

The background of the slide is a blue-tinted image of Earth from space. A bright sun flare is visible on the left side, creating a lens flare effect. The Earth's surface shows clouds and landmasses.

How to make your data open

Marialaura Vignocchi

Alma Digital Library

Multimedia Center

University of Bologna

The bigger picture outside academia

Thursday 29th October 2015



open data charter

[HOME](#)

[ABOUT](#)

[CHARTER](#)

[RESOURCE CENTER](#)

[FAQ](#)

[SPREAD THE WORD](#)



Seventeen Governments Adopt the New International Open Data Charter

There is a strong societal demand for democracy, accountability, inclusion. Open government data are considered a tool for enhancing transparency, fighting corruption and driving economic growth....

Research funders policies require open access to research data and scientific journals ask for underlying data availability

HORIZON 2020

PLOS ONE Publish

SCIENTIFIC DATA

Home | Archive | About | For Authors | For Referees | Data Policies | Co

Home ► Data Policies

Data policies

Scientific Data publishes manuscripts describing scientifically valuable datasets. All data must be made available to editors and referees at the time of submission, and to the scientific community as a condition of publication. Authors will need to identify suitable data for their datasets. On this page, we provide information on the types of data that are acceptable for publication.

Data Availability

The following policy applies to all of PLOS journals, unless otherwise specified.

NIH Data Sharing Policy

“Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.”

Acceptable Data-Sharing Methods

Unacceptable Data Access Restrictions

Explanatory Notes and Guidelines

Recomm

FAQs

Decreto Direttoriale 4 novembre 2015 n. 2488

Bando PRIN 2015



Ministero dell'Istruzione, dell'Università e della Ricerca

Dipartimento per la Formazione Superiore e per la Ricerca

Direzione Generale per il coordinamento, la promozione e la valorizzazione della ricerca

Articolo 7 Open access

1. Ciascun responsabile di unità deve garantire l'accesso aperto (accesso gratuito on-line) tutte le pubblicazioni scientifiche *peer-reviewed* relative ai risultati ottenuti nell'ambito di progetti di ricerca previsti dall'art.4, commi 2 e 2 bis, del decreto legge 8 agosto 2013, n.91, convertito con modificazioni dalla legge 10 ottobre 2013, n.112.

2. Il rispetto di quanto indicato al comma 1 del presente articolo non incide su eventuali obblighi relativi alla tutela dei dati personali, ognuno dei quali resta impegnato a garantire.

3. Come eccezione, i responsabili di unità sono esentati da assicurare l'accesso aperto a parti specifiche dei propri dati di ricerca, se l'accesso aperto a tali dati dovesse compromettere il raggiungimento del principale obiettivo della ricerca stessa.

H2020 Open data pilot requirements

Art. 29.3 of the Grant Agreement states that “[r]egarding the digital research data generated in the action (‘data’), the beneficiaries must:

- ***deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate*** — free of charge for any user — the following:
 - *the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
 - *other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan';*
- ***provide information*** — via the repository — ***about tools and instruments*** at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).
- Exceptions: commercial exploitations, ethical issues, confidentiality,

Emergent trends

- Research outputs resulting from public funding should be openly available (research funding obligation)
- Underlying data (research data used as evidence to support the findings in science) should be made available along with scientific publications to validate research

Data availability fosters scientific progress



Benefits

- **Validation**, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses
- **Transparency**: malpractice and fraud detection
- **Reproducibility** of research: reduction of duplication
- **Preservation**: efforts to ensure data are archived, increasing the value of the investment made in funding scientific research
- **Reduction of the burden on authors** in unearthing old data, retaining old hard drives and answering email requests
- **Easier citation** of data as well as research articles, enhancing visibility and ensuring recognition for authors

Data availability fosters scientific progress

PLOS data policy <http://journals.plos.org/plosone/s/data-availability>

Open data: what does *open* mean?

“Open data is digital data that is made available with the **technical and legal characteristics** necessary for it to be **freely used, reused, and redistributed** by anyone, anytime, anywhere”

International Open Data Charter (Sept. 2015)

<http://opendatacharter.net/principles/>

«Intelligent openness»

Research data should be:

- ✓ **discoverable** (metadata and unique identifiers)
 - ✓ **accessible** (open licensing)
 - ✓ **intelligible** (documentation, standards)
 - ✓ **assessable** (metadata, documentation, standards)
 - ✓ **usable** beyond the original purpose for which they were created (long term preservation, data curation and persistent identification)
 - ✓ **interoperable** (interoperability standards)
-
- H2020 *Guidelines on data Management* (updated 30th October 2015) http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

How do scholars perceive open data
requirements?
A burden or an opportunity?



or a even risk?



Challenges

- Cultural
- Legal
- Technical
- Financial



Cultural Barriers

- The *Gollum* syndrome: *my precious...*
- Research communities have different sharing cultures and habits
- Data sharing is not «scientifically rewarding», reputation is still attached to publications
- ...



Legal, Ethical and Security Issues

- Re-use of third-party owned data needs proper copyright clearance if the data have to be openly shared
- Informed consent must be obtained in research involving human participants and data must be handled and published as not to compromise personal privacy, sensitive data
- Data disclosure must not pose security risks

The Case of Text and Data Mining

These activities raise legal issues when performed on commercial copyrighted databases:

A database may be protected by both copyright and database right. For database right to apply, the database must be the result of substantial intellectual investment in obtaining, verifying or presenting the content. The database right is an automatic right and protects databases against the unauthorised extraction and re-use of the contents.

<http://www.data-archive.ac.uk/create-manage/copyright/duration-rights>

Italian Law: http://www.interlex.it/testi/l41_633.htm#102-bis

Technical Challenges: Data Diversity

- Data may be **anything**: statistics, the results of measurements, experiments, observations, of surveys, audio or video recordings of interviews, images of objects...
- Data may be **any format**: .txt, .csv, .jpg, .mp3, .mp4...
- Data may have **any volume**: big and small data
- Data may **change status** in the course of the research: raw, processed, uncleaned, cleaned, unvalidated, validated...

More on Technical Challenges

- Data processing and transformation for publication may require anonymization, compression, conversion to open formats...
- Technological infrastructures for dissemination and preservation must ensure:
 - versioning track
 - discoverability
 - interoperability
 - persistent identification
 - data quality and integrity, security and curation
 - long term preservation and data reusability

Costs: who pays?

- Research funding policies generally cover the **costs for proper data management and processing** provided that projects respond to technical requirements and keep a **Data Management Plan**
- **Archiving infrastructures (repositories)** for dissemination, long term preservation, and curation of data are generally provided by institutions or community-based collaborative initiatives . How can they be sustainable in the long run? Research institutions need to coordinate their efforts to support these initiatives in the long term in order to keep them independent from private interests.

How to make your data open: **practical tips**

- ✓ Address data management issues right from the start and prepare a **Data Management Plan**
- ✓ **Deposit** your data and related metadata in a trustworthy public data repository for dissemination and preservation
- ✓ If data underpin a scientific publication provide **cross-references** by citing persistent IDs
- ✓ Seek advice and support from **your library** staff!

Library support

Libraries may have a role, they may:

- provide **intellectual property rights and licensing** advice
- provide **metadata services** for research data
- promote **data set citation**
- support research data life cycle by mediating services for **storage, discovery and permanent access** in collaboration with IT services (**Data Repository**)
- offer research data management support, **including data management plans for grant applications**

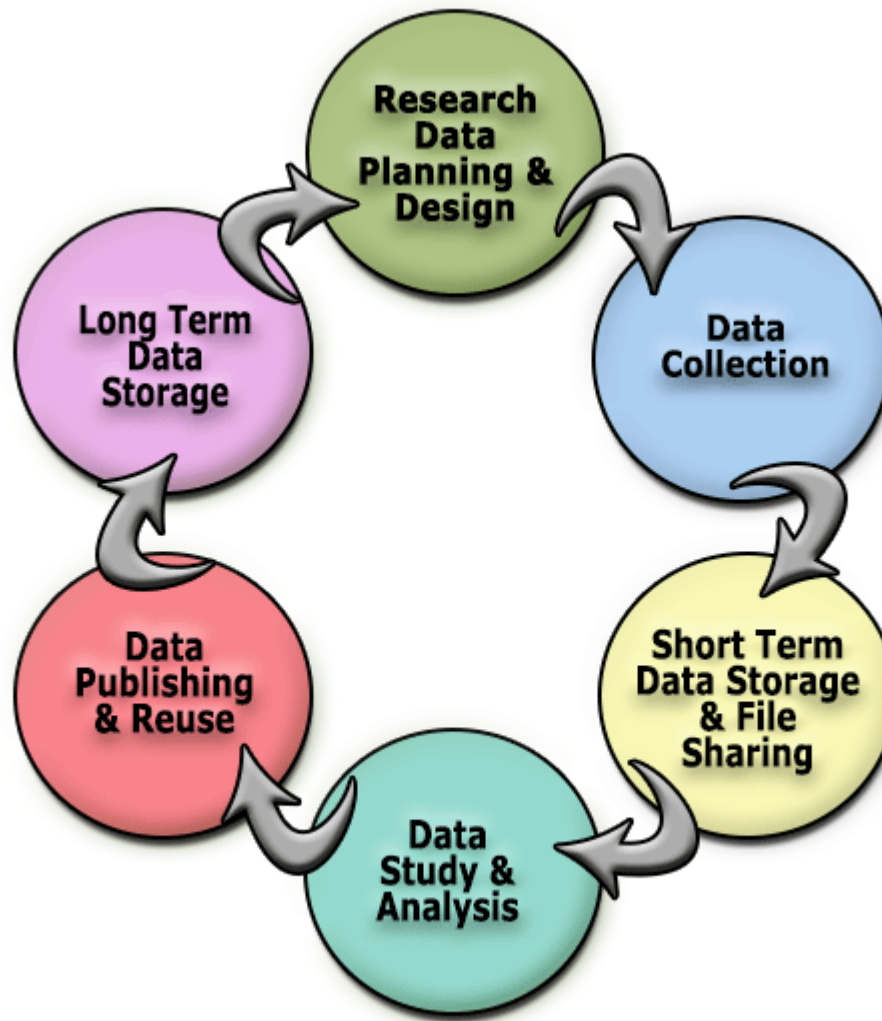
Write a Data Management Plan

- It helps you focus on:
 - data management responsibilities
 - data identification and description
 - adoption of standards and metadata
 - sharing and licensing
 - archiving and preservation

DMPs focus on the right issues



DMPs are live documents and must be updated in the course of the research project





Upcoming DCC workshop - ['Advanced DMPonline: how to customise the tool for your institution'](#),

23 November 2015, Glasgow

Home

About

Roadmap

Help

Welcome.

DMPonline has been developed by the Digital Curation Centre to help you write data management plans.

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

Sign in

[Forgot your password?](#)

Deposit data in a public repository

- Repositories may be:
 - ✓ subject-based, community recognized ([arXiv](#), [RePEc](#), [PubMed Central](#)...)
 - ✓ general-purpose, multidisciplinary ([zenodo](#), [Dryad](#), [figshare](#)...)
 - ✓ institutional ([AperTO](#) (UniTO), [AMS Acta](#) (UniBO)...

How to find a repository

- Lists of recommended repositories: PLOS, Nature...
- Data repository registries:
 - <http://service.re3data.org/browse/by-subject/>



How to choose a DR: quality requirements

- Data repositories must meet requirements for data access, preservation and integrity, ideally they should:
 - provide versioning track
 - be able to mint DOIs
 - use licenses that allow data reusability (CC)
 - ensure proper citations
 - have long term preservation policies
 - **be compliant with interoperability standard ([OpenAIRE](#) plus)**
 - be reliable and trustworthy (seals, certification...)
 - offer integration services with scientific journals (confidential peer review)
 - guarantee data integrity and fixity (*universal numeric fingerprints*)
 - state clear metadata and data policies, allowing reuse and mining
- See Austin, Claire; Brown, Susan; Fong, Nancy; Humphrey, Chuck; Leahey, Amber; Webster, Peter, *Research Data Repositories: Review of current features, gap analysis, and recommendations for minimum requirements*. IASSIST Quarterly (IQ) Preprint (2015.05.19)

<http://www.rdc-drc.ca/wp-content/uploads/Review-of-Research-Data-Repositories-2015.pdf>

General purpose repository compliant with OpenAIRE



Search

Communities

Browse ▼

Upload

Get started ▼

➔ Sign in

Filter by types

<https://zenodo.org>


Publications (17572)

Books (2020) Book sections (0) Conference papers (726) Journal articles (10438) Patents (2) Preprints (175)

Project Deliverables (45) Project Milestones (0) Proposals (15) Reports (501) Theses (198) Technical notes


The value of data

Current trends are forcing a **reevaluation of research data** as recognized and citable research products:

- ✓ Scientific journals data availability policies
- ✓ Research funders open data policies
- ✓ Data Journals: e.g. [*Scientific Data*](#) (npj) 
- ✓ Data papers: complementary to scientific publications
- ✓ [*Joint declaration of data citation principles*](#) (Force11); [*Data Citation*](#) by Dataverse Project

Data **quality** will be a key issue in the near future and comprehensive **metadata** and dependable repository infrastructures will play a strategic role in guaranteeing it.

Institutions and scientific communities should engage in developing standards and infrastructures to cope with challenges and issues raised by data management.

A view of Earth from space, showing the blue and white clouds of the planet against the black background of space. A bright sun is visible on the left side, creating a lens flare effect.

Thank you for your attention!

Wherever you are in the outer space

Always ask a librarian!

❖ marialaura.vignocchi@unibo.it