Perché i dati vanno curati?



È l'incubo del data steward:

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

Perché occuparci di dati?

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

2013

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



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

DATA AB INITIO

AGING RESEARCH DATA RIGHT, FROM THE START

K.Birney, 2015

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

and, as you don't want to be stuck with



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

Thomas Herndon* Michael Ash Robert Pollin

April 15, 2013

Herndon, 2013

JEL CODES: E60, E62, E65

Abstract

We replicate Reinhart and Rogoff (2010a and 2010b) and find that coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period. Our finding is that when properly calculated, the average real GDP growth rate for countries carrying a public-debt-to-GDP ratio of over 90 percent is actually 2.2 percent, not −0.1 percent as published in Reinhart and Rogoff. That is, contrary to RR, average GDP growth at public debt/GDP ratios over 90 percent is not dramatically different than when debt/GDP ratios are lower.

We also show how the relationship between public debt and GDP growth varies significantly by time period and country. Overall, the evidence we review contradicts Reinhart and Rogoff's claim to have identified an important stylized fact, that public debt loads greater than 90 percent of GDP consistently reduce GDP growth.

Perché occuparci di dati?



1995

WaveLab and Reproducible Research

Jonathan B. Buckheit and David L. Donoho

Stanford University, Stanford CA 94305, USA

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

UN ARTICOLO SENZA I DATI È SOLO LA PUBBLICITÀ DELLA RICERCA



2015



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

Parliamo di dati

«pezzi»
di conoscenza osservabili

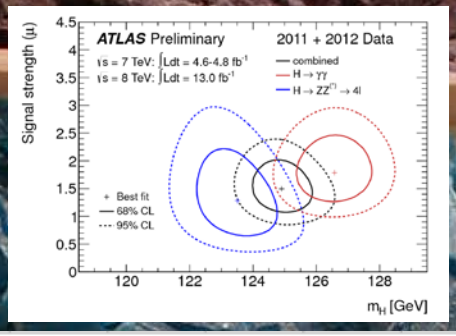




Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample from the cohort.

OTU	A01_TP1	A01_TP2	A01_TP3	A03_TP1	A03_TP3	A04_TP1	A04_TP2	A04_TP3	A05_T
OTU_1	261	76	1206	523	2131	25707	64473	60665	
OTU_2	49	52	117	43035	206	119	1152	539	
OTU_9	148	162176					22858	1898	
OTU_6	21	17					1457	29	
							19	85	
							646	214	
							292	37	
							18	170	
							6	4	

Gaucelm Faidit

I.

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





Wilma van Wezenbeek
@wvanwezenbeek

Following

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

Traduci il Tweet

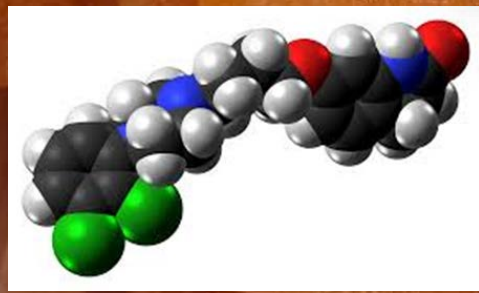
12:51 - 13 mar 2018

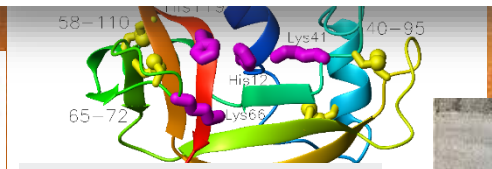
3 Retweet **1** Mi piace

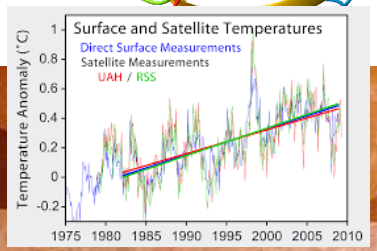
https://twitter.com/wvanwezenbeek/status/973527086685093893



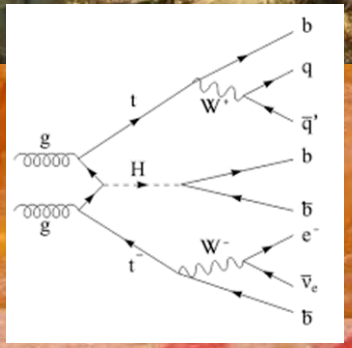




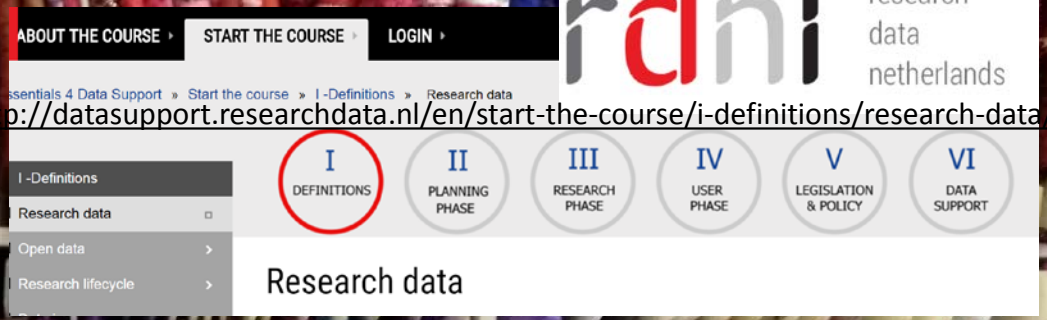








Un po' di glossario



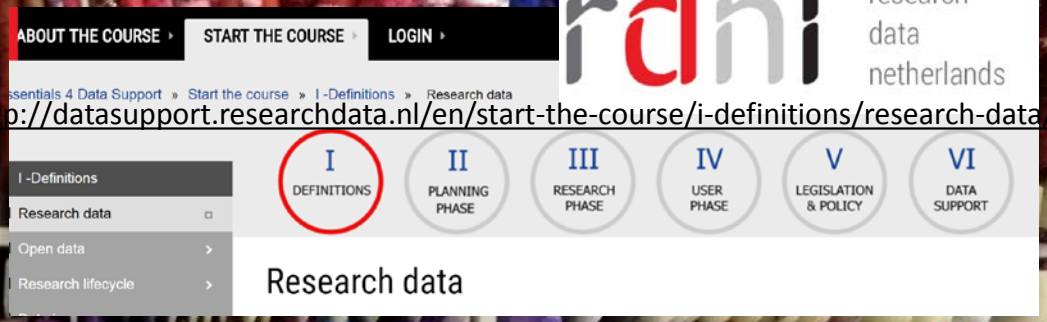
rdnl research data netherlands

http://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-data/

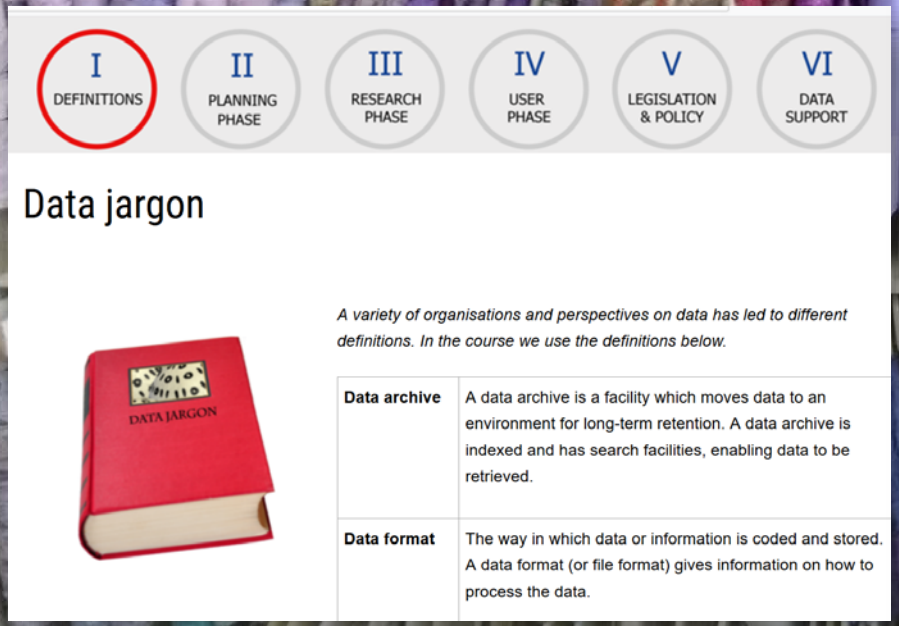
1 2 3 4

Research data is the material underpinning a research assertion. (4)

Un po' di glossario



http://datasupport.researchdata.nl/en/start-the-course/i-definitions/research-data/



RDF	RDF is a standard model for data interchange on the Web (see http://www.w3.org/RDF/).
Research data	Data are facts, observations or experiences on which an argument or theory is based. (see http://ands.org.au/guides/what-is-research-data.pdf).
Resolver	A system that brings about the link between a persistent identifier and the location where the object is currently situated.
Text- and data mining	The computer-based process of deriving or organising information from text or data. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns, trends and hypotheses or by providing the means to organise the information mined. (see www.ipo.gov.uk/ipreview-doc-t.pdf).

Un po' di glossario

5 modi per pensare i dati:

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



- **The way the data is collected**.
  - By experimenting, simulations, observations, derived data, reference data.
- **The data forms**.
  - For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, slides, artefacts, models, scripts, databases, metadata, etc.
- **The formats for electronic storage of the research data**.
- **The size (volume) of the data files**.
- **The *research lifecycle* phase the data is in.**

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

Part I

Five Ways To Think About Research Data

Science has progressed by 'standing on the shoulders of giants' and for centuries research and knowledge has been shared through the publication and disemmination of books, papers and scholarly communications. Moving forward much of our understanding builds on (large scale) data sets which have been collected or generated as part of this scientific process of discovery. How will this be made available for future generations? How will we ensure that, once collected or generated, others can stand on the shoulders of the data we produce?

Deciding on how to look after data depends on what your data looks like and what needs to be done with it. You should find out if your discipline already has standard practices and use them. We hope that this brief introduction will give some templates of what is already being done in a few disciplines and enable you to start thinking about what you might do with your research data to make it accessible to others.

Further University of Southampton guidance can be found on the library's web site `http://library.soton.ac.uk/researchdata`. Any research data management questions can be emailed to `researchdata@soton.ac.uk`.

This part of the guide introduces five ways of looking at research data.

1 Research data collection

The first way of thinking about research data is where it comes from (Research Information Network, 2008). Each of the case studies in Part II illustrates one of these categories.

Reference data: *Example: the reference human genome sequence in Case Study 1*
A data set that can be used for validation, comparison or information lookup.

Scientific experiments: *Example: materials engineering fatigue test in Case Study 2*
Data generated by, e.g. instruments during a scientific experiment.

Models or simulations: *Example: CFD helicopter rotor wake simulation in Case Study 3*
Data generated on computer by an algorithm, mathematical model, or the simulation of an experiment. A computer simulation can help when experiments are too expensive, time consuming, dangerous or even impossible to perform.

Derived data: *Example: chemical structures in chemistry in Case Study 4*
A data set created by taking existing data and performing some manipulation to it. Each data set requires careful curation because the original data may be needed to understand the new data.

Observations: *Example: archaeological dig in Case Study 5*
Data generated by recording observations of a specific, possibly unrepeatable, event at a specific time or location.

2 Types of research data

Research can come in many different forms, some electronic and some physical. Here are some examples:

- Electronic text documents, e.g. text, PDF, Microsoft Word files
- Spreadsheets
- Laboratory notebooks, field notebooks and diaries
- Questionnaires, transcripts and codebooks
- Audiotapes and videotapes
- Photographs and films
- Examination results
- Specimens, samples, artefacts and slides
- Digital objects, e.g. figures, videos
- Database schemas
- Database contents
- Models, algorithms and scripts
- Software configuration, e.g. case files
- Software pre-process files, e.g. geometry, mesh
- Software post-process files, e.g. plots, comma-separated value data (CSV)
- Methodologies, workflows, standard operating procedures and protocols
- Experimental results
- Metadata (data describing data), e.g. environmental conditions during experiment
- Other data files, e.g. literature review records, email archives

3 Electronic storage

The third way to think about research data is how it is stored on a computer. Here are some of the categories of electronic data:

Textual, e.g.:
- Flat text files
- Microsoft Word
- PDF
- RTF

Numerical, e.g.:
- Excel
- CSV

Multimedia, e.g.:
- Image (JPEG, TIFF, DICOM)
- Movie (MPEG, AVI)
- Audio (MP3, WAV, OGG)

Structured, e.g.:
- Multi-purpose (XML)
- Relational (MySQL database)

Software code, e.g.:
- Java
- C

Software specific, e.g.:
- Mesh
- Geometry
- 3D CAD
- Statistical model

Discipline specific, e.g.:
- Flexible Image Transport System (FITS) in astronomy
- Crystallographic Information File (CIF) in chemistry

Instrument specific, e.g.:
- Olympus Confocal Microscope Data Format
- Carl Zeiss Digital Microscopic Image Format (ZVI)

Data can be born digitally, such as a simulation, or ingested into a computer, such as scanning a photograph. Some data can remain in a non-digital format.

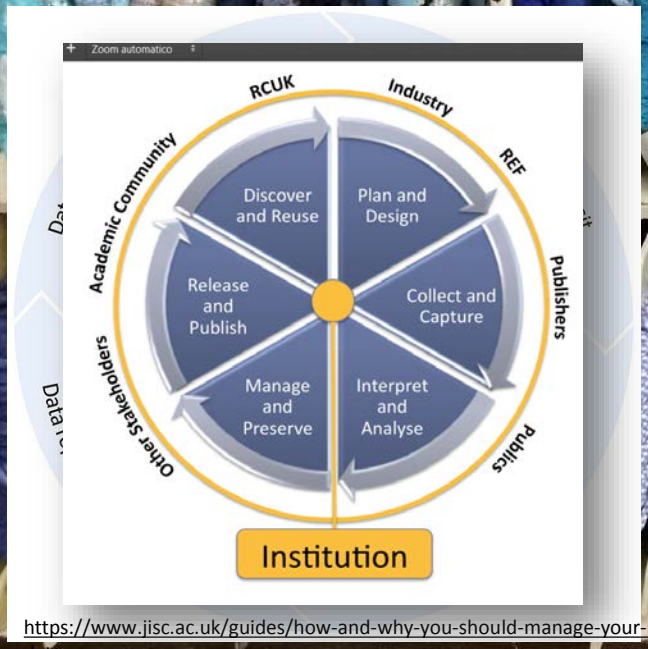
Un po' di glossario

DATA CURATION: CONSERVAZIONE SUL LUNGO PERIODO

DATA MANAGEMENT: GESTIONE LUNGO TUTTO IL CICLO DI VITA

DATA STEWARD: ESPERTO DELLA MATERIA E DELLA GESTIONE DEI DATI

Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle.



https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data



http://www.dcc.ac.uk/resources/curation-lifecycle-model

Un po’ di glossario

DCC because good research needs good data

The digital curation lifecycle

Digital curation and data preservation are ongoing processes, requiring considerable thought and the investment of adequate time and resources. You must be aware of, and undertake, actions throughout the data lifecycle.

The digital curation lifecycle comprises th

Conceptualise: conceive and plan the cr capture methods and storage options.

Create: produce digital objects and assig and technical archival metadata.

Access and use: ensure that designated on a day-to-day basis. Some digital objec others may be password protected.

Appraise and select: evaluate digital obj term curation and preservation. Adhere to legal requirements.

Dispose: rid systems of digital objects no preservation. Documented guidance, policies and legal requirements may require the secure destruction of these objects.

Ingest: transfer digital objects to an archive, trusted digital repository, data centre or similar, again adhering to documented guidance, policies and legal requirements.

Preservation action: undertake actions to ensure the long-term preservation and retention of the authoritative nature of digital objects.

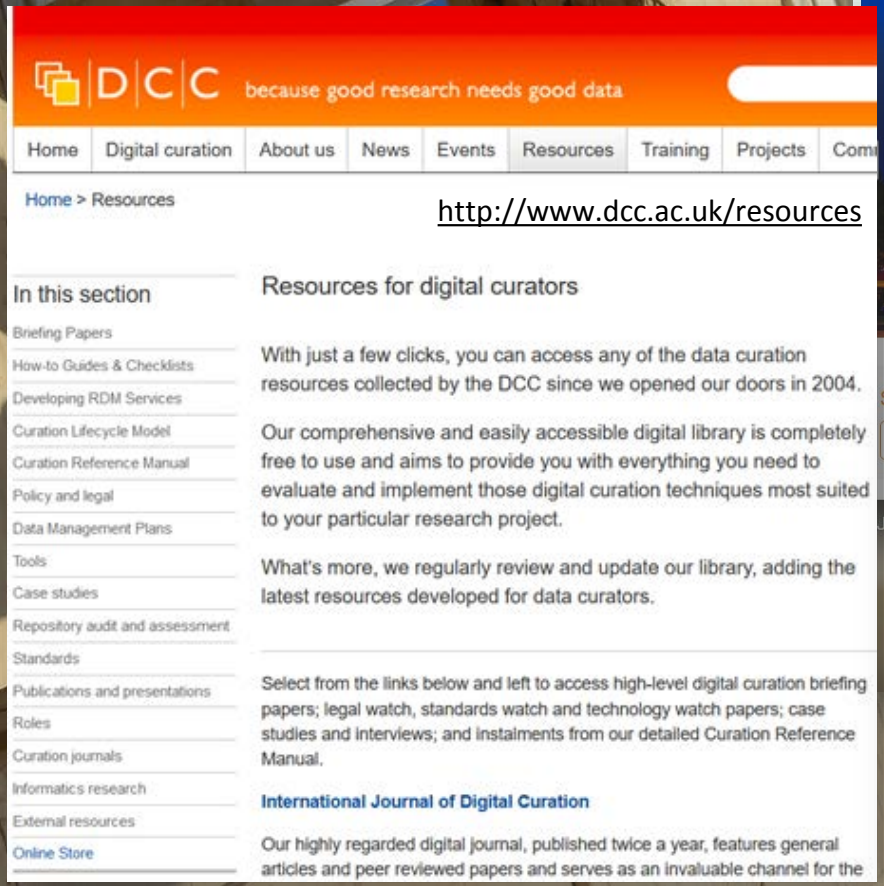
Reappraise: return digital objects that fail validation procedures for further appraisal and reselection.

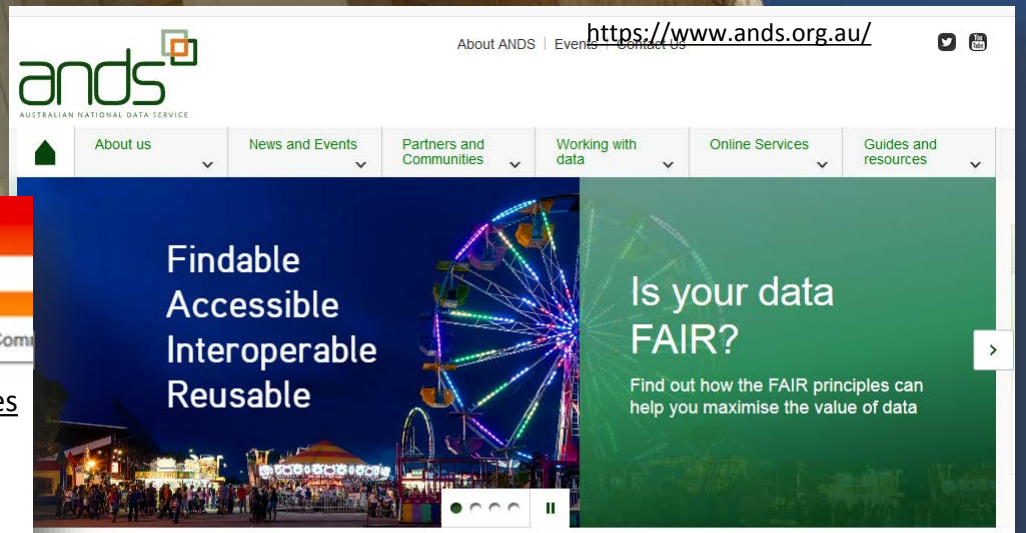
Store: keep the data in a secure manner as outlined by relevant standards.

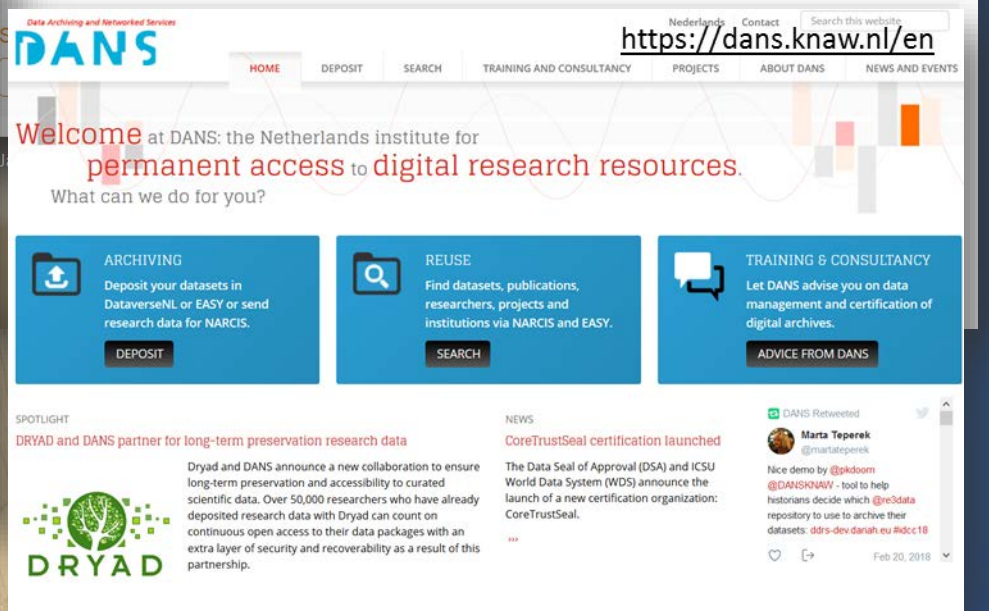
Access and reuse: ensure that data are accessible to designated users for first time use and reuse. Some material may be publicly available, whilst other data may be password protected.

Transform: create new digital objects from the original, for example, by migration into a different form.

Due pilastri, anzi tre





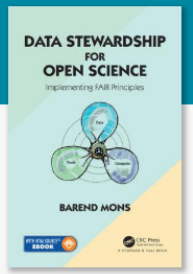


… e un maestro



Taylor & Francis Group
an informa business

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

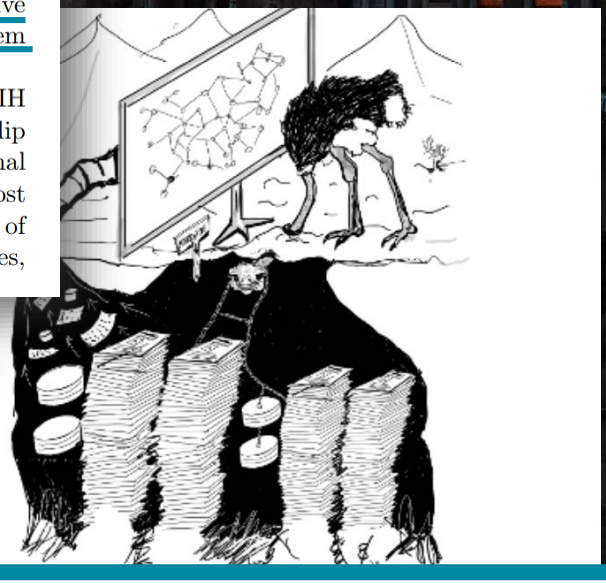


Data Stewardship for Open Science
Implementing FAIR Principles

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

Dati persi

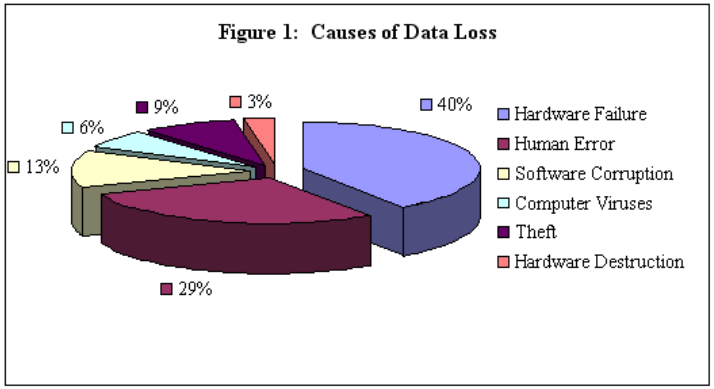
perdita dei dati

La causa di perdita di dati e informazioni può dipendere da:
- eventi distruttivi, naturali o artificiali;
- guasti ai sistemi;
- malfunzionamenti o degrado dei componenti elettronici;
- incuria o disattenzione.

Il rischio di perdita di dati è anche rappresentato da:
- comportamenti sleali e fraudolenti;
- virus informatici;
- furto di strumenti contenenti dati.

(fonte: Guida alla sicurezza dei dati in azienda)

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



D.M. Smith The cost of lost data, 2003

Dati persi

il backup: definizione (meno seria)

Il backup è quella cosa che andava fatta prima.

(fonte: Proverbio cinese)

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



COPYRIGHT-Italia.it
Avv. Simone Aliprandi, Ph.D. – Copyright-Italia.it / Array Law Firm
www.copyright-italia.it – www.aliprandi.org – www.array.eu
ARRAY

sicurezza delle reti



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

Istituto Deledda (Lecce), 25 gennaio 2017 – Sicurezza dati e privacy (le norme)

… i dati sono fragili

Scientists losing data at a rapid rate

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

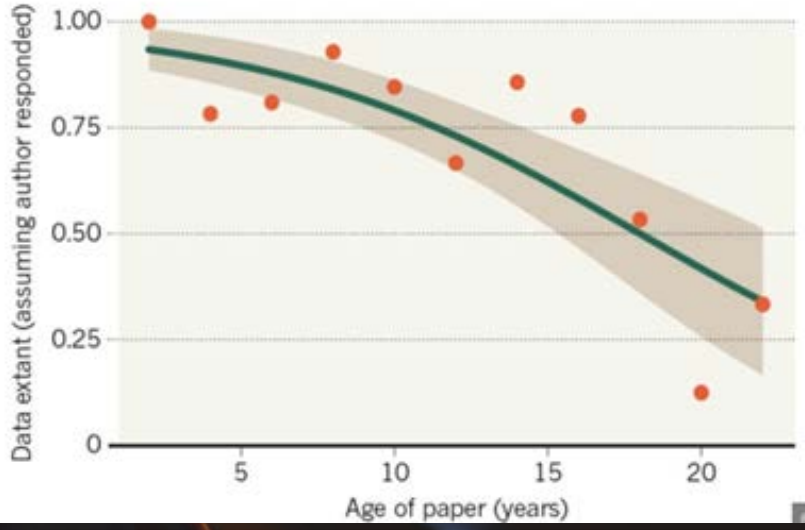
Elizabeth Gibney & Richard Van Noorden

19 December 2013

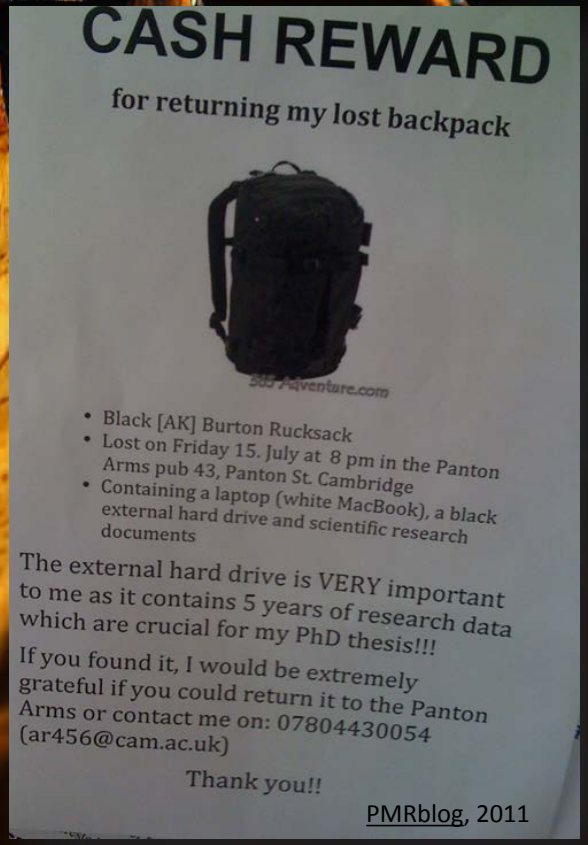
Rights & Permissions

MISSING DATA

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



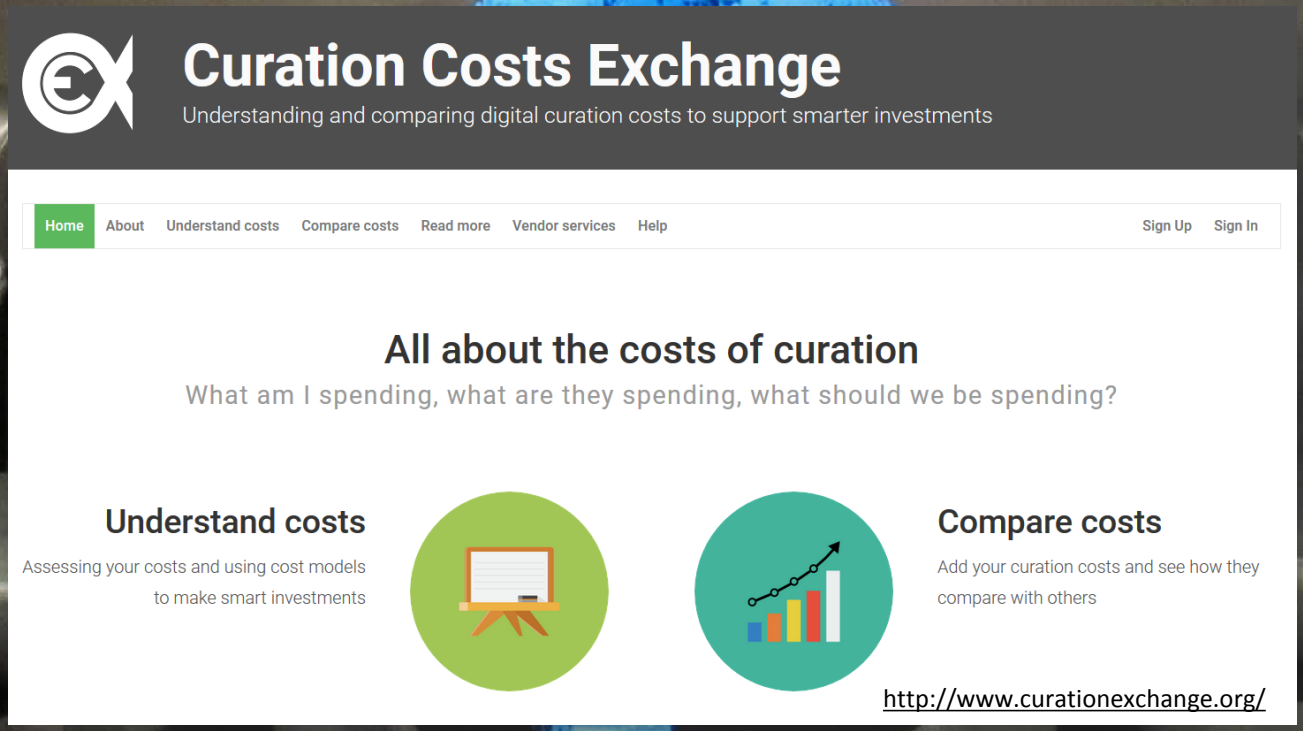
http://www.nature.com/news/scientists-losing-data-at-a-rapid-rate-1.14416



PMRblog, 2011

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

I costi



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

1. i dati vanno curati

DCC because good research needs good data

Why preserve digital data?

Digital data preservation should be a key aspect of all research projects. Some research data are unique and cannot be replaced if destroyed or lost, yet only by referring to verifiable data can your research be judged as sound.

What's more, it is recognised good practice for institutions and researchers to manage and retain their research data, and sometimes they are legally required to do so for many years after project funding has ceased. So, putting in place adequate data preservation initiatives should be top of your list when planning any new research project.



What is digital curation?

Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle.

The active management of research data reduces threats to their long-term research value and mitigates the risk of digital obsolescence. Meanwhile, curated data in trusted digital repositories may be shared among the wider UK research community.

As well as reducing duplication of effort in research data creation, curation enhances the long-term value of existing data by making it available for further high quality research.

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 communicat
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where
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.

«ACCESIBLE»
NON
SIGNIFICA «APERTI»
possono anche essere chiusi, purché si sappia dove sono e a chi chiedere permesso

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

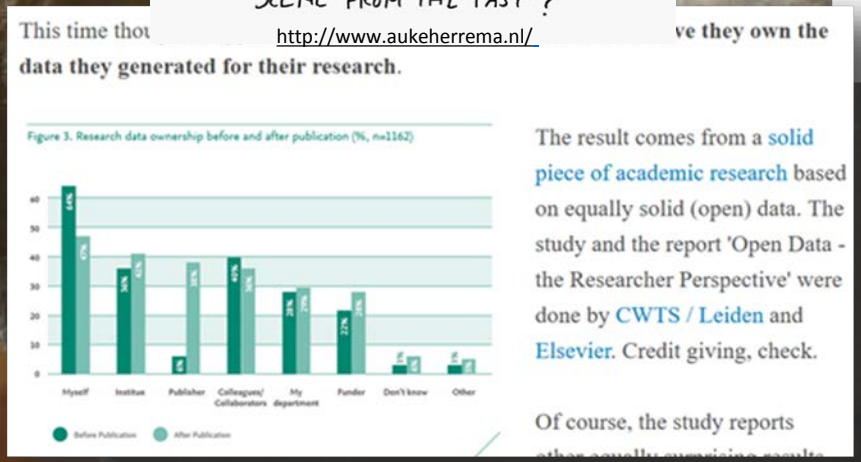
[il fondamento]



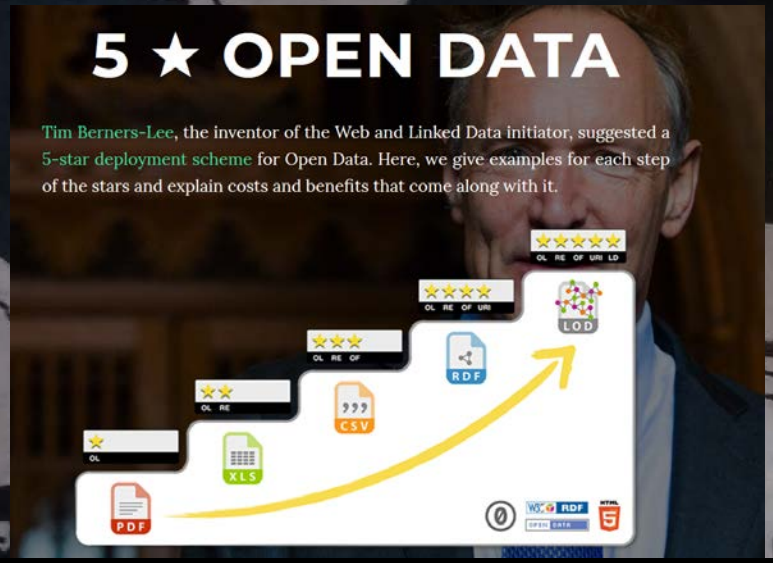
http://www.aukeherrema.nl/



Lusoli, Apr.2017



3. i dati POSSONO essere aperti



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

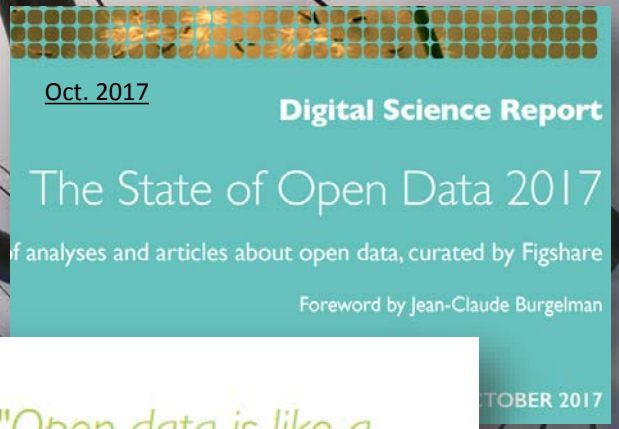
★★ make it available as structured data (e.g., Excel instead of image scan of a table)[2]

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

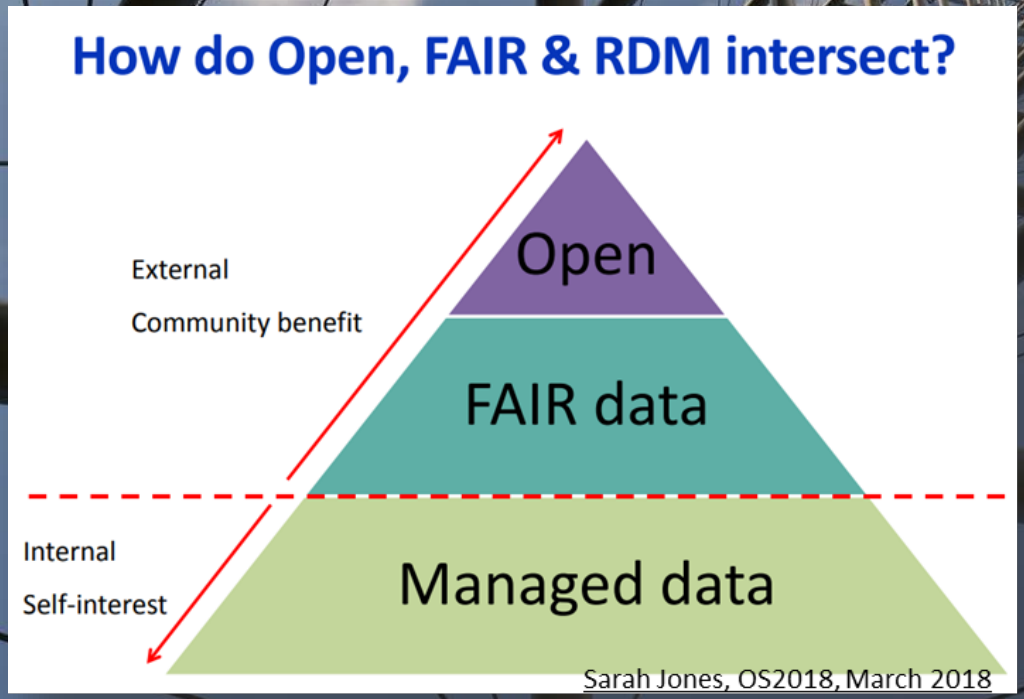
★★★★ use URIs to denote things, so that people can point at your stuff[4]

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

Open data



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



Prendersi cura dei dati: come fare



CONTENTS



ENDORSEMENTS iii

FOREWORD 1

SHARING YOUR DATA - WHY AND HOW 2
- Why share research data 3
- How to share your data 4

DATA MANAGEMENT PLANNING 5
- Roles and responsibilities 7
- Costing data management 7

DOCUMENTING YOUR DATA 8
- Data documentation 9
- Metadata 10

FORMATTING YOUR DATA 11
- File formats 12
- Data conversions 13
- Organising files and folders 13
- Quality assurance 14
- Version control and authenticity 14
- Transcription 15

STORING YOUR DATA 17
- Making back-ups 18
- Data storage 18
- Data security 19
- Data transmission and encryption 20
- Data disposal 21
- File sharing and collaborative environments 21

ETHICS AND CONSENT 22
- Legal and ethical issues 23
- Informed consent and data sharing 23
- Anonymising data 26
- Access control 27

COPYRIGHT 28

STRATEGIES FOR CENTRES 31
- Data management resources library 32
- Data inventory 33

REFERENCES 34

DATA MANAGEMENT CHECKLIST 35

A CRUCIAL PART OF MAKING DATA USER-FRIENDLY, SHAREABLE AND WITH LONG-LASTING USABILITY IS TO ENSURE THEY CAN BE UNDERSTOOD AND INTERPRETED BY ANY USER. THIS REQUIRES CLEAR AND DETAILED DATA DESCRIPTION, ANNOTATION AND CONTEXTUAL INFORMATION.

DATA DOCUMENTATION

Data documentation explains how data were created or digitised, what data mean, what their content and structure are and any data manipulations that may have taken place. Documenting data should be considered best practice when creating, organising and managing data and is important for data preservation. Whenever data are used sufficient contextual information is required to make sense of that data.

Good data documentation includes information on:

- the context of data collection: project history, aim, objectives and hypotheses
- data collection methods: sampling, data collection process, instruments used, hardware and software used, scale and resolution, temporal and geographic coverage and secondary data sources used
- dataset structure of data files, study cases, relationships between files
- data validation, checking, proofing, cleaning and quality assurance procedures carried out
- changes made to data over time since their original creation and identification of different versions of data files
- information on access and use conditions or data confidentiality

At the data-level, documentation may include:

- names, labels and descriptions for variables, records and their values
- explanation or definition of codes and classification schemes used
- definitions of specialist terminology or acronyms used
- codes of, and reasons for, missing values
- derived data created after collection, with code, algorithm or command file
- weighting and grossing variables created
- data listing of annotations for cases, individuals or items

Data-level descriptions can be embedded within a data file itself. Many data analysis software packages have facilities for data annotation and description, as variable attributes (labels, codes, data type, missing values), data type definitions, table relationships, etc.

Other documentation may be contained in publications, final reports, working papers and lab books or created as a data collection user guide.

10 regole per la cura dei dati

10 Simple Rules for the Care and Feeding of Scientific Data

https://arxiv.org/pdf/1401.2134v1.pdf

Alyssa Goodman[1], Alberto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[4], Rosanne Di Stefano[1], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. Hogg[3], Vinay Kashyap[1], Ashish Mahabal[8], Aneta Siemiginowska[1], Aleksandra Slavkovic[9]

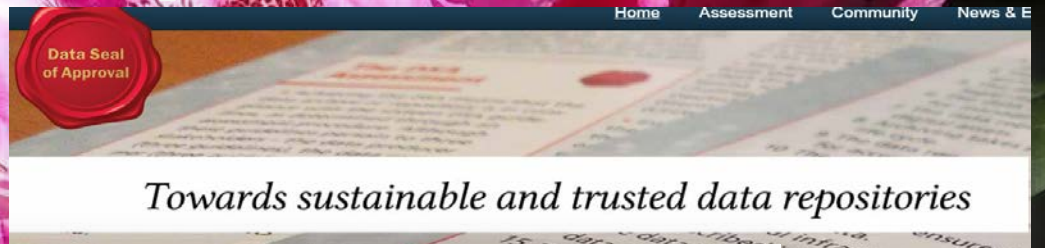
Rule 1. Love your data, and help others love it too.

Data management is a repeat-play game. If you take care to make your data easily available to others, others are more likely to do the same–eventually. While we wait for this new sharing-equilibrium to be reached, you can take two important actions. First, cherish, document, and **publish your data**, preferably using the robust methods described in Rule 2. Get started now, as: better tools and resources for data management are becoming more numerous; universities and research communities are moving toward bigger investments in data repositories (Rule 8); and more librarians and scientists are learning data management skills (Rule 10). At the very least, loving your own data available will serve *you*: you'll be able to find and reuse your own data if you treat them well. Second, enable and **encourage others to cherish, document, and publish their data**. If you are a research scientist, chances are that not only are you an author, but also a reviewer for a specialized journal or conference venue. As a reviewer, **request that the authors of papers you review provide documentation and access to their data** according to the rules set out in the remainder of this article. While institutional approaches are clearly essential (Rules 8 and 10), changing minds one scientist at a time is effective as well.

Rule 2. Share your data online, with a permanent identifier.

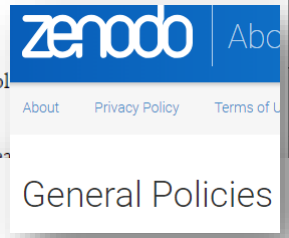
Nothing really lasts forever, so "permanent" actually just means long-lasting. For example, your personal web site is unlikely to be a good option for long-term data storage (yet, in the very short run, putting your data on your site is better than doing nothing at all!). In general, although many papers include URLs to give access to datasets, most become inaccessible within a few years [5]. The best option for releasing your data with long-term guarantee is to **deposit them in whatever data archive is the "go to" place for your field**. A proper, trustworthy, archive will: (1) assign an identifier such as a "handle" (hdl) or "digital object identifier" (doi); (2) require that you provide adequate documentation and metadata; and (3) manage the "care and feeding" of your data by employing good curation practices. If no such archive exists in your field, there are also generic (non-domain-specific) online services that can host your data and issue persistent identifiers (see Rule 8). Pointers to a few generic repositories are listed in the Resources section (A), and longer compilations of such services are at the links in Resources (B).

[trusted repositories]



The Core Trustworthy Data Repository Requirements

1. The repository has an explicit mission to provide access to and preserve data in its domain.
2. The repository maintains all applicable licenses covering data access and use and monitorscompliance.
3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compl
ethical norms.
5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clea
13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.
15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.



Longevity

- **Versions**: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and record are preserved.
- **Replicas**: All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period**: Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Functional preservation**: Zenodo makes no promises of usability and understandability of deposited objects over time.
- **File preservation**: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- **Fixity and authenticity**: All data files are stored along with a MD5 checksum of the file content. Files are regularly checked against their checksums to assure that file content remains constant.
- **Succession plans**: In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

Data Seal of Approval

Data S

Certifying your repository against t
chief among them that the repositor

DSA

2014-2017

After the Data Seal of Approval (DS
on the repository's Web site. The Bo
DSA logo and a link to the organizat

The DSA Board then places the appr
site, using the name of the specific r

Data Sea

10 regole per la cura dei dati

10 Simple Rules for the Care and Feeding of Scientific Data https://arxiv.org/pdf/1401.2134v1.pdf

Alyssa Goodman[1], Alberto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[4], Rosanne Di Stefano[4], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. Hogg[3], Vinay Kashyap[4], Ashish Mahabal[8], Aneta Siemiginowska[4], Aleksandra Slavkovic[9]

Rule 3. Conduct science with a particular level of reuse in mind.

Data from others are hard to use without context describing what the data are and how they were obtained. The W3C Provenance Group defines information *provenance* as the sum of all of the processes, people (institutions or agents), and documents (data included!) that were involved in generating or otherwise influencing or delivering a piece of information. Perfect documentation of provenance is rarely, if ever, attained in scientific work today. The higher the quality of provenance information, the higher the chance of enabling data reuse. In general, data reuse is most possible when: 1) data; 2) metadata (information describing the data); and 3) information about the process of generating those data, such as code, are all provided. In trying to follow the Rules listed in this article, you will do best if you plan in advance for ways to provide all three kinds of information. **In carrying out your work, consider what level of reuse you realistically expect and plan accordingly.** Do you want your work to be fully *reproducible*? If so, then provenance information is a must (e.g., working pipeline analysis code, a platform to run it on, and verifiable versions of the data). Or do you just want your work to be *inspectable*? If so, then intermediate data products and pseudo-code may be sufficient. Or maybe your goal is that your data is *usable* in a wide range of applications? If so, **consider adopting standard formats and metadata standards early on**. At the very least, **keep careful track of versions of data and code**, with associated dates. Taking these steps as you plan and carry out projects will

10 regole per la cura dei dati

10 Simple Rules for the Care and Feeding of Scientific Data https://arxiv.org/pdf/1401.2134v1.pdf

Alyssa Goodman[1], Alberto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[4], Rosanne Di Stefano[4], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. Hogg[8], Vinay Kashyap[4], Ashish Mahabal[8], Aneta Siemiginowska[4], Aleksandra Slavkovic[9]

Rule 4. Publish workflow as context.

Publishing a description of your processing steps offers essential context for interpreting and re-using data. As-such, scientists typically include a "methods" and/or "analysis" section(s) in a scholarly article, used to describe data collection, manipulation, and analysis processes. Computer and information scientists call the combination of the collection methods and analysis processes for a project its "workflow," and they consider the information used and captured in workflow to be part of the "provenance" of the data. In some cases (mostly in genomics), scientists can use existing workflow software in *running* experiments and in *recording* what was done in those experiments, e.g. Gene Pattern. In that best-case scenario, the workflow software, its version, and settings used can be published alongside data using the other rules laid out here. But, it is rare outside of genomics to see the end-to-end process described in a research paper run, orchestrated, and/or recorded by a single software package. In a plausible utopian future, automated workflow documentation could extend to all fields, so that an electronic provenance record could link together all the pieces that led to a result: the data citation (Rule 2), the pointer to the code (Rule 6), the workflow (this Rule), and a scholarly paper (Rule 5). But what can you do now? **At a minimum, provide, alongside any deposit of data, a simple sketch of data flow across software, indicating how intermediate and final data products and results are generated. If it's feasible and you are willing to deal with a higher level of complexity, also consider using an online service to encapsulate your workflow (see Resources (C) for a list of services).** Keep in mind that even if the data used are not "new," in that they come from a well-documented archive, it is still important to document the archive query that produced the data you used, along with all the operations you performed on the data after they were retrieved. Keeping better track of workflow, as context, will likely benefit you and your collaborators enough to justify the loftier, more altruistic, goals espoused here.

10 regole per la cura dei dati

Rules for the Care and Feeding of ata https://arxiv.org/pdf/1401.2134v1.pdf

berto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], ne Di Stefano[4], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. [4], Ashish Mahabal[8], Aneta Siemiginowska[4], Aleksandra Slavkovic[9]

Rule 5. Link your data to your publications as often as possible.

Whether your "data" include tables, spreadsheets, images, graphs, databases and/or code, you should make as much of it as possible available *with* any paper that presents it. **If it's practical and helpful, share your data as early as possible in your research workflow: as soon as you are done with the analysis, even before you write any article(s) about it.** Your data can even be cited before (or without) its inclusion in a paper (see Rule 7). Many journals now offer standard ways to contribute data to their archives and link it to your paper, often with a persistent identifier. Whenever possible, **embed citations (links) to your data and code, each with its own persistent identifier, right into the text of your paper, just like you would reference other literature.** If a journal hosting your paper doesn't offer a place for your data, and or an identifier for it, use a repository (Rule 8) and get your own identifier (Rule 2). At a *minimum*, you can post, and refer to, a package of files (data, codes, documentation on parameters, metadata, license information, and/or lists of links to such) with a persistent online identifier (Rule 2). And, if your domain's journals' policies do not allow for good data-literature interlinking, try to effect change (see Rules 1 and 10).

Rule 6. Publish your code (even the small bits).

Did you write any code to run your analysis? **No matter how buggy and insignificant you may find it, publish it.** Many easy-to-use source code repositories exist, which allow not only hosting of software but also facilitate collaboration and version tracking (see Resources, D). Your code, even the shortest script (whether or not you are proud of its quality), can be an important component for understanding your data and how you got your results [6]. Software plays several roles in relation to data and scientific research, and norms around its publication are still evolving and different across disciplines [7]. In some cases, software is the primary data product (e.g., new algorithms). In some cases, software is the primary data product (e.g., new algorithms). In some other cases, data are the primary research products, yet the best way to document their provenance is to publish the software that was used to generate them as "metadata." In both cases, publishing the source code and its version history is crucial to enhance transparency and reproducibility. The use of open source software when possible reduces barriers for subsequent users of your software related data products. [8] The same best practices discussed above in relation to data and workflow also apply to software materials: cite the software that you use and provide unique, persistent identifiers (Rule 2) to the code you share.

10 regole per la cura dei dati

Rule 7. Say how you want to get credit.

10 Simple Rules for the Care and Feeding of Scientific Data https://arxiv.org/pdf/1401.2134v1.pdf

Alyssa Goodman[1], Alberto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], Merce Crosas[4], Rosanne Di Stefano[4], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. Hogg[8], Vinay Kashyap[4], Ashish Mahabal[8], Aneta Siemiginowska[4], Aleksandra Slavkovic[9]

Chances are that you want to get credit for what you share. The attribution system used for articles, accomplished via citations, often breaks in the case of data and software. When other authors reuse or cite your data or code, you may get an acknowledgement or an incoming link. If you and your colleagues have gone to the trouble to write a "data paper," whose main purpose is to describe your data and/or code, you may also get a citation. [9] But, "data paper" writing is not always desirable, or relevant. So, how do you go about getting the full credit you deserve for your data and code? **The best way is to simply describe your expectations on how you would like to be acknowledged.** If you want, **you can also release your data under a license and indicate explicitly in the paper or in the metadata how you want others to give you credit.** But, while legal mechanisms have advantages, they can also inadvertently lead to limitations on the reuse of the data you are sharing. In any case, **make information about you (e.g. name, institution), about the data and/or code (e.g. origin, version, associated files and metadata), and about exactly how you would like to get credit, as clear as possible.** Easy-to-implement licenses, many of which offer the advantage of being machine-readable, are offered by the Creative Commons organization, as are other similar options, such as those offered by Open Data Commons. Resources section, G, provides more information.

Rule 8. Foster and use data repositories.

Sometimes the hardest and most time-consuming step of sharing data and code is finding and deciding where to put them. Data-sharing practices vary widely across disciplines: in some fields data sharing and reuse are essential and commonplace, while in others data sharing is a "gift exchange" culture [10]. **If your community already has a standard repository, use it. If you don't know where to start looking, or you need help choosing amongst relevant repositories, ask an information specialist, such as a data scientist or a librarian working in your field** (and consult the directories of data repositories listed in Resources, B). When choosing amongst repositories, try to find the one offering the best combination ease-of-deposit, community uptake, accessibility, discoverability, value-added curation, preservation infrastructure, organizational persistence, and support for the data formats and standards you use. **Remember that even if your field has no domain-based repository, your institution may have one**, and your local librarian or archivist can instruct you on how to use that local resource. If neither your community nor your institution has a relevant repository, try a generic repository or consider setting up your own (see Rule 2, and Resources, F).

10 regole per la cura dei dati

10 Simple Rules for the Care and Feeding of Scientific Data https://arxiv.org/pdf/1401.2134v1.pdf

Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], [1], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. habal[8], Aneta Siemiginowska[1], Aleksandra Slavkovic[9]

Rule 9. Reward colleagues who share their data properly.

Whether you do it in person at scientific meetings and conferences or by written communication when reviewing papers and grants, **reward your colleagues who share data and code. Rally your colleagues and engage your community by providing feedback on the quality of the data assets in your field. Praise those following the best practices.** The more the data created by your colleagues is accessible as an organized collection of some sort, the better your community's research capacity. The more data get shared, used, and cited, the more they improve. Besides personal involvement and encouragement, the best way to reward data sharing is by attribution: always cite the sources of data that you use. **Follow good scientific practice and give credit to those whose data you use, following their preferred reference format and according to current best practices.** Standards and practices for citing and attributing data sources are actively being developed through international partnerships. [11, 12]

Rule 10: Be a booster for data science.

As Rule 1 says, it is important not just that *you* love your own data, but that *others* love data too. An attitude that data and code are "2nd class objects," behind traditional scholarly publications is still prevalent. But, every day, as scientists try to use the frustrating but tantalizing hodgepodge of research data available via the present ad-hoc network of online systems, the value of organizing an open network of re-usable data and code is becoming more and more clear, to more and more people. **You, as a scientist, need to help organize your discipline, and your institution to move more quickly to a world of open, discoverable, reproducible data and research. One important step is to *advocate* for hiring data specialists and for the overall support of institutional programs that improve data sharing.** Make sure not only advanced researchers (e.g., postdocs) experience the pleasures of doing research with freely available data and tools: ***explain* and *show* the value of well-loved data to graduate and undergraduate researchers**. *Teach* whole courses, or mini-courses, related to caring for data and software, or incorporate the ideas into existing courses. *Form groups* specific to your discipline to foster data and code sharing. Hold birds-of-a-feather or special *sessions during large meetings* demonstrating examples where good sharing practices have led to better results and collaborations. Lead by practicing what you preach.

[appendice di strumenti utili]

10 Simple Rules for the Care and Feeding of Scientific Data https://arxiv.org/pdf/1401.2134v1.pdf

[1], Alberto Pepe[1,*], Alexander W. Blocker[4], Christine L. Borgman[2], Kyle Cranmer[3], Rosanne Di Stefano[1], Yolanda Gil[5], Paul Groth[6], Margaret Hedstrom[7], David W. Kashyap[1], Ashish Mahabal[8], Aneta Siemiginowska[1], Aleksandra Slavkovic[9]

B: Directories of Research Data Repositories

- DataBib. Databib is a tool for helping people identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search.
- re3data.org. Re3data is a global registry of research data repositories from different academic disciplines for researchers, funding bodies, publishers and scholarly institutions.
- Data repositories, Open Access Directory A list of repositories and databases for open data.
- Force 11 Catalog A dynamic inventory of web-based scholarly resources, a collection of alternative publication systems, databases, organizations and groups, software, services, standards, formats, and training tools.

C: Workflow Management Systems

- Taverna An open source and domain-independent Workflow Management System – a suite of tools used to design and execute scientific workflows and aid in silico experimentation.
- Kepler Software designed to help scientists, analysts, and computer programmers create, execute, and share models and analyses across a broad range of scientific and engineering disciplines.
- Wings A semantic workflow system that assists scientists with the design of computational experiments.
- VisTrails An open-source scientific workflow and provenance management system that supports data exploration and visualization.
- Knime A graphical workbench for the entire analysis process: data access, data transformation, initial investigation, powerful predictive analytics, visualization and reporting.

D: Source Code Repositories

- Github A web-based hosting service for software development projects that use the Git revision control system, including many open source projects.
- Git A free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.
- Mercurial A free, distributed source control management tool. It efficiently handles projects of any size and offers an easy and intuitive interface.
- BitBucket A web-based hosting service for projects that use either the Mercurial or Git revision control systems.

E: Systems to Package, Access, and Execute Data and Code

- iPython Notebooks A web-based interactive computational environment where you can combine code execution, text, mathematics, plots and rich media into a single document.

Perché i dati aperti?



-meglio basarsi sui DATI che sulla loro interpretazione
[data make up per pubblicare…]
-confrontare/dibattere con i propri dati
- creare nuova conoscenza aggiungendo i propri dati

Sharing data: good for science, good for you



https://www.youtube.com/watch?v=HJbo-OAaJ1I&feature=youtu.be



http://www.dcc.ac.uk/news/managing-research-data-video

Condividere



Dati chiusi: solo paura

Alastair Dunning
@alastairdunning

Following

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

Traduci dalla lingua originale: inglese

12:49 - 27 feb 2018

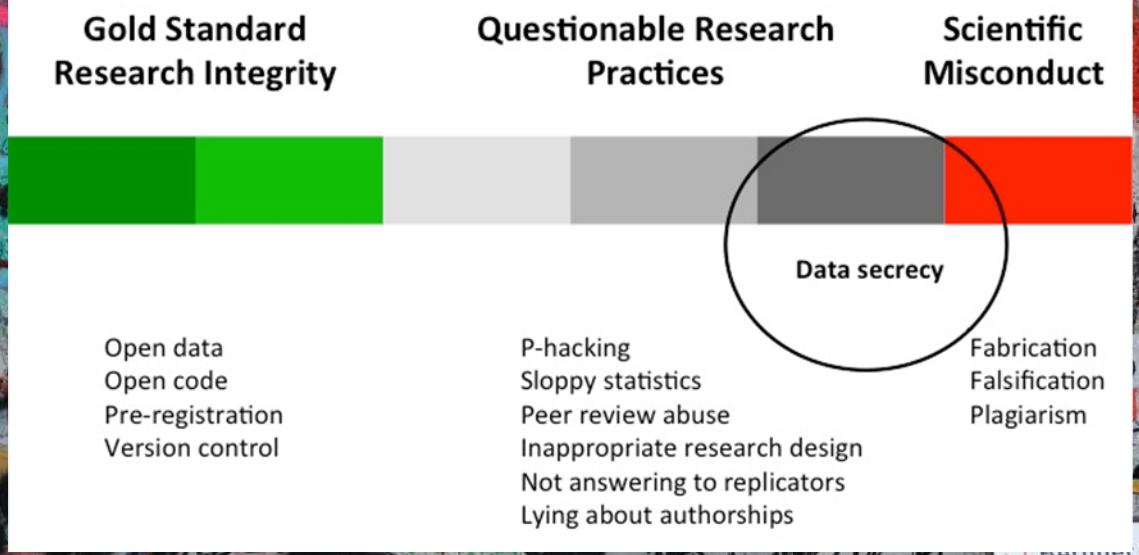
https://twitter.com/alastairdunning/status/968453078218395648

2 Retweet 8 Mi piace

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



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



People will contact me to ask about stuff

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

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

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that misinterpret it by accident will be grateful for the help." From the UK Data Archive: "Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data."

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

My data is not very interesting

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

I might want to use it in a research paper

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

C&A suggest the embargo route: "One option is to have an automatic or optional embargo; require people to archive their data at the time of creation but it becomes public after X months. You could even give the option to renew the embargo so only things that are no longer cared about become published, but nothing is lost and eventually everything can become open." Researchers like to have a say in the use of their datasets, but I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository.

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

I'm not sure I own the data

My data is too complicated.

C&A: "Don't be too smug. If it turns out it's not that complicated, it could harm your professional [standing]." I would add that if it's too complicated to share, then it's too complicated to reproduce, which means it's arguably not real scientific progress. This can be solved by more documentation.

My data is embarrassingly bad

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

It's not a priority and I'm busy

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

CARLY STRASSER

Open Science + Data Sharing Advocate

Home

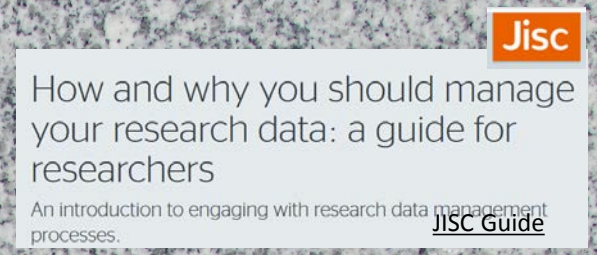
Blog Posts

Previous Research

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

Closed Data… Excuses, Excuses

Perché occuparsi dei dati?



JISC Guide

EVITARE DI PERDERLI

ORGANIZZARLI PER RENDERE PIÙ EFFICACE LA RICERCA

(SE APERTI) ESSERE PIÙ VISIBILI

ALCUNI SONO UNICI E IRRIPETIBILI (meteorologia)

PERMETTERE VALIDAZIONE E CONTROLLI

(SE APERTI) FAVORIRE COLLABORAZIONI

MIGLIORARE INTEGRITÀ DELLA RICERCA

(SE APERTI) FAVORIRE RIUSO INEDITO

ESSERE RIPRODUCIBILI

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

Perché occuparsi dei dati?

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

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

I dati creano ponti fra le discipline



Perché i dati negativi?



…pubblicando anche i dati negativi si evitano duplicazioni inutili…

…torniamo ai nostri FAIR data



https://youtu.be/jFekfemq7qU



https://youtu.be/_OZbYz-GJLk

LIBER Webinar: Turning FAIR Data Into Reality

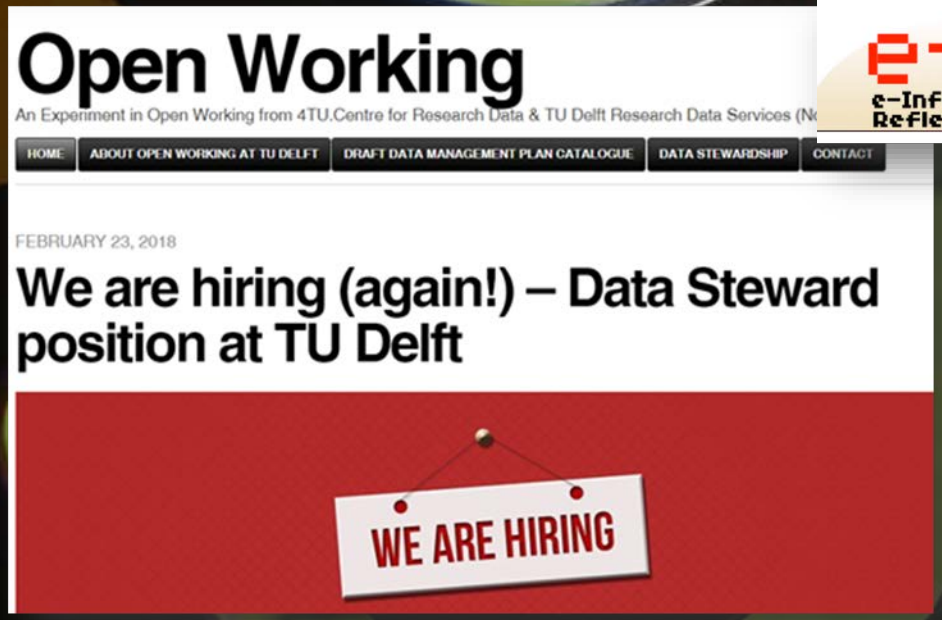
This is a recording of a LIBER webinar, made on 23 April 2018. It focuses …

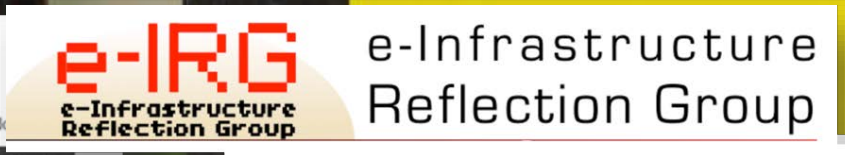
…servono profess

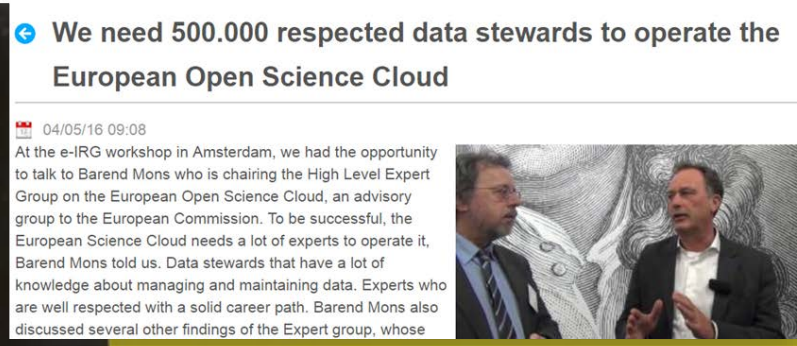


Report, 2016

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



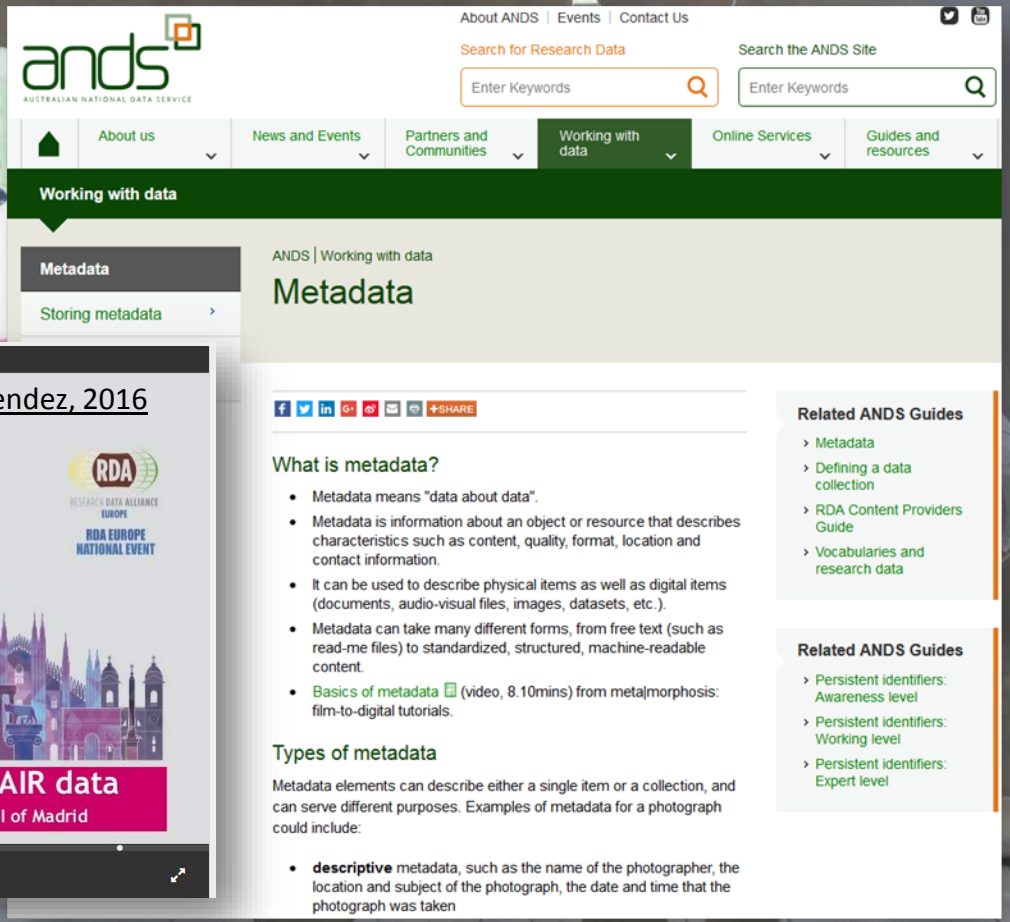




F = findable: I metadati



Eva Mendez, 2016



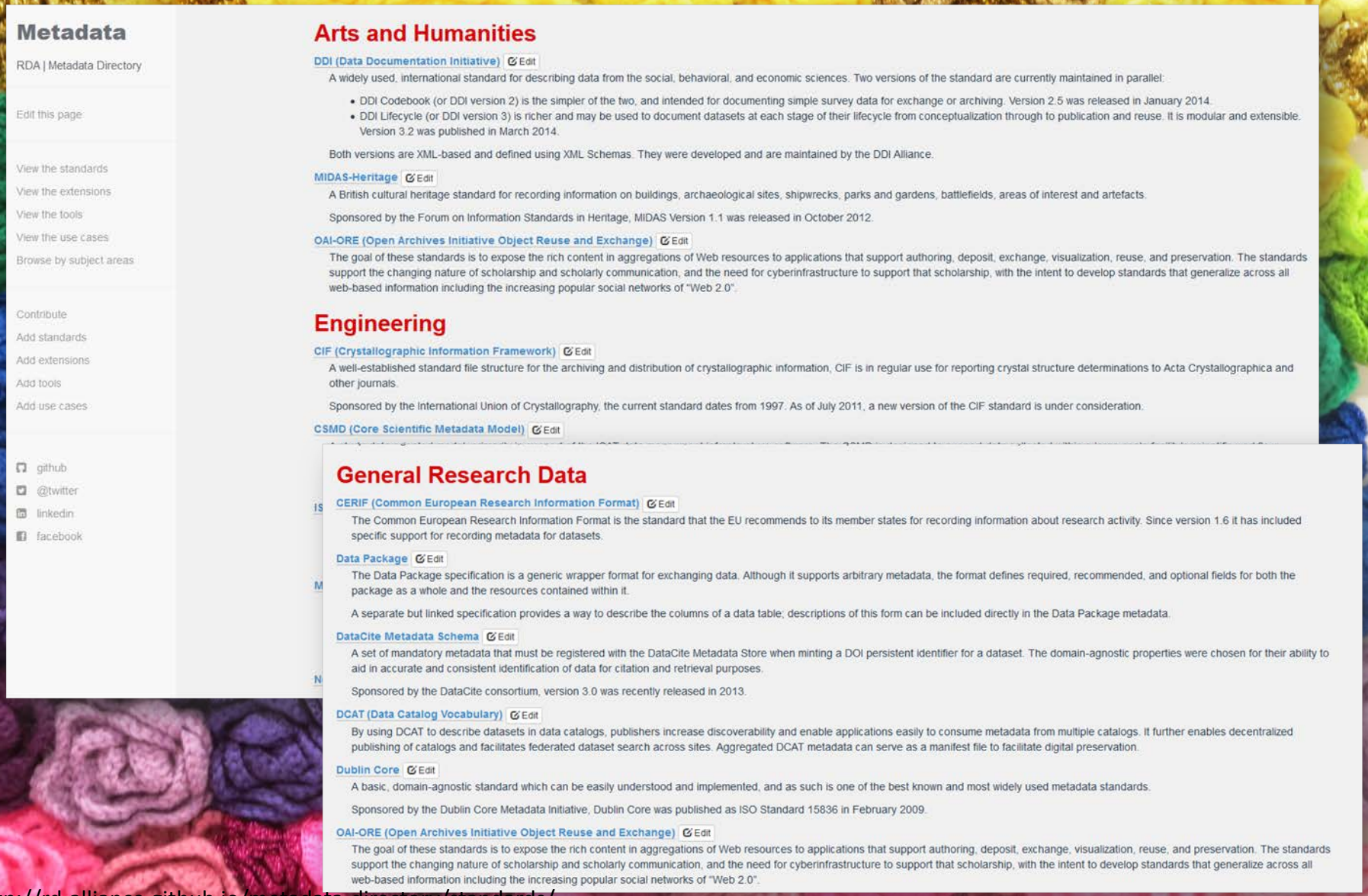
https://www.ands.org.au/working-with-data/metadata

F = findable. Tipi di metadati

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

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

F = findable. Metadata standards



Metadata

RDA | Metadata Directory

Edit this page

View the standards
View the extensions
View the tools
View the use cases
Browse by subject areas

Contribute
Add standards
Add extensions
Add tools
Add use cases

github
@twitter
linkedin
facebook

Arts and Humanities

DDI (Data Documentation Initiative) Edit

A widely used, international standard for describing data from the social, behavioral, and economic sciences. Two versions of the standard are currently maintained in parallel:

- DDI Codebook (or DDI version 2) is the simpler of the two, and intended for documenting simple survey data for exchange or archiving. Version 2.5 was released in January 2014.
- DDI Lifecycle (or DDI version 3) is richer and may be used to document datasets at each stage of their lifecycle from conceptualization through to publication and reuse. It is modular and extensible. Version 3.2 was published in March 2014.

Both versions are XML-based and defined using XML Schemas. They were developed and are maintained by the DDI Alliance.

MIDAS-Heritage Edit

A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts.

Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.

OAI-ORE (Open Archives Initiative Object Reuse and Exchange) Edit

The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of "Web 2.0".

Engineering

CIF (Crystallographic Information Framework) Edit

A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals.

Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.

CSMD (Core Scientific Metadata Model) Edit

General Research Data

CERIF (Common European Research Information Format) Edit

The Common European Research Information Format is the standard that the EU recommends to its member states for recording information about research activity. Since version 1.6 it has included specific support for recording metadata for datasets.

Data Package Edit

The Data Package specification is a generic wrapper format for exchanging data. Although it supports arbitrary metadata, the format defines required, recommended, and optional fields for both the package as a whole and the resources contained within it.

A separate but linked specification provides a way to describe the columns of a data table; descriptions of this form can be included directly in the Data Package metadata.

DataCite Metadata Schema Edit

A set of mandatory metadata that must be registered with the DataCite Metadata Store when minting a DOI persistent identifier for a dataset. The domain-agnostic properties were chosen for their ability to aid in accurate and consistent identification of data for citation and retrieval purposes.

Sponsored by the DataCite consortium, version 3.0 was recently released in 2013.

DCAT (Data Catalog Vocabulary) Edit

By using DCAT to describe datasets in data catalogs, publishers increase discoverability and enable applications easily to consume metadata from multiple catalogs. It further enables decentralized publishing of catalogs and facilitates federated dataset search across sites. Aggregated DCAT metadata can serve as a manifest file to facilitate digital preservation.

Dublin Core Edit

A basic, domain-agnostic standard which can be easily understood and implemented, and as such is one of the best known and most widely used metadata standards.

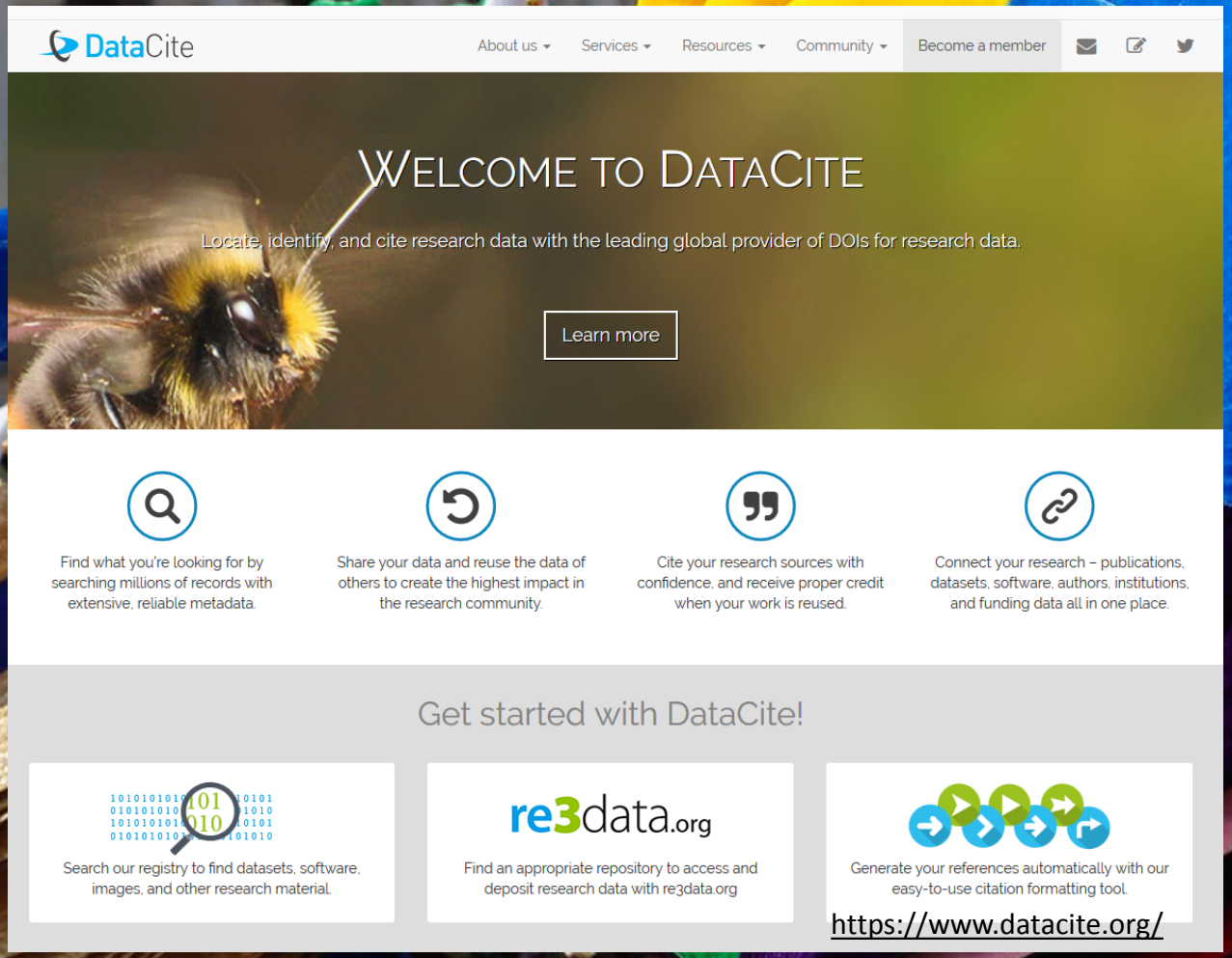
Sponsored by the Dublin Core Metadata Initiative, Dublin Core was published as ISO Standard 15836 in February 2009.

OAI-ORE (Open Archives Initiative Object Reuse and Exchange) Edit

The goal of these standards is to expose the rich content in aggregations of Web resources to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. The standards support the changing nature of scholarship and scholarly communication, and the need for cyberinfrastructure to support that scholarship, with the intent to develop standards that generalize across all web-based information including the increasing popular social networks of "Web 2.0".

http://rd-alliance.github.io/metadata-directory/standards/

F = findable. Dati riconoscibili e citabili [identificativi persistenti]



https://www.datacite.org/

… [per chi non può usare il DOI]

FORCE11
The Future of Research Communications and e-Scholarship

ABOUT | COMMUNITY | GROUPS | RESOURCES | NEWS + BLOGS | EVENTS | PUBLICATIONS

FORCE11 » Community News » Introducing a new standard for the citation of research data

INTRODUCING A NEW STANDARD FOR THE CITATION OF RESEARCH DATA

Posted by Jennifer McLennan | May 8, 2018 | Sign In or Join Now! to post comments

Rules, registry and recommendations

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

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

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

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

SCIENTIFIC DATA

Altmetric: 20 May 8, 2018 More detail »

Article | OPEN

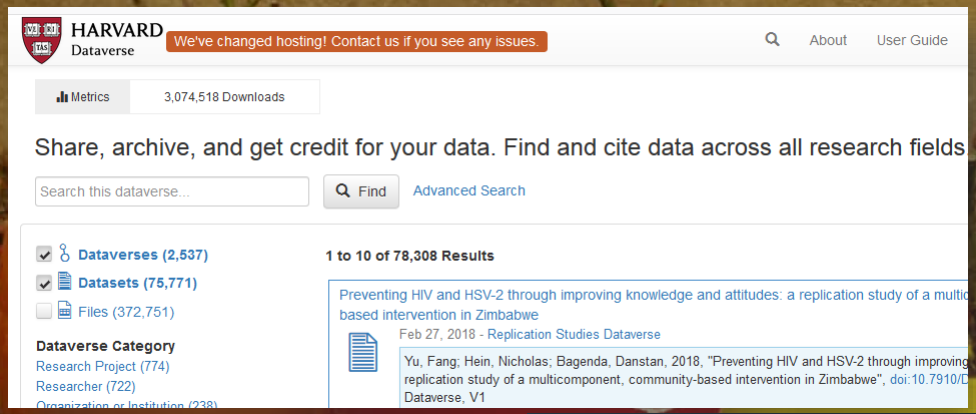
Uniform resolution of compact identifiers for biomedical data

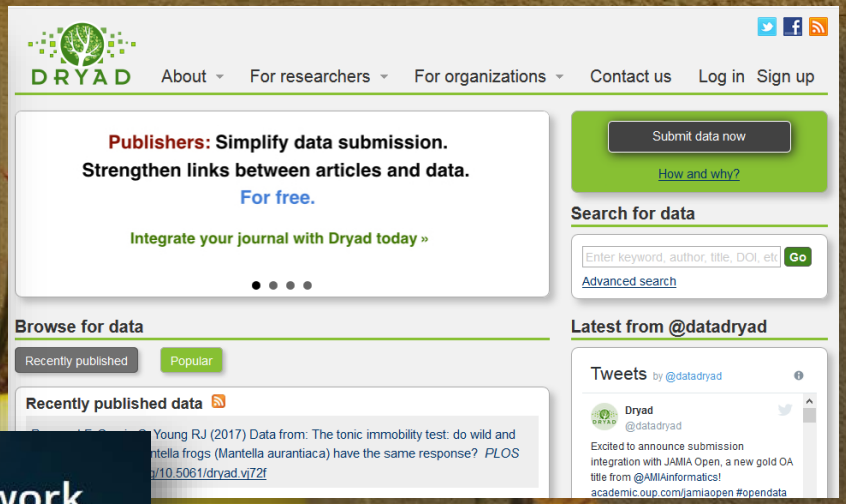
Sarala M. Wimalaratne, Nick Juty, John Kunze, Greg Janée, Julie A. McMurry, Niall Beard, Rafael Jimenez, Jeffrey S. Grethe, Henning Hermjakob, Maryann E. Martone & Tim Clark

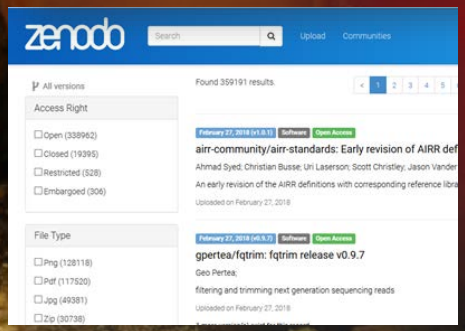
tember 2017
uary 2018
y 2018

A reasonable solution to the identifiers problem is to assign Digital Object Identifiers (DOIs) to identify datasets. DOIs are already widely used in the publishing world as persistent identifiers for scholarly publications. They have been adopted by generalist data repositories such as Dryad, FigShare, Zenodo and Dataverse, as well as by domain data repositories outside of biomedicine. Handles[14], which underlie the DOI system, may also be used directly. The DataCite consortium provides a robust central means for assigning DOIs to data.

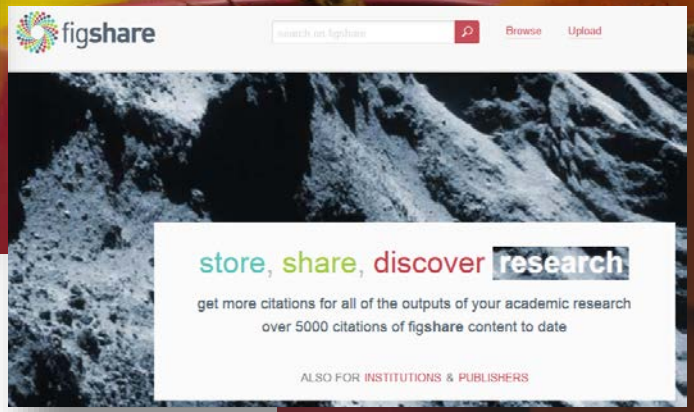
However, DOIs are not commonly used for biomedical data, which is partitioned across over 600 autonomous repositories that are independently funded. Instead, in biomedicine there has been a











General depositories for research data

The following depositories are of interest to researchers in all domains:

- Zenodo (not-for-profit, hosted by CERN): https://zenodo.org:
- Dryad (not-for-profit membership organisation): http://www.datadryad.org
- Figshare (free service provided by private company): https://figshare.com
- Open Science Framework (not-for-profit, developed and maintained by the Center for Open Science[1]): https://osf.io
- Harvard Dataverse (not-for-profit, hosted by the Institute for Quantitative Social Studies IQSS at Harvard University): https://dataverse.harvard.edu

A = accessible. Cercate un data repository?



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
Search... Search

re3data.org

Filter

Subjects
Content Types
Countries
AID systems
API
Certificates
Data access
Data access restrictions
Database access
Database access restrictions
Database licenses
Data licenses
Data upload
Data upload restrictions
Enhanced publication
Institution responsibility type
Institution type
Keywords
Metadata standards
PID systems
Provider types
Quality management
Repository languages
Software
Syndications
Repository types
Versioning

genetic

← Previous 1 2 3 4 Next →

Found 87 result(s)

GESDB
Genetic Epidemiology Simulation Database
Subject(s)
Content type(s)
Country
GESDB is a platform for sharing simulation data and discussion manuscripts. The forum provides a platform for Q&A for the sim

Drosophila Genetic Reference Panel 2
DGRP2
Subject(s)
Content type(s)
Country
The Drosophila Genetic Reference Panel (DGRP) is a populatic complex traits, and a community resource for whole genome as

Online Mendelian Inheritance in Man
OMIM
Subject(s)

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

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

Read more

Three new DOI Fabrica features to simplify account management

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

Read more

One step closer towards instant DOI search results

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

Read more

https://www.re3data.org/

A = accessible. I data journals

Data Journals

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

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

Dataset Description

Object Name

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

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

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

Publication Date
05/02/2012

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

License
Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. http://dx.doi.org/ org/10.5284/100 dx.doi.org/10.528 to the fact that th cal. The ASP data locations, dates a ally in the databas structures and ter

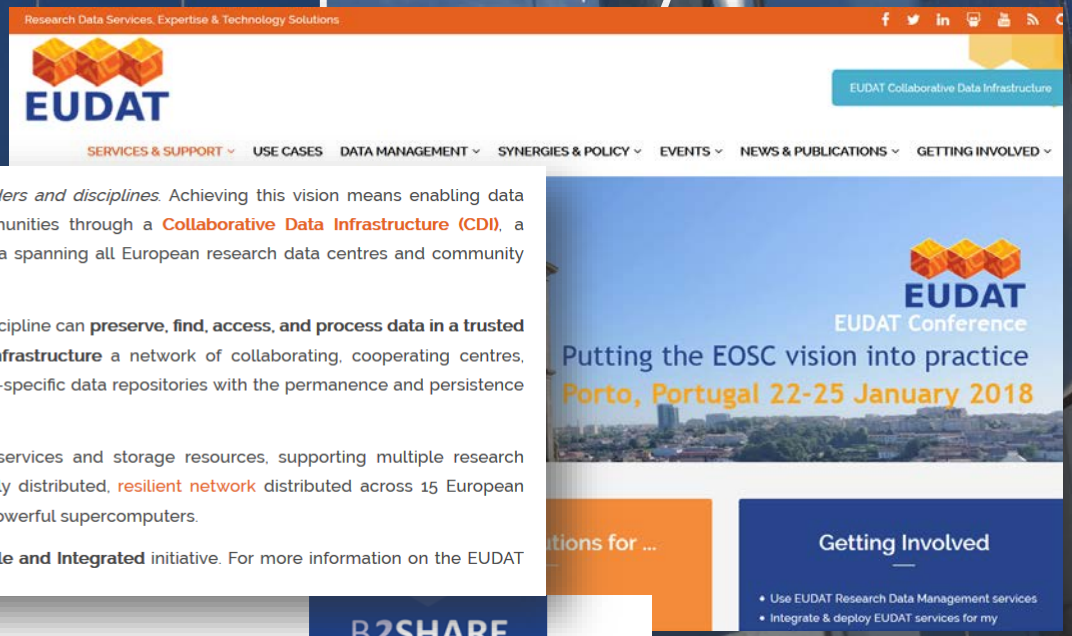
Data journals

Panayiota Polydoratou

Alexander Technological Educational Institute of Thessaloniki

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

A = accessibile. Data repository+servizi



EUDAT's vision is *Data is shared and preserved across borders and disciplines*. Achieving this vision means enabling data stewardship within and between European research communities through a **Collaborative Data Infrastructure (CDI)**, a common model and service infrastructure for managing data spanning all European research data centres and community data repositories.

European researchers and practitioners from any research discipline can **preserve, find, access, and process data in a trusted environment, as part of the EUDAT Collaborative Data Infrastructure** a network of collaborating, cooperating centres, combining the richness of numerous generic and community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres.

EUDAT offers heterogeneous research data management services and storage resources, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network distributed across 15 European nations and data is stored alongside some of Europe's most powerful supercomputers.

EUDAT is a Service-oriented, Community driven, Sustainable and Integrated initiative. For more information on the EUDAT services, check out the Service Catalogue.

What is B2SHARE?

B2SHARE

B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store and share small-scale research data from diverse contexts.

A winning solution to:

- **Store**: facilitates research data storage
- **Preserve**: guarantees long-term persistence of data
- **Share**: allows data, results or ideas to be shared worldwide

B2SHARE features

- integrated with the **EUDAT collaborative data infrastructure**
- **free upload and registration** of stable research data
- data assigned a **permanent identifier**, which can be retraced to the data owner
- data **owner defines access** policy
- **community-specific metadata** extensions and user interfaces
- **openly accessible and harvestable** metadata
- representational **state transfer application programming interface** (REST API) for integration with community sites
- data **integrity ensured by checksum** during data ingest
- **professionally managed storage** service – no need to worry about hardware or network
- EUDAT **user support**
- **monitoring** of availability and use

Use B2SHARE | Help Desk

Services

- B2HANDLE
- B2ACCESS
- B2DROP
- B2SHARE
- B2STAGE
- B2SAFE
- B2FIND

About EUDAT

- What is EUDAT?
- EUDAT Partners

https://www.eudat.eu/

A = accessible. I formati preferiti

Data Archiving and Networked Services
DANS
HOME DEPOSIT

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

https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats

A = accessible. I formati preferiti

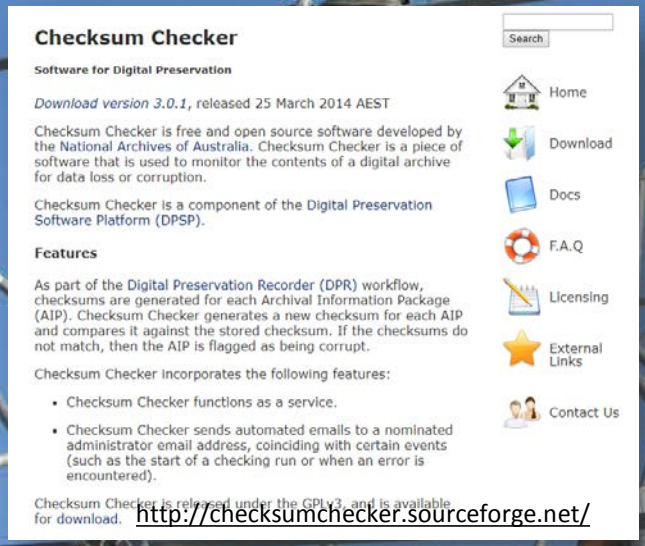
Data Archiving and Networked Services
DANS
HOME DEPOSIT

Type	Preferred format(s)	Non-preferred format(s)
Vector images	• SVG (.svg)	• Illustrator (.ai) • EPS (.eps)
Audio	• BWF (.bfw) • MXF (.mxf) • Matroska (.mka) • FLAC (.flac)	• WAVE (.wav) • MP3 (.mp3) • AAC (.aac, .m4a) • AIFF (.aif, .aiff) • OGG (.ogg)
Video	• MXF (.mxf) • Matroska (.mkv)	• MPEG-4 (.mp4, .m4a, .m4v) • MPEG-2 (.mpg, .mpeg, .m2v, mpg2) • AVI (.avi) • QuickTime (.mov, .qt)
Computer Aided Design (CAD)	• AutoCAD DXF v. R12 (.dxf)	• AutoCAD other versions (.dwg, .dxf)
Geographical Information (GIS)	• GML (.gml) • MIF/MID (.mif/.mid)	• ESRI Shapefiles (.shp & related files) • MapInfo (.tab & related files) • KML (.kml)
Geo referenced images	• GeoTIFF (.tif, .tiff)	• TIFF World File (.tfw & .tif)
Raster GIS	• ASCII GRID (.asc, .txt)	• ESRI GRID (.grd & related files)
3D	• WaveFront Object (.obj) • X3D (.x3d)	• COLLADA (.dae) • Autodesk FBX (.fbx)
RDF	• W3C standards	

https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats

A = accessible. Come conservare

CONSERVAZIONE A BREVE, MEDIO E LUNGO PERIODO hanno principi e strumenti diversi



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/

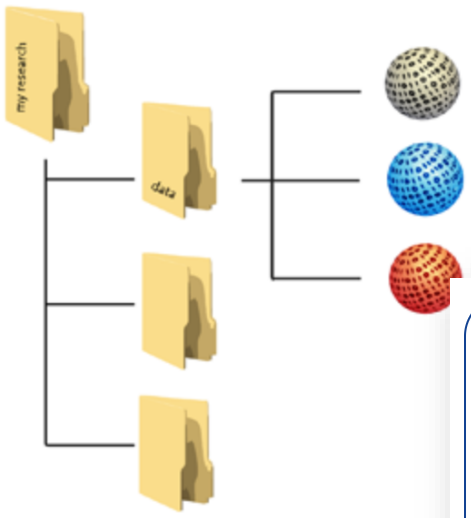
Storage Solutions	Advantages	Disadvantages	Suitable for
Personal Computer & Laptop	*Always available* *Portable*	*Drive may fail* *Laptop may be stolen*	*Temporary storage*
Networked drives File servers managed by your university, research group or facilities like a NAS-server	*Regularly backed up* *Stored 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).

A = accessible. Come conservare

*"Not one system is going to put structure in your data, because it doesn't understand your research. The structure comes from you." - **Patrick Vandewalle***



Folder structure

Main points

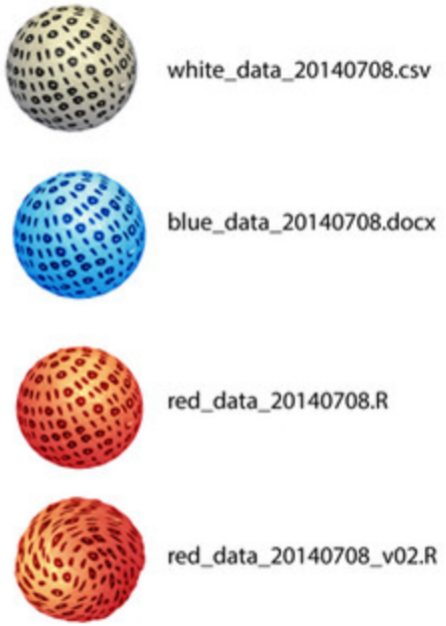
If you want your research data to be easily traced and interpreted, it is vital to store it in a structured and consistent way with appropriate data documentation (metadata). The **folder structure** used for the data and the file names for the data files should be logical (see box).

Meaningful file names

Below are tips on meaningful and consistent file names. Read more in *'Naming files and folders'*.[2]

- Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD of yymmdd).
- Do not use strange characters like ?\!@*%{[<> in the file name.
- Use traceable file names, such as Project_Instrument_locatie_YYYYMMDD.ext.
- Make sure to only use each file once in the folder structure. If you store a file in more than one place, several versions of the same file can unwillingly be created.
- See also version management.

It is good practice to note the file naming and its meaning in a readme.txt.



Meaningful file names

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

A = accessible.

UNIVERSITY OF LEICESTER

Naming files and folders

Naming conventions are rules which enable the titling of electronic and physical folders, d
consistent and logical way. This ensures that the correct records can be located, identified
in a timely fashion, and that they are stored in an appropriate secure location. Ideally, the
structure the documents and directories you create is at the start of a project.

The principles of naming conventions can equally be applied to electronic and physical files/folde

Benefits of naming conventions

Naming records consistently, logically and in a predictable way will distinguish similar records fro
doing so will facilitate the storage and retrieval of data. Through consistency and the application
secure storage, and the ability to locate and access information.

File identifiability

Good practice dictates that all information (files, datasets, documents, or records) should be iden
achieved by following good practices by applying referencing to all documents/files.

Document/file references will include:

- File name, or full file path including file name
- Name/role of file author(s) or originator(s)
- Date of creation, edit or event which is the subject of the document/file
- Version number if applicable

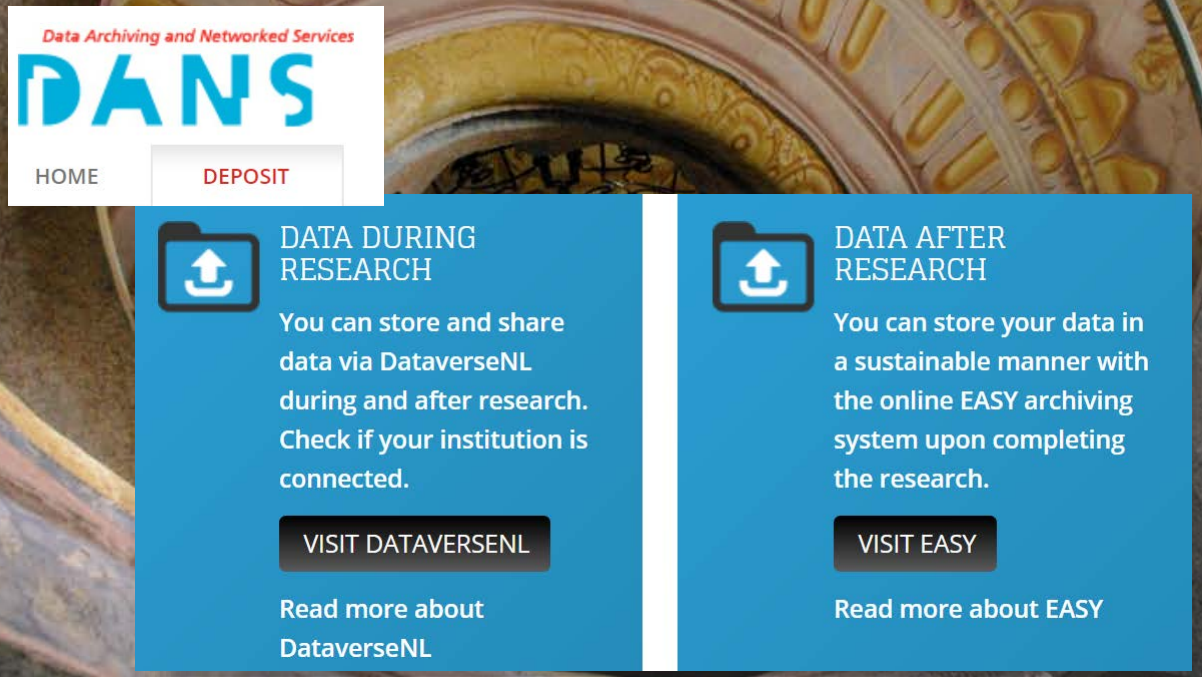
Good Practice: Remove spaces from file names or use punctuation such as unders
words e.g. "AHRC_TechnicalApp_Response20120925.docx" or "AHRC-TechnicalApp-Re
"what we got back from funders about the data stuff.docx"



Suggested file and folder naming conventions

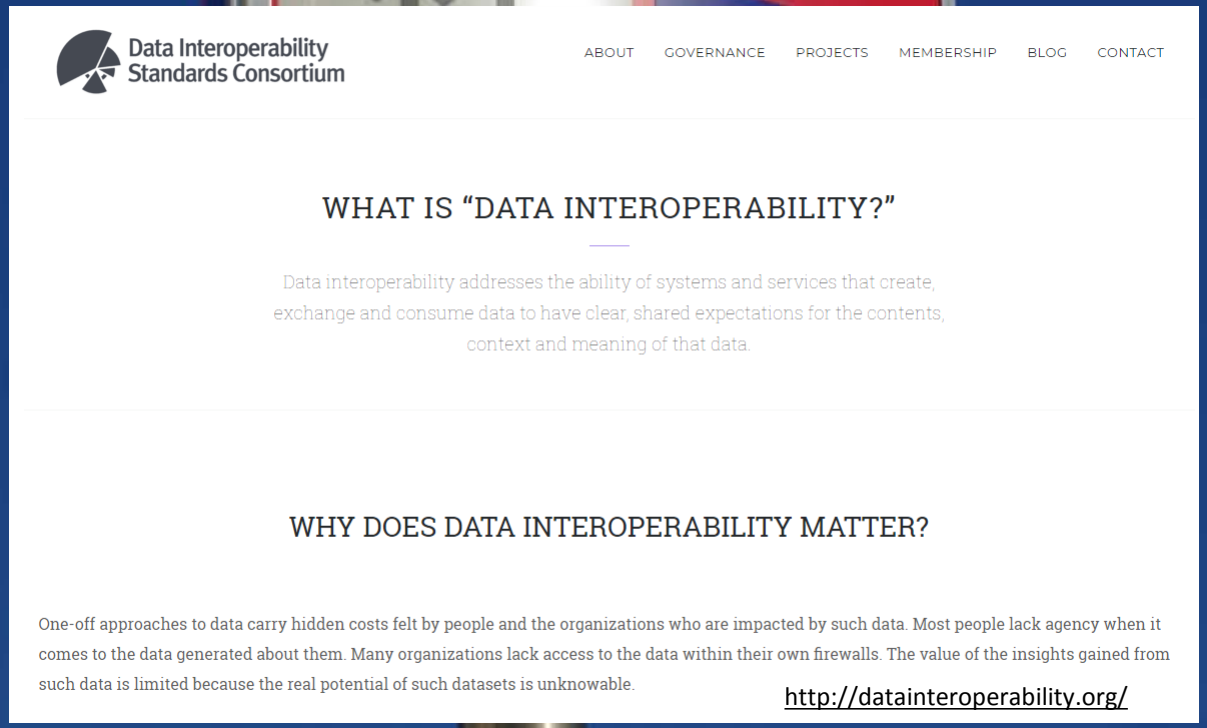
1. Keep file and folder names short, but meaningful.
2. Avoid unnecessary repetition and redundant words in file names and file paths.
3. Use capital letters to delimit words, not spaces.
4. When including a number in a file name always give it as a two-digit number rather than one, i.e. 01, 02 … 99, 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 'd' followed by the version number and, where applicable, 'd' indicating 'draft version'.
13. Avoid using non-alphanumeric characters in file names.

A = accessible. Prima durante e dopo



Durante il progetto e dopo il deposito ha necessità differenti.
Durante: deve essere consentito anche uso condiviso

I = Interoperability



http://datainteroperability.org/

R = reusable. Documentazione

rdnl research data netherlands — *Essentials 4 Data Support*

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

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

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

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

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

I dati per essere riusati devono avere:

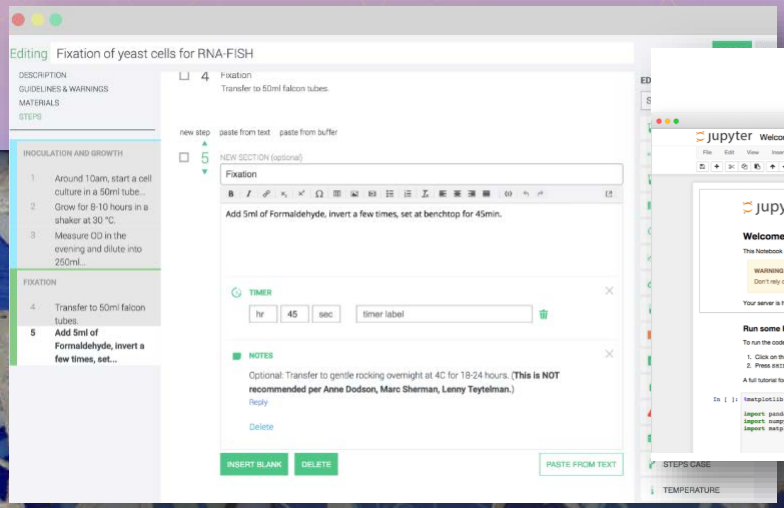
- DOCUMENTAZIONE
- LICENZE

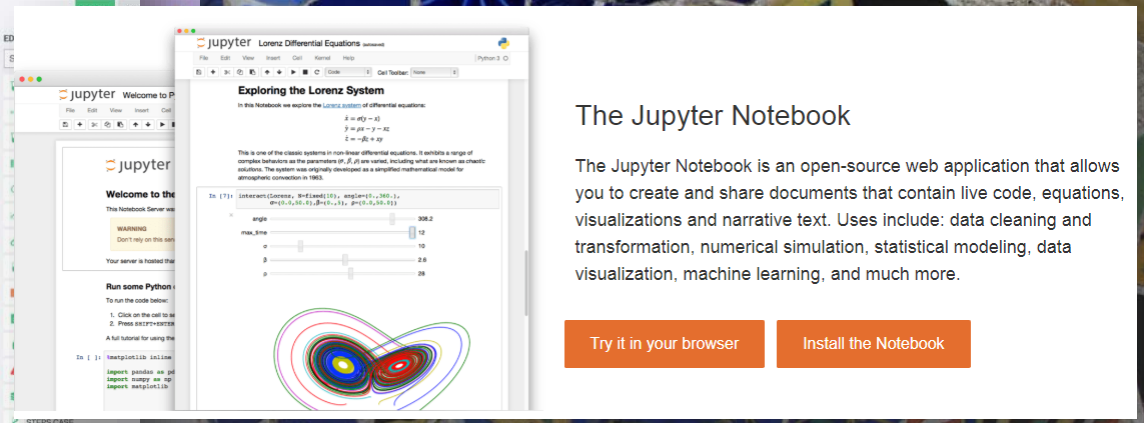
R = reusable. Documentazione

protocols.io

Make your science more reproducible

protocols.io is the #1 open access repository for science methods





…può richiedere tempo le prime volte… ma poi, oltre ad avere tutto tracciabile e riproducibile, di fatto si ha l'articolo pronto

R = reusable

Delta May 7, 2018 Spotlight | Education Science Campus Opinion Delta Lab



Don't sit on our research data; share it with the rest of academia. This was the case argued by TU Library director Wilma van Wezenbeek and Professor of Data Science Geert-Jan Houben at the Dies Natalis. "Sharing your data will make you a better student, lecturer or researcher."

Un ricercatore è come un fotografo, crea un'impressione virtuale della realtà. Solo lui/lei conosce i dettagli. Solo lui/lei può spiegare come usare/riusare

A researcher is like a photographer. Scientists collect data and use it to create a virtual impression of reality. When was the photo taken, from which angle, with which lighting and shutter speed? Only the researcher knows the details and how the photo (data) should be interpreted.

R = resuable: licenze



È fondamentale associare una licenza ai propri dati perché se no, di fatto, se ne impedisce il riuso: il potenziale utilizzatore NON SA cosa può fare/non fare (normativa complessa+progetti internazionali)

[Dati e copyright]



Che cosa tutela il diritto d'autore?

[continua] Articolo 2 – Legge 633/41

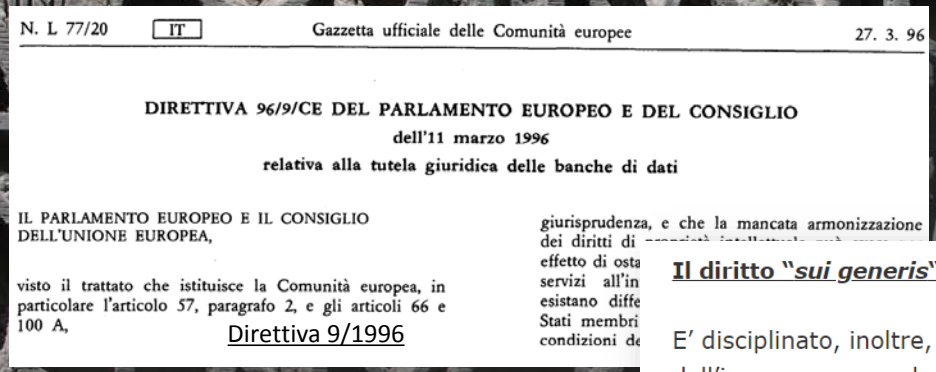
8) i programmi per elaboratore, in qualsiasi forma espressi purché originali quale risultato di creazione intellettuale dell'autore. Restano esclusi dalla tutela accordata dalla presente legge le idee e i principi che stanno alla base di qualsiasi elemento di un programma, compresi quelli alla base delle sue interfacce. Il termine programma comprende anche il materiale preparatorio per la progettazione del programma stesso

9) le banche di dati di cui al secondo comma dell'articolo 1, intese come raccolte di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo. La tutela delle banche di dati non si estende al loro contenuto e lascia impregiudicati diritti esistenti su tale contenuto.

10) Le opere del disegno industriale che presentino di per sé carattere creativo e valore artistico.

https://www.openstarts.units.it/bitstream/10077/14731/1/Aliprandi-PubblicareRicercaDidattica.pdf

[Dati e copyright: il diritto sui generis sulle banche dati]



N. L 77/20 IT Gazzetta ufficiale delle Comunità europee 27. 3. 96

DIRETTIVA 96/9/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO
dell'11 marzo 1996
relativa alla tutela giuridica delle banche di dati

IL PARLAMENTO EUROPEO E IL CONSIGLIO DELL'UNIONE EUROPEA,

visto il trattato che istituisce la Comunità europea, in particolare l'articolo 57, paragrafo 2, e gli articoli 66 e 100 A,

giurisprudenza, e che la mancata armonizzazione dei diritti di
effetto di osta
servizi all'in
esistano diffe
Stati membri
condizioni de

Direttiva 9/1996

Il diritto "*sui generis*"

R.Pellegrino, 2014

E' disciplinato, inoltre, il caso in cui il risultato finale di tale attività non sia un'opera dell'ingegno, ma un bene giuridicamente rilevante da tutelare in ragione degli ingenti investimenti finanziari, di tempo o di lavoro.

Colui che effettua tali investimenti finalizzati alla realizzazione di una banca di dati è definito dalla legge il **costitutore** (art. 102-bis, comma 1 lett. a) l.d.a.), ed è titolare di un diritto *sui generis*: il costitutore di una banca di dati, cittadino o residente abituale sul territorio dell'Unione Europea, può "*vietare le operazioni di estrazione ovvero reimpiego della totalità o di una parte sostanziale della stessa", salvi, ovviamente, i diritti già esistenti sul contenuto della raccolta o parti di esso*" (art. 102-*bis*, comma 3, l.d.a.).

Inoltre, non sono consentiti l'estrazione e il reimpiego di parti non sostanziali fatte in modo sistematico e ripetuto "*qualora presuppongano operazioni contrarie alla normale gestione della banca di dati o arrechino un pregiudizio ingiustificato al costitutore della banca di dati*" (art. 102-*bis*, comma 9, l.d.a.).

La durata del diritto del costitutore è di 15 anni, rinnovabile in caso di modifiche o integrazioni sostanziali apportate alla raccolta, decorrenti dal 1° gennaio dell'anno successivo alla data del

[Dati e copyright: il diritto sui generis sulle banche dati]



2014

N. L 77/20 IT Gazzetta ufficiale delle Comunità europee 27. 3. 96

DIRETTIVA 96/9/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO
dell'11 marzo 1996
relativa alla tutela giuridica delle banche di dati

IL PARLAMENTO EUROPEO E IL CONSIGLIO DELL'UNIONE EUROPEA

visto il trattato che istit
particolare l'articolo 57,
100 A,

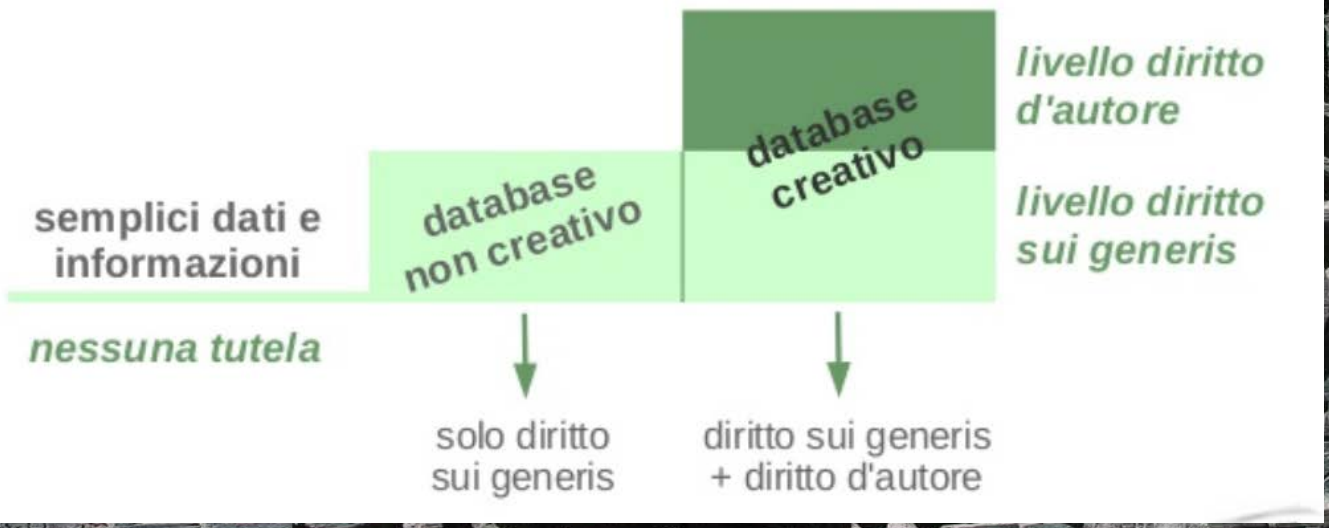
giurisprudenza, e che la mancata armonizzazione

Simone Aliprandi

QUALI DIRITTI SUI DATI?

2014

diversi livelli di tutela



[Dati e copyright: il diritto sui generis sulle banche dati]



2014

Simone Aliprandi

QUALI DIRITTI SUI DATI?

2014

Utili quesiti da porsi

Per capire se un database sia o meno tutelato, dovrei quindi chiedermi...

0) Innanzitutto... siamo in ambito europeo?

1) Ho a che fare con un database (in senso proprio)?

2) Questo database ha o non ha carattere creativo?

3) La costituzione di questo database ha richiesto un rilevante investimento?

4) Sto facendo un'attività di estrazione o re-impiego di parti sostanziali del database?

norme di riferimento

- direttiva 96/9/CE → definizioni e principi generali (comuni a tutti i paesi UE)
- articoli 64 quinquies e sexies, legge 633/1941 → livello diritto d'autore
- articoli 102 bis e ter, legge 633/1941 → livello diritto sui generis

Dati e licenze: come



Le linee guida dell'AgID sull'open data e il tema delle licenze (capitolo 8)

Si vuole qui ribadire l'importanza di associare ai dati pubblici una licenza aperta [...] che consenta di rispettare requisiti di
I) interoperabilità, anche transfrontaliera (non limitandosi a selezionare una licenza che siavalida e nota solo entro i confini nazionali) e, per quanto possibile,
(II) di massimo riutilizzo dei dati.





https://www.slideshare.net/simonealiprandi/il-licensing-di-dati-e-le-principali-licenze-open-data

Dati e licenze: come



DCC because good research needs good data

How to License Research Data

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

http://www.dcc.ac.uk/resources/how-guides/license-research-data

Dati e licenze: come

DCC because good research needs good data

How to License Research Data

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

Creative Commons at a glance

Good for

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

Watch out for

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

ODC-By at a glance

Good for

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

Watch out for

- attribution stacking

ODC-ODbL at a glance

Good for

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

Watch out for

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

Public domain at a glance

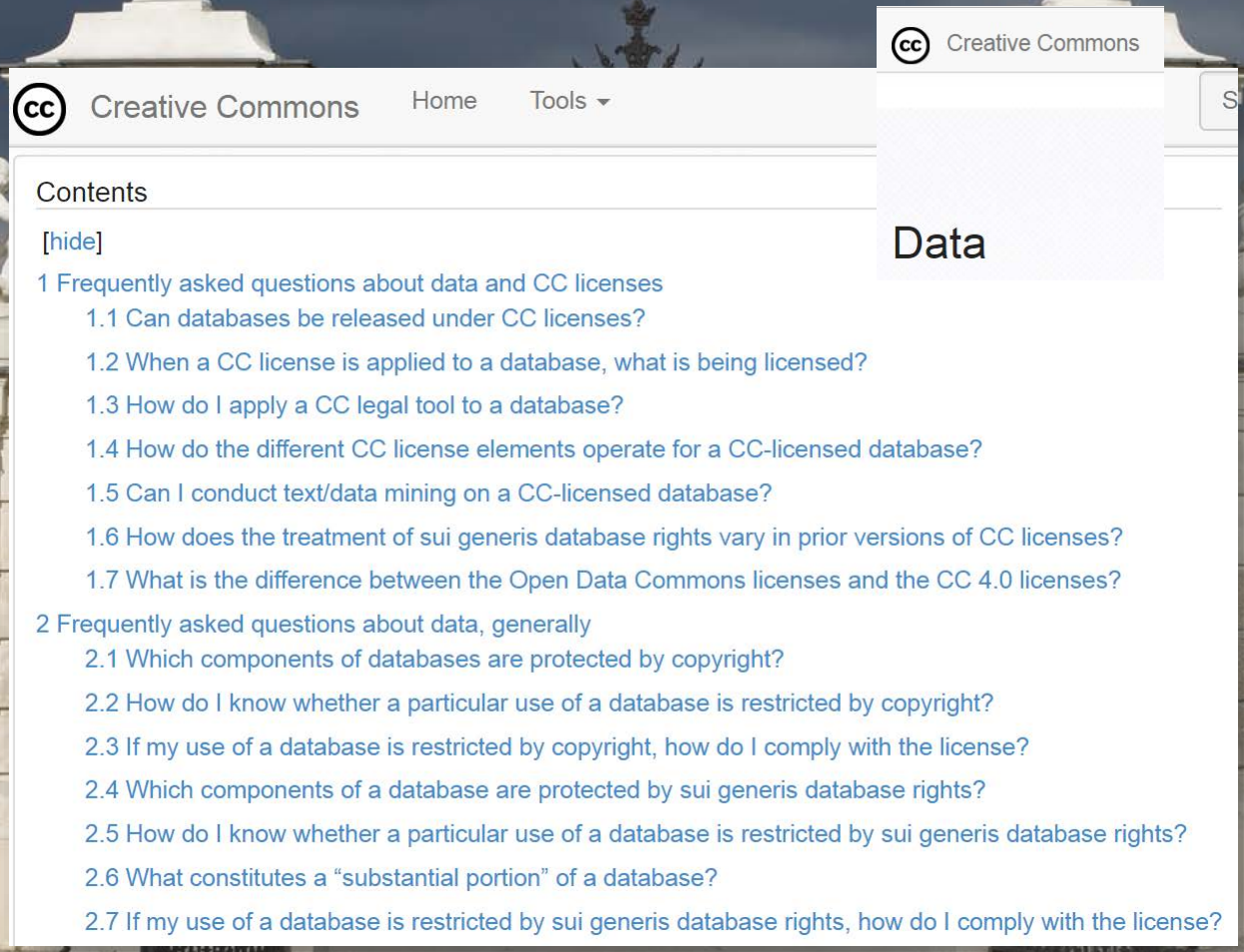
Good for

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

Watch out for

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

Dati e licenze: come



Creative Commons

Creative Commons Home Tools

Contents

[hide]

1 Frequently asked questions about data and CC licenses
- 1.1 Can databases be released under CC licenses?
- 1.2 When a CC license is applied to a database, what is being licensed?
- 1.3 How do I apply a CC legal tool to a database?
- 1.4 How do the different CC license elements operate for a CC-licensed database?
- 1.5 Can I conduct text/data mining on a CC-licensed database?
- 1.6 How does the treatment of sui generis database rights vary in prior versions of CC licenses?
- 1.7 What is the difference between the Open Data Commons licenses and the CC 4.0 licenses?

2 Frequently asked questions about data, generally
- 2.1 Which components of databases are protected by copyright?
- 2.2 How do I know whether a particular use of a database is restricted by copyright?
- 2.3 If my use of a database is restricted by copyright, how do I comply with the license?
- 2.4 Which components of a database are protected by sui generis database rights?
- 2.5 How do I know whether a particular use of a database is restricted by sui generis database rights?
- 2.6 What constitutes a “substantial portion” of a database?
- 2.7 If my use of a database is restricted by sui generis database rights, how do I comply with the license?

Creative Commons

Data

https://wiki.creativecommons.org/wiki/Data#If_my_use_of_a_database_is_restricted_by_sui_generis_database_rights.2C_how_do_I_comply_with_the_license.3F

Quale licenza? (per non sb



5) Data, dataset, databases, etc

These should be under a CC0, which **does not require any attribution** (it is in fact a waiver more than a licence) and allows the widest reuse without imposing any restriction.

(in certain cases you could use a Public Domain Mark to simply state that something is in the public domain worldwide, but if you don't know what it is, just disregard this!)

openMINTED

PERILL DE

CC0??? Senza attribuzione???
Va chiarito che è per essere machine-readable.
Le consuetudini di citazione della fonte si danno per scontate nell'etica della ricerca

Creative Com

creative commons UK

FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE V.0.1

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

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

What is Open Science?

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

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

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

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO

PUBLIC DOMAIN MARK LOGO

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

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

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

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

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

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN

CC ZERO	PUBLIC DOMAIN
'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free	'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

What is Open Science?

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

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

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

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO

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

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

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

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

PUBLIC DOMAIN MARK LOGO

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

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN

CC ZERO	PUBLIC DOMAIN
'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free	'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

nmons e Op

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

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

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

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

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

USARE CC0

- CHIEDERE CHE VENGA DATO CREDITO ALL'AUTORE
- PROPORRE GIÀ LA CITAZIONE-TIPO

(non citare la fonte è scorretto scientificamente)

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

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

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

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

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

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

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

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

I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

We recommend you avoid using a 'No Derivatives' licence. Here's why:

Similar to how a non-commercial licence might restrict meaningful reuse of your dataset, a ND licence can have the same effect: it may prevent someone from recombining and reusing your data for new research. For data to be truly Open Access, it must permit these important types of reuse.

Data Management Plan

PERMETTE DI EVITARE ERRORI DURANTE IL PROGETTO (REGOLE STABILITE ALL'INIZIO)

Lo strumento in cui si raccolgono tutte queste informazioni sui dati è il DMP, Data Management Plan

È UN MODO STRUTTURATO DI PENSARE AI PROPRI DATI: raccolta, conservazione, descrizione, condivisione



È UN «LIVING DOCUMENT», VA AGGIORNATO



Add a "version management" tab to your spreadsheet.

Now, let me expand on this idea.

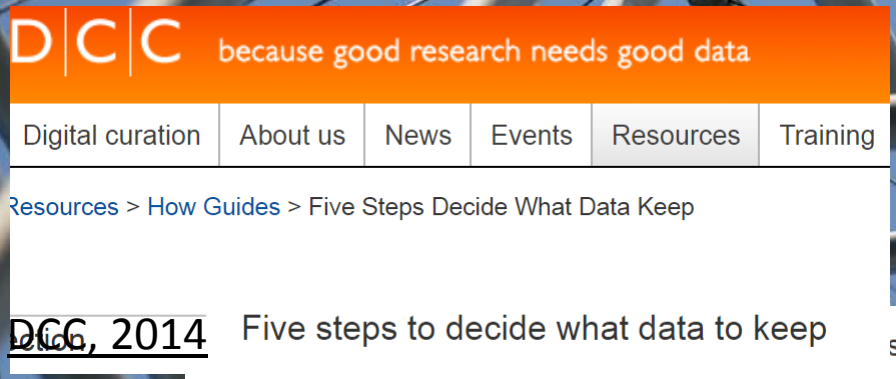
Start by adding an extra "version management" tab to a new spreadsheet. In this sheet, carefully write down a version name (name of the file, typically) in the first column, in the second column the date, and in a third column an explanation of all changes you made to the sheet. Carefully fill out this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

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

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



Cosa conservare?



DCC, 2014

Five steps to decide what data to keep

1. Consider potential **reuse purposes** - what aims could the data meet?
2. Check for indications that it must be kept considering **legal or policy compliance risks**
3. Identify which data should be kept as it may have **long-term value**
4. **Weigh up the costs -** which data management costs have already been incurred and therefore contribute to its value, and how much more is planned and affordable? Where will the funds to pay these costs come from? Considering these questions will give you the cost element of your data appraisal and should help identify any need for external advice, e.g., on how to deal with any shortfall in the budget.
5. **Complete your data appraisal** - this will list what data must, should or could be kept to fulfil which potential reuse purposes. The appraisal should also summarise any actions needed to prepare the data for deposit, or the justification for not keeping it.

Le domande-tipo

DCC

Checklist for a Data Management Plan, v4.0

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

DCC Checklist	DCC Guidance and questions to consider
Administrative Data	
ID	A pertinent ID as determined by the funder and/or institution
Funder	St
Grant Reference Number	Er
Project Name	If
Project Description	Q - V - V - F G Br fo
PI / Researcher	Ni
PI / Researcher ID	E.
Project Data Contact	Ni
Date of First Version	D
Date of Last Update	D
Related Policies	Q - / - [- [- [- [- / G Li m re to
Data Collection	
What data will you collect or create?	Q - V - [- / G Gi w ch of
Documentation and Metadata	
What documentation and metadata will accompany the data?	**Questions to consider:** - What information is needed for the data to be to be read and interpreted in the future? - How will you capture / create this documentation and metadata? - What metadata standards will you use and why? **Guidance:** Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed. Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards.
Ethics and Legal Compliance	
How will you manage any ethical issues?	**Questions to consider:** - Have you gained consent for data preservation and sharing? - How will you protect the identity of participants if required? e.g. via anonymisation - How will sensitive data be handled to ensure it is stored and transferred securely? **Guidance:** Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.
How will you manage copyright and Intellectual Property Rights (IPR) issues?	**Questions to consider:** - Who owns the data? - How will the data be licensed for reuse? - Are there any restrictions on the reuse of third-party data? - Will data sharing be postponed / restricted e.g. to publish or seek patents? **Guidance:** State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.
Storage and Backup	
How will the data be stored and backed up during the research?	**Questions to consider:** - Do you have sufficient storage or will you need to include charges for additional services? - How will the data be backed up? - Who will be responsible for backup and recovery? - How will the data be recovered in the event of an incident?

gesis Leibniz Institute for the Social Sciences

Research data management questions

	Yes	Partly	No
Planning			
• Have you identified responsibilities within the research team and your institution for collection, organization, protection, distribution and storage of data?			
• Have you assigned responsibilities?			
• Have you identified gaps for the provision of research data management beyond the cost of normal research procedures? Are these costs included in your budget?			
Intellectual property			
• Have you established ownership of data?			
• Are you clear about what you can and cannot do with the data?			
Consent and ethics			
• Have you identified any ethical issues that make your research problematic?			
• Does your consent process avoid language explicitly preventing archiving? Does it include consent to archive and share data?			
• What measures are considered in the context of the collection and processing of personal data? Is there a strategy in place to replace disclosive identifiers of an individual or entity from the data?			
Data archiving			
• Do you know if you are required to archive data as a condition of funding or to comply with institutional data policies?			
• Have you identified a suitable archive for your data? Have you contacted them to discuss data management and archiving?			
Documentation and metadata			
• Does your documentation and metadata provide sufficient contextual information for others to understand, evaluate, and replicate the data without requiring additional information from the original researcher?			
• Is the structure of the data evident, including variable names, coding, abbreviations, and specialist terms?			
File formats			
• Have you checked formats and software you use against an archive list of recommended formats for long-term preservation?			

https://www.gesis.org/fileadmin/upload/trainingcenter/images__documents/Research_data_management_questions.pdf

DMP: Il modello inglese

DMP ONLINE

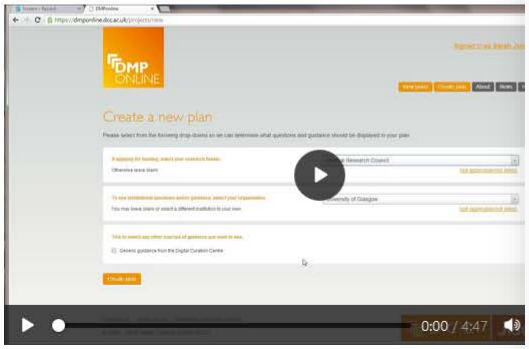
Home | About | Future plans | Help | Change language

Welcome.

DMPonline helps you to create, review, and share data management
meet institutional and funder requirements. It has been jointly develo
Digital Curation Centre (DCC) and the University of California Cura
(UC3).

Sign in

Screencast on how to use DMPonline



Veteran tapes

Project Details | Plan overview | Write Plan | Share | Download

expand all | collapse all

13/13 answered

Data Collection (2 / 2)

What data will you collect or create?

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

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

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

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

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

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

Save

Guidance | Comments (1)

DCC | DCC guidance

Guidance
Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

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

Lo stile italiano

Documento del GdL Dati della ricerca – rilasciato 15.05.2017[1]

Griglia per l'elaborazione del piano di gestione dei dati della ricerca

L'elaborato consiste di due parti, la prima più estesa contiene una griglia che illustra in modo dettagliato tutti gli aspetti relativi all'elaborazione di un piano di gestione dei dati e riflette i requisiti richiesti dalla Commissione Europea e dai principali finanziatori della ricerca, la seconda contiene alcune definizioni.

Ciascuna voce relativa al DMP è in inglese, d'altronde i DMP richiesti dai finanziatori internazionali sono prevalentemente in lingua inglese, mentre le spiegazioni sono riportate in italiano. Per ciascuna voce sono stati inclusi link a risorse informative correnti quando è stato possibile.

Le sezioni della prima parte del documento sono le seguenti:

- Sezione relativa a dettagli amministrativi del progetto
- Sezione relativa alla descrizione dei dataset
- Sezione relativa agli standard e ai metadati
- Sezione relativa alla sicurezza e alla confidenzialità dei dati
- Sezione relativa alla condivisione e all'accesso ai dati
- Sezione relativa al data management, alla documentazione e alla curation dei dati
- Sezione relativa alle responsabilità
- Sezione relativa alle politiche istituzionali sulla condivisione e sicurezza dei dati

IOSSG

IOSSG

Italian Open Science Support Group

DATASET DESCRIPTION	Descrivere i dati esistenti o che si intendono creare, indicandone provenienza, natura e ordine di grandezza. Motivare la creazione di nuovi dataset e relativo valore aggiunto.
Provenance of data (content)	*Descrivere se i dati provengono da interviste, indagini, sono estratti da archivi disciplinari, banche dati e/o da altri progetti (in tal caso, segnalare il titolo dei progetti)*
Provenance of metadata	*Idem*
Type of data	*Descrivere se sono qualitativi, quantitativi*
Nature and formats	*Descrivere natura e formato dei dati (meglio se di formato non proprietario), ad esempio:* *a) documenti testo (DOC, ODF, PDF, TXT, etc);* *b) immagini (JPG, GIF, SVG, PNG, TIFF);* *c) video/film (MPEG, AVI, WMV, MP4);* *d) registrazioni audio (MP3, WAV, AIFF, OGG, etc);* *e) dati strutturati (HTML, JSON, TEX, XML, RDF);* *f) tabelle (CSV, ODS,TSV, XLS, SAS, Stata, SPSS portable);* *g) codici sorgente (C, CSS, JavaScript, Java, etc);* *h) configuration data (INI, CONF, etc)* *i) database (MS Access, MySql, Oracle, ect)* *Max info: "Recommended File formats"*
Amount of data	*Stimare l'ordine di grandezza dell'intero dataset (GB, MB, TB, PB);*
Requirements for software	*Specificare l'eventuale utilizzo di software specifici*
Requirements for hardware	*Specificare l'eventuale utilizzo di hardware specifici*

https://sites.google.com/view/iossg/materiali-prodotti?authuser=0

nt Plan H2020

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata identifiable and locatable by means of a standard identification mechanism (e.g persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in you discipline, please outline what type of metadata will be created and how.

3. Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agree conditions).

Who will be responsible for data management in your project?

Are the resources for long term preservation discussed (costs and potential who decides and how what data will be kept and for how long)?

4. Data security

What provisions are in place for data security (including data recovery as w secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for *opting out*.

How will the data be made accessible (e.g. by deposition in a repository)?

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)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

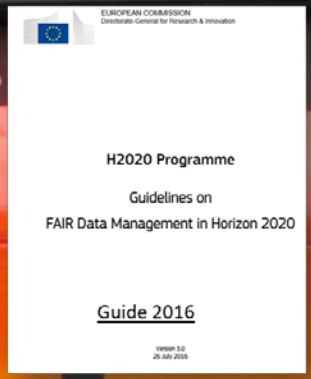
Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?


EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

Guide 2016

Version 3.0
26 July 2016

Data Management Plan ERC


European Research Council
Scientific Council
Established by the European Commission

Open Research Data and Data Management Plans
Information for ERC grantees
by the ERC Scientific Council

Feb. 23, 2018

Version 1.0
23 February 2018

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

European Research Council
Established by the European Commission

Project Acronym	Project Number

Template for the ERC Open Research Data Management Plan (DMP)[1]. Th
how you plan to make the project data Findable, Accessible, Interoperab
Each of the following five issues should be addressed with a level of deta

SUMMARY *(dataset[2] reference and name; origin and expected size of the*
and formats)

1. MAKING DATA FINDABLE *(dataset description: metadata, persistent and unique identifiers e.g., DOI)*

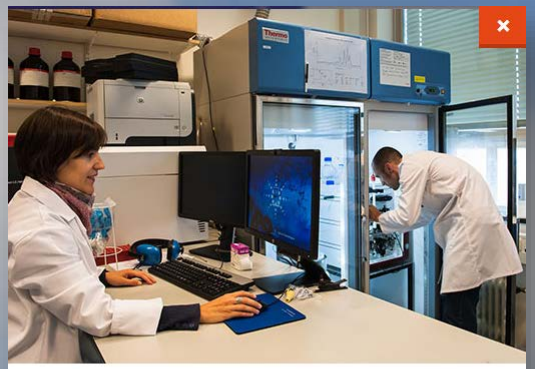
2. MAKING DATA OPENLY ACCESSIBLE *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)*

3. MAKING DATA INTEROPERABLE *(which standard or field-specific data and metadata vocabularies and methods will be used)*

4. INCREASE DATA RE-USE *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)*

5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)*

Be open



Research Data Management

Prepare your data for publication

Decide what research data you want to make openly accessible for whom and whether you have the rights to do so. Think for example about the following topics:

• Resolve issues around intellectual property, ethics and privacy if you haven't covered issues surrounding data ownership and data

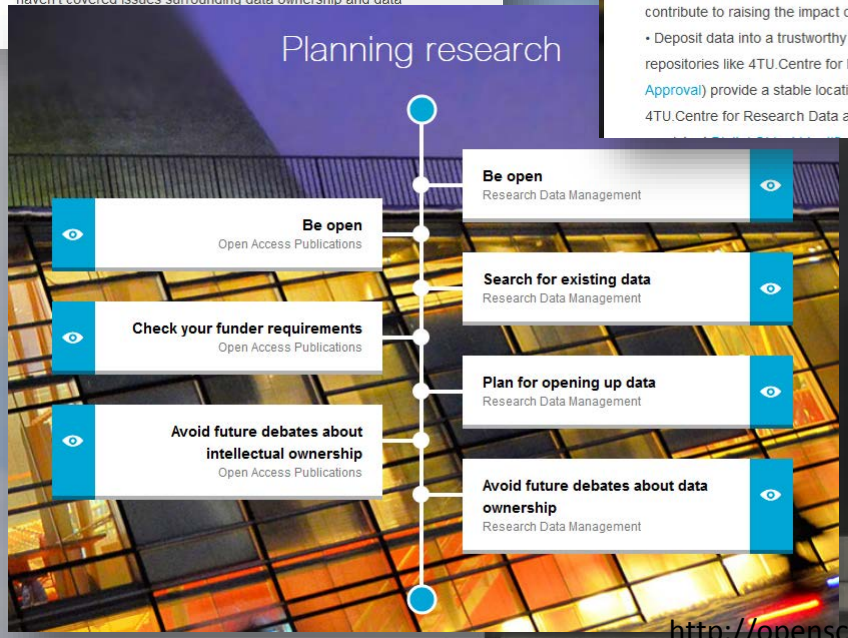


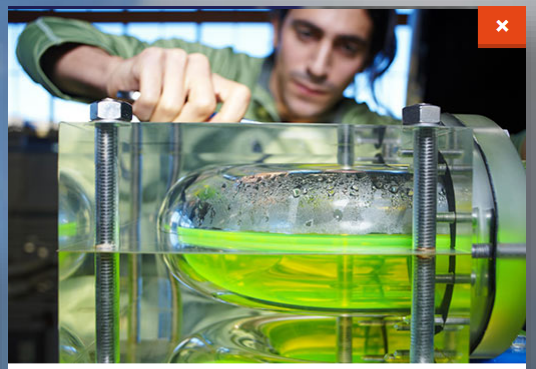
Research Data Management

Raise impact

Many of the tools available for measuring the impact of data are at an early stage of development. Here is what you can personally contribute to raising the impact of your data:

• Deposit data into a trustworthy repository. Trustworthy data repositories like 4TU.Centre for Research Data (with a Data Seal of Approval) provide a stable location for your data. Every dataset in the 4TU.Centre for Research Data archive is provided with a unique,





Research Data Management

Keep your data safe

TU Delft offers its employees several options for storing and exchanging research data safely. Solutions range from basic storage to tailored solutions like an integrated workflow for processing and sharing dynamic geodata in a tailored data lab.

The table below summarizes the current possibilities. The solution that fits your research data depends on your specific (security) needs. Contact Faculty IT management if you want advice on the software solution and protective measures (like encryption). If you want to set up an environment specifically designed to share research data – like a Dataverse or data lab – contact 4TU.Centre for Research Data for assistance.

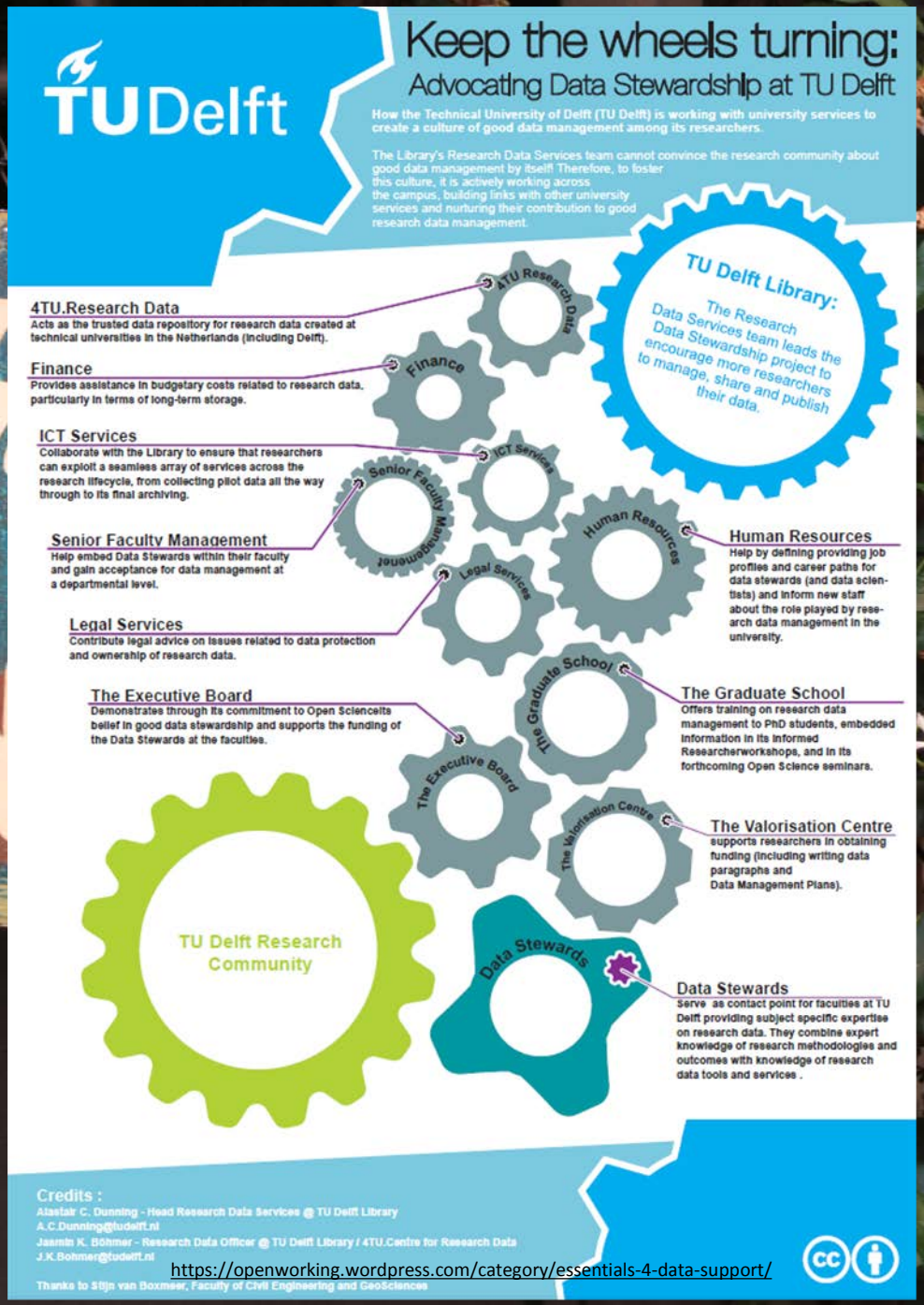
Goal	Solution
I want to easily store data and share it selected others (in and outside of TU Delft).	SURFdrive allows you to store, synchronise and exchange 100 GB of research data safely and easily with third parties. You
I want to store and backup personal data.	Personal data storage offers a H-drive with standard 8 GB and back-up.
I want to store, share and backup data with faculty colleagues.	Group data storage is a data storage facility for multiple users in the same tutorial, department or research group. The standard storage space for a group folder is 50 GB and back-up.
I want to store raw data.	With bulk data storage large quantities of temporary or measuring data can be stored default without backup: contact your Faculty IT manager for the possibilities and costs of back-ups).

http://openscienceguide.tudelft.nl/

Come dovrebbe







https://openworking.wordpress.com/category/essentials-4-data-support/

Preparing for the General Data Protection Regulation (GDPR) 12 steps to take now

1 Awareness
You should make sure that decision makers and key people in your organisation are aware that the law is changing to the GDPR. They need to appreciate the impact this is likely to have.

2 Information you hold

7 Consent
You should review how you seek, record and manage consent and whether you need to make any changes. Refresh existing consents now if they don't meet the GDPR standard.

ico. Information Commissioner's Office
ico.org.uk

https://iapp.org/resources/article/preparing-for-the-gdpr-12-steps-to-take-now/

Understanding the Implications of the GDPR on Research

UK • DATA ARCHIVE

Dr Scott Summers
University of Essex

Ensuring Compliance with the GDPR in Higher Education
1st February 2018

http://www.insidegovernment.co.uk/uploads/2018/02/Presentation-Scott-Summers-Final.pdf

https://www.insight.mrc.ac.uk/2018/04/16/gdpr-research-changes/

MRC Medical Research Council

Stories about the people, science and research of the Medical Research

Main MRC website | Insight home | About | Moderation | Contact us

GDPR: What researchers need to know

by Guest Author on 16 April 2018

The EU General Data Protection Regulation (GDPR) and new Data Protection Act come into force on 25 May. Both apply in the UK and will influence research involving personal data. So what's changing and how should you, as a researcher, prepare? ***Sarah Dickson****, Head of the MRC Regulatory Support Centre, is here to help.*

Regulatory Support Centre

What is GDPR?

The EU General Data Protection Regulation (GDPR), along with the new UK Data Protection Act, will govern the processing (holding or using) of personal data in the UK.

Although the new regulations haven't been designed specifically for research, we'll need to make some changes to research practice. The Information Commissioner's Office (ICO) is the UK regulator. The Health Research Authority (HRA), in collaboration, is providing official guidance for people working in health and social care research. We're working with both organisations.

What counts as 'personal data'?

This is data about living people from which they can be identified. As well as data containing obvious

Strumenti

IDCC18 | *Research Paper* Jan. 22, 2018

Data Stewardship – addressing disciplinary data management needs

Marta Teperek
Research Data Services, TU Delft Library
TU Delft, Delft, Netherlands

Maria J. Cruz
4TU.Centre for Research Data,
TU Delft Library, TU Delft, Delft,
Netherlands

- **Analyse data management needs** - through undertaking a mixture of semi-structured qualitative interviews, as was as quantitative surveys;¶
- **Provide advice and consultancy** - meet with researchers, discuss their data management practices, make suggestions for possible improvements; become the trusted person for any questions about data management;¶
- **Liaise with key faculty stakeholders** - ensure that the various faculty service providers (such as contracts managers or faculty information coordinators) are aware of good data stewardship and that requirements of good data stewardship are aligned with their workflows (for example, budgeting for data management in grant applications);¶
- **Train and inspire** - advocate for good data management, deliver information sessions, analyse training needs, develop and deliver workshops to ensure that researche
- **Help com** drafting t and advis
- **Develop** consultati the different faculty-level stakeholders; drive policy implementation, evaluation and revision;¶
- **Prepare the faculty for the future** - keep the faculty up to date with new developments and policy changes related to data stewardship; keep abreast of new developments in the faculty's research area to ensure that researchers have the right skills to manage their data, despite of evolving research methodologies;¶
- **Liaise with the Data Stewardship Coordinator and other Stewards** - liaise with other members of the Data Stewardship programme to exchange practice and to discuss relevant issues;¶
- **Deliver regular reports** - regularly evaluate, monitor and report on data management practices within the faculty.¶

In addition, we believed that disciplinary expertise, reflected in a PhD degree (or equivalent experience) in the area of faculty's research, was necessary for the Stewards to provide relevant and tailored advice to their communities.¶

http://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

Strumenti

Jan. 2018

Science Europe Guidance Document

Presenting a Framework for Discipline-specific Research Data Management

January 2018

Privacy

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

Intellectual Property Rights

- **Copyright Acts** regulate the rights of the creator of a work. One distinguishes betwe exploitation rights and personal intellectual rights ('moral rights').
- The **Database Rights Act** recognises the investments made in creating and/or compiling database. It is based on European Directive 96/9/EC.[11]

Introduction 5

A Framework for Research Data Management 6
Domain Data Protocols 6
The Framework 7
Proof of Concept 7

A Framework for Domain Data Protocols 8
Components of the Framework 9
1. Formal Minimum Conditions 10
2. Applicable Laws and Regulations 11
3. FAIR Principles 13
4. Applicable Standards 14
5. Templates and Examples 15
6. Support Resources 15

Proofs of Concept from different Communities 18
Humanities (general): DARIAH 20
Humanities – Archaeology: PARTHENOS/ARIADNE 22
Linguistics – Language Data: CLARIN 24
Social Sciences – Survey data: CESSDA 26
Social/Behavioural Sciences – Psychology 28
Social Sciences – Family of Studies on Longitudinal Ageing 30
Life Sciences – Bio-informatics: ELIXIR and Force11/RDA FAIRSharing 32
Plant Sciences: ERA-CAPS 35
Climate Research: ICOS 38

Set minimo che deve essere presente in ogni progetto

Strumenti

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

RIO Home Articles About About Pensoft Books E-Books Blog

Project Report | Research Ideas and Outcomes 4: e26439 https://doi.org/10.3897/rio.4.e26439 | May 9, 2018 | Reviewable v1

Support Your Data: A Research Data Management Guide for Researchers

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

Abstract

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

Suppl. material 5: Draft Guide - Preparing doi

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

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

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

Suppl. material 6: Draft Guide - Analyzing doi

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

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

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

Suppl. material 7: Draft Guide - Sharing doi

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Sharing and publishing your data" row of the

http://www.scienceeu

23 cose



RESEARCH DATA ALLIANCE UNITED STATES

23 Things: Libraries for Research Data

An overview of practical, free, online resources and tools that you can begin using today to incorporate research data into your practice of librarianship.

Research Data Sharing Without Barriers

Learning Resources

Librarians are learning how to apply the principles of library science to solve problems and to provide new services related to research data.

1. A **"top ten" list of recommendations** for libraries to get started with research data from LIBER, http://bit.ly/1qUvKG3
2. Relevant concepts are presented and mapped in the **e-Science Thesaurus**, http://bit.ly/1LEo4h8
3. Understanding the life of research data with the **DCC Curation Lifecycle Model**, http://bit.ly/1MoGGGv
4. **MANTRA** online training modules for librarians, http://bit.ly/1RRvVju
5. Read the most current literature

Learning Resources
Data Reference and Outreach
Data Management Plans
Data Literacy
Citing Data
Data Licensing and Privacy
Digital Preservation
Data Repositories
and a Community of Practice

..to help librarians engage in research data!

Data Reference & Outreach

Librarians are answering questions about data from patrons and conducting outreach to assess the data needs of their researchers and students.

7. Begin a conversation with a researcher about data by **Conducting a Data Interview**, http://bit.ly/1ERuuyc
10. Questions about data answered by experts on the **DataQ** forum, http://bit.ly/1MoH4Vg

Data Management Plans

Librarians are becoming familiar with funder requirements and consulting with researchers to help them write and implement effective data management plans.

11. One example is the **DMPTool** that lists funder requirements in the United States and builds a plan by asking the researcher to answer a series of questions. Other countries such as the U.K. and Canada have similar tools, http://bit.ly/1LuNZMH

Data Literacy

Librarians are including data in their information literacy instruction, to recognize when they have a need for data and have the ability to locate, evaluate, and use data.

https://www.rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_RDA.pdf

Le politiche sui dati

SPARC Europe

Setting the Default to Open

Feb. 2018

Who We Are | What We Do | NEWS | Membe

News

Home › Uncategorised › Update to Analysis of Open Data Policies finds new activity around OA and OD policies in multiple countries

Update to Analysis of Open Data Policies finds new activity around OA and OD policies in multiple countries

28th February 2018 | News, Uncategorised

Open Data policy development in Europe is constantly evolving. In an effort to stay abreast of these changes on behalf of the community, SPARC Europe, together with the Digital Curation Centre (DCC), recently released an update o our analysis of Open Data policies in Europe.

he report was first released in May 2017. This latest version has been updated to eflect Open Data policy changes that have surfaced between June and November 017, and encompasses the 28 EU member states as well as several non-EU embers of the European Research Area, such as Iceland, Norway and witzerland.

ew activity around national approaches to open data and open science has been entified in several countries, notably the Czech Republic, Hungary, Serbia and weden.

mong other key findings:

- 11 of 28 EU mem place
- The majority of th research funders; recipients.
- Policies generally activity during rec

TABLE OF CONTENTS

1. Introduction
2. Executive summary
3. Overview

Overview table: Status of national policies

4. State by state

4.1 Summary

Overview table: countries with national policies in place

4.2 Member states with existing national policies (10/28)

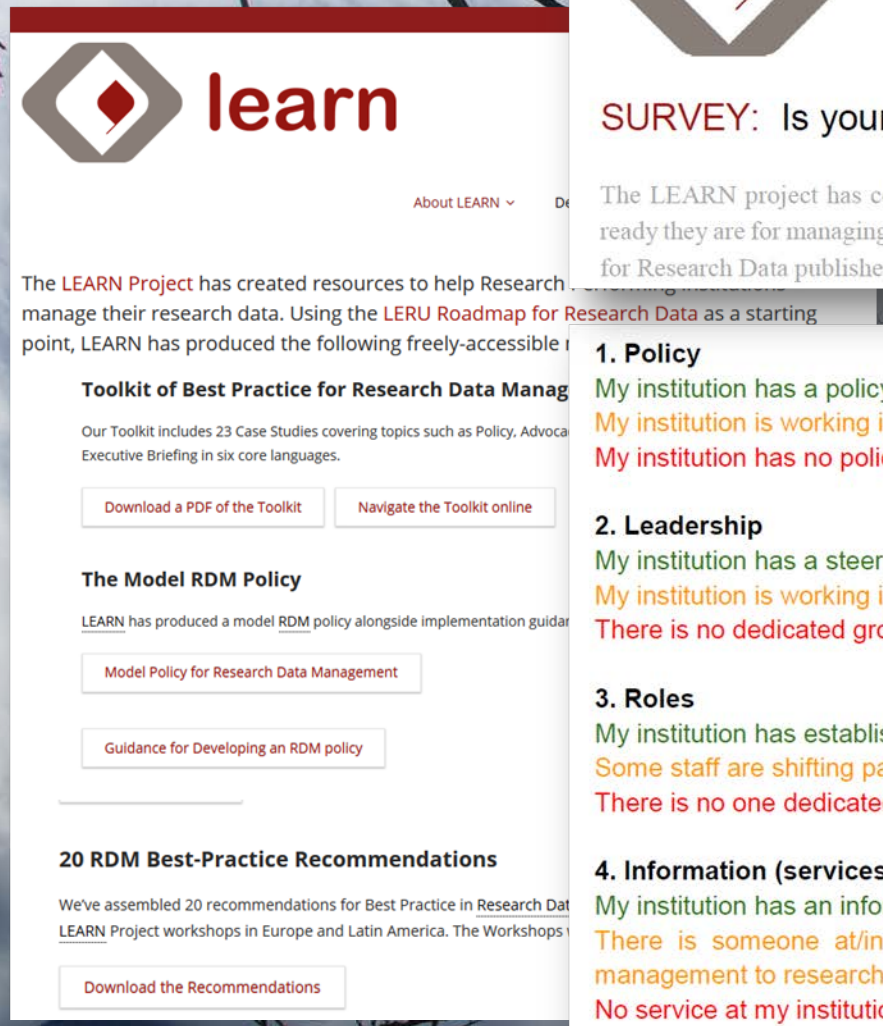
BELGIUM (BE)
CYPRUS (CY)
DENMARK (DK)
FINLAND (FI)
FRANCE (FR)
GERMANY (DE)
LITHUANIA (LT)
NETHERLANDS (NL)
PORTUGAL (PT)
UNITED KINGDOM (UK)

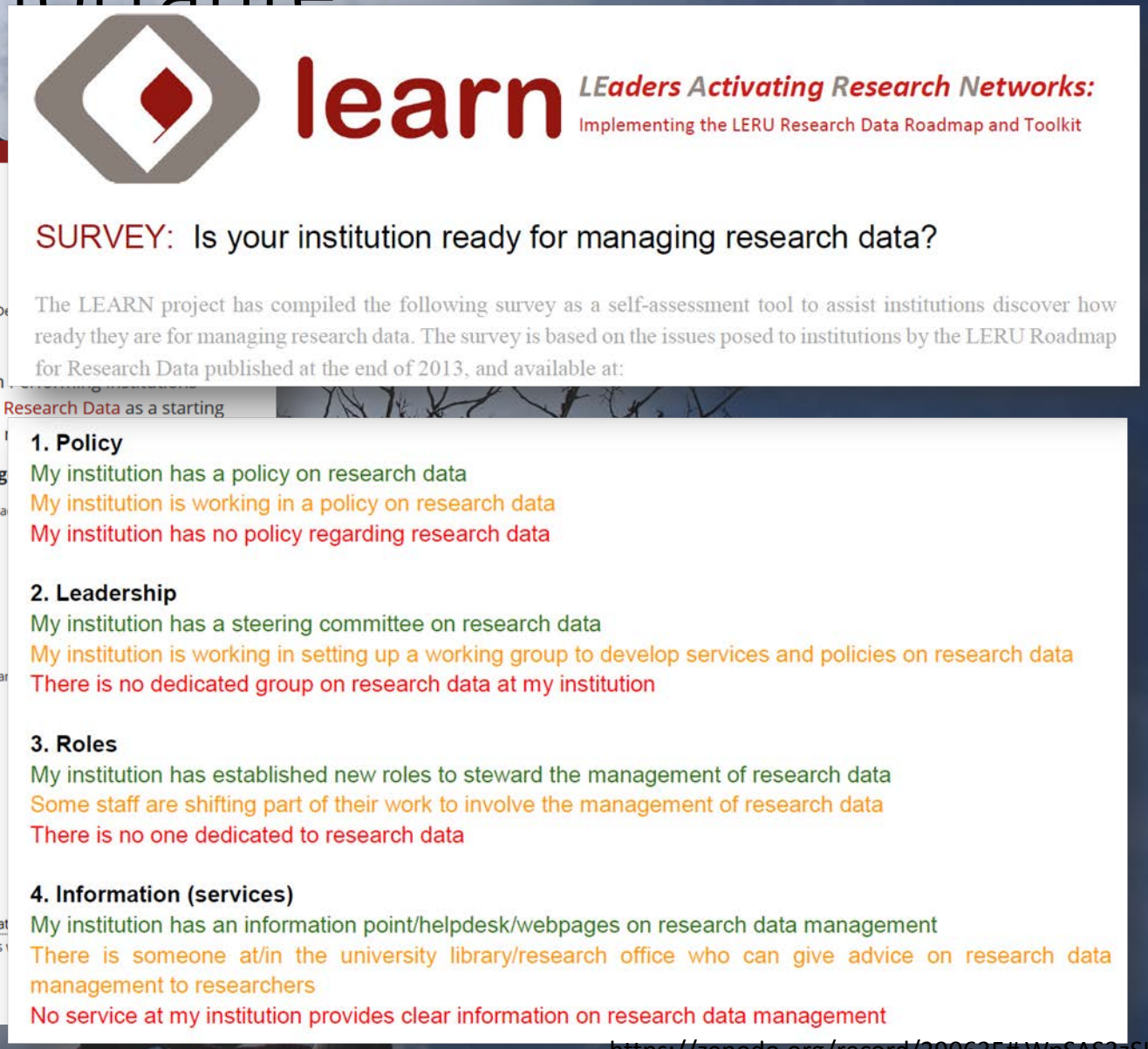
4.3 Member states with no national policy but which are active in this space (12/28)

AUSTRIA (AT)
BULGARIA (BG)
CROATIA (HR)
CZECH REPUBLIC (CZ)
ESTONIA (EE)
HUNGARY (HU)
IRELAND (IE)
ITALY (IT)
POLAND (PL)
SLOVENIA (SI)
SPAIN (ES)
SWEDEN (SE)

4.4 Member states with no policy or activity (6/28)

Le politiche: un panorama sconfortante





https://zenodo.org/record/290635#.WpSAS3zSKUl

Il modello internazionale



learn

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

1. PREAMBLE

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

2. JURISDICTION

This policy for the management of research data applies to all researchers active at the [*name of research institution*]. The policy was approved by the [*dean/commission/authority*] on [*date*]. In cases when research is funded by a third party, any agreements made with that party concerning intellectual property rights, access rights and the storage of research data take precedence over this policy.

3. INTELLECTUAL PROPERTY RIGHTS

Intellectual property rights (IPR) are defined in the work contract between a researcher and his or her

http://discovery.ucl.ac.uk/1546606/1/25_Learn_Model%20Policy_133-136.pdf

Lo stile italiano

IOSSG

Italian Open Science Support Group

Documento del GdL Dati della ricerca* – rilasciato 27.03.217

Policy sulla gestione dei dati della ricerca

1) **Premessa**

L'Università ……. riconosce l' importanza fondamentale dei dati prodotti durante l'attività di ricerca come un valido risultato scientifico. Pertanto riconosce l'importanza della loro gestione per il mantenimento dei valori di qualità ed integrità della ricerca scientifica e si impegna ad applicare i più elevati standard per la loro raccolta, archiviazione e conservazione.

L'Università …… riconosce che dati della ricerca affidabili e facilmente recuperabili sono alla base di ogni progetto di ricerca e sono altresì necessari per la verifica di attendibilità e correttezza dei processi e dei risultati del progetto e per la sua replicabilità.

L'Università …….. riconosce che i dati della ricerca, anche una volta che un progetto è stato terminato, costituiscono un patrimonio dell'istituzione universitaria, nonché una risorsa, anche a lungo termine, per la ricerca scientifica, la didattica universitaria e il progresso della società.

Ai fini della presente policy si considera la definizione di "dati della ricerca" e di "afferenti all'università ……." così come da allegato 1

2) **Ambiti di applicazione**

La presente policy si applica a tutti gli afferenti all'Università ……..(quali docenti, ricercatori, dottorandi, assegnisti, borsisti, personale tecnico amministrativo coinvolto nelle attività di ricerca). Nel caso la ricerca sia stata finanziata da parti terze ed esistano accordi specifici relativi al controllo dei dati, al loro accesso e conservazione, gli accordi prevalgono sulla presente policy.

Nel rispetto della vigente normativa in materia di protezione dei dati personali e di proprietà intellettuale, nonché delle disposizioni contenute nello Statuto e nei regolamenti di Ateneo e fatti salvi gli specifici accordi per il finanziamento della ricerca stipulati con terze parti, i dati della ricerca sono archiviati, e resi [illegible] tracciabile e

https://sites.google.com/view/iossg/materiali-prodotti?authuser=0

Politiche: cosa offre l'Italia

UNIVERSITÀ DEGLI STUDI DI MILANO

Open Research Data

L'Università supporta i principi e le azioni favorevoli alla gestione aperta dei dati della ricerca, presupposto indispensabile della riproducibilità e dell'accesso aperto ai risultati della scienza.

La piena adesione dell'Ateneo all'Open Research Data è stata formalizzata nella **Policy sulla gestione dei dati**, con cui l'Università degli Studi di Milano, riconoscendo la rilevanza della gestione dei dati per il mantenimento dei valori di qualità e integrità della ricerca scientifica, si impegna ad applicare i più elevati standard per la loro raccolta, archiviazione e conservazione.

Oltre all'indicazione degli ambiti di applicazione e della disciplina del trattamento dei dati della ricerca, la Policy stabilisce precise responsabilità, diritti e doveri dell'Ateneo e di chi fa ricerca in Università, proponendo un modello di Data Management Plan (DMP) che risponda alle richieste a livello europeo di uno strumento per una corretta gestione dei dati durante l'intero processo della ricerca: dalla raccolta, documentazione e archiviazione all'accesso, uso e conservazione (o distruzione) dei dati.

DOCUMENTI SCARICABILI

Policy sulla gestione dei dati

Policy sulla gestione dei dati 620.93kb

Research Data Management Policy 398.89kb

Publishing research without data is simply advertising, not science

G.Steel, 2013

Conclusion

"Give a scientist data/tools, and you feed the science world for a day. Teach them openness, and you feed the science world for a lifetime" – Jonathan Eisen, 2011