Foundational Data Stewardship Workshop

Outline for this week

- Day 1
  - Intro to RDM, open research, DMPs and FAIR (S. Venkataraman)
- Day 2
  - Open and responsible research (Louise Bezuidenhout)
- Day 3
  - Practical implementation (Joy Davidson)

Introduction to Research Data Management and Open Research

S. Venkataraman PhD, Research Data Specialist, Digital Curation Centre

s.venkataraman@ed.ac.uk

Data Stewardship Workshop, 17th May 2021, University of Botswana (Virtual)

About the DCC

Data Schools

- ○ Established in 2004.
- ○ Based in Edinburgh and Glasgow.
- ○ Works at national and international levels.
- ○ One of leading organisations in the world specialising in training, consultancy, policy making and advocacy in digital data management best practice and services provision.
- ○ Involved in many international consortia and schools.
- ○ (We do not curate any data ourselves!)

Learning outcomes

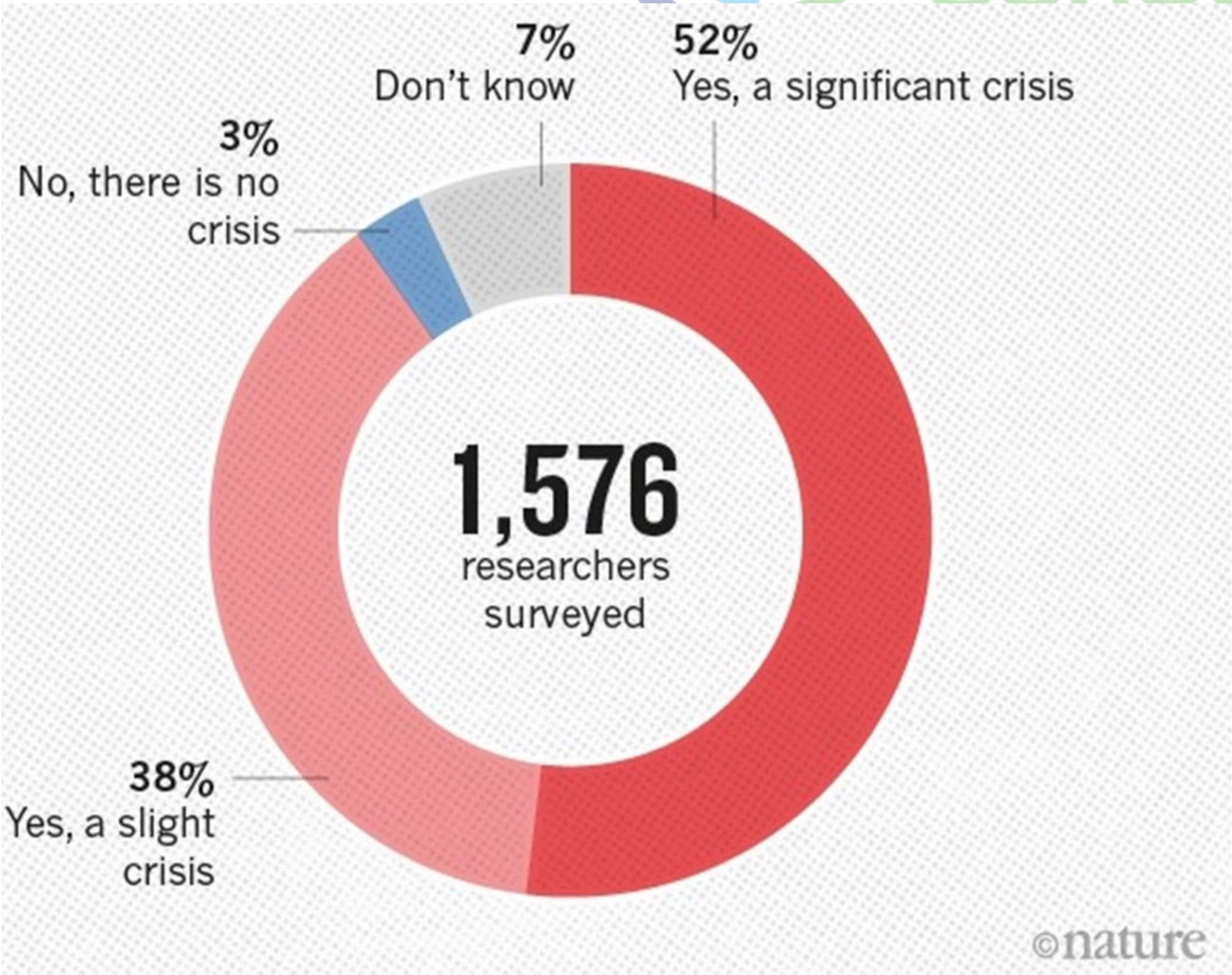
- Be familiar with the curation lifecycle.
- Understand the standardisation methods and principles available to add value to your data.
- Learn about resources to aid your workflows.
- Increase/encourage your level of openness.
- Learn about data management plans and the value in implementing them.
- Understand how data stewards integrate this knowledge

Is there a reproducibility crisis?

Baker, M. “1,500 scientists lift the lid on reproducibility” *Nature* 533: 452-454 (2016).

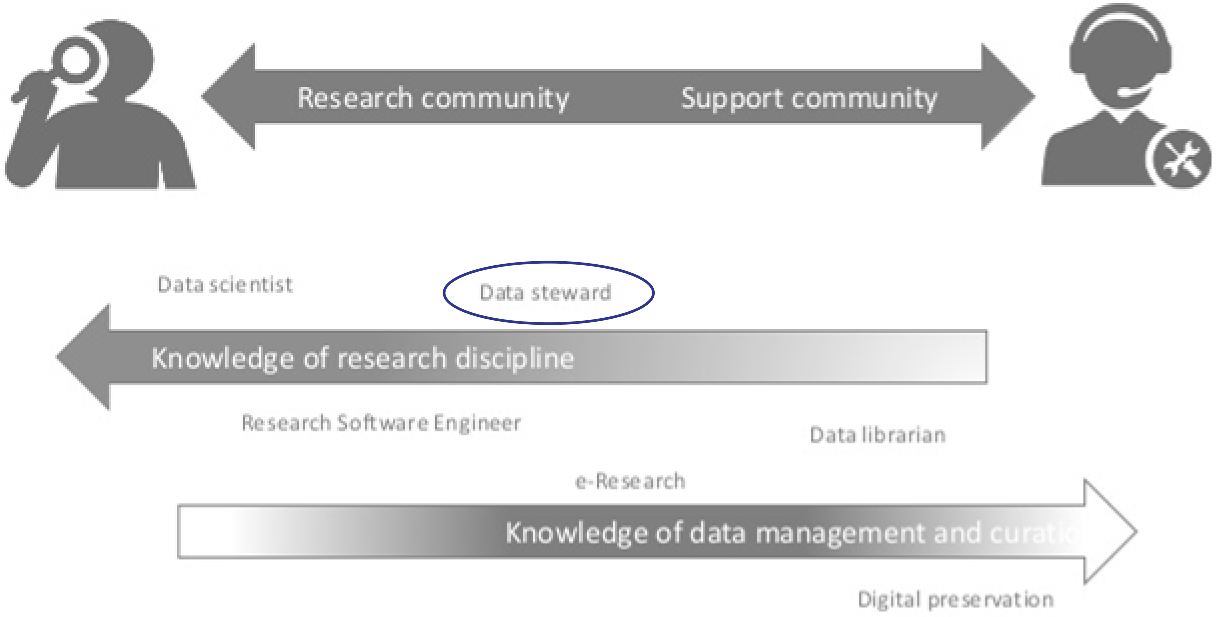
http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

doi:10.1038/533452a



Where do data stewards fit in existing landscape?

- Increase in RDM policies, DMPs, and awareness of best practices.
- Not enough people with knowledge in data stewardship to meet demand.
- Formal training even less.
- The data steward is at the boundary between researchers and support community.



Research community

Support community

Data scientist

Data steward

Knowledge of research discipline

Research Software Engineer

Data librarian

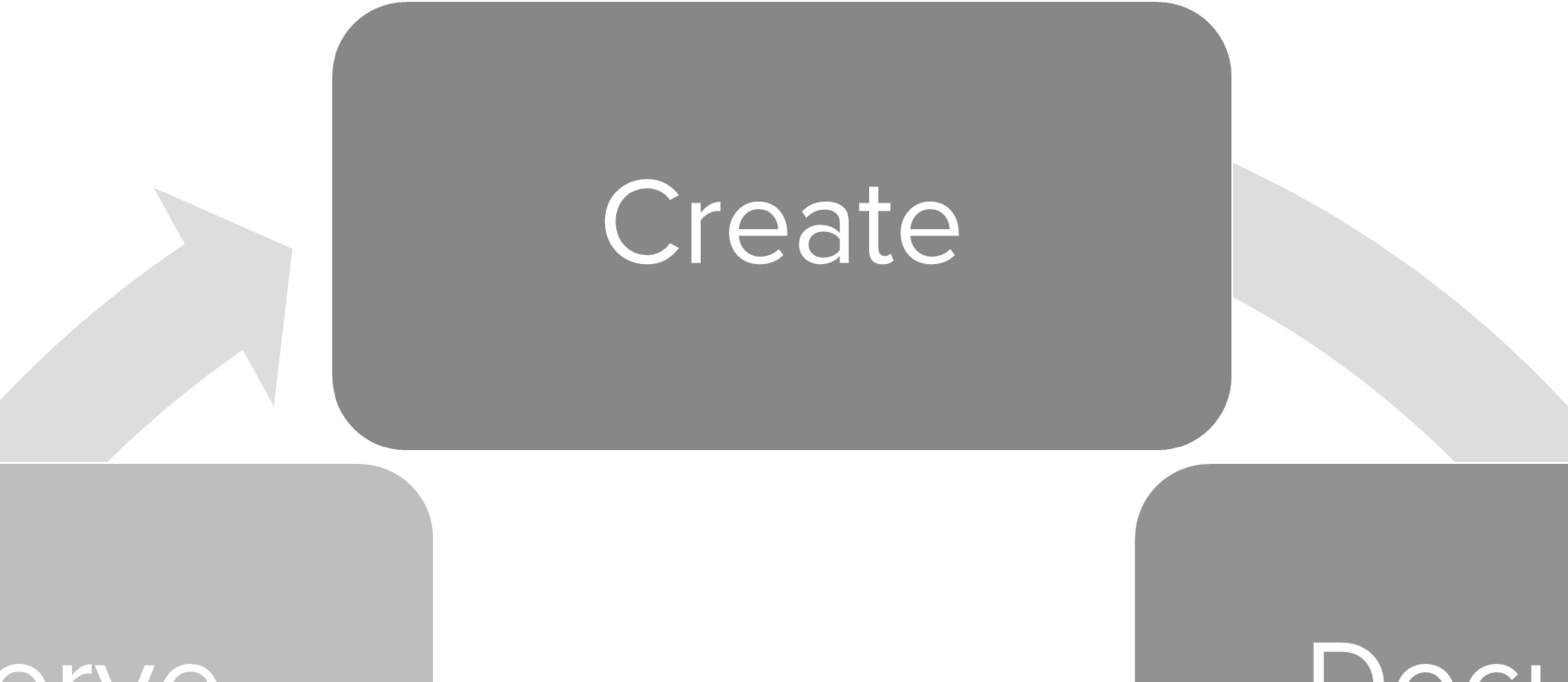
e-Research

Knowledge of data management and curati

Digital preservation

Data Schools

RDM & the Data Lifecycle



Create

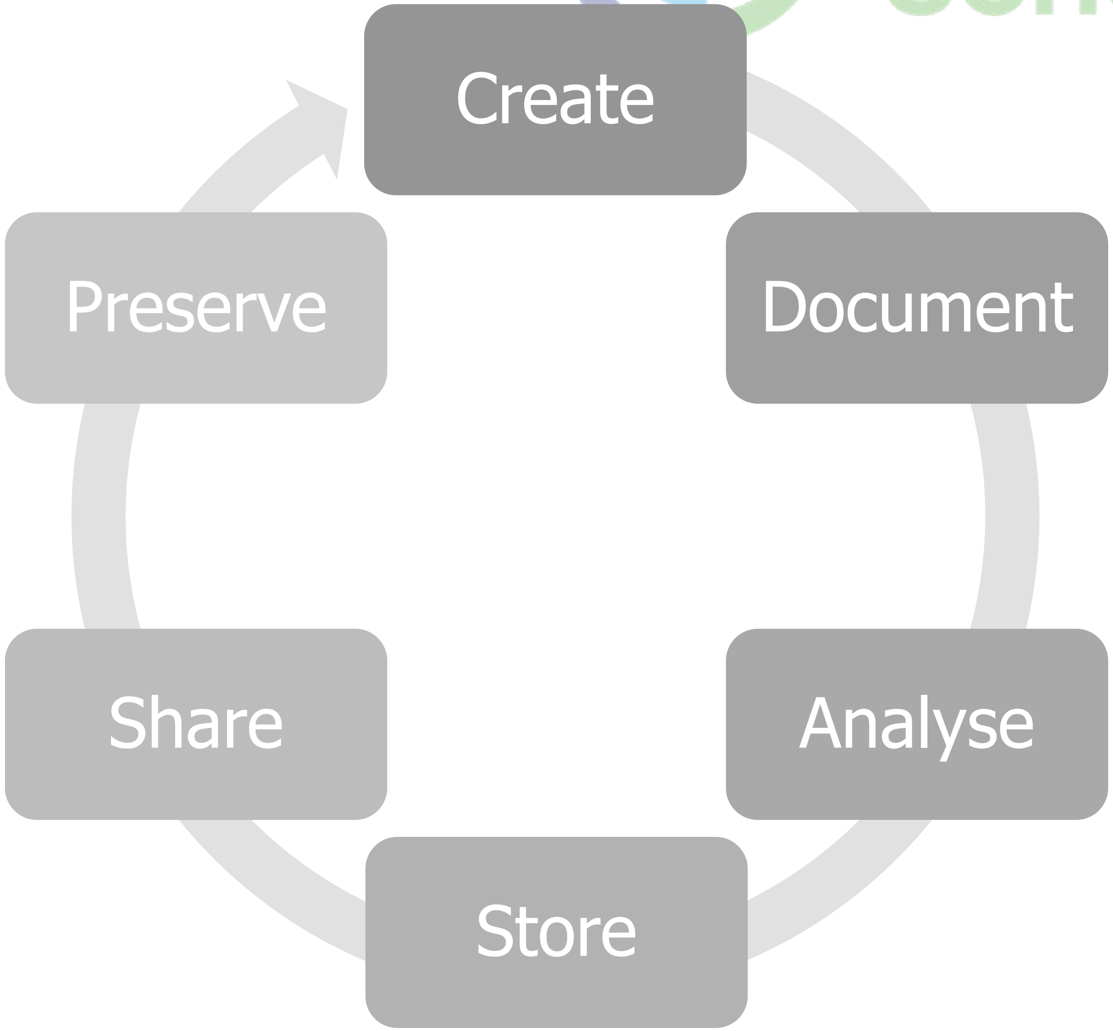
erve

Dec

What is Research Data Management?

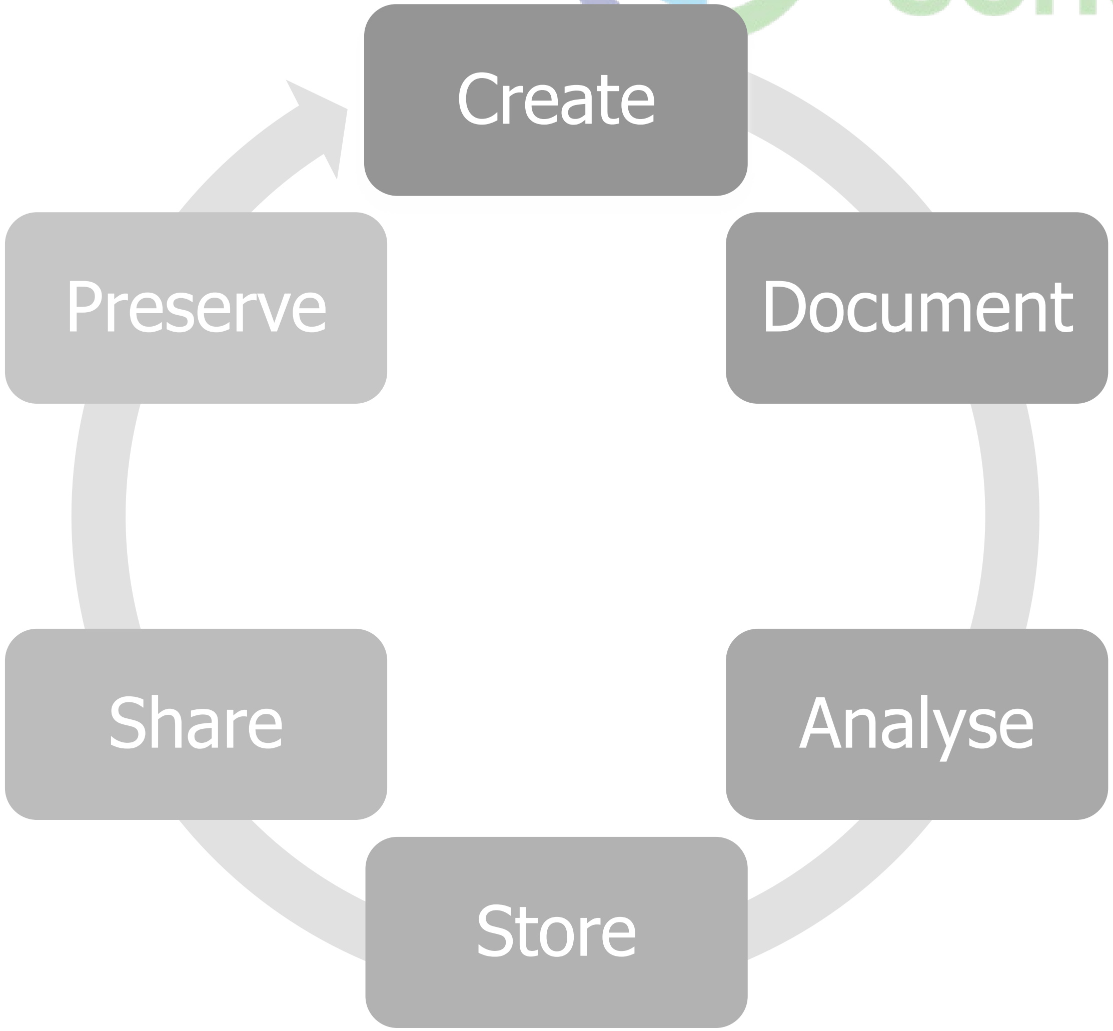
“the active management and appraisal of data over the lifecycle of scholarly and scientific interest”

Data management is part of good research practice.



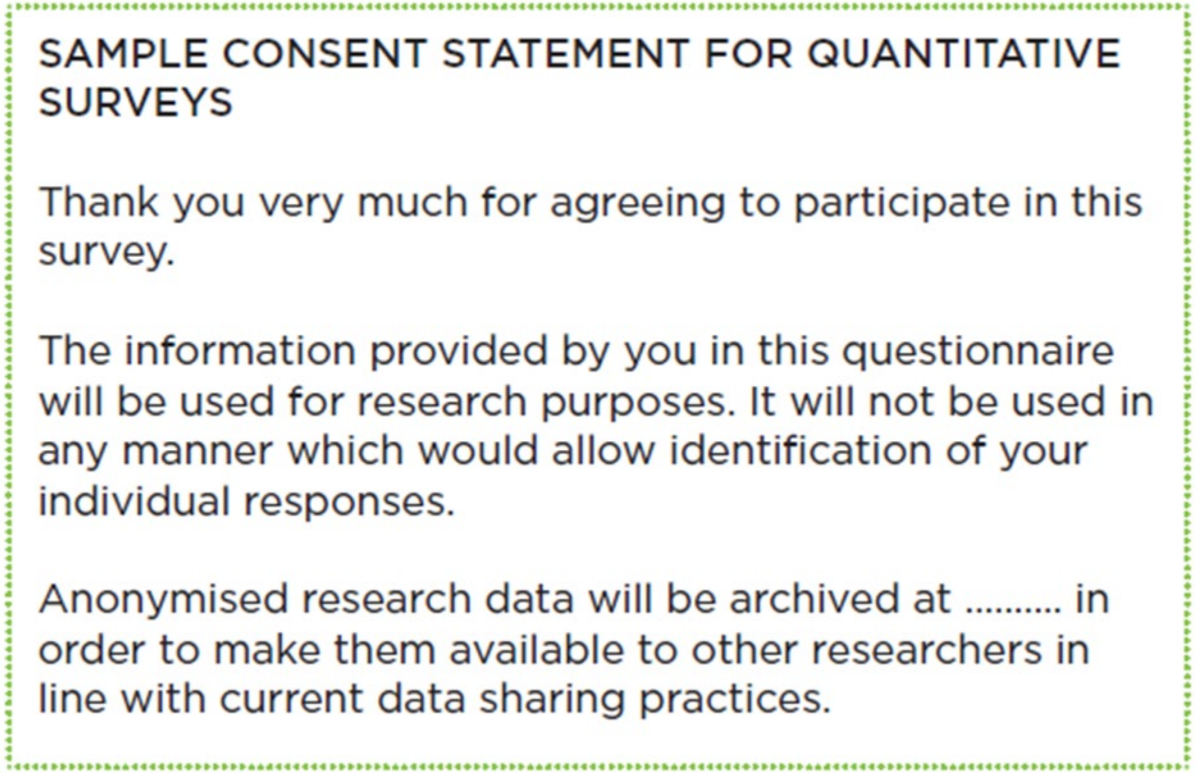
Data creation tips

- Ensure consent forms, licences and agreements don’t restrict opportunities to share data.
- Choose appropriate formats.
- Adopt a file naming convention.
- Create metadata and documentation as you go.



Ask for consent for data sharing

If not, data centres won’t be able to accept the data – regardless of any conditions on the original grant.



SAMPLE CONSENT STATEMENT FOR QUANTITATIVE SURVEYS

Thank you very much for agreeing to participate in this survey.

The information provided by you in this questionnaire will be used for research purposes. It will not be used in any manner which would allow identification of your individual responses.

Anonymised research data will be archived at .... in order to make them available to other researchers in line with current data sharing practices.

Choose appropriate file formats

- Different formats are good for different things.
- *open*, *lossless* formats are more sustainable e.g. rtf, xml, tif, wav.
- proprietary and/or compressed formats are less preservable but are often in widespread use e.g. doc, jpg, mp3.
- One format for analysis then convert to a standard format.
- Data centres may suggest preferred formats for deposit.

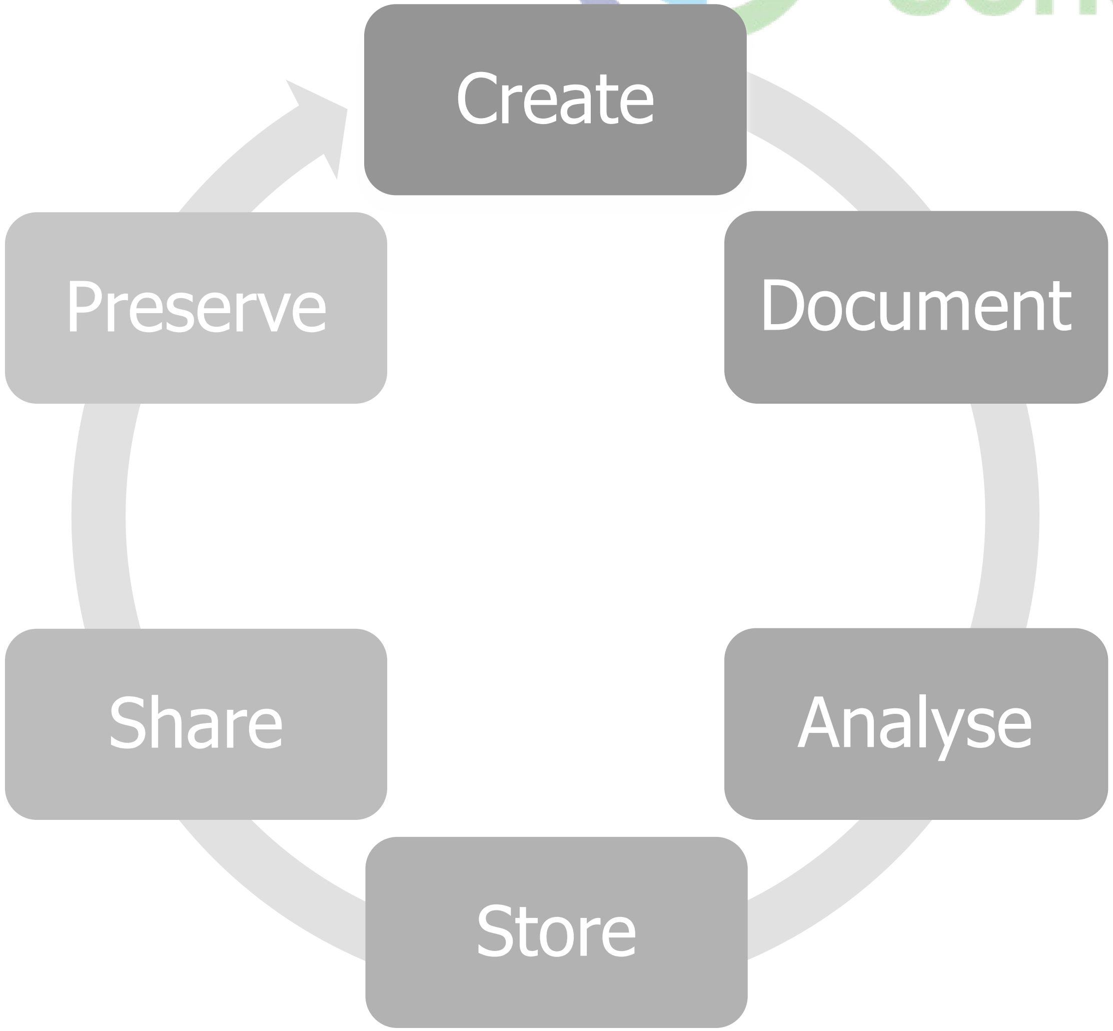
Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages
Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0

https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats

Documentation

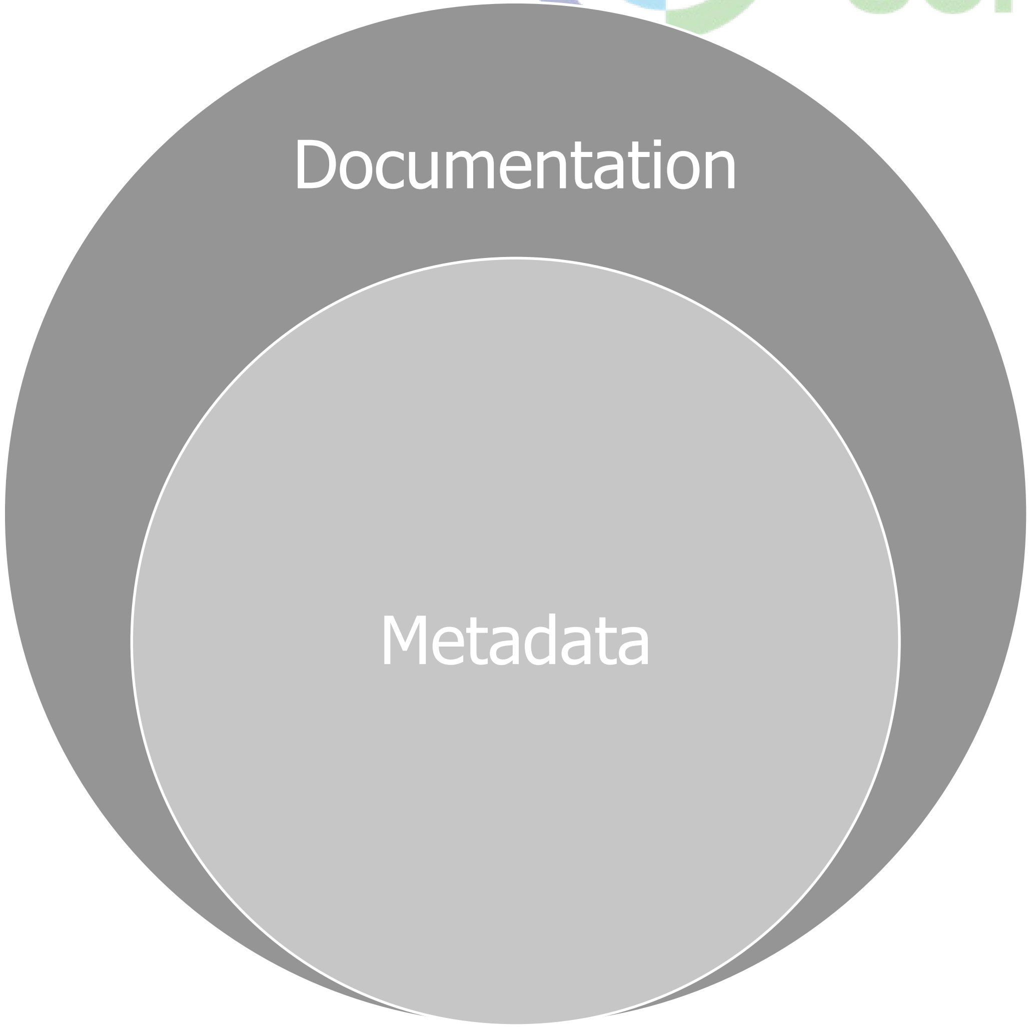
Think about what is needed in order to evaluate, understand, and reuse the data.

- Why was the data created?
- Have you documented what you did and how?
- Did you develop code to run analyses? If so, this should be kept and shared too.
- Important to provide wider context for trust.



What are metadata?

- Metadata
  - Standardised
  - Structured
  - Machine and human readable
- Metadata helps to cite and disambiguate data.
- Documentation aids reuse.



Documentation

Metadata

Metadata standards

These can be general – such as Dublin Core

Or discipline specific:

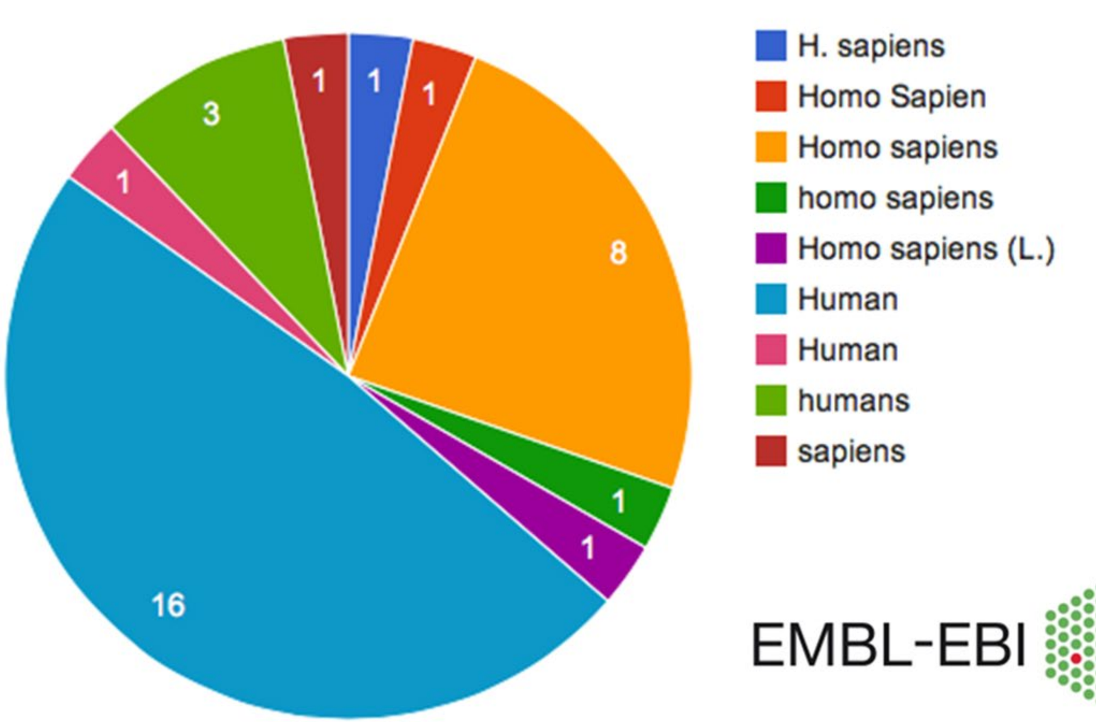
- Data Documentation Initiative (DDI) – social science
- Ecological Metadata Language (EML) - ecology
- Flexible Image Transport System (FITS) – astronomy

Search for standards in catalogues like:

- http://rd-alliance.github.io/metadata-directory/
- https://rdamsc.dcc.ac.uk/

Controlled vocabularies

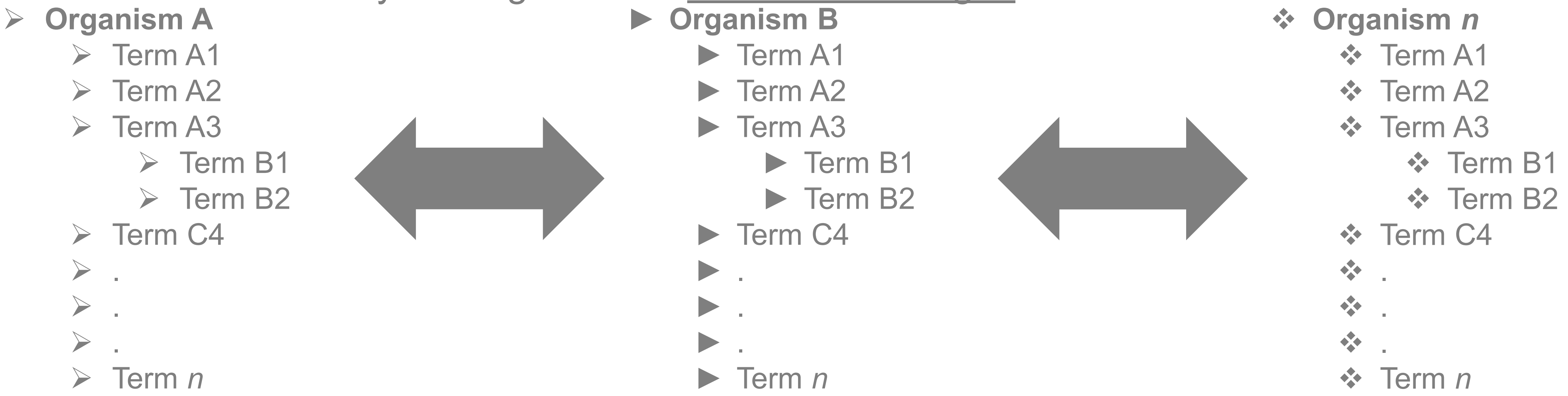
“MTBLS1: A metabolomic study of urinary changes in type 2 diabetes in……”



Example courtesy of Ken Haug, European Bioinformatics Institute (EMBL-EBI)

...and ontologies?

- e.g. SNOMED CT (clinical terms) or MeSH
- Defined terms + taxonomy.
- Useful for selecting keywords to tag datasets.
- You can find many ontologies in the BARTOC catalogue and elsewhere.



- **Organism A**
  - Term A1
  - Term A2
  - Term A3
    - Term B1
    - Term B2
  - Term C4
  - .
  - .
  - .
  - Term *n*

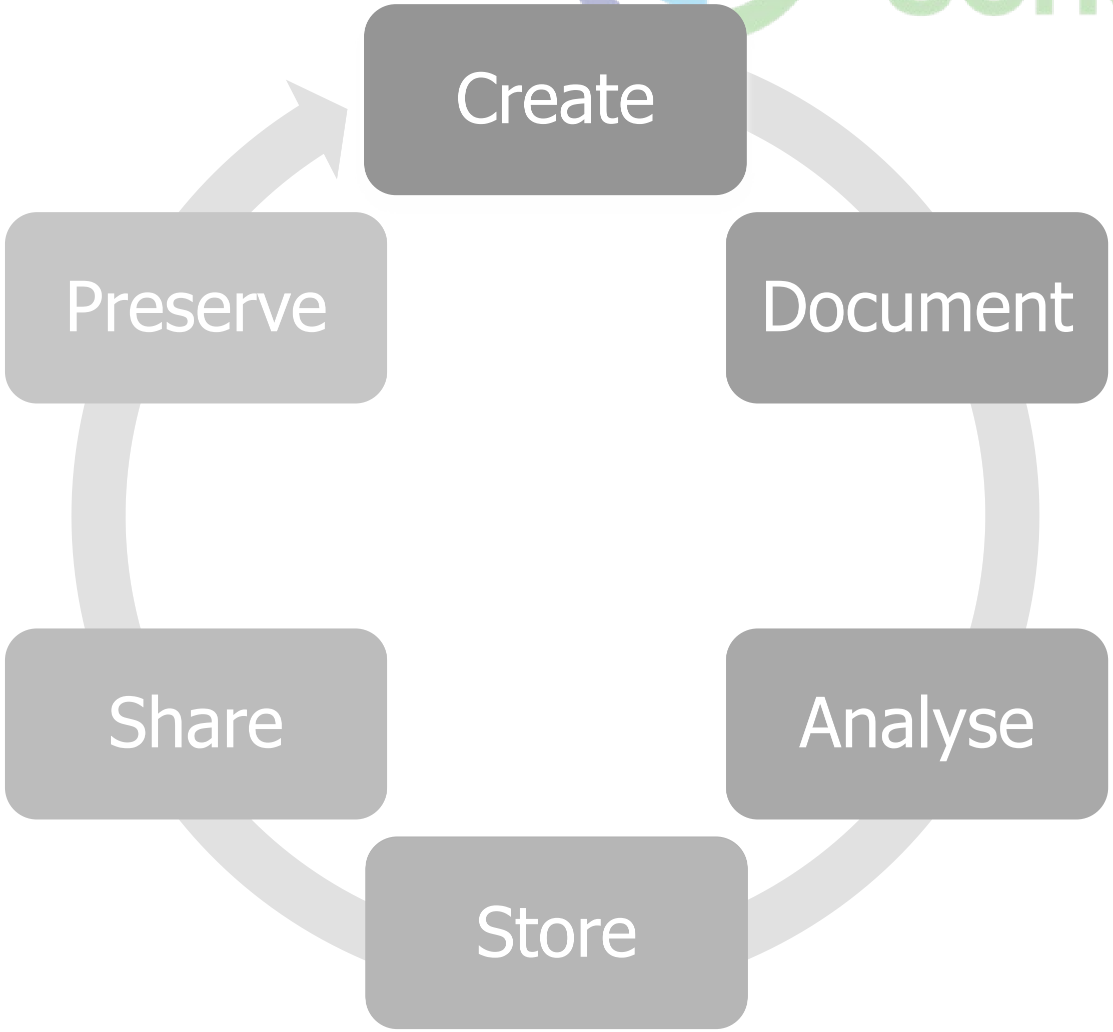
- **Organism B**
  - Term A1
  - Term A2
  - Term A3
    - Term B1
    - Term B2
  - Term C4
  - .
  - .
  - .
  - Term *n*

- **Organism *n***
  - Term A1
  - Term A2
  - Term A3
    - Term B1
    - Term B2
  - Term C4
  - .
  - .
  - .
  - Term *n*

Where will you store the data?

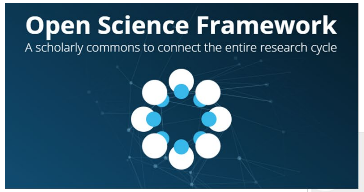
- Your own device (laptop, flash drive, server etc.)
  - And if you lose it? Or it breaks?
- Departmental drives or university servers.
- “Cloud” storage.
- Do they care as much about your data?

The decision will be based on how sensitive your data are, how robust you need the storage to be, and who needs access to the data and when.



Collaborative platforms and third-party tools

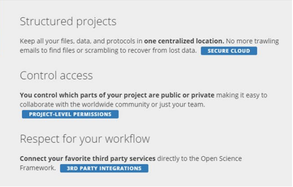
- OSF - open platform for sharing data in active phase with fellow researchers and others in secure environment.
- Third-party - commercial (e.g. Dropbox, G Drive, OneDrive) or open source (e.g. ownCloud)



Open Science Framework

A scholarly commons to connect the entire research cycle

https://osf.io



Structured projects

Keep all your files, data, and protocols in **one centralized location.** No more trawling emails to find files or scrambling to recover from lost data. SECURE CLOUD

Control access

You control which parts of your project are public or private making it easy to collaborate with the worldwide community or just your team. PROJECT-LEVEL PERMISSIONS

Respect for your workflow

Connect your favorite third party services directly to the Open Science Framework. 3RD PARTY INTEGRATIONS

https://owncloud.org

Backup vs. preservation

Backups

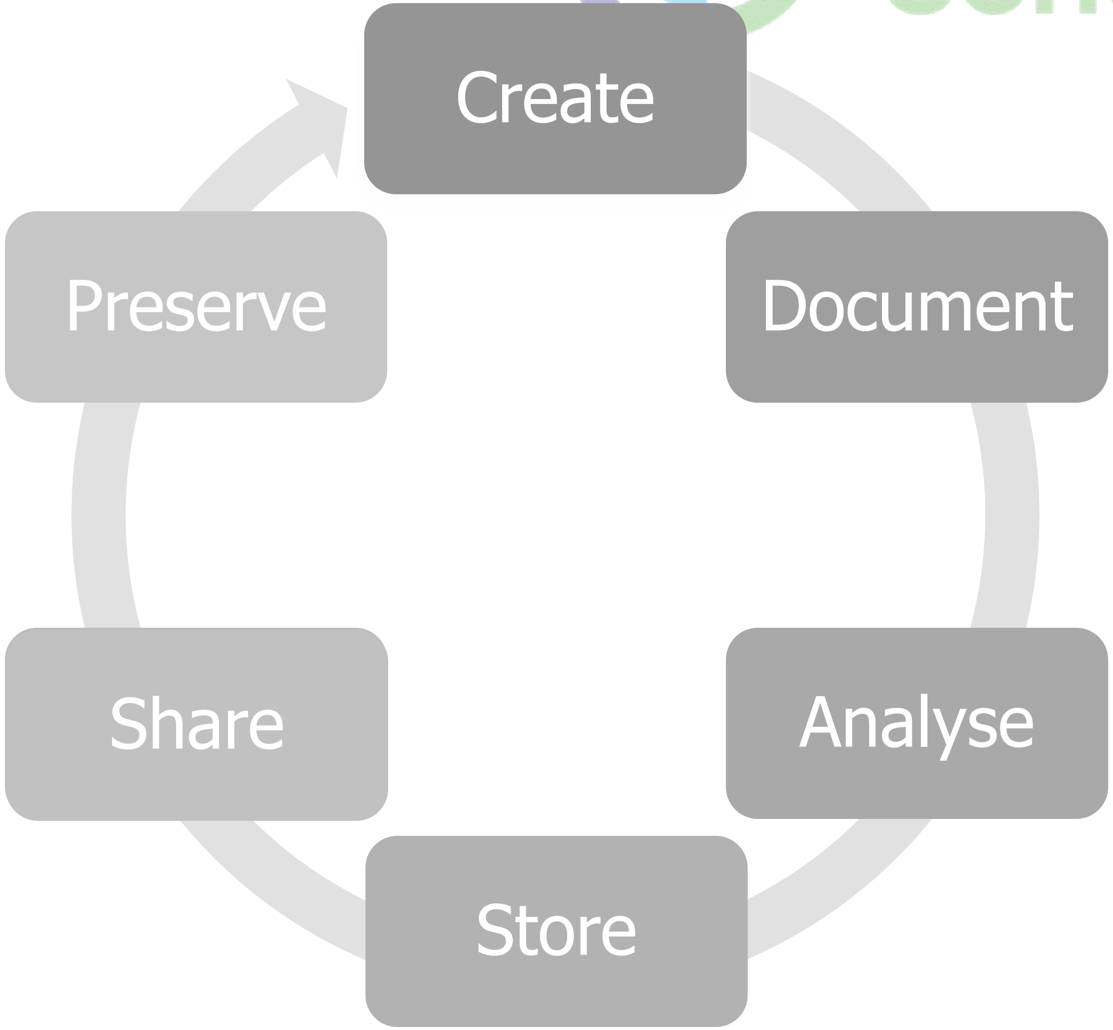
- Used to take periodic snapshots of data in case the current version is destroyed or lost.
- Backups are copies of files stored for short or near-long-term.
- Often performed on a somewhat frequent schedule.

Archiving

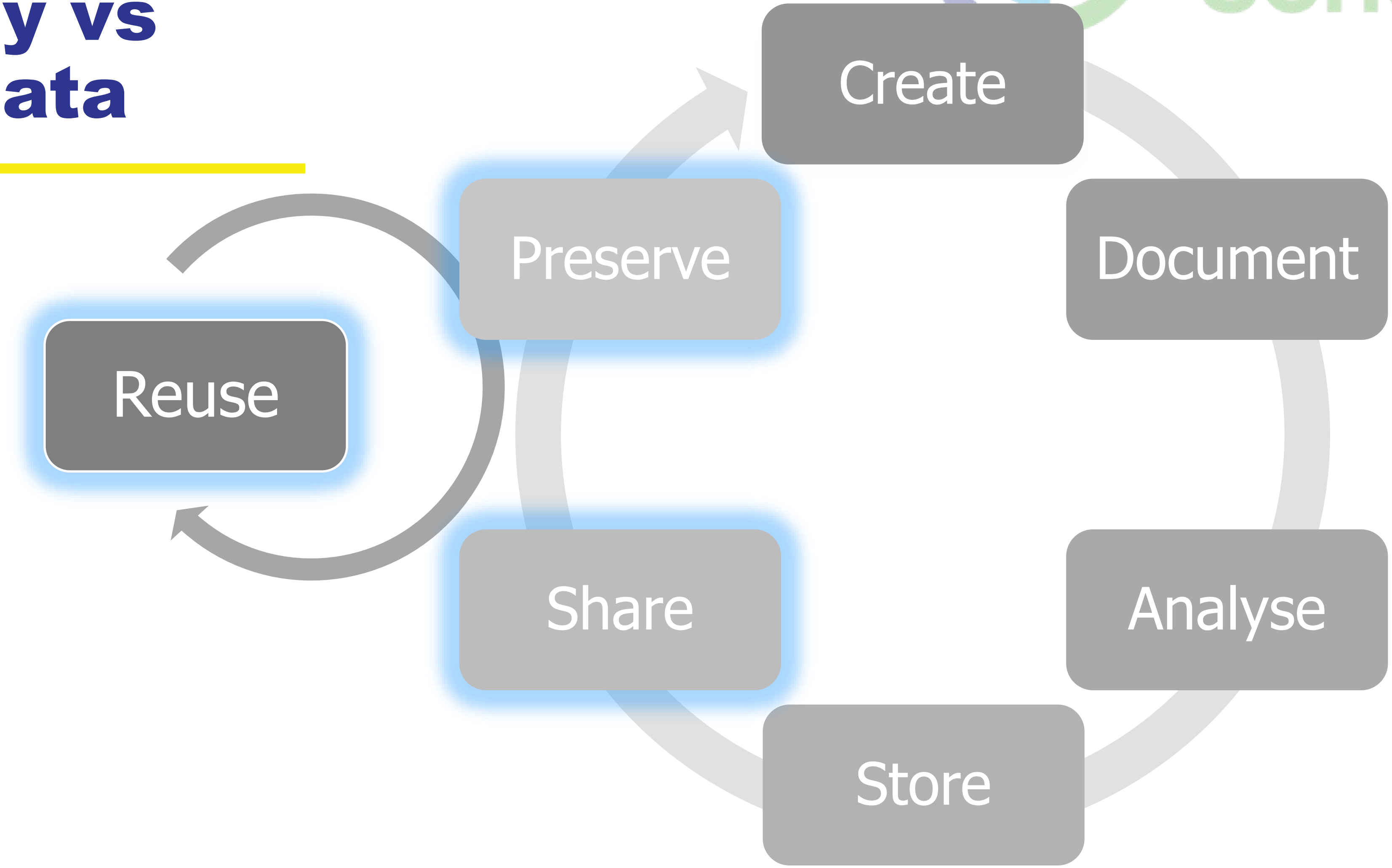
- Used to preserve data for historical reference or potentially during disasters.
- Archives are usually the final version, stored for long-term, and generally not copied over.
- Often performed at the end of a project or during major milestones.

How will you allow others to use your data?

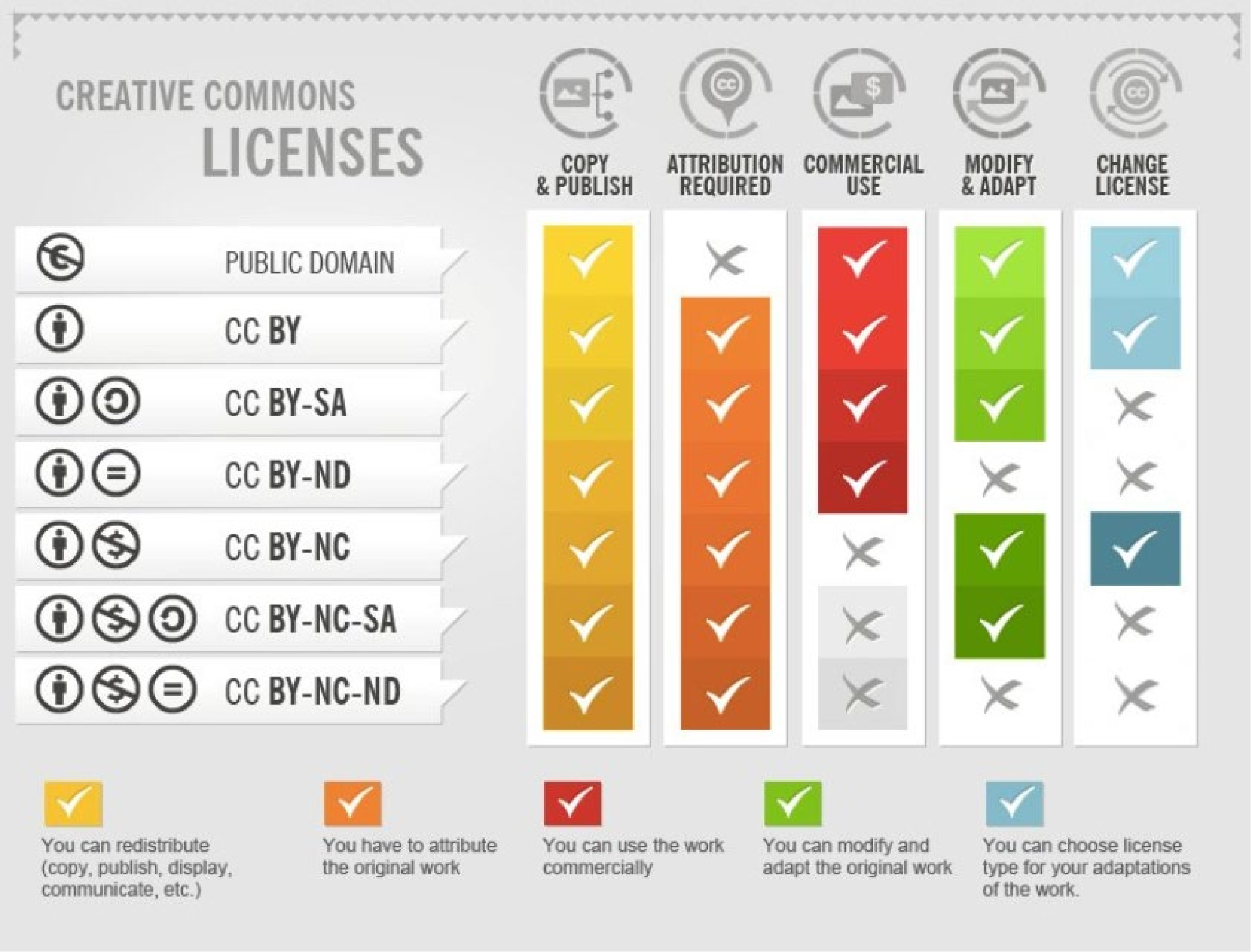
Apply licences to disambiguate reuse restrictions.



Secondary vs primary data



License research data openly



CREATIVE COMMONS LICENSES

	COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
PUBLIC DOMAIN	✓	✗	✓	✓	✓
CC BY	✓	✓	✓	✓	✓
CC BY-SA	✓	✓	✓	✓	✗
CC BY-ND	✓	✓	✓	✗	✗
CC BY-NC	✓	✓	✗	✓	✓
CC BY-NC-SA	✓	✓	✗	✓	✗
CC BY-NC-ND	✓	✓	✗	✗	✗

- 🟨✓ You can redistribute (copy, publish, display, communicate, etc.)
- 🟧✓ You have to attribute the original work
- 🟥✓ You can use the work commercially
- 🟩✓ You can modify and adapt the original work
- 🟦✓ You can choose license type for your adaptations of the work.

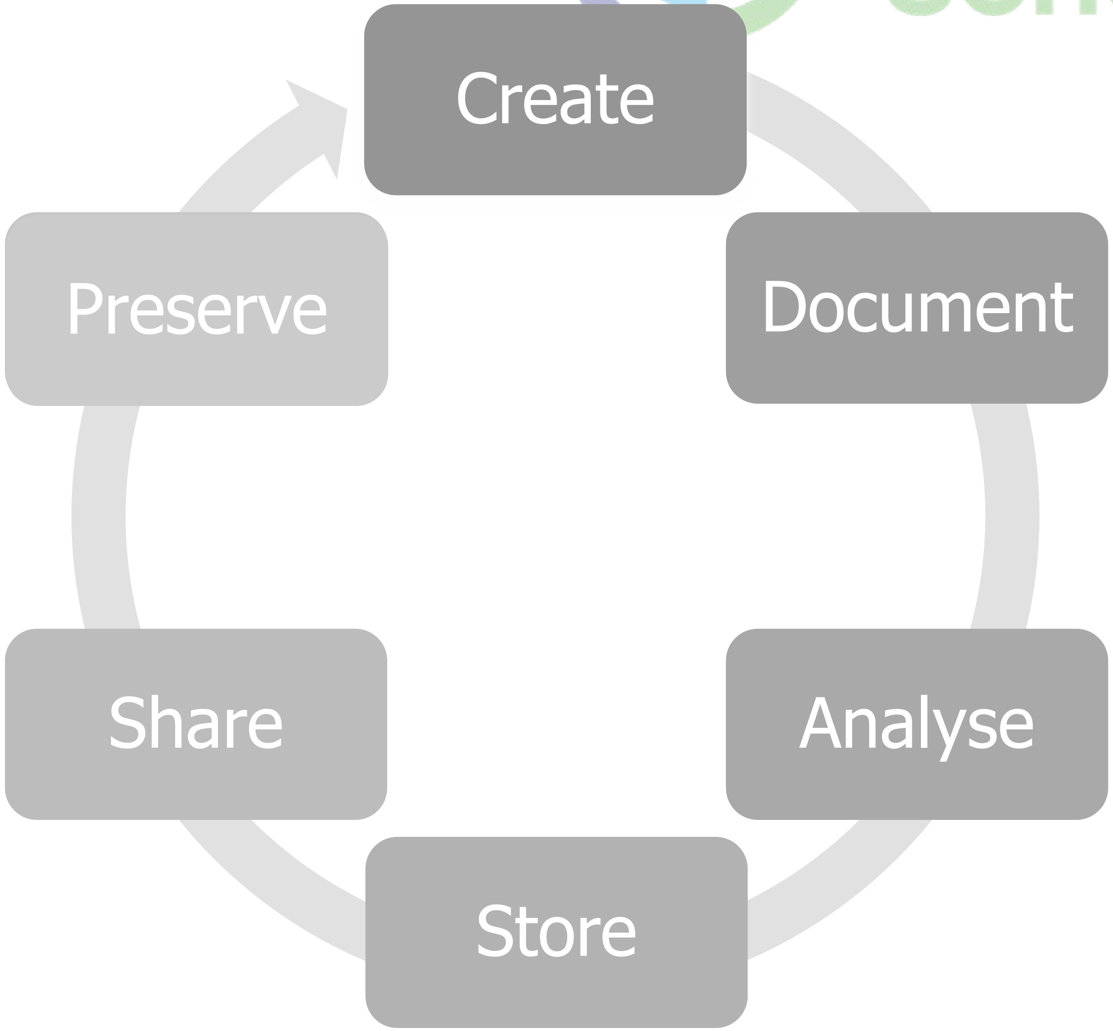
Part of How To Attribute Creative Commons Photos by Foter, licensed CC BY SA 3.0

Try the EUDAT online licence wizard:

https://ufal.github.io/public-license-selector/

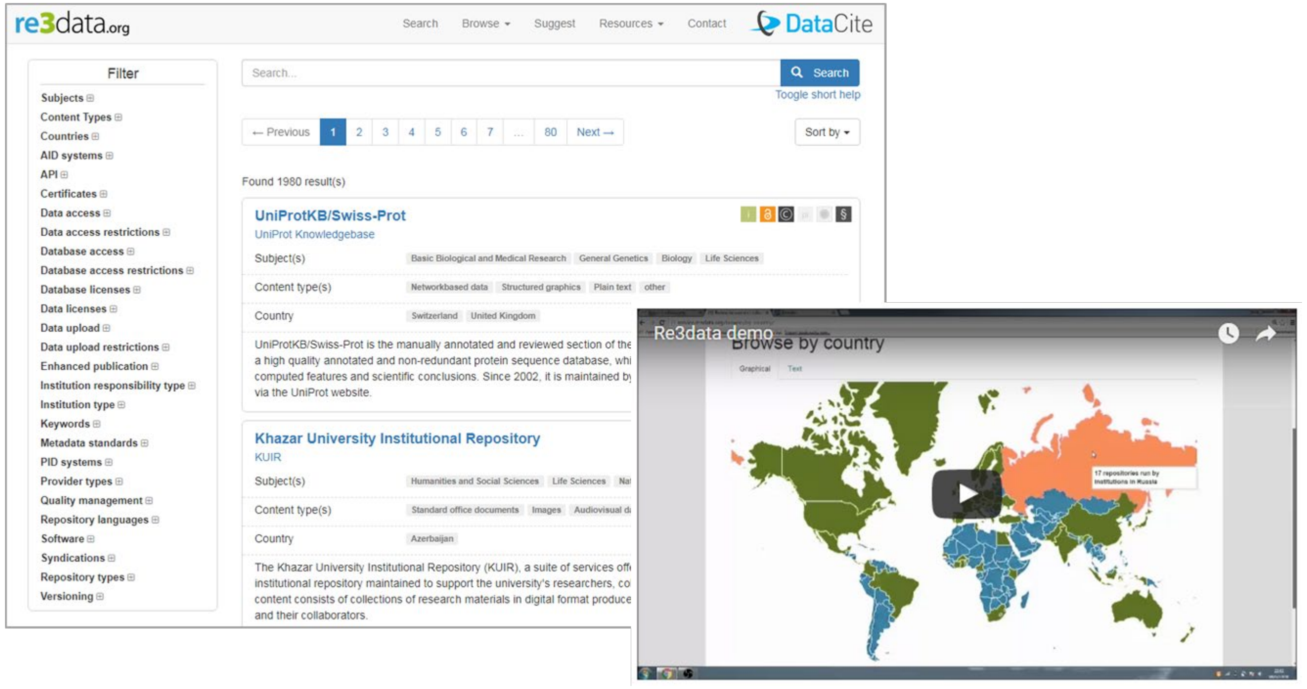
Deposit in a data repository

Long-term preservation of data.



Deposit in a data repository

- The Re3data catalogue can be searched to find a home for data.
- www.fosteropenscience.eu/content/re3data-demo
- Better to use a domain specific repository if available.
- Check they match particular data needs e.g. formats accepted, mixture of Open and Restricted Access.
- Do they assign a persistent and globally unique identifier for sustainable citations and to links back to particular researchers and grants?
- Look for certification as a '*Trustworthy Digital Repository*' with an explicit ambition to keep the data available in long term.



Found 1980 result(s)

UniProtKB/Swiss-Prot
UniProt Knowledgebase

Subject(s): Basic Biological and Medical Research, General Genetics, Biology, Life Sciences

Content type(s): Networkbased data, Structured graphics, Plain text, other

Country: Switzerland, United Kingdom

UniProtKB/Swiss-Prot is the manually annotated and reviewed section of the
a high quality annotated and non-redundant protein sequence database, whi
computed features and scientific conclusions. Since 2002, it is maintained by
via the UniProt website.

Khazar University Institutional Repository
KUIR

Subject(s): Humanities and Social Sciences, Life Sciences, Na

Content type(s): Standard office documents, Images, Audiovisual d

Country: Azerbaijan

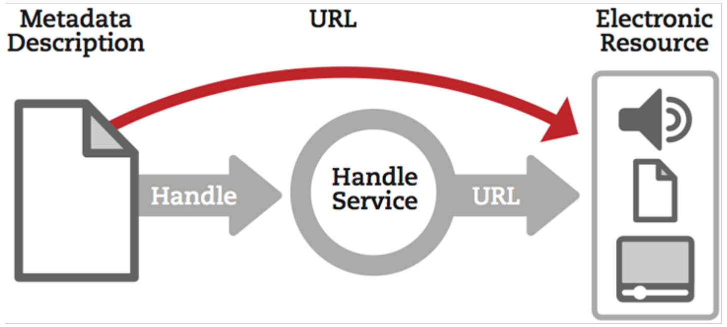
The Khazar University Institutional Repository (KUIR), a suite of services off
institutional repository maintained to support the university's researchers, co
content consists of collections of research materials in digital format produce
and their collaborators.

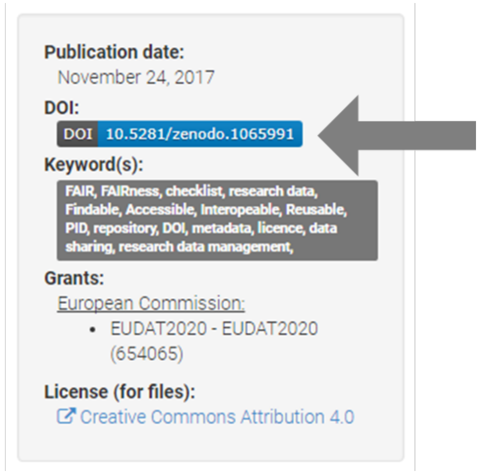
www.re3data.org

What is a Persistent Identifier (PID)?

a long-lasting reference to a document, file or other object

- PIDs come in various forms e.g. ORCID, DOI, ISBN...
- Typically they’re actionable i.e. type it into web browser to access.
- Many repositories will assign them on deposit.
- Important for **provenance**.




Publication date:
November 24, 2017

DOI:
DOI 10.5281/zenodo.1065991

Keyword(s):
FAIR, FAIRness, checklist, research data, Findable, Accessible, Interopeable, Reusable, PID, repository, DOI, metadata, licence, data sharing, research data management,

Grants:
European Commission:
- EUDAT2020 - EUDAT2020 (654065)

License (for files):
Creative Commons Attribution 4.0

www.re3data.org

Data Schools

The FAIR Principles

Global efforts for alignment

Findable **A**ccessible **I**nteroperable **R**eusable

Findable

- Metadata
- PIDs
- Repositories

Accessible

- Metadata
- Open file formats and software

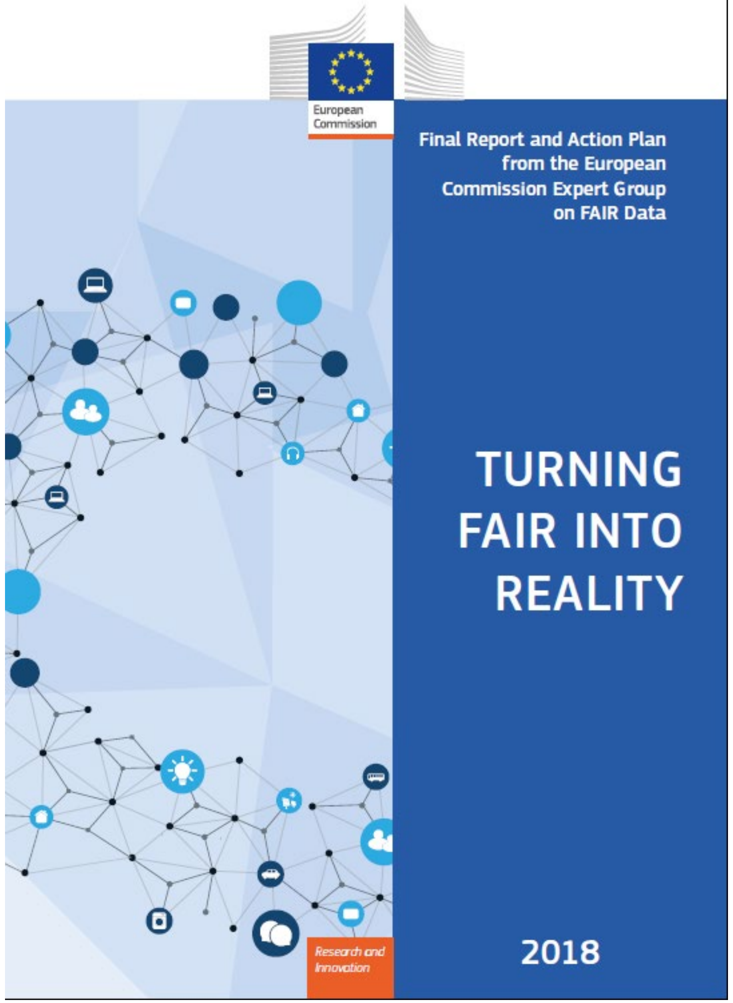
Interoperable

- Metadata
- Ontologies
- Repositories

Reusable

- Metadata
- Licences

European perspective...



European Commission

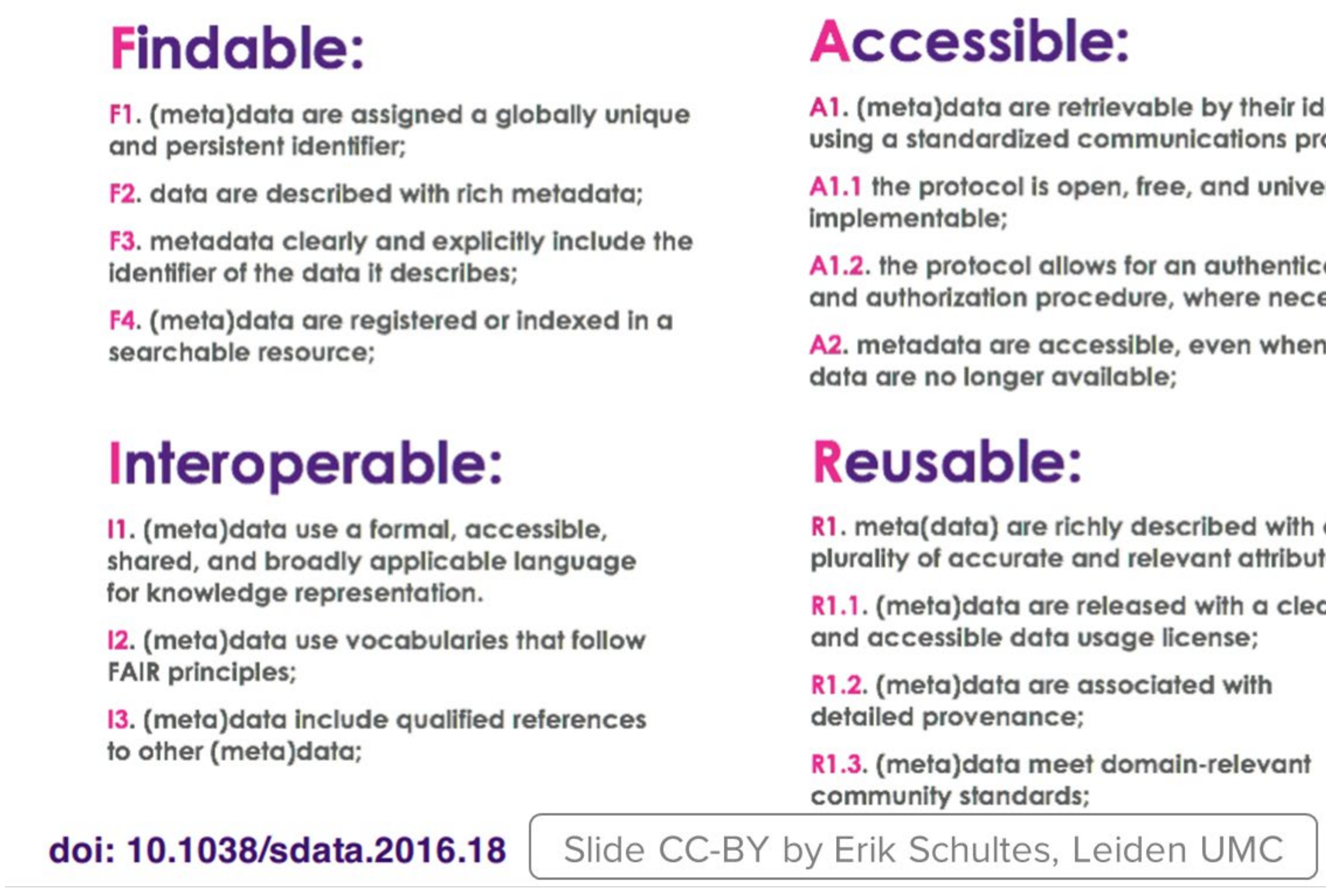
Final Report and Action Plan from the European Commission Expert Group on FAIR Data

TURNING FAIR INTO REALITY

2018

https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283

What FAIR means: 15 principles



Findable:

F1. (meta)data are assigned a globally unique and persistent identifier;

F2. data are described with rich metadata;

F3. metadata clearly and explicitly include the identifier of the data it describes;

F4. (meta)data are registered or indexed in a searchable resource;

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;

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;

Reusable:

R1. meta(data) are richly described with 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 detailed provenance;

R1.3. (meta)data meet domain-relevant community standards;

doi: 10.1038/sdata.2016.18

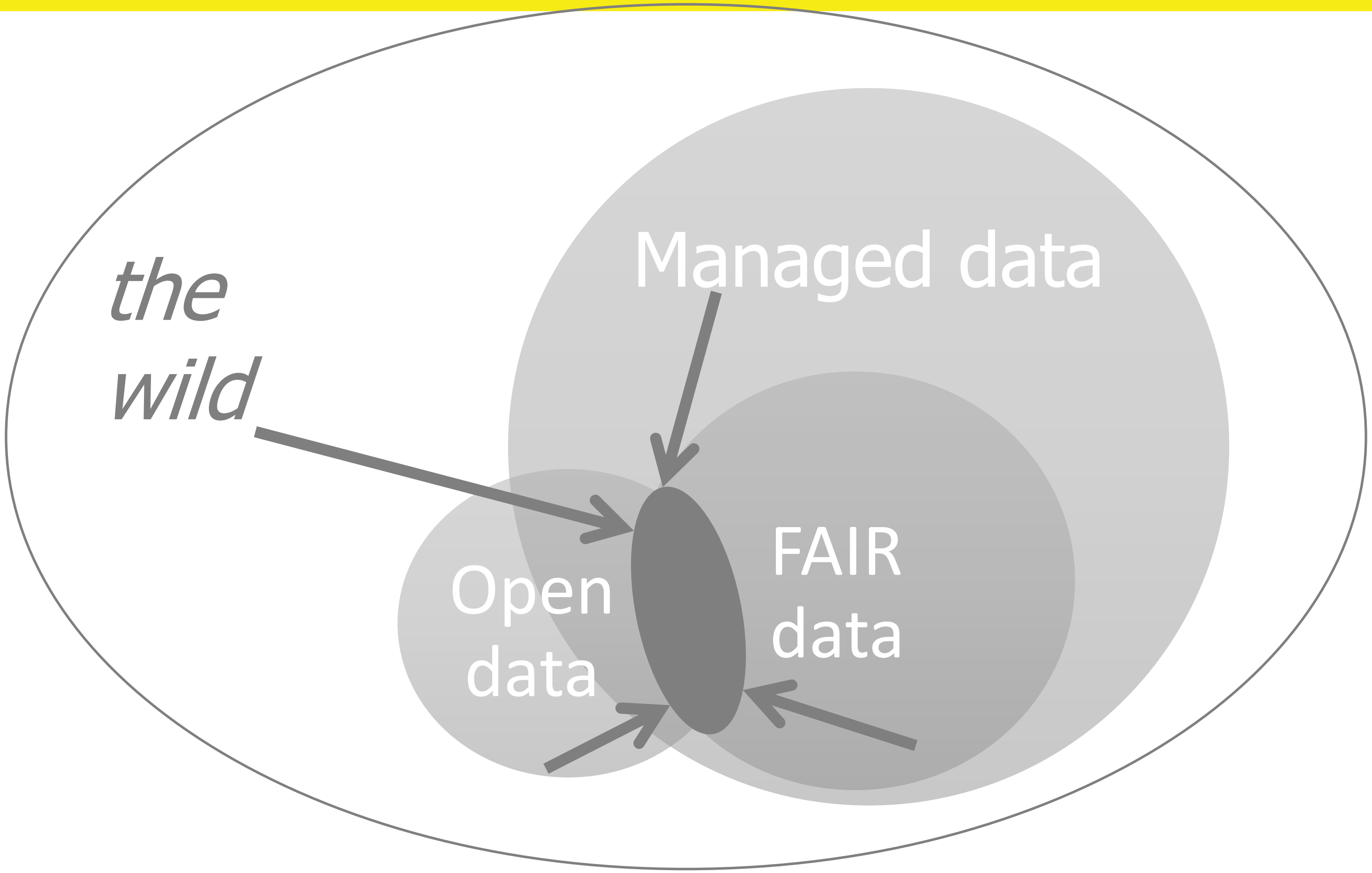
Slide CC-BY by Erik Schultes, Leiden UMC

Comprehensive descriptions can be found at
https://www.go-fair.org/fair-principles/

Common misconceptions

- FAIR data does not have to be open.
- The principles do not specify particular technologies or implementations e.g. semantic web.
- FAIR is not a standard to be followed or strict criteria – it’s a spectrum/continuum.
- It doesn’t only apply to the life sciences.

Increasing that which is FAIR & open



FAIR ≠ Open

as open as possible, as closed as necessary

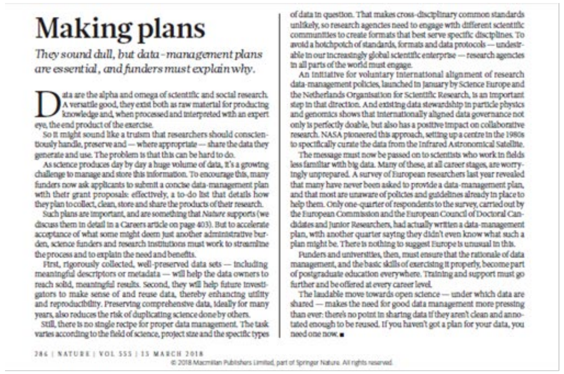
Image: ‘Balancing rocks’ by Viewminder CC-BY-SA-ND www.flickr.com/photos/light_seeker/7780857224

Data Schools

Data Management Plans

Bringing together what you’ve learnt

- Make informed decisions to anticipate and avoid problems.
- Avoid duplication, data loss and security breaches.
- Develop procedures early on for consistency.
- Ensure data are accurate, complete, reliable and secure.
- Save time and effort to make your life easier!
- Useful both to researchers and institutions



Making plans

They sound dull, but data-management plans are essential, and funders must explain why.

Data are the alpha and omega of scientific and social research. A versatile good, they exist both as raw material for producing knowledge and, when processed and interpreted with an expert eye, the end product of the exercise.

So it might sound like a truism that researchers should conscientiously handle, preserve and — where appropriate — share the data they generate and use. The problem is that this can be hard to do.

As science produces day by day a huge volume of data, it's a growing challenge to manage and store this information. To encourage this, many funders now ask applicants to submit a concise data-management plan with their grant proposals: effectively, a to-do list that details how they plan to collect, clean, store and share the products of their research.

Such plans are important, and are something that *Nature* supports (we discuss them in detail in a Careers article on page 403). But to accelerate acceptance of what some might deem just another administrative burden, science funders and research institutions must work to streamline the process and to explain the need and benefits.

First, rigorously collected, well-preserved data sets — including meaningful descriptors or metadata — will help the data owners to reach solid, meaningful results. Second, they will help future investigators to make sense of and reuse data, thereby enhancing utility and reproducibility. Preserving comprehensive data, ideally for many years, also reduces the risk of duplicating science done by others.

Still, there is no single recipe for proper data management. The task varies according to the field of science, project size and the specific types of data in question. That makes cross-disciplinary common standards unlikely, so research agencies need to engage with different scientific communities to create formats that best serve specific disciplines. To avoid a hotchpotch of standards, formats and data protocols — undesirable in our increasingly global scientific enterprise — research agencies in all parts of the world must engage.

An initiative for voluntary international alignment of research data-management policies, launched in January by Science Europe and the Netherlands Organisation for Scientific Research, is an important step in that direction. And existing data stewardship in particle physics and genomics shows that internationally aligned data governance not only is perfectly doable, but also has a positive impact on collaborative research. NASA pioneered this approach, setting up a centre in the 1980s to specifically curate the data from the Infrared Astronomical Satellite.

The message must now be passed on to scientists who work in fields less familiar with big data. Many of these, at all career stages, are worryingly unprepared. A survey of European researchers last year revealed that many have never been asked to provide a data-management plan, and that most are unaware of policies and guidelines already in place to help them. Only one-quarter of respondents to the survey, carried out by the European Commission and the European Council of Doctoral Candidates and Junior Researchers, had actually written a data-management plan, with another quarter saying they didn't even know what such a plan might be. There is nothing to suggest Europe is unusual in this.

Funders and universities, then, must ensure that the rationale of data management, and the basic skills of exercising it properly, become part of postgraduate education everywhere. Training and support must go further and be offered at every career level.

The laudable move towards open science — under which data are shared — makes the need for good data management more pressing than ever: there's no point in sharing data if they aren't clean and annotated enough to be reused. If you haven't got a plan for your data, you need one now. ■

ADAPTED FROM GETTY

CAREERS

PERSONAL ETHICS How a vegetarian biologist balances his beliefs with his work **p.405** | **BLOG** Personal stories and careers counsel **http://blogs.nature.com/naturejobs** | **NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**

DATA MANAGEMENT

For the record

Making project data freely available is vital for open science.

BY QUIRIN SCHIERMEIER

When Marjorie Etique learnt that she had to create a data-management plan for her next research project, she was not sure exactly what to do.

The soil chemist, a postdoc at the Swiss Federal Institute of Technology (ETH) in Zurich, studies the interaction of trace elements in sediments and water. While preparing a grant proposal for the Swiss National Science Foundation last October, she learnt of the funder's new data rules. These require applicants to provide a written plan for the organization and long-term storage of their research data, to help minimize the risk of data loss and provide guidance for other scientists on how to use the data in the future.

Etique found the task daunting. "Data management is really not my primary skill," she says. "I had absolutely no idea how to go about it." She was able to get advice from her supervisor and from ETH's digital library service. Other researchers might not be so lucky, and might not even know what a data-management plan is — let alone why they would need one and how to produce it. Here, we answer these questions.

WHAT ARE DATA-MANAGEMENT PLANS?

A data-management plan explains how researchers will handle their data during and after a project, and encompasses creating, sharing and preserving research data of any type, including text, spreadsheets, images, recordings, models, algorithms and software. It does not matter whether the data are generated by large pieces of research equipment, such as imaging tools or particle accelerators, or from straightforward field observation.

Many funders are asking grant applicants to provide data plans. Requirements vary from one discipline to another. But in general, scientists will need to describe — before they begin any research — what data they will generate; how the data will be documented, described, secured and curated; and who will have access to those data after the research is completed. They must also explain any data sharing and reuse restrictions, such as legal and confidentiality issues. Researchers can consult their funder and their host institute's digital library services for assistance. Colleagues who have previously produced data plans may also be able to help (see 'Keeping stock').

WHO NEEDS THEM?

Data management is one example of the way in which public research sponsors and research institutions are implementing 'open science', the push to make scientific research and data freely accessible. Many funding agencies have made data-management plans mandatory for grant applicants in the past decade or so. All US federal agencies, including the National Science Foundation and the National Institutes of Health, have such policies. Data-management plans must also now be included in grant proposals to the European Research Council and other European Union-funded research programmes. And many national funding agencies in Europe — including the UK research councils and the London-based Wellcome Trust, world's largest biomedical research charity — also ask for data plans.

Many scientists already practise data management by default. Astronomers, for example, have been doing so for decades when calibrating their observations and archiving huge amounts of telescope-survey data in standardized, machine-readable catalogues for reuse.

Geneticists, too, use special data repositories to archive the vast amounts of DNA and genome-sequencing data (see go.nature.com/2omlrbe). But less data-intensive fields of science and social research also benefit from data management. For example, geochemists analysing soil bacteria and mineral products in different environments can use it to ▶

Schiermeier, Q. “Data management made simple” *Nature* **555**, 403-405 (2018).
https://www.nature.com/articles/d41586-018-03071-1
doi: 10.1038/d41586-018-03071-1

DCC Checklist for a DMP

- The DCC assessed existing funder requirements, DMP templates and other best practice to see what should be included in plans. This was synthesised down into common themes and questions.
- 13 questions on what’s asked across the board.
- Prompts/pointers to help researchers get started.
- Guidance on how to answer.

DCC Checklist for a Data Management Plan

DCC

www.dcc.ac.uk/sites/default/files/documents/resource/DMP_Checklist_2013.pdf

Example plans

- Plans from several funders and disciplines via DCC www.dcc.ac.uk/resources/data-management-plans/guidance-examples

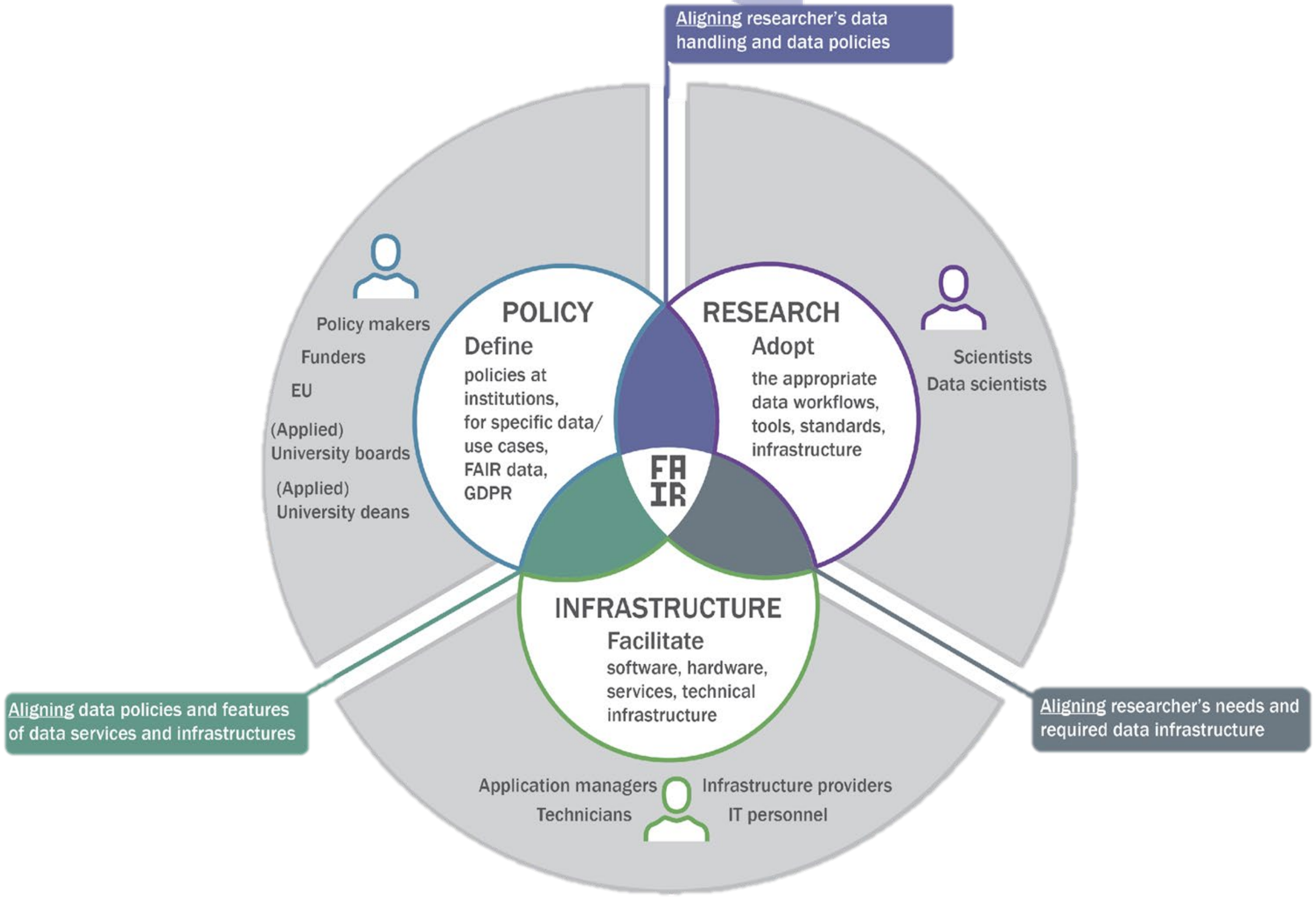
- Scientific DMPs submitted to the NSF (USA) provided by DataOne https://www.dataone.org/data-management-planning

- DMPs published in RIO journal http://riojournal.com/browse_user_collection_documents.php?collection_id=3&journal_id=17

- Share yours! - www.dcc.ac.uk/share-DMPs

The different roles of a data steward

- Unlike most other roles, data stewards traverse the researcher-service provider barrier.
- Need knowledge from both perspectives.
- Three key areas requiring data steward training: **Policy**, **Research** and **Infrastructure**.
- **At the heart of these are the FAIR principles.**
- NB. For **Infrastructure**, please refer to e.g. the RISE tool.



Aligning researcher's data handling and data policies
Policy makers
Funders
EU
(Applied) University boards
(Applied) University deans
POLICY
Define
policies at institutions, for specific data/ use cases, FAIR data, GDPR
RESEARCH
Adopt
the appropriate data workflows, tools, standards, infrastructure
Scientists
Data scientists
FA IR
INFRASTRUCTURE
Facilitate
software, hardware, services, technical infrastructure
Aligning data policies and features of data services and infrastructures
Aligning researcher's needs and required data infrastructure
Application managers
Technicians
Infrastructure providers
IT personnel

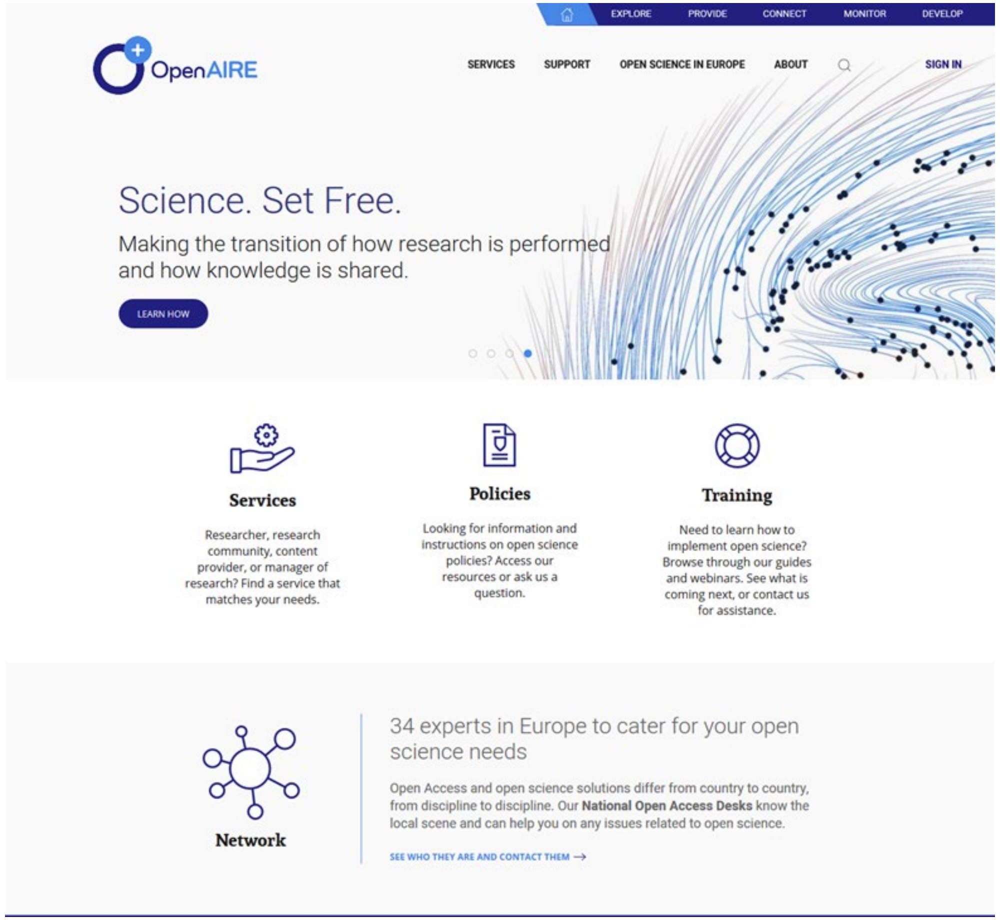
Staiger et al. *Data stewards function landscape and its stakeholders.* (2019) Zenodo. http://doi.org/10.5281/zenodo.3460552

FOSTER Open Science

What is Open Science?	Best Practice in Open Research	Open Access Publishing	Open Peer Review	Sharing Preprints
Data Protection & Ethics	Open Source Software & Workflows	Managing & Sharing Research Data	Open Science & Innovation	Open Licensing

https://www.fosteropenscience.eu/toolkit

OpenAIRE



Science. Set Free.

Making the transition of how research is performed and how knowledge is shared.

Services

Researcher, research community, content provider, or manager of research? Find a service that matches your needs.

Policies

Looking for information and instructions on open science policies? Access our resources or ask us a question.

Training

Need to learn how to implement open science? Browse through our guides and webinars. See what is coming next, or contact us for assistance.

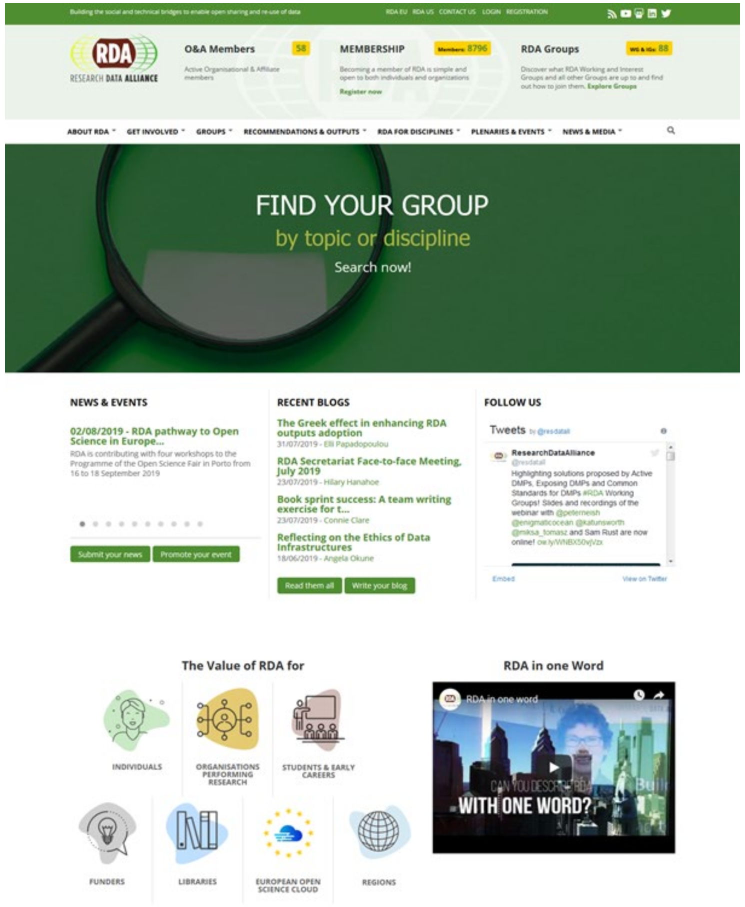
Network

34 experts in Europe to cater for your open science needs

Open Access and open science solutions differ from country to country, from discipline to discipline. Our **National Open Access Desks** know the local scene and can help you on any issues related to open science.

https://www.openaire.eu/

Research Data Alliance



O&A Members 58

Active Organisational & Affiliate members

MEMBERSHIP Members: 8796

Becoming a member of RDA is simple and open to both individuals and organizations

Register now

RDA Groups WG & IGs: 88

Discover what RDA Working and Interest Groups and all other Groups are up to and find out how to join them. Explore Groups

FIND YOUR GROUP

by topic or discipline

Search now!

NEWS & EVENTS

02/08/2019 - RDA pathway to Open Science in Europe...

RDA is contributing with four workshops to the Programme of the Open Science Fair in Porto from 16 to 18 September 2019

RECENT BLOGS

The Greek effect in enhancing RDA outputs adoption
31/07/2019 - Elli Papadopoulou

RDA Secretariat Face-to-face Meeting, July 2019
23/07/2019 - Hilary Hanahoe

Book sprint success: A team writing exercise for t...
23/07/2019 - Connie Clare

Reflecting on the Ethics of Data Infrastructures
18/06/2019 - Angela Okune

FOLLOW US

Tweets by @resdatall

ResearchDataAlliance
@resdatall
Highlighting solutions proposed by Active DMPs, Exposing DMPs and Common Standards for DMPs #RDA Working Groups! Slides and recordings of the webinar with @peterneish @enigmaticocean @katunsworth @miksa_tomasz and Sam Rust are now online! ow.ly/WNBX50vjVzx

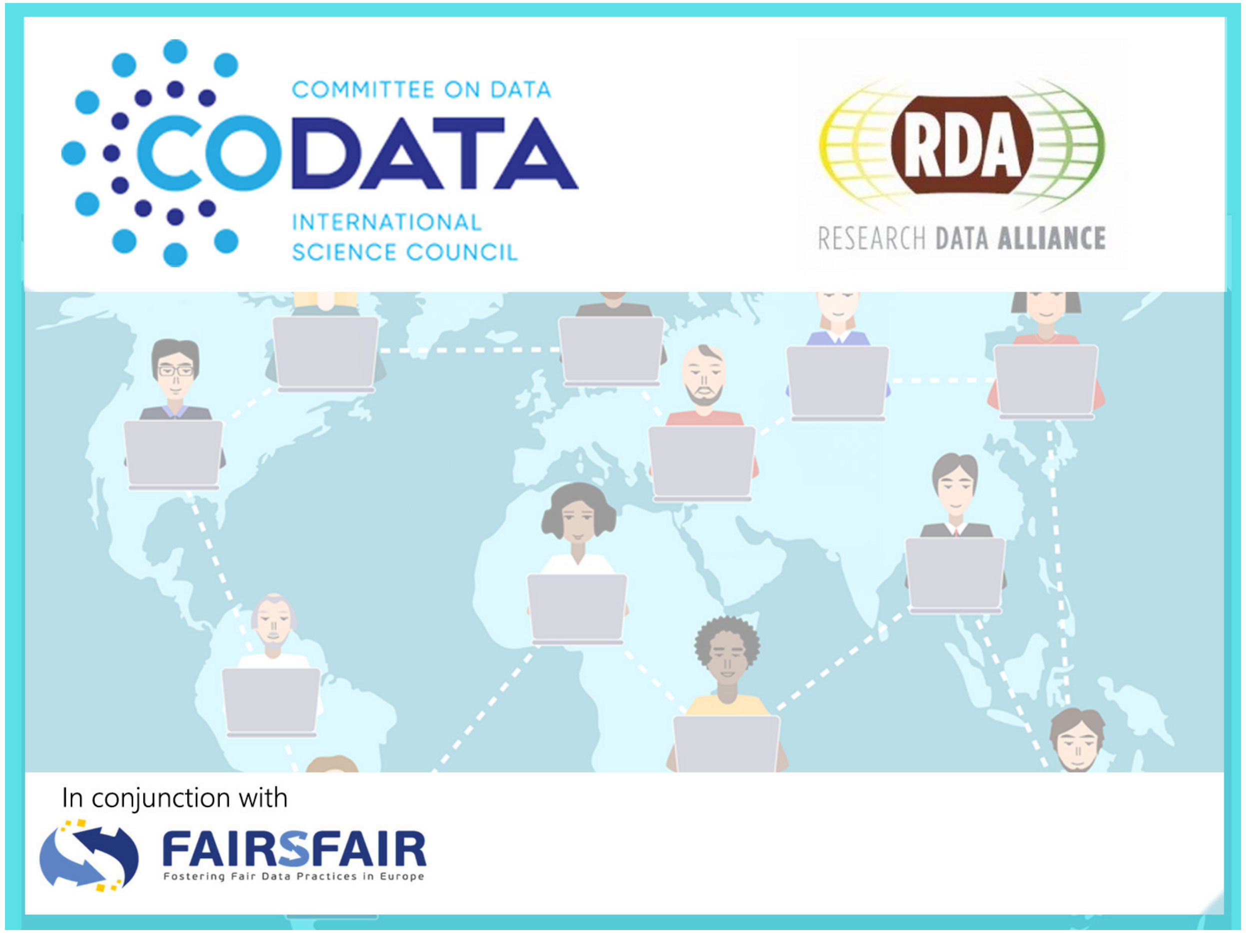
The Value of RDA for

RDA in one Word

https://www.rd-alliance.org

Data Schools

Acknowledgements



Homework exercise

Imagine you are a biologist who is doing microscopy experiments imaging tissue specimens. The data captured by the imaging is 100s of GB in size and is then cleaned and analysed to produce derivatives of the original captured data. Some of these derivatives may eventually be published. In preparation for publication, the data will also be segmented and annotated using standard ontologies. Documentation will also include metadata standards that will sufficiently describe the experimental procedure to allow reproducibility. Publication of the data is mandatory due to funder policy and must be deposited in a repository within 3 years of data production and must use an open licence without restrictions on reuse.

Now…please split into groups and see if you can answer the following questions using the tools and guidelines that have been described:

- What **file format(s)** should data be captured/preserved in?
- Which **metadata standard(s)** should be used?
- What **ontology(ies)** should be used?
- Which **licence(s)** should be used?
- Which **repository** would be the best fit for these data?
- Do you foresee any problems with the data?
- (Hint: not all the questions can be answered definitively! – but why not?)
- **Please use the FAIR Data Forum to post answers and discuss!**