OPEN SCIENCE – OPEN ACCESS
4 - DATI FAIR E EOSC

Elena Giglia
Università di Torino

elena.giglia@unito.it
@egiglia

Impareremo in questo modulo

1. Cosa succede in Europa
2. Cosa significa FAIR nella ricerca quotidiana
3. come scrivere un DMP

MESSAGGI CHIAVE

- FAIR è il futuro (se no si resta tagliati fuori)
- È più facile di quanto sembri, basta provarci
- I dati FAIR sono una ENORME opportunità di lavoro

«make science fit for the 21th century»

Transition to open science is a multidimensional and multistage process. There is value and risk of being a first mover, but there is higher risk of being a follower. The European Commission has taken

RISCHI A ESSERE I PRIMI, RISCHI MAGGIORI A ESSERE GLI ULTIMI

Open science will make science more efficient, reliable, and responsive to societal challenges. The European Commission

E ARTICLE Provisionally accepted The full-text will be published soon Notify me
Front. Big Data | doi: 10.3389/fdata.2019.00043

Open science, open data and open scholarship: European policies to make science fit for the 21st century

Jean-Claude Burgelman[1*], Corina Pascu[1*], Katarzyna Szkuta[1], Rene Von Schomberg[1], Athanasios Karalopoulos[1], Konstantinos Repanas[1] and Michel Schouppe[1]

Open science (or in fact, open scholarship) has shifted the prime focus of researchers away from publishing toward knowledge sharing.

and access will be maximized. In Horizon Europe, research data will be open by default while taking into account the need to balance openness and protection of scientific information, commercialization and Intellectual Property Rights, privacy concerns and security, following the principle "as open as possible, as closed as necessary." Data management plans (DMP) will become mandatory, even if not making research data open. The requirement for responsible data management will be separated from the requirement for providing open access to research data. Emphasis will be placed on supporting as much as possible the proliferation of data that are findable, accessible, interoperable, and re-usable (FAIR). Finally, the use of trusted or certified repositories and infrastructures like the European Open Science Cloud (EOSC) will be required for research data in some Horizon Europe work programs.

SINTESI DELEL POLITICHE DI APERTURA DEGLI ULTIMI 15 ANNI

Changing the reward and incentive system for researchers is a key open science challenge and a broader issue for which primarily the responsibility lies in the scientific community (universities and funders). This includes making open science practices rewardable and fundable as well as the employment of specific indicators for researchers' engagement with open science. A change of the reward and incentive system can only be stakeholders-driven, and it has to be bottom-up. This change also includes changing mind-sets of researchers to open up and share data and "seduction" to make open science easy, useful, and affordable[3].

The European Open Science agenda contain the ambition to make FAIR data sharing the default for scientific research by 2020. To

… as Open as pos

Carlos Moedas ✓
@Moedas

Segui

2/4 "Open as possible, as closed as necessary" is the new principle for all #data from publicly
ccess

Iryna Kuchma @irynakuchma · 18 nov 2015
#Openscience is about making sure that science serves innovation & growth – Günther Oettinger & Carlos Moedas

HORIZON 2020

TESTI E DATI OPEN BY DEFAULT (GUIDA)

HORIZON2020: AS OPEN AS POSSIBLE
HORIZON EUROPE: ANCORA MAGGIORE APERTURA + UNITED NATIONS SUSTAINABLE DEVELOPMENT GOALS

Open Science across the programme

Open Science — Better dissemination and exploitation of R&I results and support to active engagement of society

Mandatory Open Access to publications: beneficiaries shall ensure that they or the authors retain sufficient intellectual property rights to comply with open access requirements

Open Access to research data ensured: in line with the principle "as open as possible, as closed as necessary"; Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Re-usable) and Open Research Data

- Support to researcher skills and reward systems for open science
- Use of European Open Science Cloud

May 2019 | Version 25

European Commission

Horizon Europe

THE NEXT EU RESEARCH & INNOVATION INVESTMENT PROGRAMME (2021 – 2027)

2019

Our vision

A sustainable, fair and **prosperous** future for **people** and **planet** based on European values.

- Tackling **climate change** (35 % budgetary target)
- Helping to achieve **Sustainable Development Goals**
- Boosting the Union's **competitiveness and growth**

May 2019 | Version 25

European Commission

...Open Access by default in 2020...

12. AGREES to further promote the mainstreaming of open access to scientific publications by continuing to support a transition to immediate open access as the default by 2020, using the various models possible and in a cost-effective way, without embargoes or with as short as possible embargoes, and without financial and legal barriers, taking into account the diversity in research systems and disciplines, and that open access to scientific publications should be the principle that no researcher should be prevented from ission, Member States and relevant stakeholders, including o catalyse this transition; and STRESSES the importance of reements.¶

Council of the European Union

Brussels, 27 May 2016
(OR. en)

9526/16

RECH 208
TELECOM 100

OUTCOME OF PROCEEDINGS

From:	General Secretariat of the Council
To:	Delegations
No. prev. doc.:	8791/16 RECH 133 TELECOM 74
Subject:	The transition towards an Open Science system - Council conclusions (adopted on 27/05/2016)

European Council
Council of the European Union

The European Council | The Council of the EU | Topics | Policies | Meetings | Documents & Publications

Home > Meetings > Competitiveness Council, 26-27/05/2016

Competitiveness Council, 26-27/05/2016

> Indicative programme - Competitiveness Council of 26-27/05/2016
> Background brief

Research and Innovation

Following a debate on **open science**, the Council adopted conclusions on the transition towards an open science system.

"Open Science is a topic which is very dear to our hearts. During the Netherlands presidency, we have aimed at bringing Europe to the forefront of global change and at leading the transition to a new way of doing research and science based on openness, big data and cloud computing."

Sander Dekker, State Secretary of Education, Culture and Science of the Netherlands

It also adopted conclusions on the lessons learnt from the **7th research framework programme and the future outlook** and on the creation of a friendly regulatory **environment for research and innovation**.

Chairing the Council, Sander Dekker, State Secretary of Education, Culture and Science of the Netherlands, made the following statement: "Open Science is a topic which is very dear to our hearts. During the Netherlands presidency, we have aimed at bringing Europe to the forefront of global change and at leading the transition to a new way of doing research and science based on openness, big data and cloud computing. Open Science breaks down the barriers around universities and ensures that society benefits as much as possible from all scientific insights. In that way we maximize the input of researchers, universities and knowledge institutions".

Highlights of the Competitiveness Council, taking place on 27 May in Brussels.

Today, building on work done during recent months, particularly at the April conference when we approved the "Amsterdam Call for Action on Open Science", I can say that we have made a major step forward".

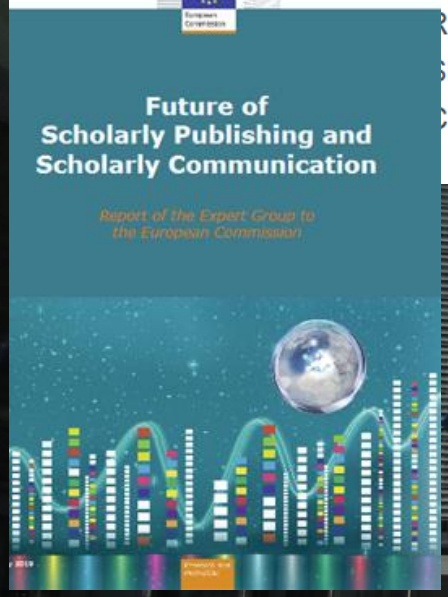
...sulla ...cience

- Rewards and Incentives
- Research Indicators and Next-Generation Metrics
- Future of Scholarly Communication
- European Open Science Cloud
- FAIR Data
- Research Integrity
- Skills and Education
- Citizen Science

Integrated advice of the Open Science Policy Platform on 8 prioritised Open Science ambitions

2018

2019



Amsterdam Call for Action on Open Science

2016

2017

Apr. 25, 2018

Report on OS and careers, July 2017

…sulla via della Open science

Open Science for its own sake has never been the goal. While a focus on Open Science as a mechanism must be emphasised in any transition, Open Science must ultimately be embedded as part of a larger more systemic effort to foster all practices and processes that enable the creation, contribution, discovery and reuse of research knowledge more reliably, effectively and equitably. Research cannot be 'excellent' without such attributes at its core.

to help co-create, develop and maintain a 'Research System based on shared knowledge' by 2030. As a start, we commit to working together to implement a system with the five attributes outlined below.

1. An academic career structure that fosters outputs, practices and behaviours to maximise contributions to a shared research knowledge system. To this

2. A research system that is reliable, transparent and trustworthy. To achieve this, Member States should

3. A research system that enables innovation. Five key elements were identified as necessary to facilitate such a research system:

4. A research culture that facilitates diversity and equity of opportunity. To enable such a culture to

5. A research system that is built on evidence- based policy and practice. To enable this, we recommend

Raccomandazione 790/2018

Open access to scientific publications

1. Member States should set and implement clear policies (as detailed in national action plans) for the dissemination of and open access to scientific publications resulting from publicly funded research. Those policies and action plans should provide for:

- researchers, when entering into contractual agreements with scientific publishers, retain the necessary intellectual property rights, inter alia, to comply with the open access policy requirements. This concerns in particular self-archiving and re-use (notably through text and data mining);

2. Member States should ensure that research funding institutions responsible for managing public research funding and academic institutions receiving public funding implement the policies and national action plans referred to in point 1 at national level in a coordinated way by:

- setting institutional policies for the dissemination of and open access to scientific publications, and establishing implementation plans;

- including requirements for open access as a condition to give out grant agreements or to provide other financial support for research, together with mechanisms for monitoring compliance with these requirements and follow up actions to correct cases of non-compliance;

- making the necessary funding available for dissemination (including open access and re-use) in a transparent and non-discriminatory manner allowing for different channels, including digital infrastructures where appropriate, as well as new and experimental methods of scholarly communication;

- providing guidance to researchers on how to comply with open access policies, and supporting them to do so, especially regarding the management of their intellectual property rights to ensure open access to their publications;

- conducting joint negotiations with publishers to obtain transparent and the best possible terms for access to publications, including use and re-use;

Open Access policy nazionale

Mantenere i diritti

Open Access policy per Ateneo

- Legate alla valutazione
- Disseminazione in canali diversi
- Formazione e supporto

Raccomandazione 790/2018


EUROPEAN COMMISSION

790/2018

Brussels, 25.4.2018
C(2018) 2375 final

COMMISSION RECOMMENDATION

of 25.4.2018

on access to and preservation of scientific information

(12) The move towards open access is a worldwide endeavour. Member States have been part of this endeavour and should be supported in enhancing an open, collaborative research environment based on reciprocity at a global level. Open science is a key feature of Member States' policies for responsible research and for open innovation. As new digital technologies become available, research and funding policies should adapt to this new environment.

Management of research data, including open access

3. Member States should set and implement clear policies (as detailed in national action plans) for the management of research data resulting from publicly funded research, including open access. Those policies and action plans should provide for:

- research data that results from publicly funded research becomes and stays findable, accessible, interoperable and re-usable ("FAIR principles") within a secure and trusted environment, through digital infrastructures (including those federated within the European Open Science Cloud (EOSC), where relevant), unless this is not possible or is incompatible with the further exploitation of the research results ("as open as possible, as closed as necessary"). This could be for reasons, in particular, of privacy, trade secrets, national security, legitimate commercial interests and to intellectual property rights of third parties. Any data, know-how and/or information whatever its form or nature which is held by private parties in a joint public/private partnership prior to the research action should not be affected by these policies or national action plans;

4. Member States should ensure that research funding institutions responsible for managing public research funding and academic institutions receiving public funding implement the policies and national action plans referred to in point 3 at national level in a coordinated way by:

- providing guidance to researchers on how to comply with research data management policies, and supporting them to do so, especially regarding the development of sound data management planning skills and digital infrastructures that support access to and preservation of research data;

Data policy
(a livello nazionale)

Dati FAIR per EOSC

Data policy
(livello istituzionale)

Competenze
Infrastrutture

Bruxelles non sta mai f

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

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

del 20 giugno 2019

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

I DATI DELLA RICERCA
ORA RIENTRANO
NELLA DIRETTIVA DATI SETTORE PUBBLICO

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

EUROPEAN COMMISSION

Brussels, 19.2.2020
COM(2020) 66 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

A European strategy for data

In order to release Europe's potential we have to find our European way, balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards.

I DATI SONO POSSONO ESSERE RIUSATI
SENZA PERDERE VALORE.

3. The vision

The Commission's vision stems from European values and fundamental rights and the conviction that the human being is and should remain at the centre. The Commission is convinced that businesses and the public sector in the EU can be empowered through the use of data to make better decisions. It is all the more compelling to seize the opportunity presented by data for social and economic good, as data – unlike most economic resources – can be replicated at close to zero cost and its use by one person or organisation does not prevent the simultaneous use by another person or organisation. That potential should be put to work to address the needs of individuals and thus create value for the economy and society. To release this potential, there is a need to ensure better access to data and its responsible usage.

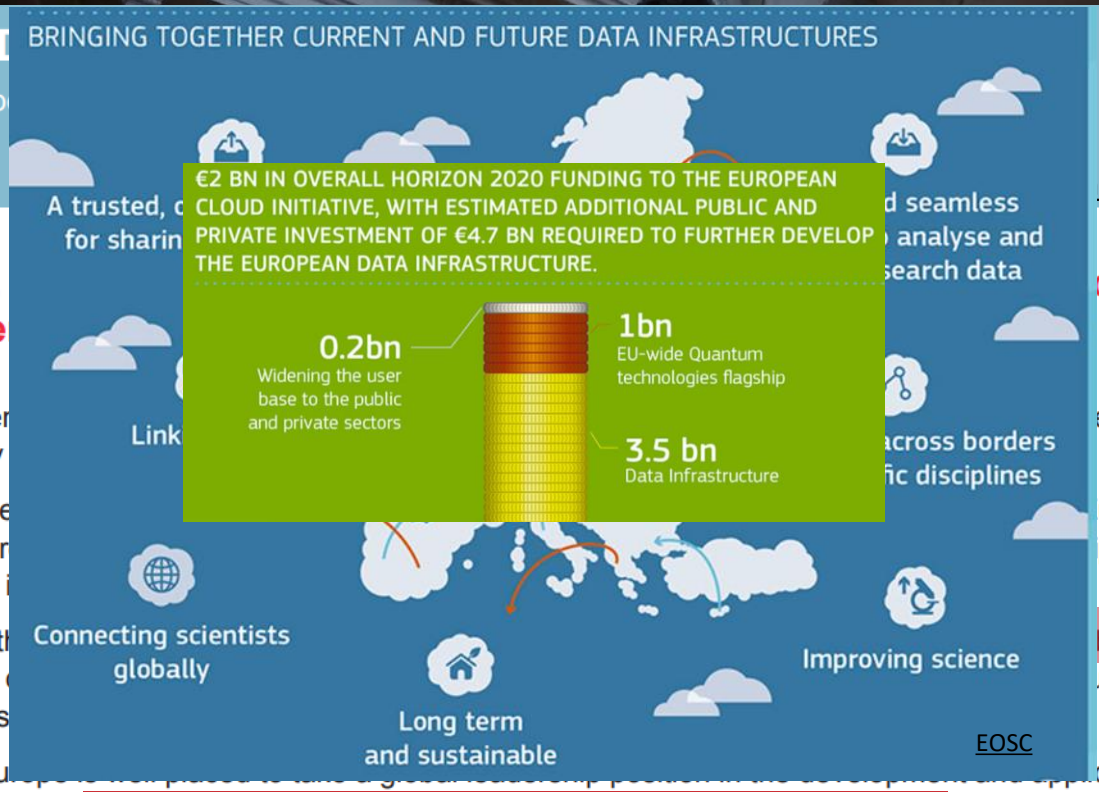
The EU should create an attractive policy environment so that, by 2030, the EU's share of the data economy – data stored, processed and put to valuable use in Europe - at least corresponds to its economic weight, not by *fiat* but by choice. The aim is to create a single European data space – a genuine single market for data, open to data from across the world – where personal as well as non-personal data, including sensitive business data, are secure and businesses also

EOSC! 23 novembre 2018

The Vienna D

Vienna, 23 Novemb

na, Nov.23, 2018



We, Ministers, European Ope

1. Recall the challe Brussels on 10 July

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

3. Recognise that t iterative and based consensus among s

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

5. Recall that the Council

SEAMLESS ACCESS TO OPEN BY DEFAULT
FAIR DATA

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

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

[EOSC – cosa?]

UN AMBIENTE CHE SUPPORTA LA OPEN SCIENCE E NON UN «OPEN CLOUD» PER LA SCIENZA

THE EUROPEAN OPEN SCIENCE CLOUD? SOME NUANCES AND DEFINITIONS

Imagine a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes. Imagine that this all operates under well-defined and trusted conditions, supported by a sustainable and just value for money model. This is the environment that must be fostered in Europe and beyond to ensure that European research and innovation contributes in full to knowledge creation, meet global challenges and fuel economic prosperity in Europe. This we believe encapsulates the concept of the European Open Science Cloud (EOSC), and indeed such a federated European endeavour might be expressed as the European contribution to an Internet of FAIR Data and services.

The European Open Science Cloud is a supporting environment for Open Science and not an 'open Cloud' for science.

The EOSC aims to accelerate the transition to more effective Open Science and Open Innovation in a Digital Single Market by removing the technical, legislative and human barriers to the re-use of research data and tools, and by supporting access to services, systems and the flow of data across disciplinary, social and geographical borders. The term European Open Science Cloud requires some reflection to dispel incorrect associations and clarify boundaries; in fact the term 'cloud' is a metaphor to help convey the idea of seamlessness and a commons.

EOSC Declaration

- necessario cambiamento culturale e formazione
- NESSUNA DISCIPLINA, NESSUNA ISTITUZIONE E NESSUN PAESE DEVE ESSERE LASCIATO INDIETRO

Data culture and FAIR data

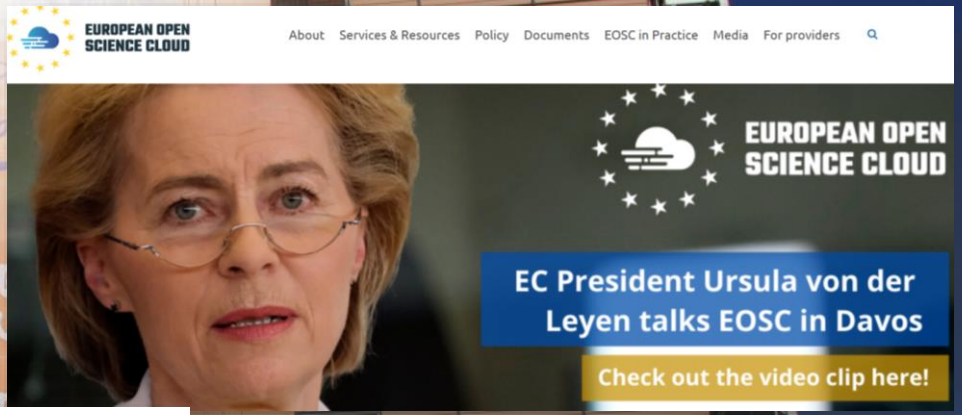
- [**Data culture**] European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout and after the period conducting the research. Only a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities: no disciplines, institutions or countries must be left behind.

- [**Open access by-default**] All researchers in Europe must enjoy access to an open-by-default, efficient and cross-disciplinary research data environment supported by FAIR data principles. Open access must be the default setting for all results of publicly funded research in Europe, allowing for proportionate limitations only in duly justified cases of personal data protection, confidentiality, IPR concerns, national security or similar (e.g. 'as open as possible and as closed as necessary').

- [**Skills**] The necessary skills and education in research data management, data stewardship and data science should be provided throughout the EU as part of higher education, the training system and on-the-job best practice in the industry. University associations, research organisations, research libraries and other educational brokers play an important role but they need substantial support from the European Commission and the Member States.

Oct. 2017

EOSC



sustainable and a data economy. Data is a renewable resource as much as sun and wind. Every 18 months we double the amount of data we produce. Industrial and commercial data, 85% of which is never used.

This is not sustainable. Within those data, there are hidden treasures and untapped opportunities for business and society. Europe is going to co-create a framework to allow the use of these data. It should consist of a trusted pool of non-personal data that governments, businesses and other stakeholders can contribute to. This pool will be a resource f open innovation, and bring new solutions to the market. And our scientists are already beginning to do this.

We are creating a European Open Science Cloud now. It is a trusted space for researchers to store their data and to access data from researchers from all other disciplines. We will create a pool of interlinked information, a 'web of research data'. Every researcher will be able to better use not only their own data, but also those of others. They will thus come to new insights, new findings and new solutions.

85% DEI DATI PRODOTTI NON VIENE USATO. INSOSTENIBILE

This is what we call the European Open Science Cloud and we are the first in the world to do that. It is being developed in Europe for Europe and for European researchers. The idea is that once we have the rules of the game ready, then we will open this up to the broader public sector and to business as well. So that companies can come in, store the data and use the data. And the idea is that it will also open up to international players.

A QUESTO SERVE EOSC.
LA STIAMO CREANDO **ADESSO**

EOSC

EC proposal for FAIR building blocks

Slide courtesy of Jean Claude Burgelman

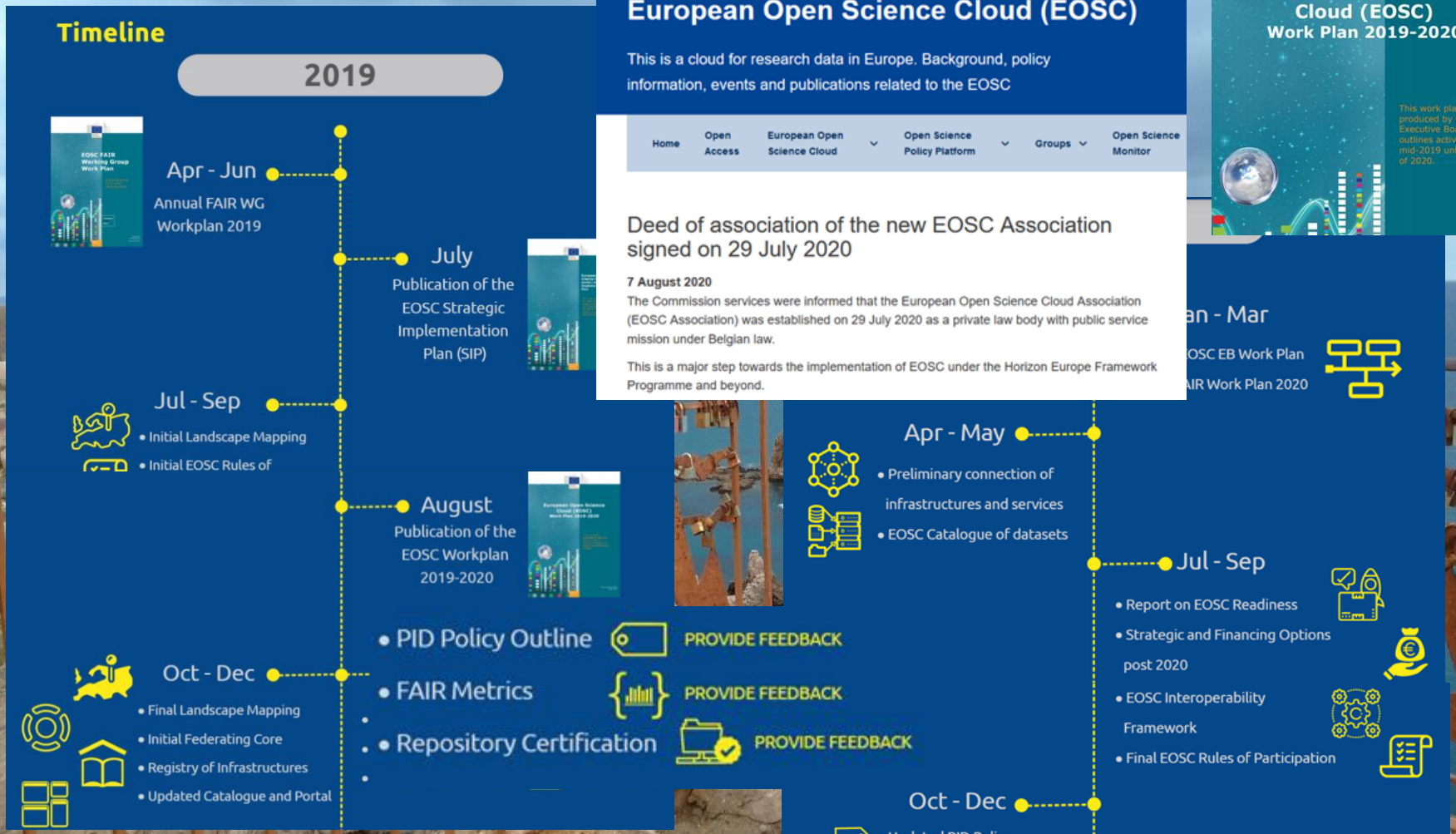
Europe's decision to develop the European Open Science Cloud reflects the willingness to embrace change, but also to empower 1.7 million European researchers and 70 million professionals in science and technology. The ultimate goal is to achieve a fundamental transformation of the whole research lifecycle and to make it more credible with increased integrity, more efficient, collaborative and more responsive to societal challenges.

I am convinced that the Cloud will allow a new generation of scholars to find, combine and analyse data and discoveries in a way that supersedes anything we have ever seen before. It will accelerate the transition to Open Science and Open Innovation and bring science and research closer to societal needs.

Carlos Moedas,

Commissioner for Research, Science and Innovation.

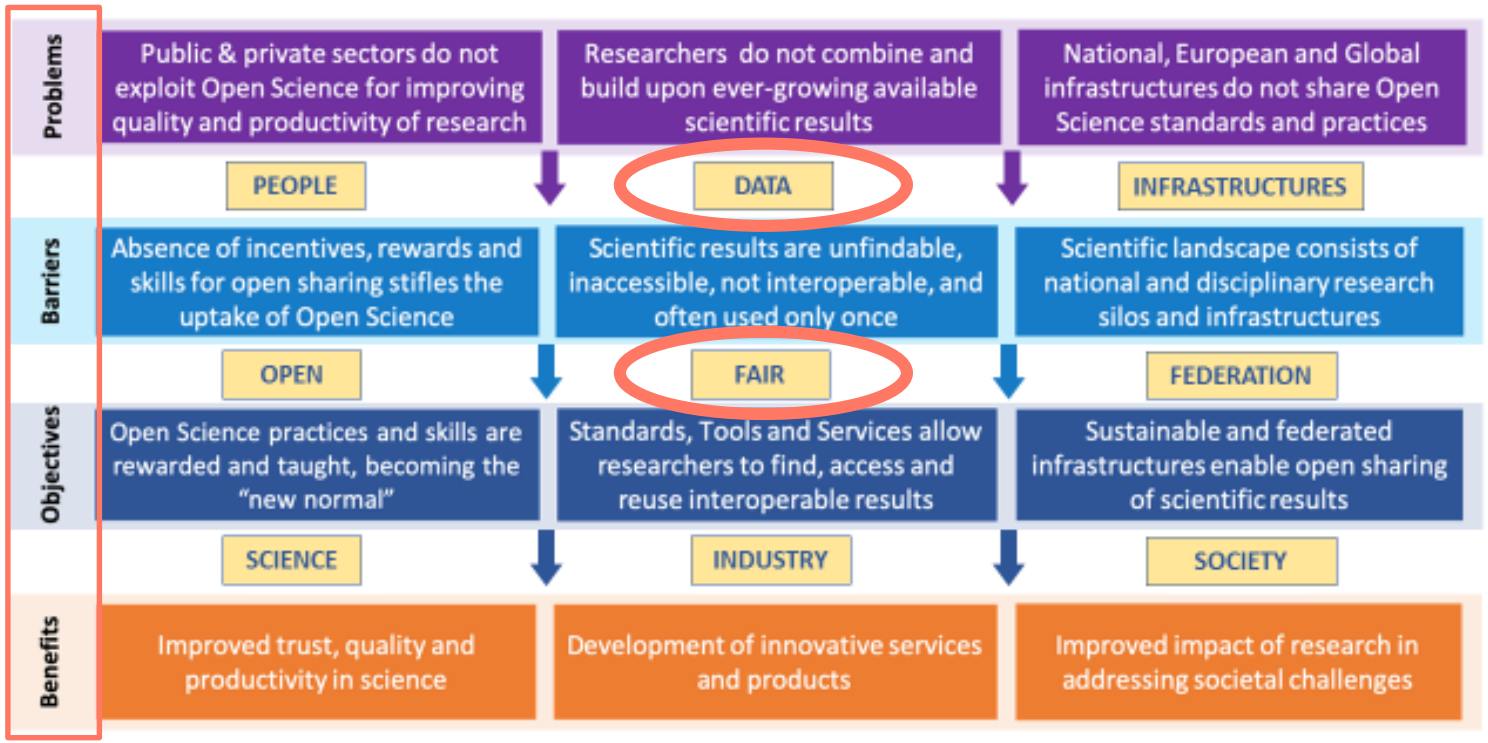
EOSC timeline



EOSC, albero degli obiettivi

EOSC Hub week May 18 2020, C. Stover

The EOSC Objective Tree



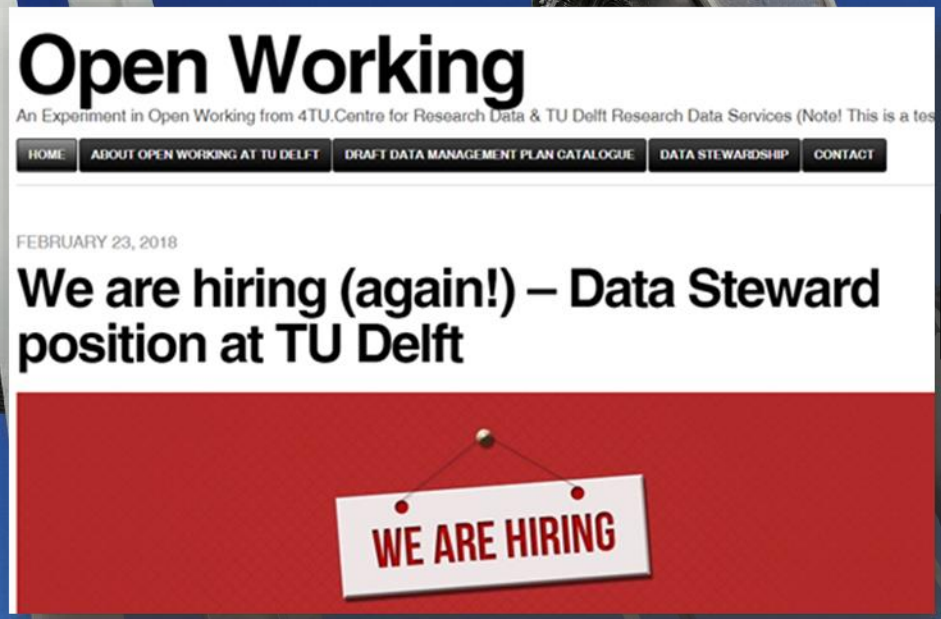
EUROPEAN OPEN SCIENCE CLOUD

[EOSC e I data stewards]

Report, 2016

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

≠ DATA SCIENTIST



≠ DATA ANALYST

[competenze]

KØBENHAVNS UNIVERSITET 06/04/202 0

Education core content

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

Preliminary Content

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

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

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

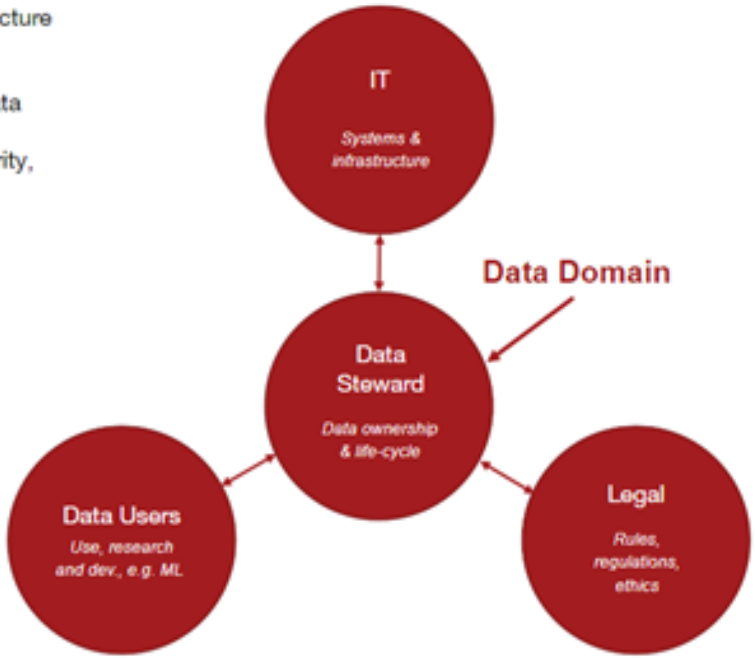
Competence Profile

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

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

Competence profile examples

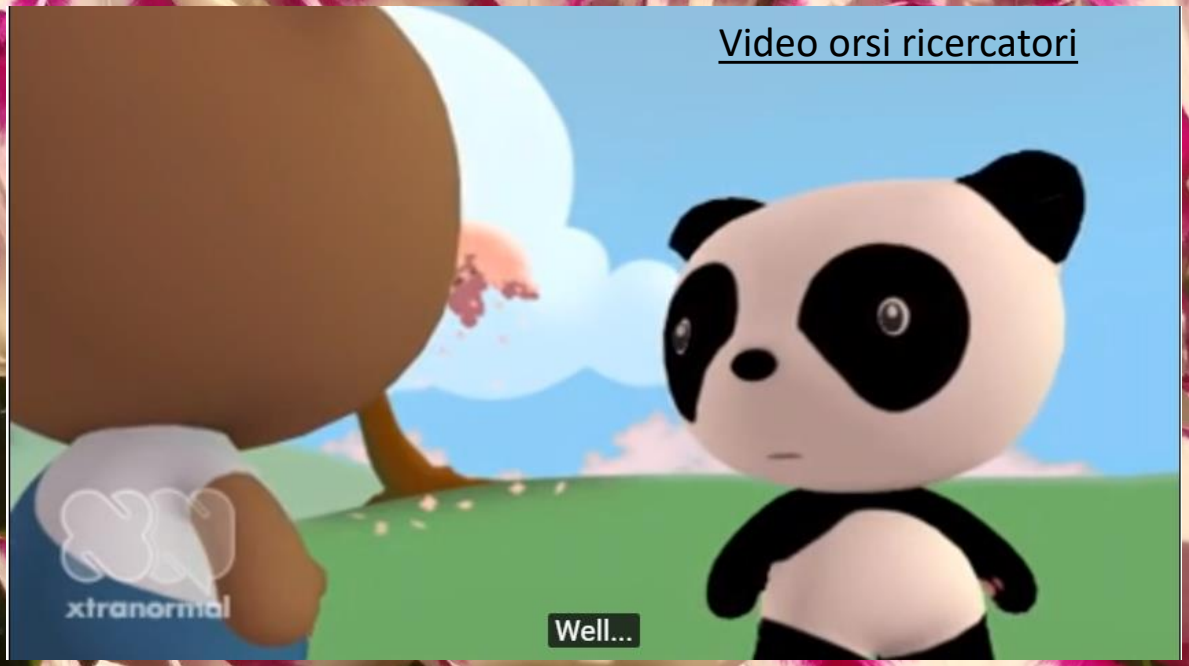
- Domain-specific data understanding
- Ability to ensure that structured and unstructured data and meta data is modelled, harvested, stored, and maintained in a documented, and regulated fashion with focus and findability, accessibility, interoperability, and reusability.
- Competences to facilitate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



Copenhagen Univ. June 17 2020

Perché occuparci dei dati?

[Video]



… È L'INCUBO DEL DATA STEWARD:

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

… E IN PIÙ:

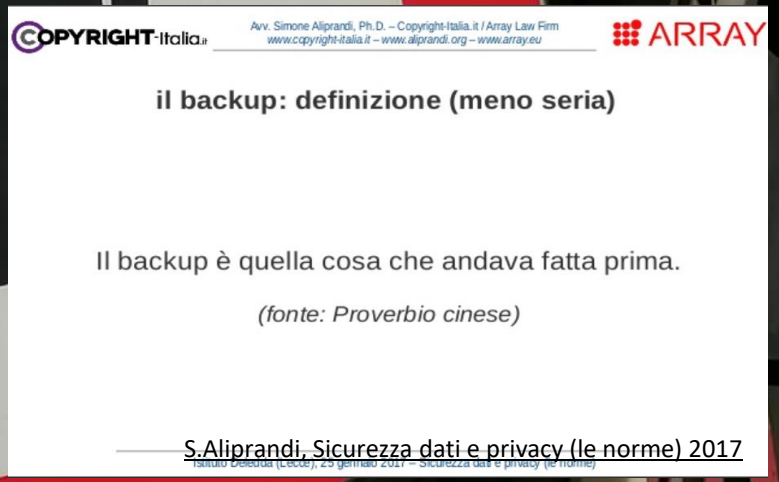
- DATI PRODOTTI CON FONDI PUBBLICI
- PUBBLICATI SU SCIENCE CHE LI RICHIEDE
- UTILI A UNA RICERCATRICE DI AREA DIVERSA

...parliamo di dati

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

QUANTI DI VOI AVREBBERO
PROBLEMI SE PERDESSERO ORA
LA CHIAVE USB?

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



Perché occuparci dei dati?

Jisc

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

An introduction to engaging with research data management processes.

Jisc Guide

EVITARE DI PERDERLI

ORGANIZZARLI PER RENDERE PIÙ EFFICACE LA RICERCA

(SE APERTI) ESSERE PIÙ VISIBILI

ALCUNI SONO UNICI E IRRIPETIBILI (meteorologia)

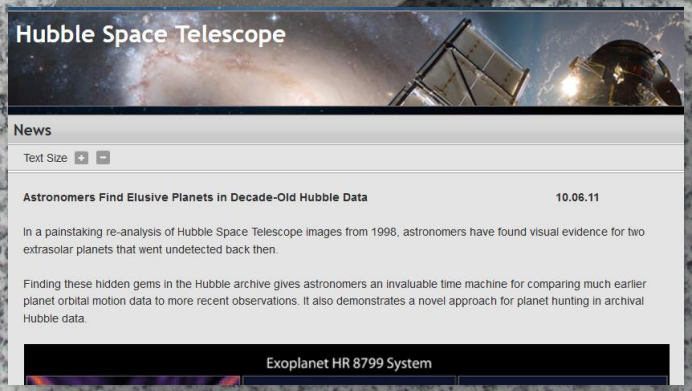
PERMETTERE VALIDAZIONE E CONTROLLI

(SE APERTI) FAVORIRE COLLABORAZIONI

MIGLIORARE INTEGRITÀ DELLA RICERCA

(SE APERTI) FAVORIRE RIUSO INEDITO

ESSERE RIPRODUCIBILI



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

Perché occuparci di dati?

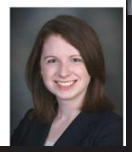
Il debito pubblico deprime la crescita? Il clamoroso errore di Carmen Reinhart e Kenneth Rogoff

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

2013

DATA AB INITIO

K.Birney, 2015



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,

PER SCOPRIRE ERRORI

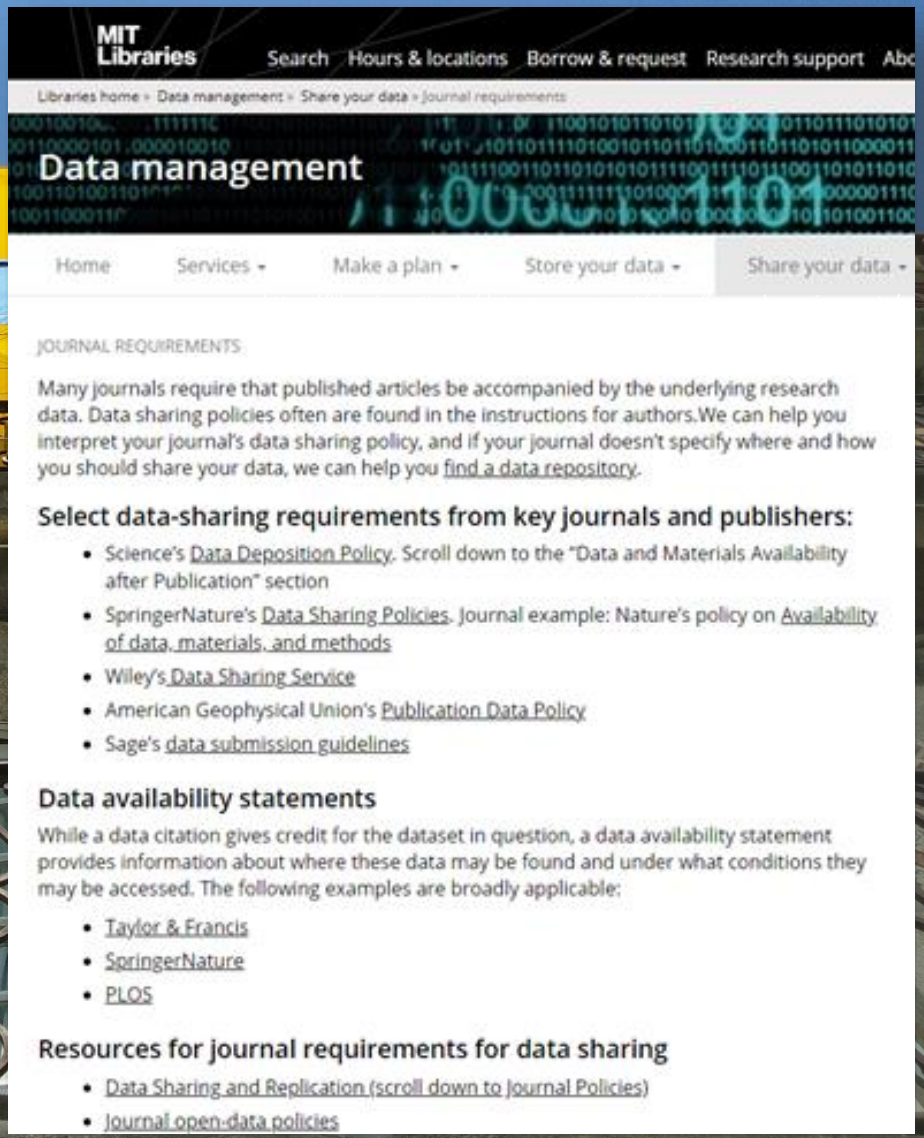
1. **l'esclusione selettiva di alcune osservazioni nei dati;**
2. **uno schema di bilanciamento dei dati non convenzionale;**
3. **un errore di codice nel foglio di calcolo originale utilizzato per selezionare i dati.**

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

Perché occuparci dei dati?

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

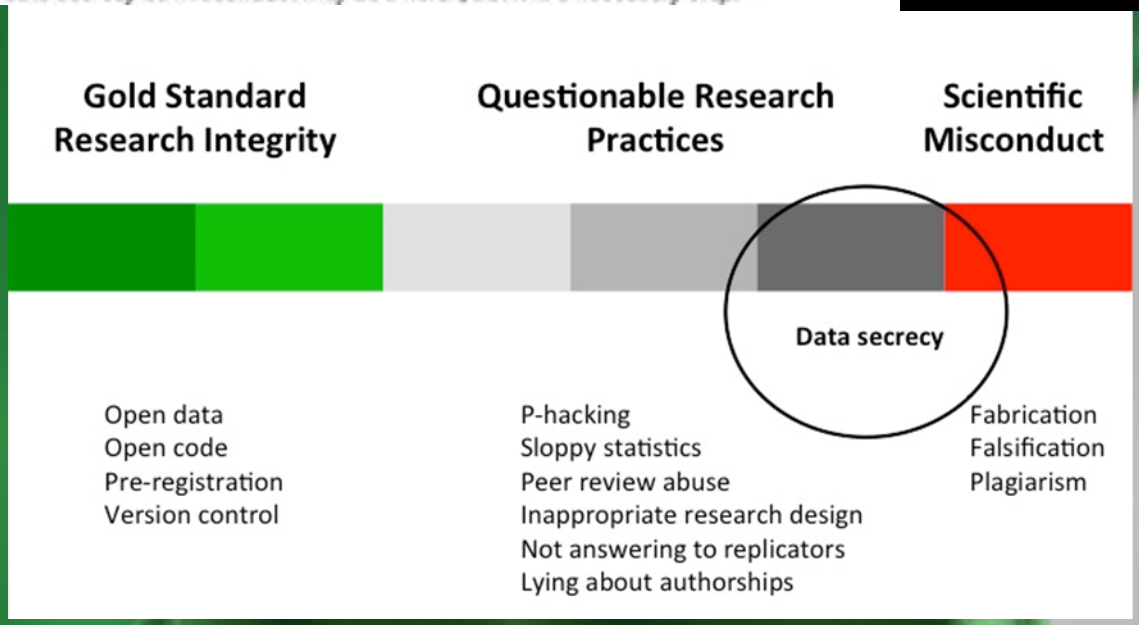
E UN NUMERO CRESCENTE DI ENTI FINANZIATORI LO RICHIEDE



Nessun dato?

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

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



I dati creano ponti

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

World Economic Forum 2012,

I DATI CREANO PONTI FRA LE DISCIPLINE…
…E NON È INDIFFERENTE NEL CONTESTO MISSION-ORIENTED DI HORIZON EUROPE…

Missions in Horizon Europe

Mission areas

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

Mission area: Adaptation to climate change including societal transformation

Mission area: Cancer

Mission area: Climate-neutral and smart cities

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

Mission area: Soil health and food

I dati rendono la ricerca più trasparente

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

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

PLOS | MEDICINE 2014

OPEN ACCESS

ESSAY

How to Mak

John P. A. Ioannidis

Published: October 21, 2

PLOS | MEDICINE 2005

OPEN ACCESS

ESSAY

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005

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

The Preregistration Revolution

Brian Nosek, Charles Ebersole, Alexander DeHaven, David Mellor

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

ALLEA All European Academies

integrity

The European Code of Conduct for Research Integrity

REVISED EDITION

Research Integrity

nature human behaviour

Altmetric: 2520 Citations: 63

Perspective | OPEN

A manifesto for reproducible science

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

Manifesto, Jan 2017

[i tre passi fondamentali]

Dati Open

Dati FAIR

Dati gestiti correttamente

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

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

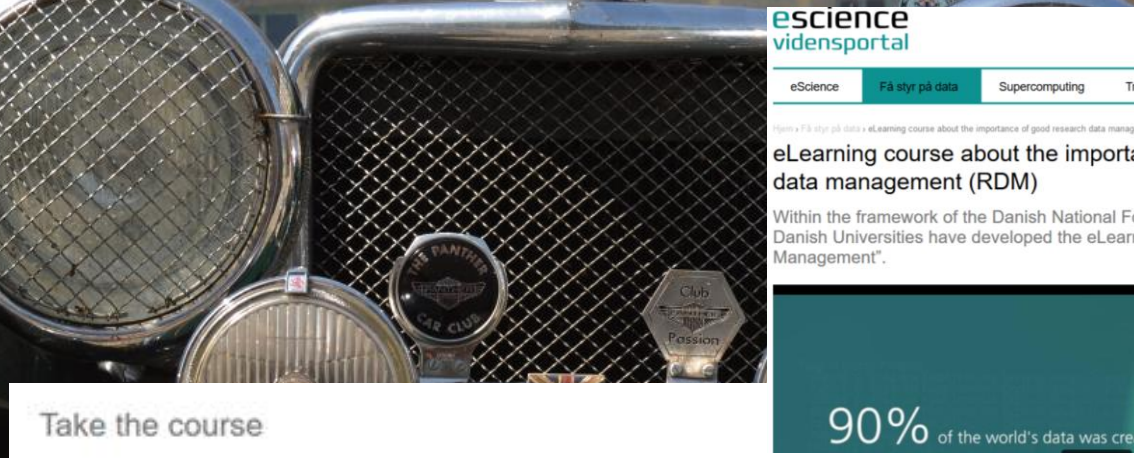
3. MA SE I DATI NON SONO CORRETTAMENTE GESTITI, RENDERLI «FAIR» COSTA TROPPO TEMPO E DENARO

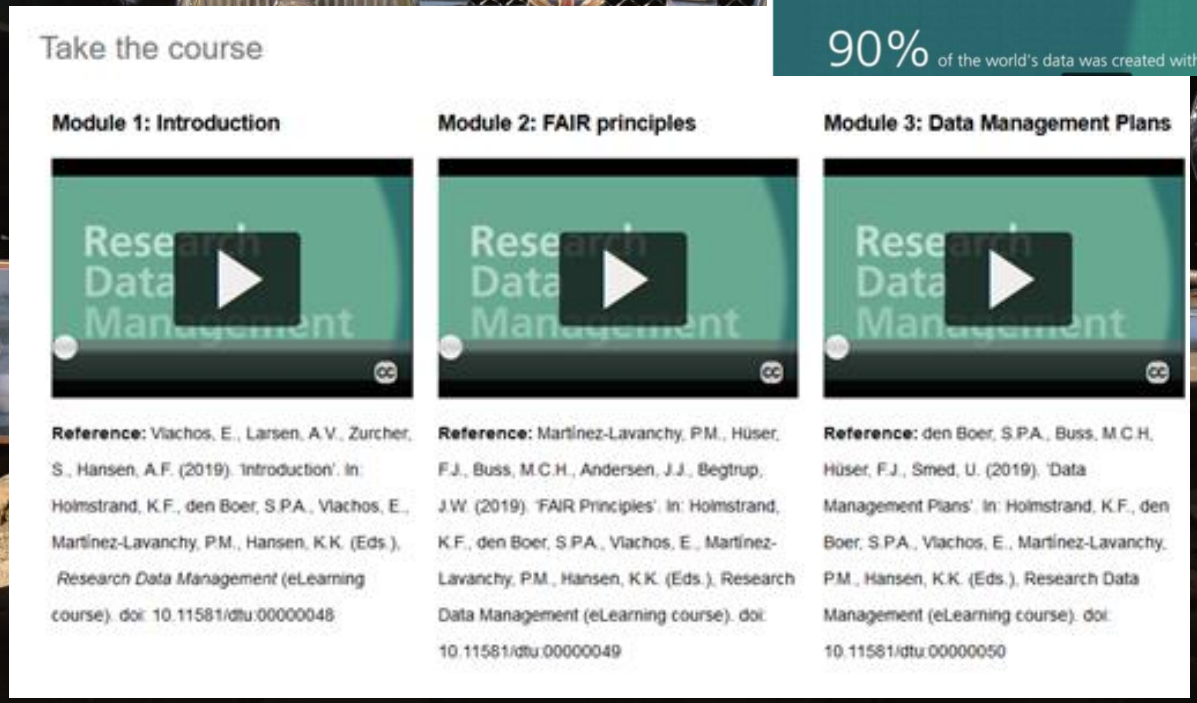
E GESTIRE I DATI CORRETTAMENTE È NELL'INTERESSE PRIMARIO DEL RICERCATORE, PERCHÉ L'INTERA RICERCA SCORRE PIÙ FLUIDA

1. I dati vanno gestiti

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

…una via veloce al data management





…formazione più strutturata

cessda TRAINING

https://www.cessd

Data Management Expert Guide

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

Plan

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

Organise & Document

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

Process

Adapt your DMP: part 6

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

Deposit your data

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

[dati personali]

Legal Basis

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

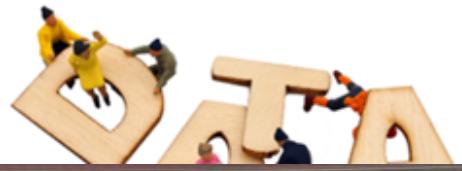
- consent of the data subject
- necessary for the performance of a contract
- legal obligation placed upon the data controller
- necessary to protect the vital interests of the data subject
- carried out in the public interest or in the exercise of official authority (public task)
- legitimate interest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

The research exemption

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

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

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



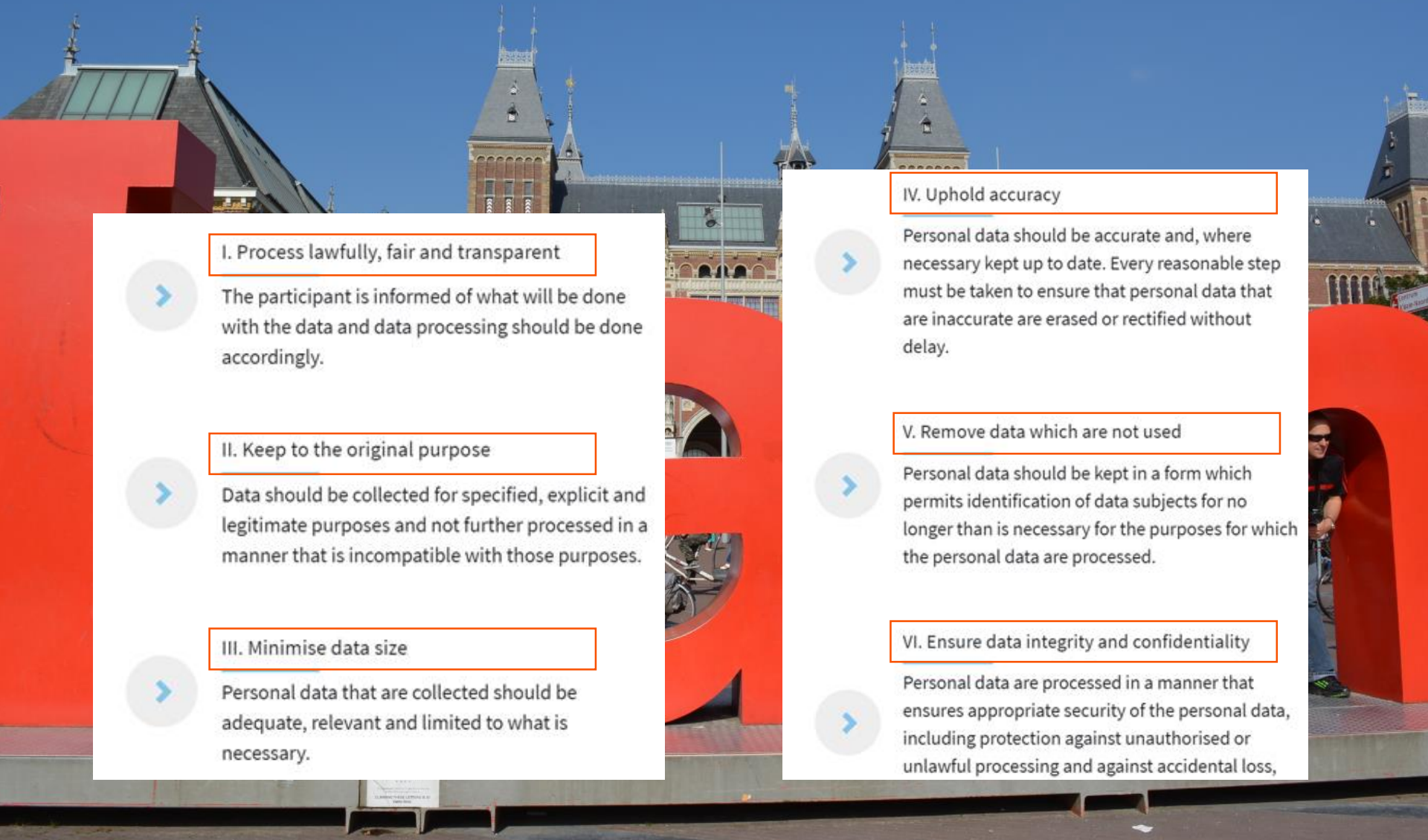
ART. 89 ECCEZIONI PER LA RICERCA MA SEMPRE SU UNA BASE LEGALE

cessda TRAINING

CESSDA guide
Data Management Expert Guide

[dati personali]

CESSDA guide
Data Management Expert Guide



I. Process lawfully, fair and transparent

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

II. Keep to the original purpose

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

III. Minimise data size

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

IV. Uphold accuracy

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

V. Remove data which are not used

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

VI. Ensure data integrity and confidentiality

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

2. I dati devono essere FAIR

To be Findable:

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

TO BE ACCESSIBLE:

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

TO BE INTEROPERABLE:

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

TO BE RE-USABLE:

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

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

FAIR/Open

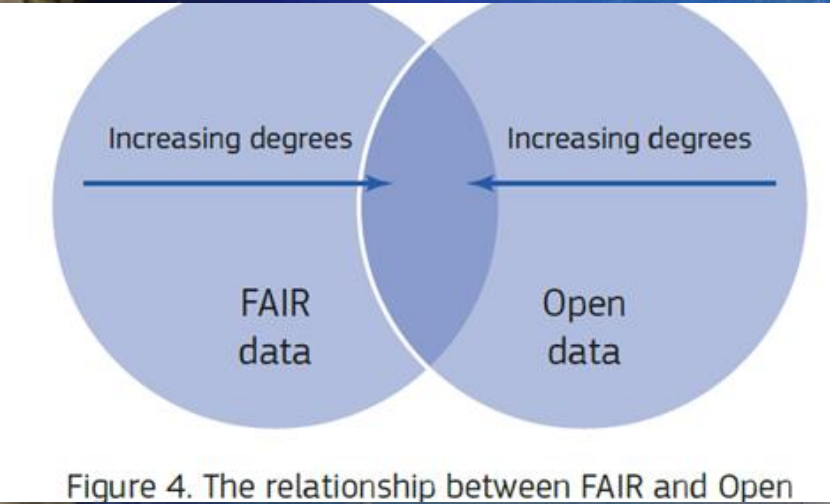


Figure 4. The relationship between FAIR and Open

Turning FAIR into reality, 2018

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

…FAIR SIGNIFICA
[anche e soprattutto per le macchine]


SCIENTIFIC DATA

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

Open Access | Published

The FAIR Guiding Principles for scientific data management and stewardship

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

FAIR guide, Nature, March 2016

FINDABLE

- IDENTIFICATIVI
- METADATI

ACCESSIBLE

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

INTEROPERABLE

- STANDARDS
- ONTOLOGIE

REUSABLE

- LICENZE D'USO
- DOCUMENTAZIONE

IL TUTTO, LEGGIBILE DALLE MACCHINE

FAIR in sintes

FAIR data training
- Findable
- Accessible
- Interoperable
- Reusable

FAIR for Developers

FAIR data self-assessment tool

+SHARE

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

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

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different leve

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

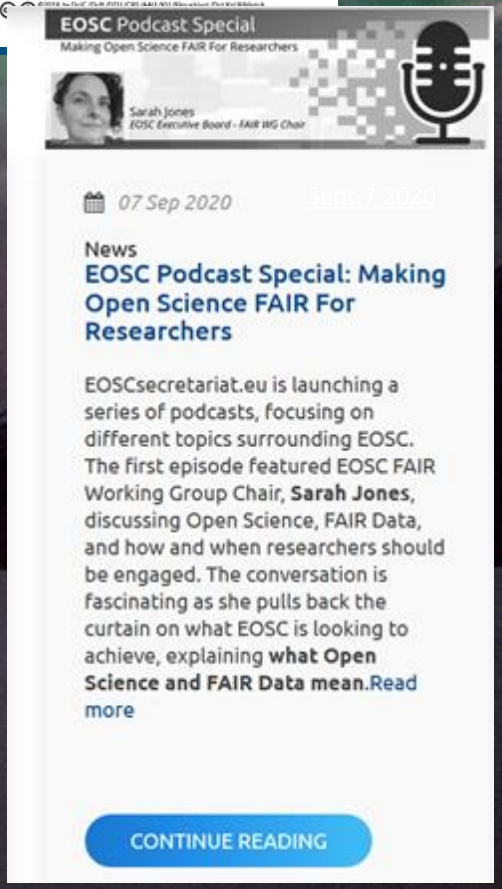
It is also a provider of services for minting persistent identifiers of many di of the data being identified):

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

Complementary to the assignment of persistent identifiers is their proper

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

I miti su FAIR



… sfumature di FAIR

Findable

Does the dataset have any identifiers assigned? — No identifier

Is the dataset identifier included in all metadata records/files describing the data? — No

How is the data described with metadata? — The data is not described

What type of repository or registry is the metadata record in? — The data is not described in any repository

Accessible

How accessible is the data? — No access to data or metadata

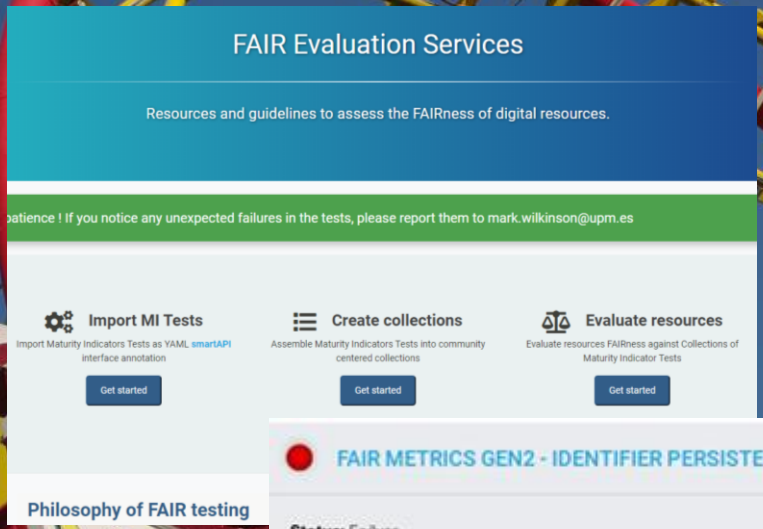
Is the data available online without requiring specialised protocols or tools once access has been approved? — No access to data

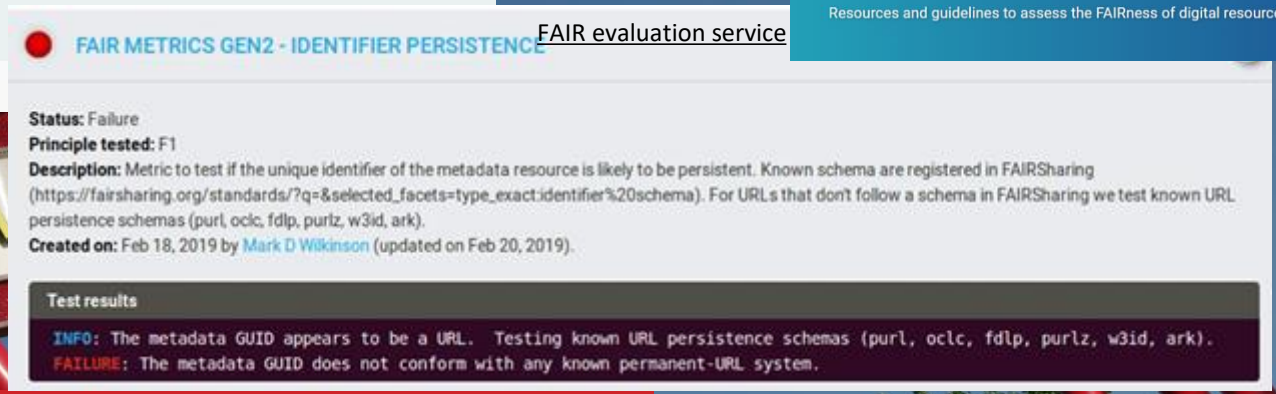
Will the metadata record be available even if the data is no longer available? — Unsure

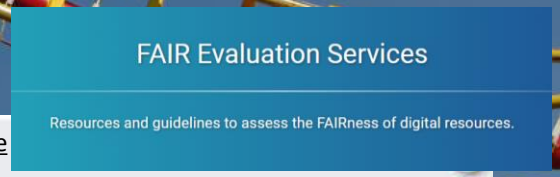
UTILISSIMO PER PORSI LE DOMANDE GIUSTE… MA SOGGETTIVO…

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

FAIR maturity evaluator







FAIR evaluation service

Sept. 20, 2019

- OGGETTIVO
- LEGGIBILE DALLE MACCHINE – COME I DATI FAIR

FAIR Data management wizard

ELIXIR
Data Stewardship Wizard

Smart Data Management Plans for FAIR Open Science

For serious researchers and data stewards

Data Stewardship Wizard

Data integration 7

Data interpretation 3

Information and insight 14

Is there any pre-existing data?

Are there any data sets av

Data Stewardship for C

No

Yes

Will reference data be cr

Will any of the data that y

thers)?

Data Stewardship for C

No

Yes

Will you be storing sample

Data Stewardship Wizard | Questionnaire Demo | Log In | Sign Up

Will you store licenses with the data?

What's up?

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

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

Do

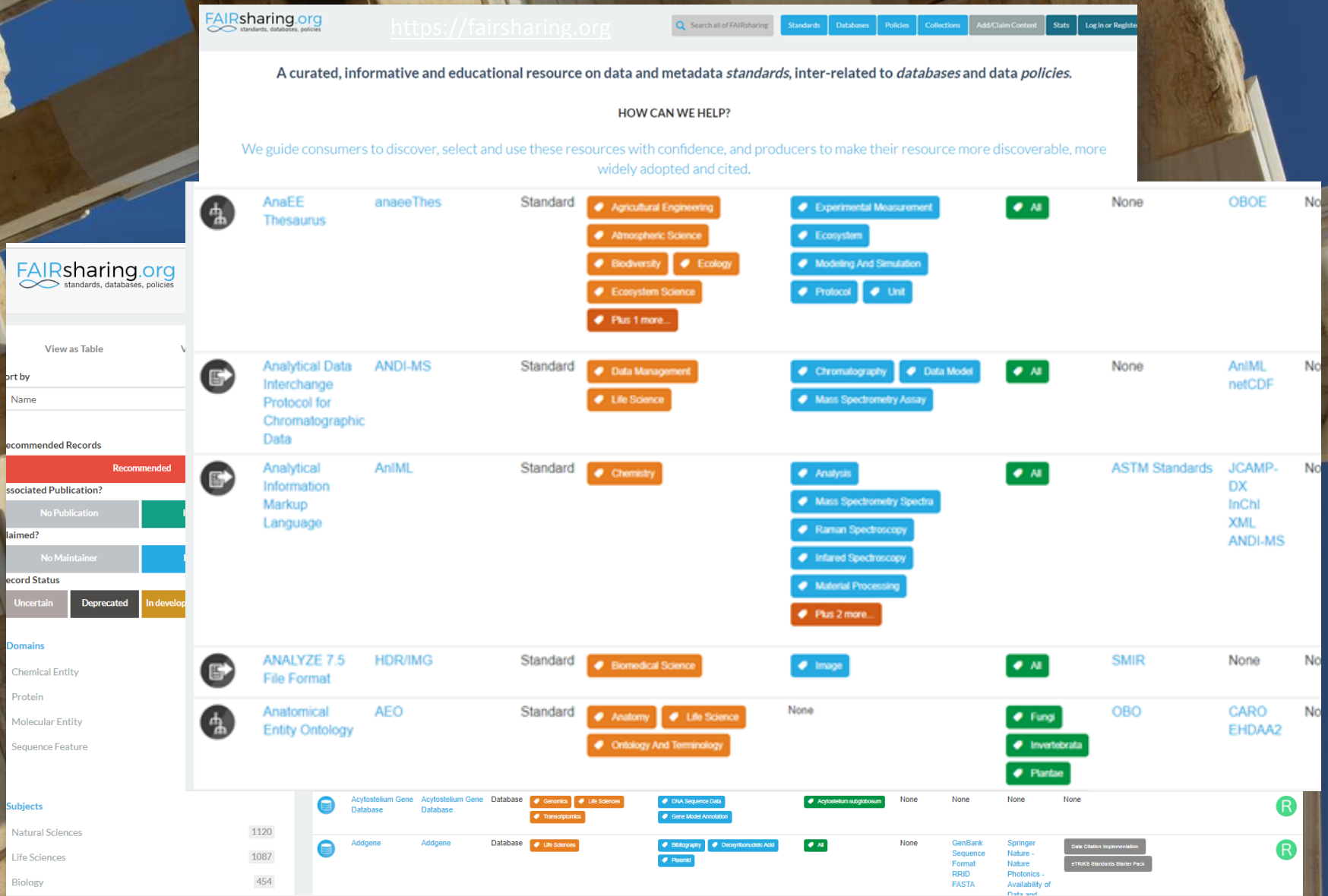
- Always carefully choose a license to be attached to your data upon publication.
- Include and clearly mark the licences PID as a concept + attributes in the metadata.
- Store and 'expose' the license as part of the metadata in Open Access environments where search engines can easily find the license,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

Don't

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

- SI COSTRUISCE IN BASE ALLE RISPOSTE
- APRE I CAPITOLO DEL MANUALE DI BAREND MONS SULLA DATA STEWARDSHIP

… un pilastro: FAIRsharing



R= Reusable. Licenze

KEEP CALM AND

RICORDA: NESSUN COPYRIGHT SUI DATI (NON CREATIVI)

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

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

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

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 11 March 1996

on the legal protection of databases

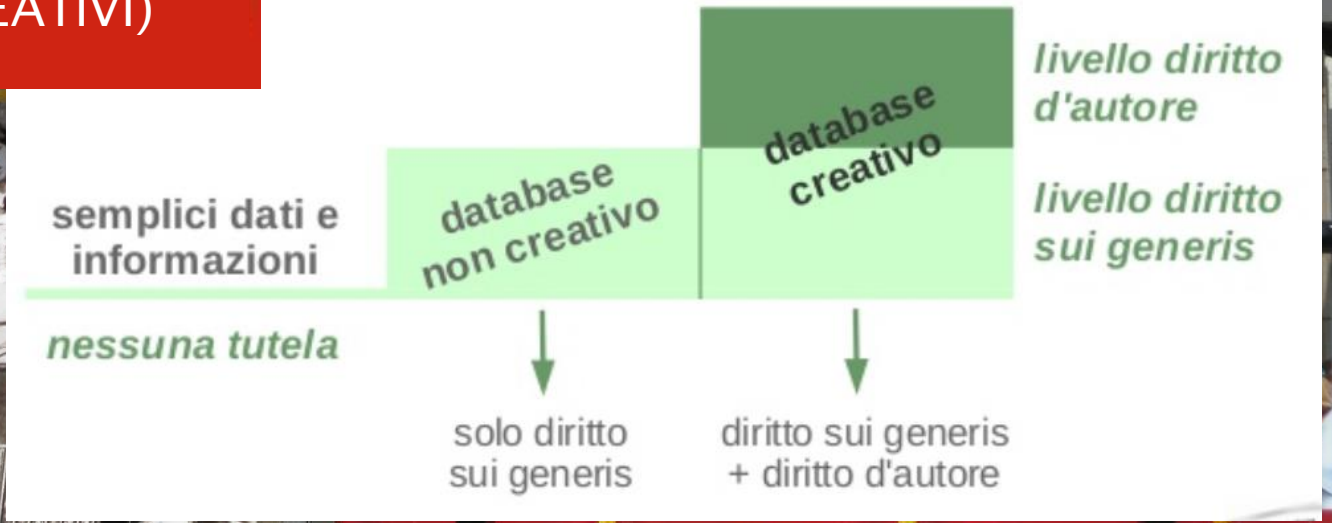
NCIL OF THE EUROPEAN UNION,

n Community, and in particular Article 57 (2), 66 and 100a thereof,

(1),

Simone Aliprandi

QUALI DIRITTI SUI DATI?

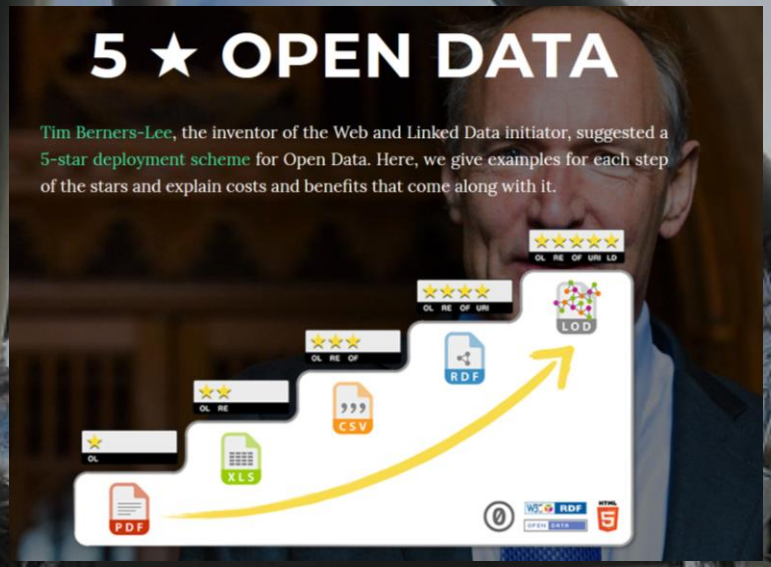


[DMP]

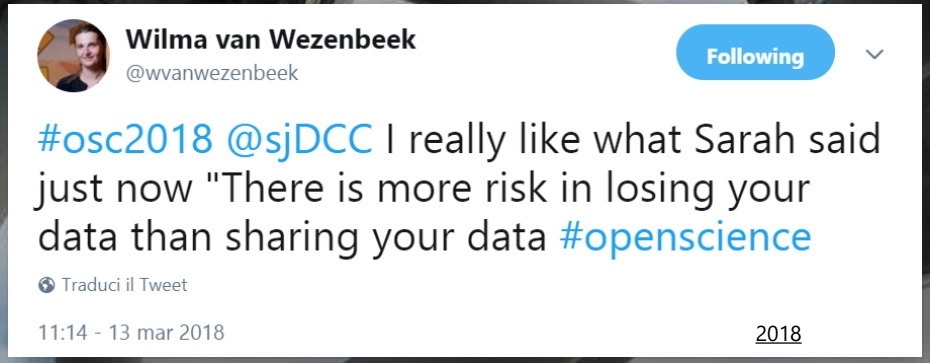
[webinar]

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

3. I dati POSSONO essere Open



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



2018

UN MODO STRUTTURATO DI PENSARE AI DATI

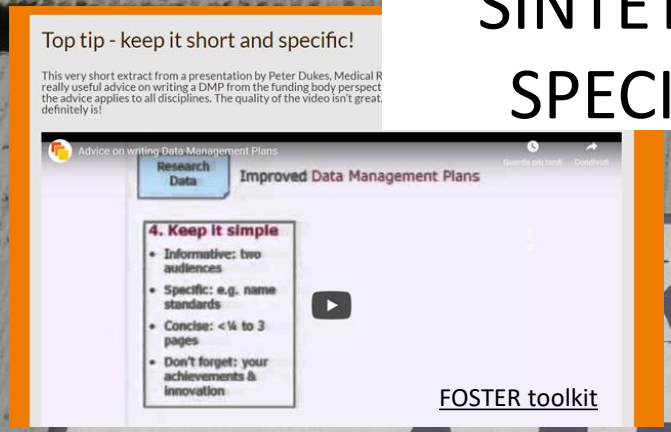
REGOLE CHIARE=MENO ERRORI DA SUBITO

È UN «LIVING DOCUMENT», CRESCE COL PROGETTO

DOVE METTERE TUTTE QUESTE INFORMAZIONI? NEL DATA MANAGEMENT PLAN

Trucchi e suggerimenti

SINTETICO E SPECIFICO



NON COPIATE

OGNI DATASET È UNICO, OGNI INFRASTRUTTURA È UNICA, OGNI RICERCA HA LA SUA IMPOSTAZIONE

- USATE TABELLE, ELENCHI PUNTATI
- SIATE SCHEMATICI E NON DILUNGATEVI

ESSERE GENERICI NON SERVE A NULLA
[we expect a huge size of data; data will be available]

CIÒ CHE DICHIARATE NEL DMP POI VA FATTO VERAMENTE… QUINDI

A) NON FATE GLI SPLENDIDI
B) NON IMPEGNATEVI A FARE COSE CHE SAPETE IMPOSSIBILI

Es. DATI PSEUDONIMYZED, non ANONIMYZED

Vantaggi di un DMP

CESSDA Guide

Benefit 3. Clarifies needed budget

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

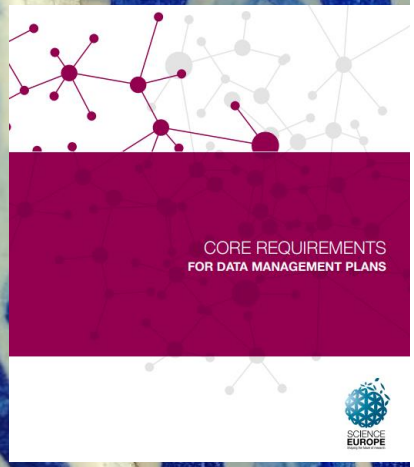
A DMP can be useful in the process of applying for funding. Grant applications should not only include time and resources for collecting, analysing, and publishing on data in their budget, time and resources for careful documentation as well as server space, backup solutions, and documentation software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

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

È FONDAMENTALE PER STIMARE I COSTI DI GESTIONE - STIMATE LA DATA STEWARDSHIP (IN-KIND?) -POSSIBILI COSTI DI STORAGE COSTI SONO RIMBORSABILI IN H2020 (6.2.D.3 AMGA)

DMP Core Requirements

Dec. 2018



CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

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

1. Data description and collection or re-use of existing data

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

b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. Documentation and data quality

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

b. What data quality control measures will be used?

3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

c. How will possible ethical issues be taken into account, and codes of conduct followed?

SEZIONI MINIME IN UN DMP

5. Data sharing and long-term preservation

a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

c. What methods or software tools will be needed to access and use the data?

d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

DMP questions

PLAN

Overview

Title of the project

Date of this plan

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

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

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

Collaborating researchers (if applicable)
- What are their contact details and their roles in the project?

Funder (if applicable)
- If funding is granted, what is the reference number of the funding granted?

Data producer
- Which organisation has the administrative responsibility for the data?

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

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

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

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

UTILI COME SPUNTO PER LE PRIME VOLTE, PER CAPIRE QUALI DOMANDE FARSI

ORGANISE & DOCUMENT

Organising and documenting your data

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

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

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

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

Folder structure and names
- How will you structure and name your folders?

File structure and names
- How will you structure and name your files?

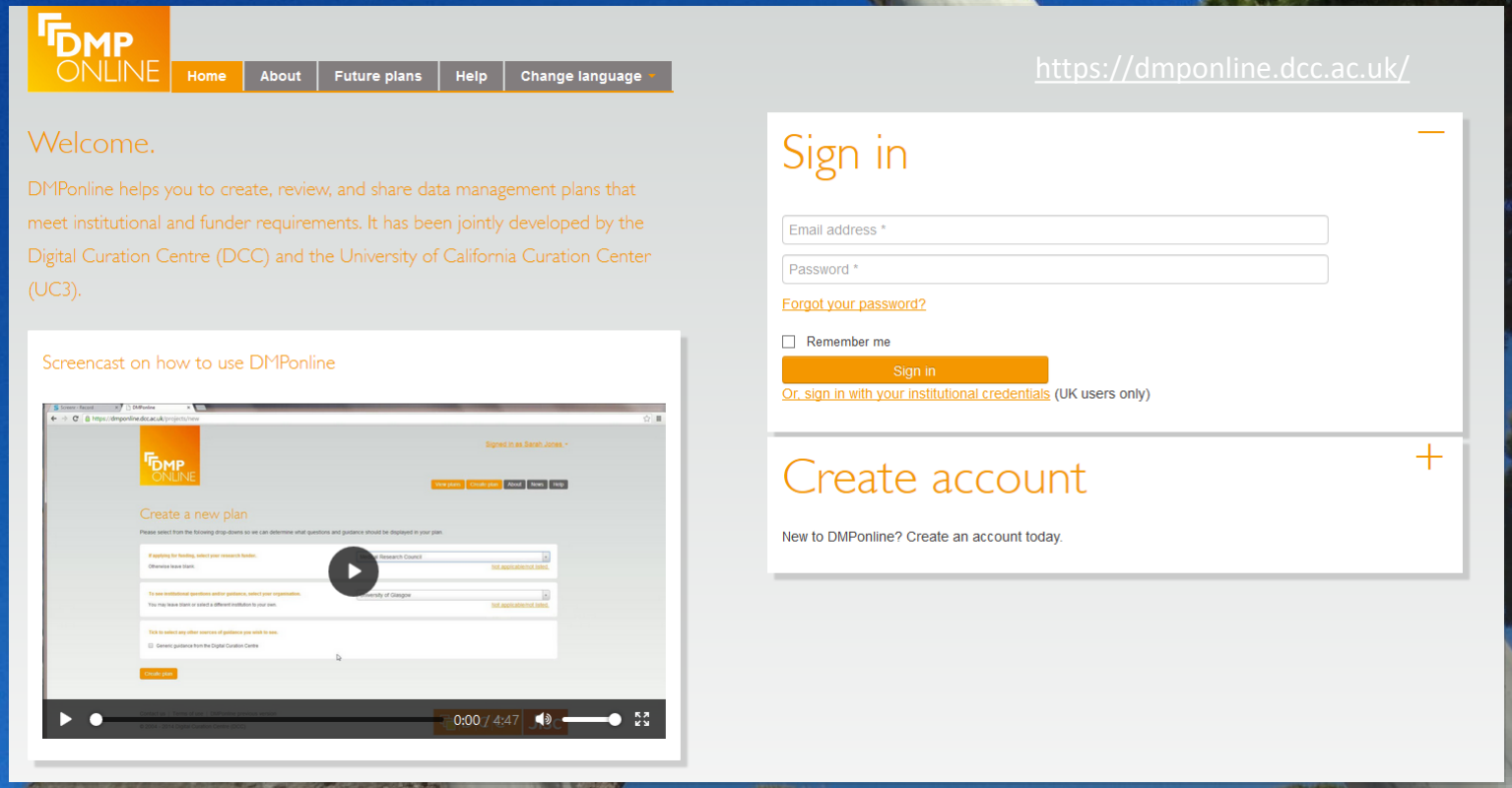
Documentation
- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

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

Metadata standard (if applicable)
- What metadata standard(s) will you use?

https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf

DMP online



REGISTRAZIONE LEZIONE SUL DATA MANAGEMENT PLAN https://doi.org/10.5281/zenodo.3900981

… ora tocca a voi!