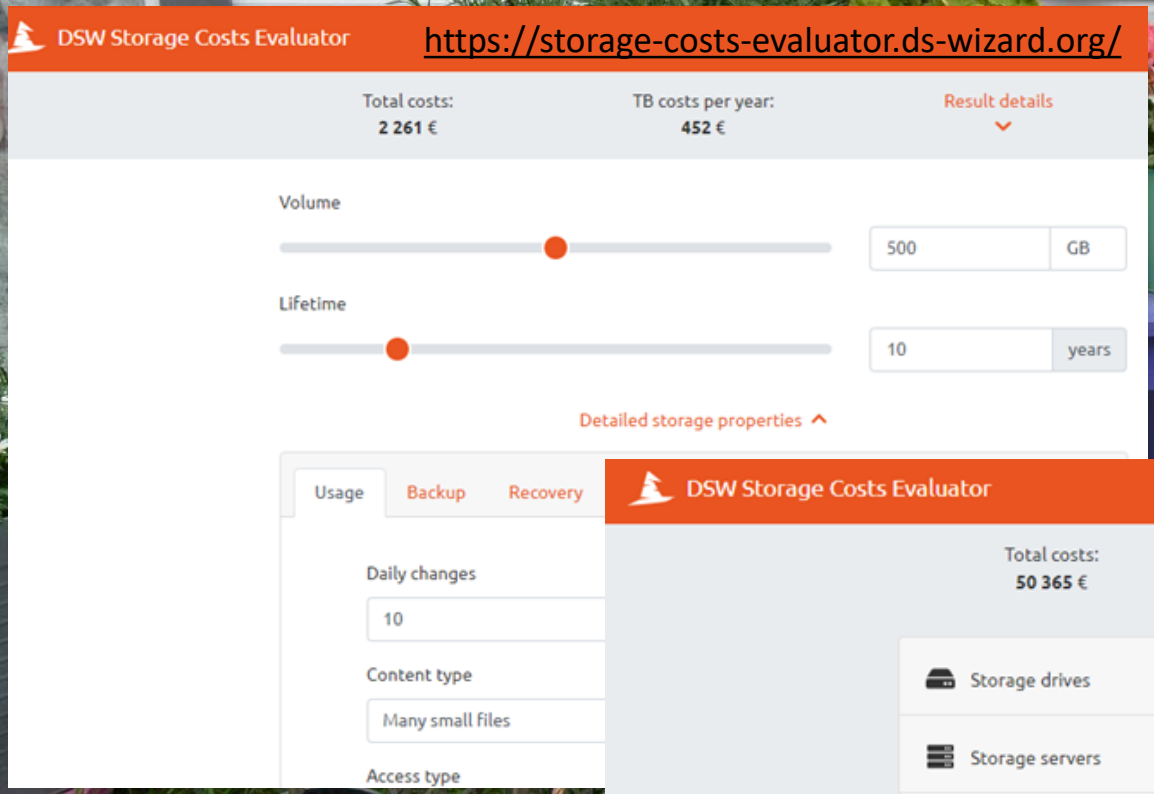
Data Management costing tool

4TU. CENTRE FOR RESEARCH DATA

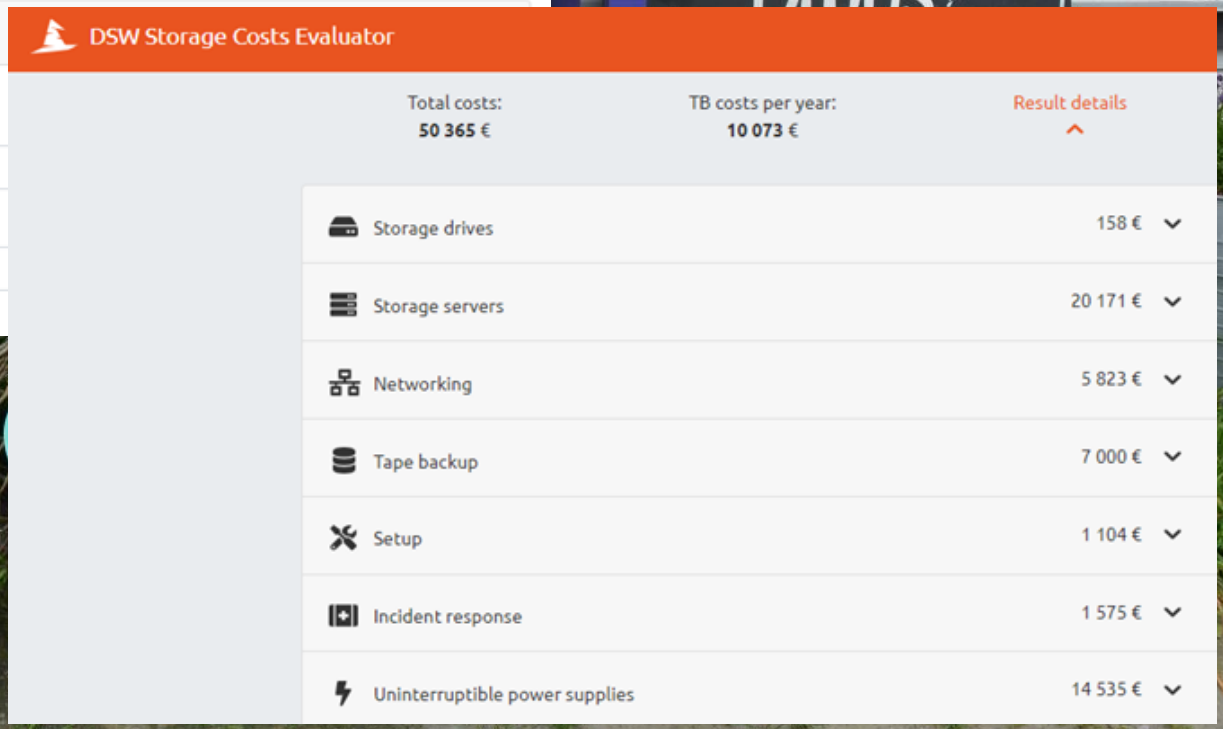
TU Delft

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.

Costs evaluator – Data Wizard



[DMP]



1. GESTIRE I DATI

… dal primo giorno

Research Data Management:
Get it right from the beginning

May 2018

CCO/ this picture needs no attribution

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

PERCHÉ RISPARMIATE
ENERGIA E LA RICERCA
SCORRE FLUIDA


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

Main Points for Good Data Management

Data acquisition

- Check the type, source of the data and how to gather/collect it
  - Data types (to help define sensitivity of
  - Data format (to help define the tools an
  - Data size (to help define storage and in
- Check the ownership of the collected and proce
  - Check with the data source about conditions (e.g. licence)
  - Check the need to make a data proce on the ownership / access control
  - Are there (own) institutional policies tha
  - Can the data be shared with other parti
- Confidentiality of the data (if applicable):
  - Register crucial information regarding d
  - Ensure security of confidential data (p data, or data that would harm society w
  - Ensure compliance with General Dat verordening gegevensbescherming whe
  - Ensure there are procedures in place to of a privacy advisor/data protection offic

Data collection

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

Data storing / backup

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

Data sharing

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

Organizational Implications

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

[ricordatevi: serv
ente/ateneo per

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

Level	Data Classification and Examples (abridged version)
5	*Information that would cause severe harm to individuals or the University if disclosed.*
	• Research information classified as Level 5 by an IRB or otherwise required to be stored or processed in a high security environment and on a computer not connected to the Harvard data networks • Certain individually identifiable medical records and genetic information, categorized as extremely sensitive
4	*Information that would likely cause serious harm to individuals or the University if disclosed.*
	• High Risk Confidential Information (HRCI) and research information classified as Level 4 by an IRB • Personally identifiable financial or medical information • Information commonly used to establish identity that is protected by state, federal, or foreign privacy laws and regulations • Individually identifiable genetic information that is not Level 5 • National security information (subject to specific government requirements) • Passwords and Harvard PINs that can be used to access confidential information
3	*Information that could cause risk of material harm to individuals or the University if disclosed.*
	• Research information classified as Level 3 by an IRB • Information protected by the Family Educational Rights and Privacy Act (FERPA) to the extent it is not covered under Level 4 including non-directory student information and directory information about students who have requested a FERPA block …mes or any other information that could identify individuals …ds (employees may discuss terms and conditions of employment with each other and third …ectory student information and directory information about students who have requested a …mes or any other information that could identify individuals …ds (employees may discuss terms and conditions of employment with each other and third …ords …tion …on protected under state, federal and foreign privacy laws not classified as Level 4 or 5
	…of which would not cause material harm, but which the University has chosen to
	…ork and intellectual property not in Level 3 or 4 …ssified as Level 2 by an IRB …work papers, drafts of research papers • Building plans and information about the University physical plant
1	*Public information.*
	• Research data that has been de-identified in accordance with applicable rules • Published research • Published information about the University • Course catalogs • Directory information about students who have not requested a FERPA block • Faculty and staff directory information

Harvard security

Serve formazione?

[DMP]

cessda TRAINING

Data Management Expert Guide

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

Plan

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

Organise & Document

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

Process

Store

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

Protect

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

Archive & Publish

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

Discover

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

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

Con un supporto pratico

[DMP]

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

cessda TRAINING

Adapt your DMP: part 6

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

⊕ Versioning

⊖ Interoperability

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

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

If applicable:

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

⊖ Deposit your data

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

CESSDA Guide

Formazione

[DMP]

rdnl research data netherlands

Essentials 4 Data Support

Essentials4data

ABOUT THE COURSE › START THE COURSE › LOGIN ›

I - A bird's-eye view

Data jargon

- DOI
- FAIR data
- GDPR
- Integrity
- Linked data
- Metadata
- Open data
- Open science
- Persistent identifier (PID)
- Preferred format

- I - A bird's-eye view ›
- II - Planning phase ›
- III - Research phase ›
- IV - Harvest phase ›
- V - Legislation and policy ›
- VI - Data support ›
- Closing remarks

23 cose

National Coordination Point Research Data Management

RDA RESEARCH DATA ALLIANCE

23 Things for Researchers and PhD Candidates

An overview of practical resources and tools that you can begin using today to incorporate research data management into your research workflows.

Contents

... to help researchers and PhD candidates engage in research data management!

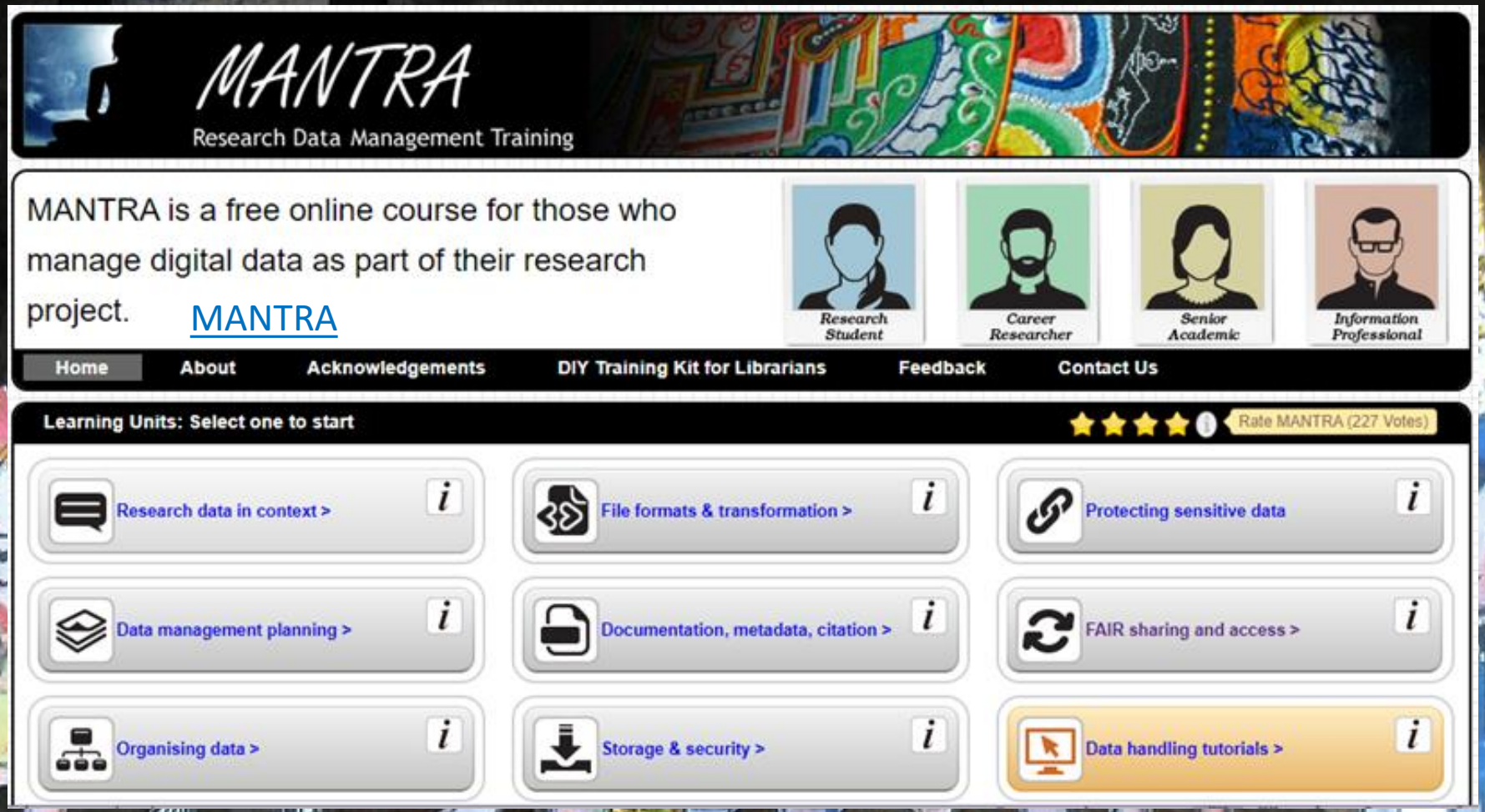
Research Data Management

Proper Research Data Management (RDM) saves time, allows the easy re-use and citation of data, and is essential for making research truly reproducible.

1. Get inspired and motivated by some visual introductions to the principles and merits of Research Data Management provided by the University of Jena, edu.nl/tr4pj

2020

Formazione


MANTRA
Research Data Management Training
MANTRA is a free online course for those who manage digital data as part of their research project.
MANTRA
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 (227 Votes)
Research data in context >
File formats & transformation >
Protecting sensitive data
Data management planning >
Documentation, metadata, citation >
FAIR sharing and access >
Organising data >
Storage & security >
Data handling tutorials >

Data management ABC – File naming

[DMP]

SCEGLIERNE UNA… ED ESSERE CONSISTENTI!

File naming conventions

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

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

File naming

[DMP]

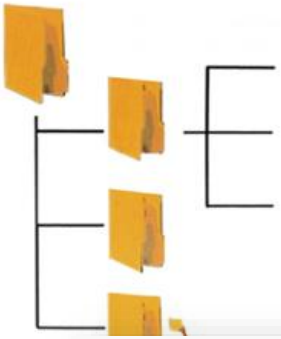
Data management ABC – File naming

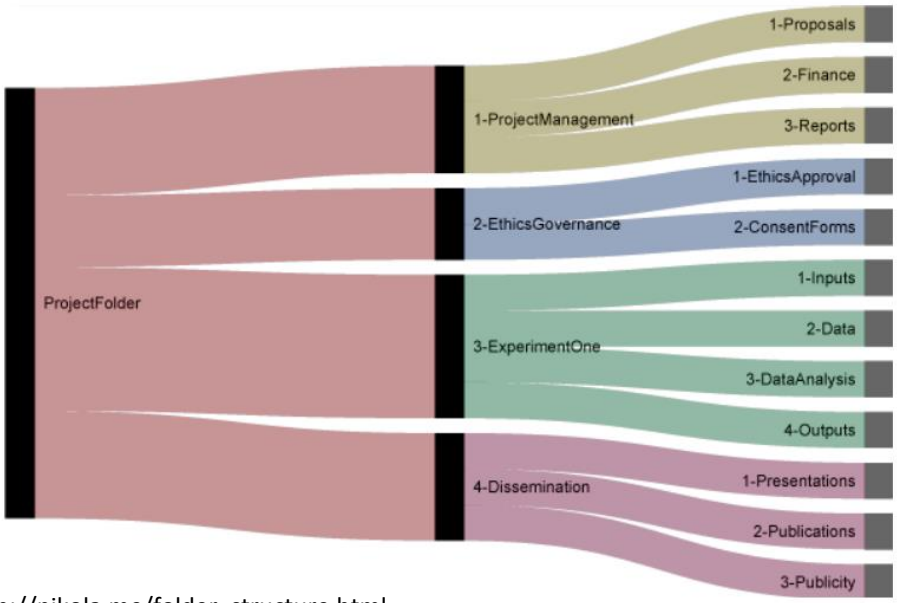
Folder structure

CESSDA training

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

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

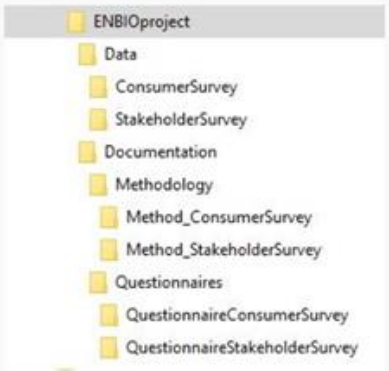




http://nikola.me/folder_structure.html

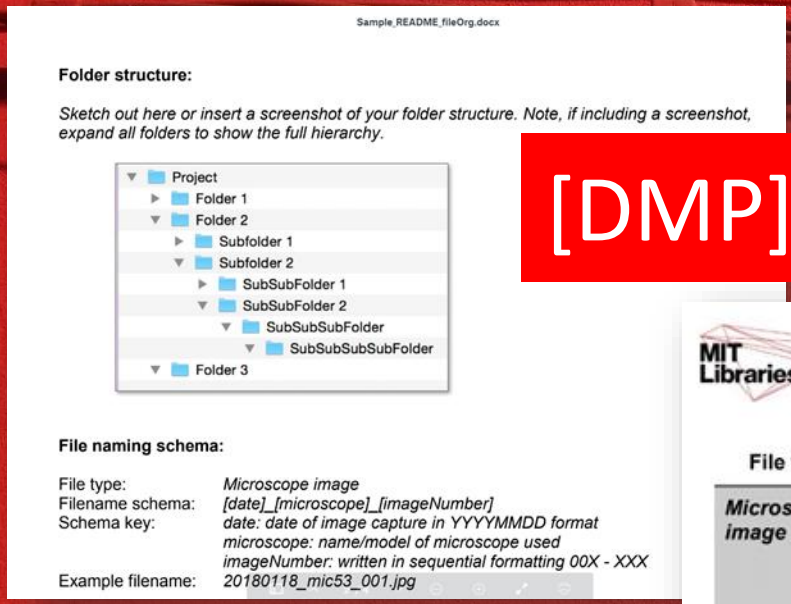
Survey data

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



Data management ABC – Readme file

[DMP]



Sample_README_fileOrg.docx

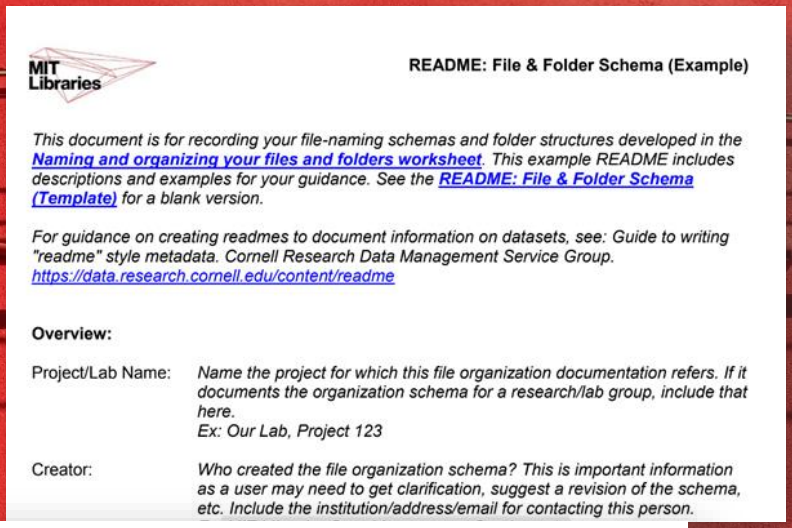
Folder structure:

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

- Project
  - Folder 1
  - Folder 2
    - Subfolder 1
    - Subfolder 2
      - SubSubFolder 1
      - SubSubFolder 2
        - SubSubSubFolder
          - SubSubSubSubFolder
  - Folder 3

File naming schema:

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



MIT Libraries

README: File & Folder Schema (Example)

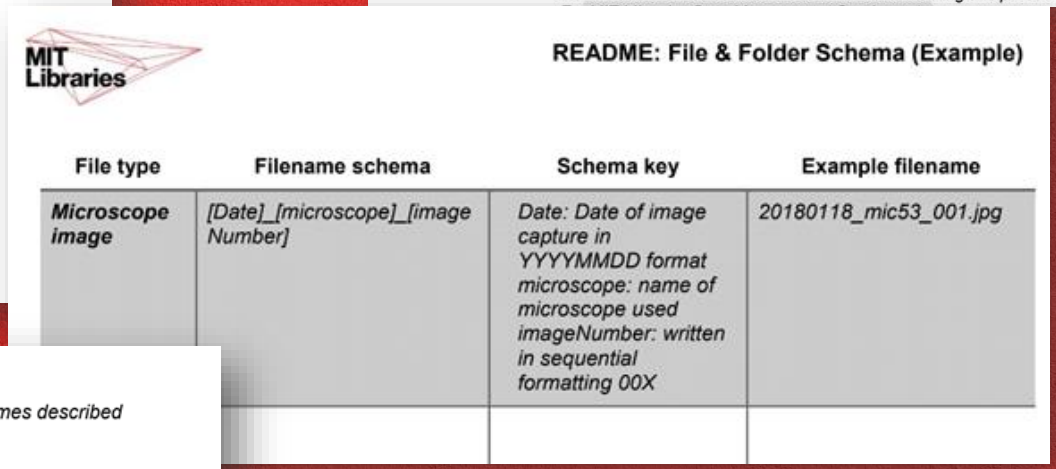
*This document is for recording your file-naming schemas and folder structures developed in the **Naming and organizing your files and folders worksheet**. This example README includes descriptions and examples for your guidance. See the **README: File & Folder Schema (Template)** for a blank version.*

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

Overview:

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

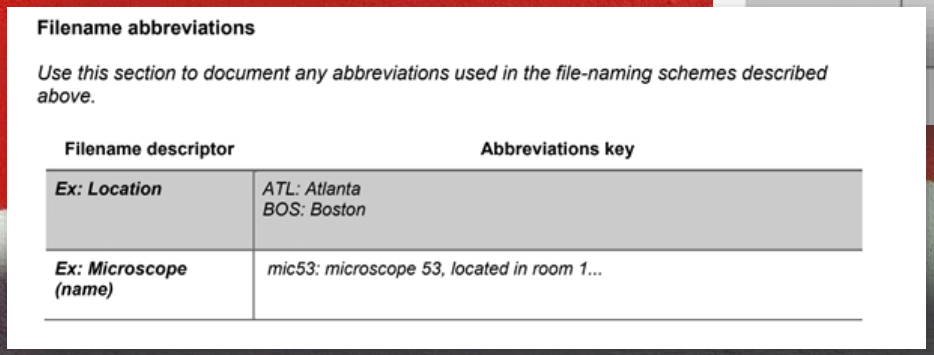
Creator: *Who created the file organization schema? This is important information as a user may need to get clarification, suggest a revision of the schema, etc. Include the institution/address/email for contacting this person.*



MIT Libraries

README: File & Folder Schema (Example)

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



Filename abbreviations

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

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

MIT data management

Data management A
Readme file

[DMP]

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

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

4TU. CENTRE FOR RESEARCH DATA

Guidelines for README

Guidelines for creating a README file

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

Best practices

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

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

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

Data management ABC – Versioning

[DMP]

Data versioning

ands AUSTRALIAN NATIONAL DATA SERVICE

About us

Working with data

SHARE

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

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

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

Why is data versioning important?

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

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

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

What tools are available for data versioning?

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

Example approaches include:

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

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

Data versioning at ArcGIS □

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

Numbering system 1

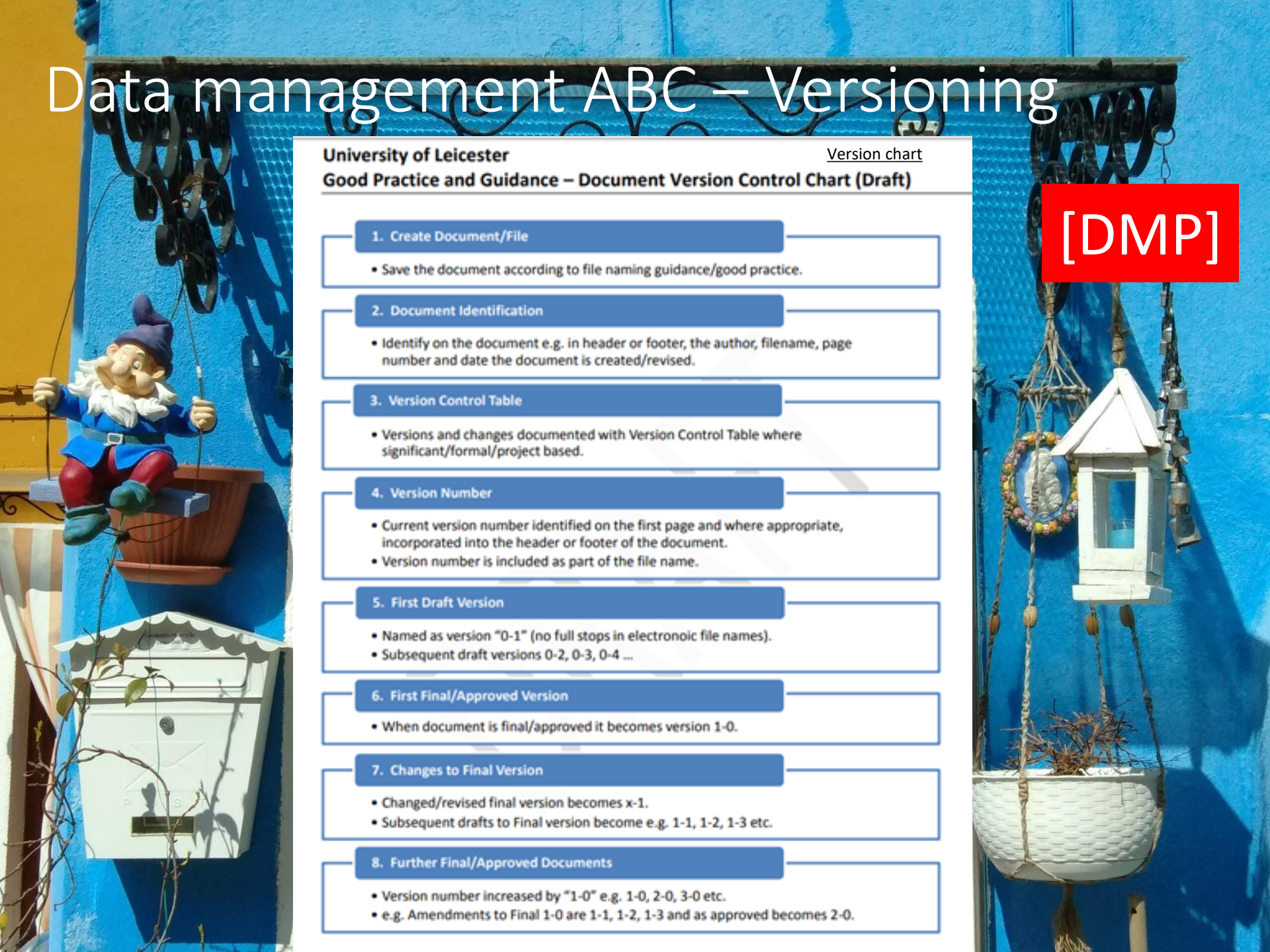
Data versioning follows a similar path to software versioning, usually applying a two-part numbering rule: Major.Minor (e.g. V2.1). Major data revision indicates a change in the formation and/or content of the dataset that may bring changes in scope, context or intended use. For example, a major revision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

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

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

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

Data management ABC – Versioning



University of Leicester Version chart

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

1. Create Document/File

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

2. Document Identification

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

3. Version Control Table

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

4. Version Number

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

5. First Draft Version

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

6. First Final/Approved Version

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

7. Changes to Final Version

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

8. Further Final/Approved Documents

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

[DMP]

Data management ABC – Versioning

[DMP]

cessda TRAINING

Version control

Version control can be done through:

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

CESSDA training

Git

Data management ABC – Data entry

[DMP]

cessda TRAINING

Data Management Expert Guide

1. Plan
2. Organise & Document
3. Process
   - **Data entry and integrity**
   - Quantitative coding
   - Qualitative coding
   - Weights of survey data
   - File formats and data conversion
   - Data authenticity
   - Wrap up: Data quality
   - Adapt your DMP: part 3
   - Sources and further reading
4. Store
5. Protect
6. Archive & Publish

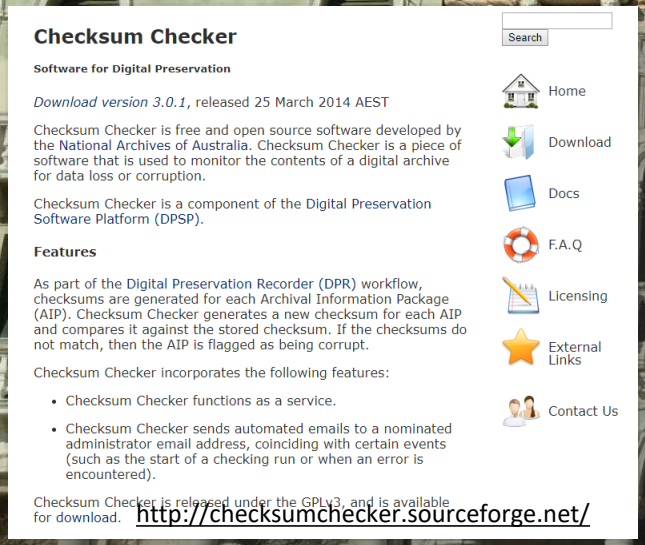
- ⊕ Check the completeness of records
- ⊕ Reduce burden at manual data entry
- ⊕ Minimise the number of steps
- ⊕ Conduct data entry twice
- ⊕ Perform in-depth checks for selected records
- ⊕ Perform logical and consistency checks
- ⊕ Automate checks whenever possible

CESSDA Guide

Data Management ABC - conservazione

[DMP]

LUNGO O BREVE TERMINE?



Checksum Checker

Software for Digital Preservation

Download version 3.0.1, released 25 March 2014 AEST

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

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

Features

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

Checksum Checker incorporates the following features:

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

Checksum Checker is released under the GPLv3, and is available for download.

Search

Home

Download

Docs

F.A.Q

Licensing

External Links

Contact Us

http://checksumchecker.sourceforge.net/

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

1 2 3 4 5 **6**

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

Data Management ABC- backup and storage

[DMP]

Portable devices | Cloud storage | Local storage | Networked driv

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

Advantages	Disadvantages/Risks	Pre (ser data
• Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field. • Low-cost solution.	• Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk. • Not robust for long-term storage or master copies of your data and files. • Possible quality control issues due to version confusion.	Use enc pass

vantages	Disadvantages/Risks	Precautions for (sensitive) personal data
• Automatic backups. • Often automatic version control.	• Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens. • Insufficient control over where the data is stored and how often it is backed up. • Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes. • Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.	• Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see 'Security' for more information on encryption; also see 'Protecting data').

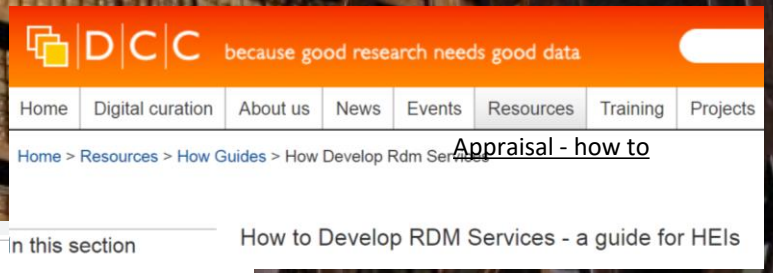
Recommendations

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

CESSDA Guide

Ci sono strumenti diversi per esigenze diverse (durante/al termine). Durante, dovete anche poterci lavorare con il team

Cosa conservare?



n this section

Establishing criteria for selection decisions

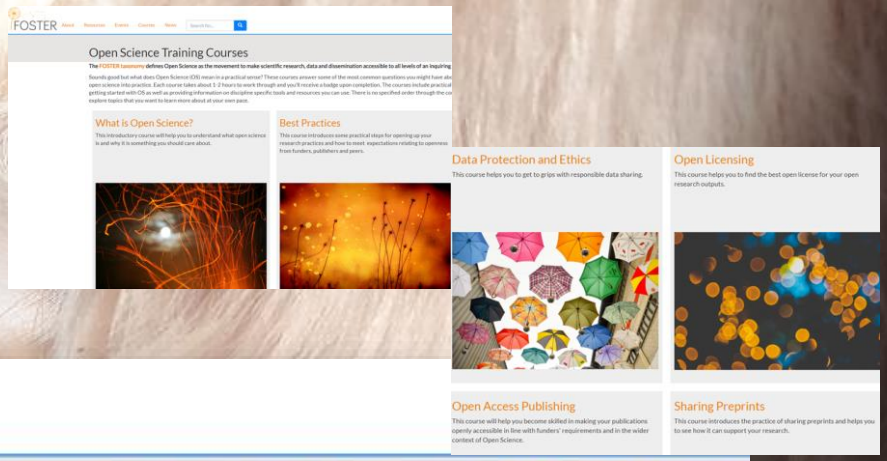
You should establish criteria to guide selection decisions. The DCC's How to Select and Appraise Research Data for Curation[56] proposes seven criteria as outlined below:

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

[DMP]

- RILEVANTI PER LA MISSIONE DELL'ENTE
- VALORE STORICO
- UNICITÀ
- POTENZIALE DI RIUSO
- NON REPLICABILI
- COSTO/BENEFICI
- DOCUMENTAZIONE COMPLETA

Imparare a gestire



FOSTER About Resources Events Courses News Search for...

Managing and Sharing Research Data

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

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

Start the Free Course

Full details

Level of knowledge: Introductory: no previous knowledge is required

Topics

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

Imparare a proteggere

What are personal data?

Click the plus sign to expand the text box

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

FOSTER About Resources Events Courses News

Data Protection and Ethics

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

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

Start the Free Course

Full details

Level of knowledge: Introductory: no previous knowle
is required

Topics

[dati personali]

Legal Basis

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

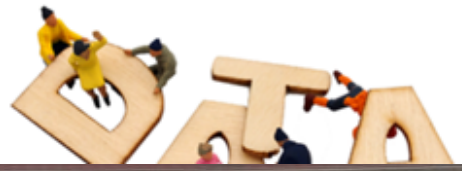
- consent of the data subject
- necessary for the performance of a contract
- legal obligation placed upon the data controller
- necessary to protect the vital interests of the data subject
- carried out in the public interest or in the exercise of official authority (public task)
- legitimate interest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

The research exemption

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

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

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



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

cessda TRAINING

CESSDA guide
Data Management Expert Guide

[dati personali]

CESSDA guide
Data Management Expert Guide

I. Process lawfully, fair and transparent

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

II. Keep to the original purpose

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

III. Minimise data size

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

IV. Uphold accuracy

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

V. Remove data which are not used

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

VI. Ensure data integrity and confidentiality

Personal data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss,

[leggi applicabili]

Science Europe 2018

Privacy

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

Intellectual Property Rights

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

Public data

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

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

Codes of Conduct/Ethical Issues

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

[Dati e GDPR]

May 22 2018

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.

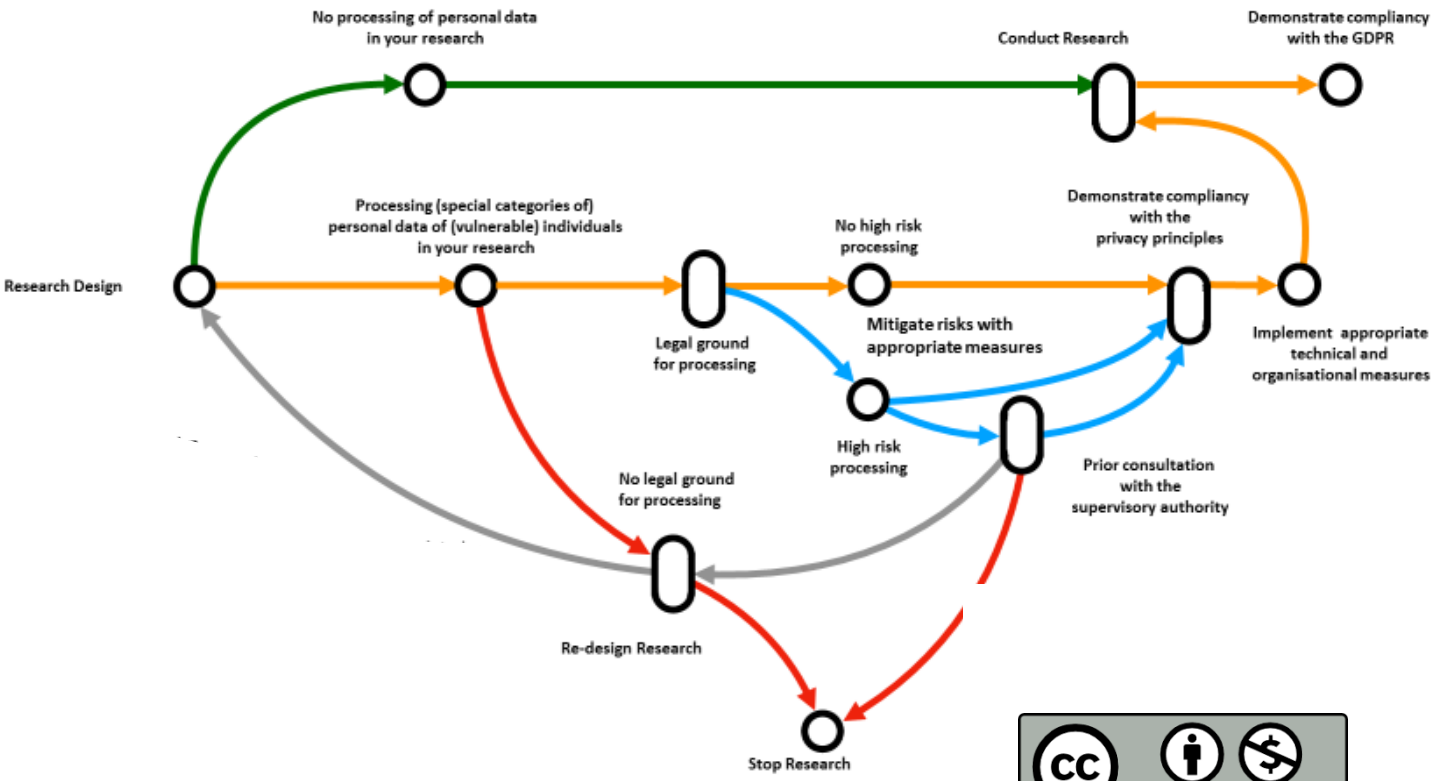
The special regime applies the usual principles such as lawfulness, purpose limitation and data subject rights, but permits some derogations from controller obligations. This includes the presumption of compatibility of processing for scientific research purposes of data collected in commercial and other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

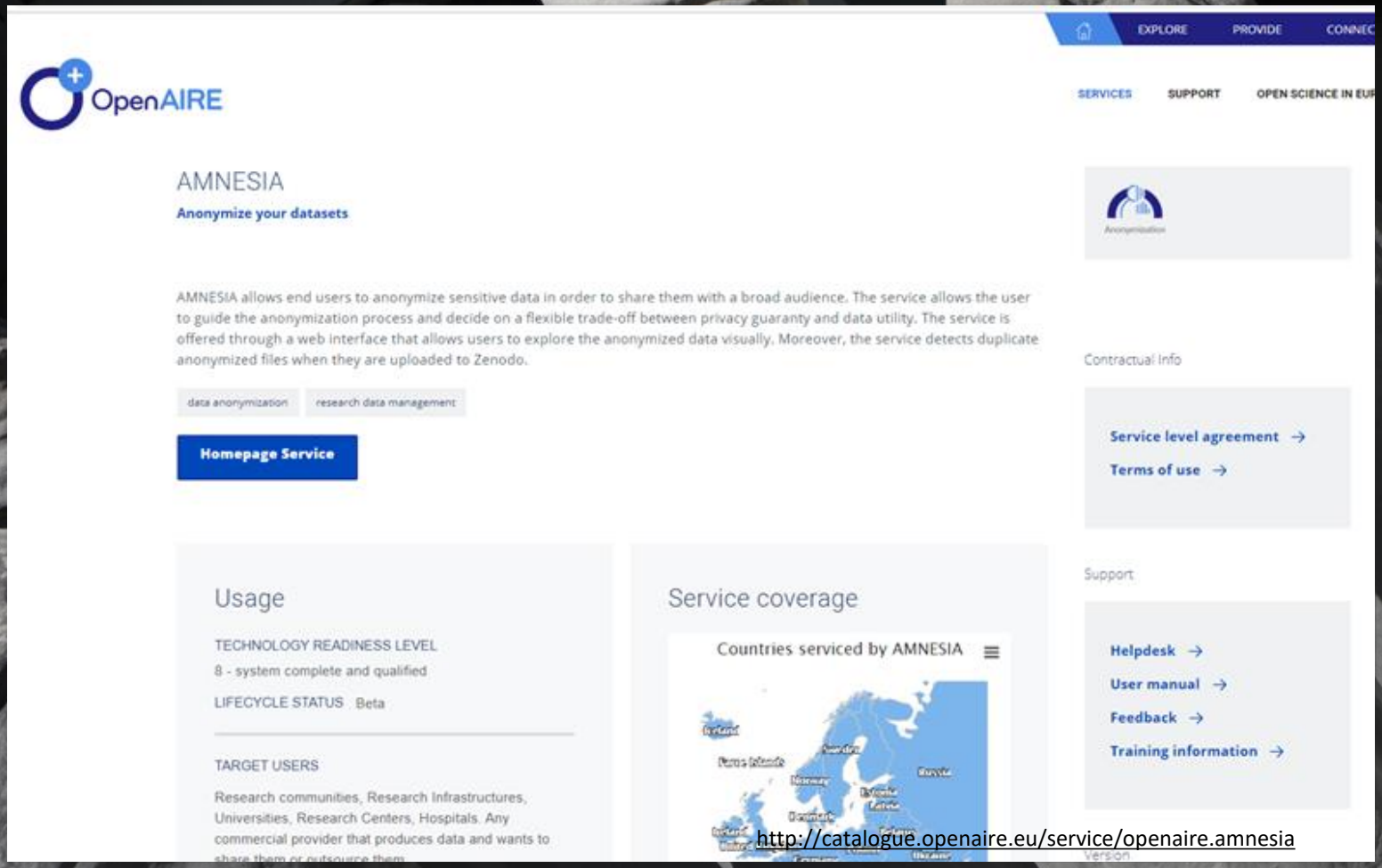
Action

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

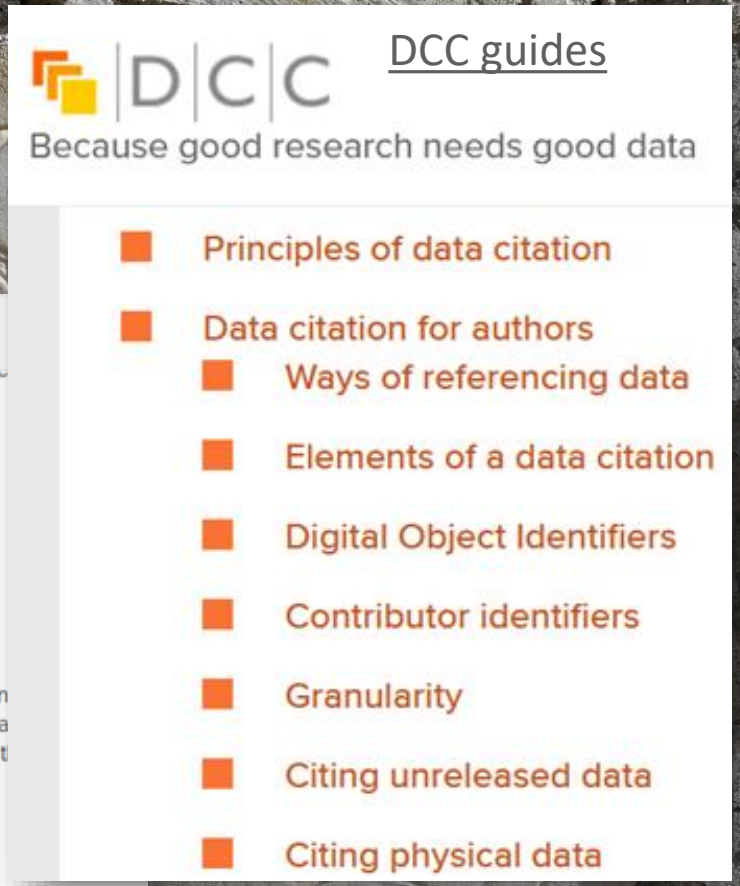
[anonimizzare]



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

...i dati vanno citati

DCC guides



Datacite How to

DataCite About us Services Resources

order to:

- support proper attribution and credit
- support collaboration and reuse of data
- enable reproducibility of findings
- foster faster and more efficient research progress, and
- provide the means to share data with future researchers

DataCite also looks to community practices that provide data citation guidance. The Joint Declaration Citation Principles is a set of guiding principles for data within scholarly literature, another dataset, or a research object (Data Citation Synthesis Group 2014). The FAIR Guiding Principles provide a guideline for t that want to enhance reuse of their data (Wilkinson 2016).

Data Citation Examples

We recognise that the challenges associated with data publication vary across disciplines, and we encourage research communities to develop citation systems that work well for them. Our recommended format for data citation is as follows:

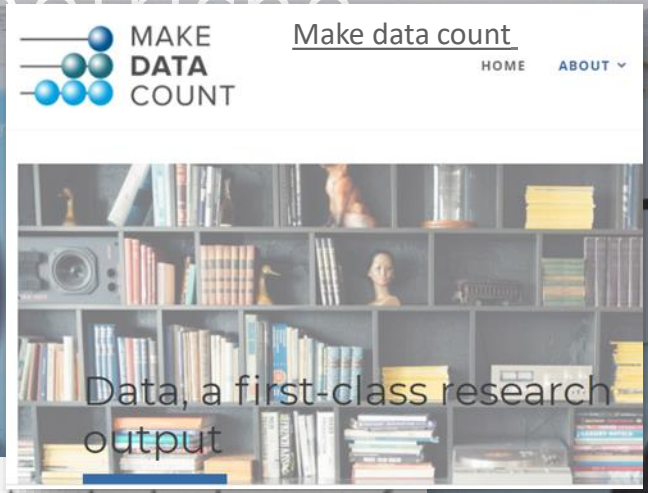
Creator (PublicationYear). Title. Publisher. Identifier

It may also be desirable to include information about two optional properties, Version and ResourceType (as appropriate). If so, the recommended form is as follows:

Creator (PublicationYear). Title. Version. Publisher. ResourceType. Identifier

… i dati devono avere metriche appropriate

OCCORRONO NUOVE METRICHE PER POTER MISURARE IL RIUSO DEI DATI



Make Data Count is a global, community-led initiative focused on the development of open data metrics. The principles of our social and technical infrastructure are rooted in transparency and accessibility. We believe that open data metrics will require broad adoption and our best path towards involvement throughout each phase of work.

B.MONS (2017) NON POSSIAMO MISURARE UNA SCIENZA NUOVA CON MISURE VECCHIE

Through the development of standards, centralized and transparent infrastructure, and reproducible bibliometrics research, we aim towards a state where researchers and research supporting communities will have properly identified indicators for data re-use that can be used for assessing research data investment ROIs and help advance scientific discovery. It is essential that these metrics are flexible, adjusting with research as the scientific landscape evolves, and that these openly reproducible metrics are broadly accepted by researchers.

GRAZIE!

Risultati Mentimeter

Dove conservate i vostri dati? E vi è mai capitato di perderli?

Mentimet

Uso il Google Drive d'istituto, con backup locali sul PC di lavoro, sul portatile, su una scheda SD e su un hard disk separato. Mai perso nulla.

Pc. Backup su hard-disk. Backup su servizio cloud a pagamento.Quasi mai perso dati

su google drive. No

Hard disk e/o cloud. Non mi è mai capitato di perderli.

hard disk (copie multiple)non ancora

NASSi diverse volte

Server interni.No.

Sul PC sincronizzati con il cloud compreso con la mail istituzionale. Per fortuna no.

Backup su Google Drive (Google suite INRIM)Alcuni dati >10 anni irrecuperabili per obsolescenza

Dove conservate i vostri dati? E vi è mai capitato di perderli?

Mentimeter

Backup su circa 5 memorie diverse. Non mi è mai capitato di perderli. Non ho neanche avuto problemi di compatibilità (dati ASCII). A volte è stato un problema ricordarsi il significato delle colonne a causa di una insufficiente descrizione.

Li conservo nel mio pc o hard disk esterno. O/E NAS del laboratorio. Sì, è capitato, che dati non "backupati" si perdessero perché la macchina si era rotta...

softwareTutti i dati >20 anni irrecuperabili per obsolescenza hardware

Google Drive, Zenodo, dischi fissi. Mai persi.

PC, hard disk, server ed eventualmente cloud. No, non mi è mai successo di perdere dati "sensibili"

due-tre differenti repositories personali (hard disks)

Dove conservate i vostri dati? E vi è mai capitato di perderli?

Mentimeter

stato un problema ricordarsi il significato delle colonne a causa di una insufficiente descrizione.

esterno. O/E: NAS del laboratorio.Si, è capitato, che dati non "backupati" si perdessero perché la macchina si era rotta...

Google Drive, Zenodo, dischi fissi. Mai persi.

PC, hard disk, server ed eventualmente cloud. No, non mi è mai successo di perdere dati "sensibili"

due-tre differenti repositories personali (hard disks)

Su Server interno. Raramente.

Se avete aperto i vostri dati, quali di queste tecniche avete adottato?