

National Initiatives for Open Science in Europe

ORDM Train-the-Trainer: The basics

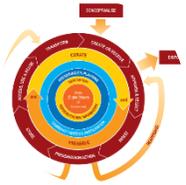
Elli Papadopoulou
Athena Research Center

elli.p@athenarc.gr





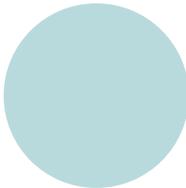
The European Open Science Framework



Open and FAIR RDM



Stakeholders and RDM



Wrap-Up

The European Open Science Framework

Open Science

SiS.net Network of National Contact Points for Science with and for Society in Horizon 2020

About National Contact Points SiS.net project SwafS Stakeholders Responsible Research & Innovation Partner Search

You are here: Front page > Responsible Research & Innovation

Responsible Research & Innovation

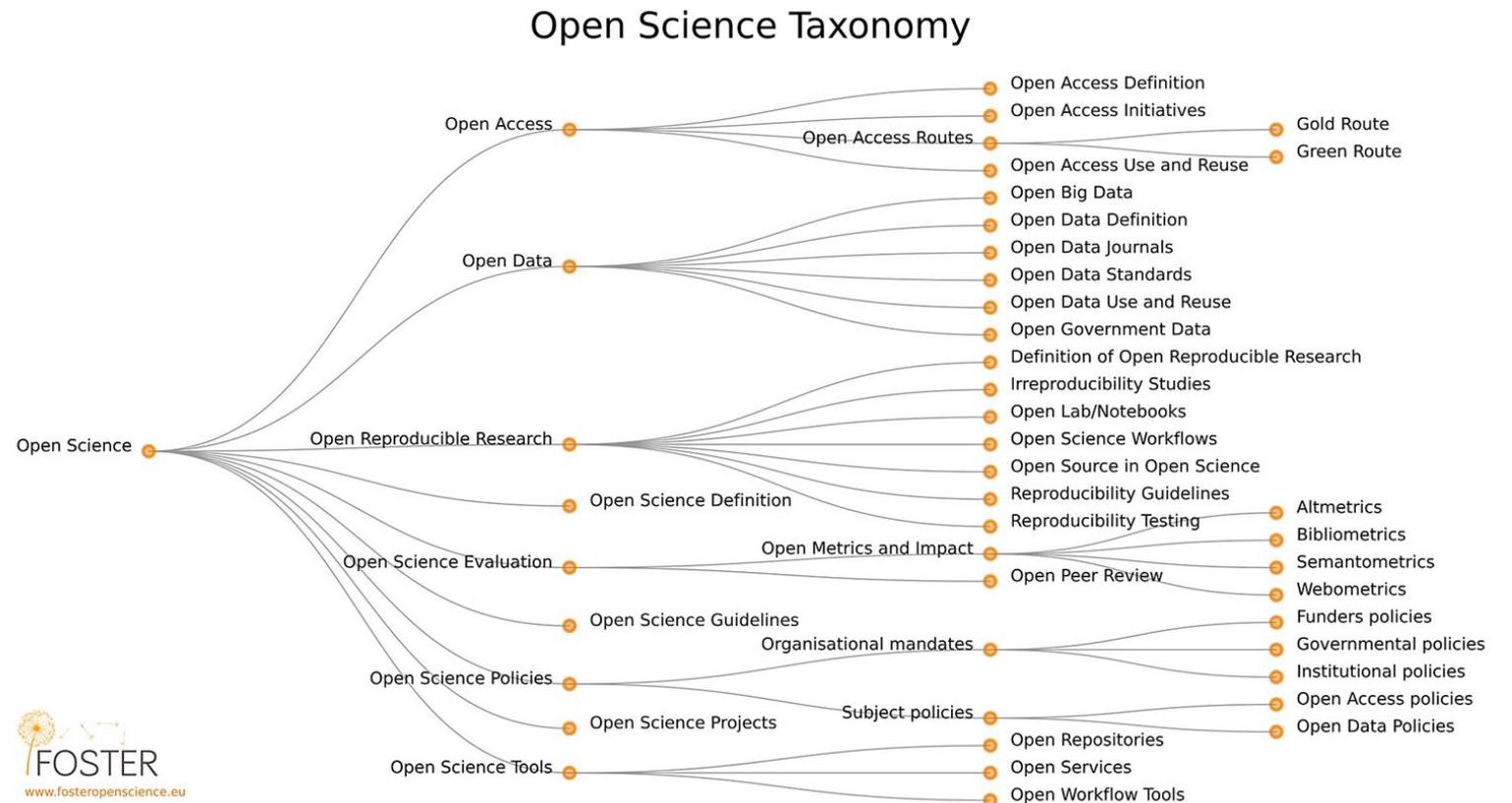
Responsible Research and Innovation (RRI) is an approach which implies that societal actors, such as researchers, citizens, policy makers, companies and civil society organisations) work together in the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society.



In practice, Responsible Research and Innovation¹² is implemented as a package that includes multi-actor and public engagement in research and innovation, enabling easier access to scientific results, the take up of gender and ethics in the research and innovation content and process, and formal and informal science education.

Responsible Research and Innovation¹² is the key action of the Science with and for Society objective. The following five RRI actions will be promoted via the Science with and for Society programme¹² of Horizon 2020¹⁰:

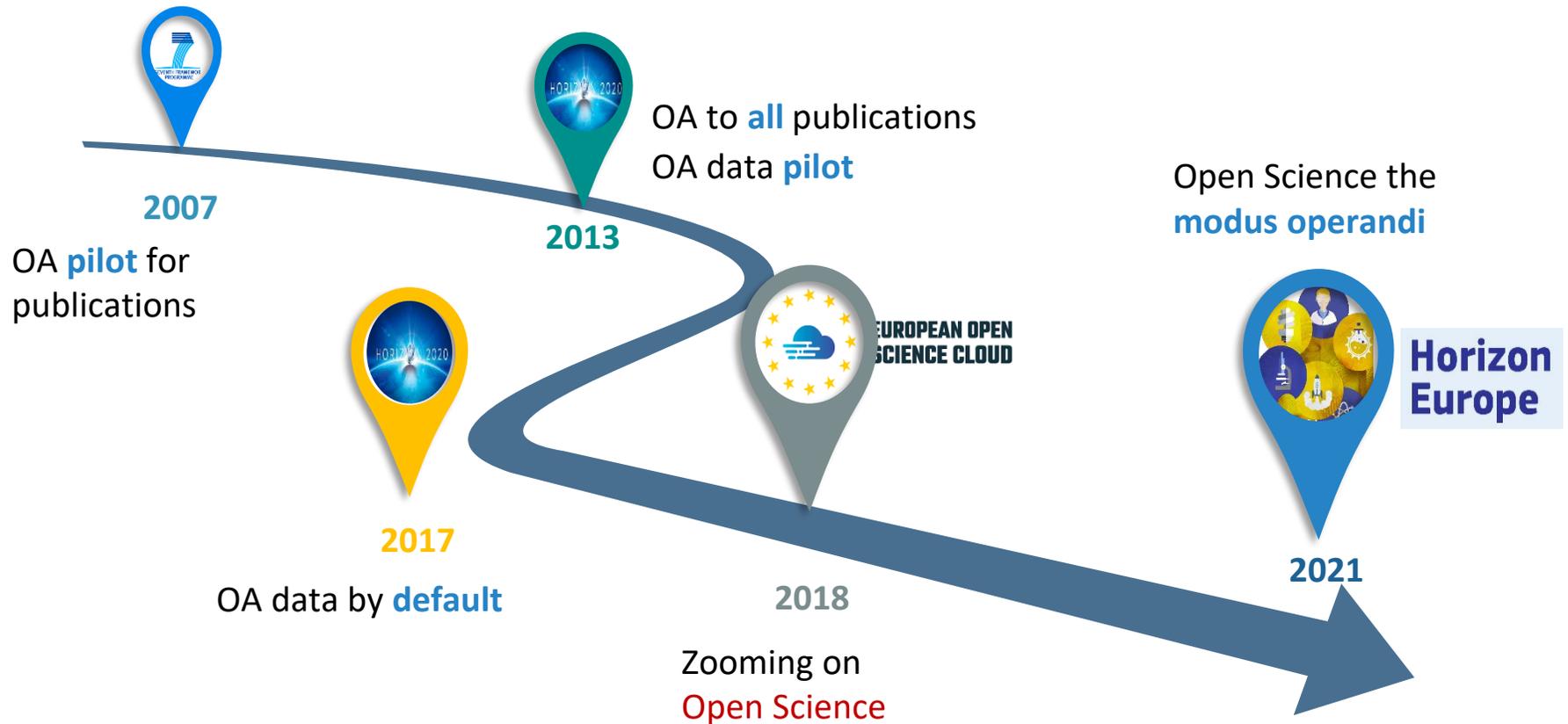
- Ethics
- Gender
- Open Science
- Public Engagement
- Science Education



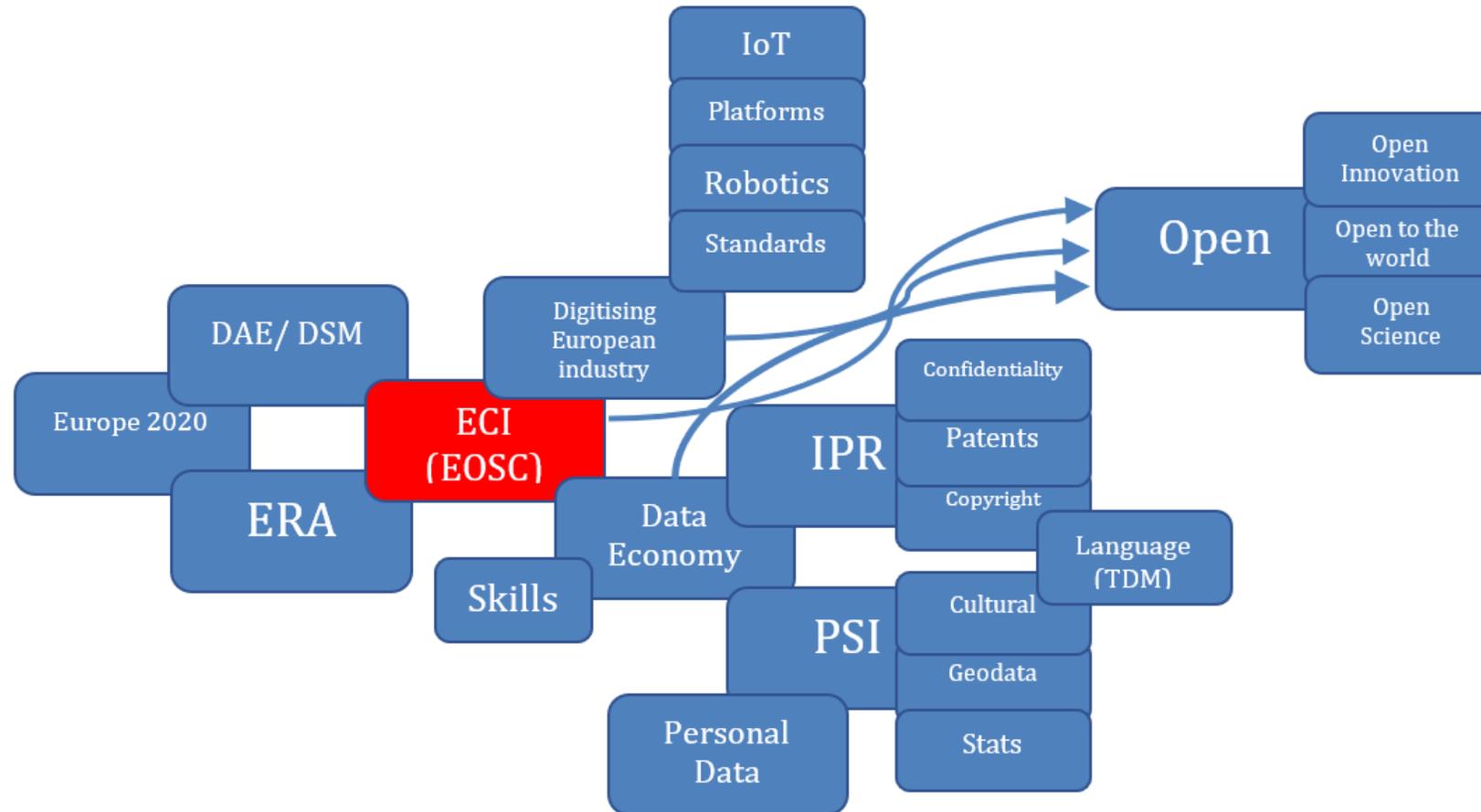
Knoth, Petr; Pontika, Nancy (2015): Open Science Taxonomy. figshare. Figure. <https://doi.org/10.6084/m9.figshare.1508606.v3>

Framework Programmes

“As open as possible, as closed as necessary”, following the FAIR principles and well-documented processes of data handling and re-use



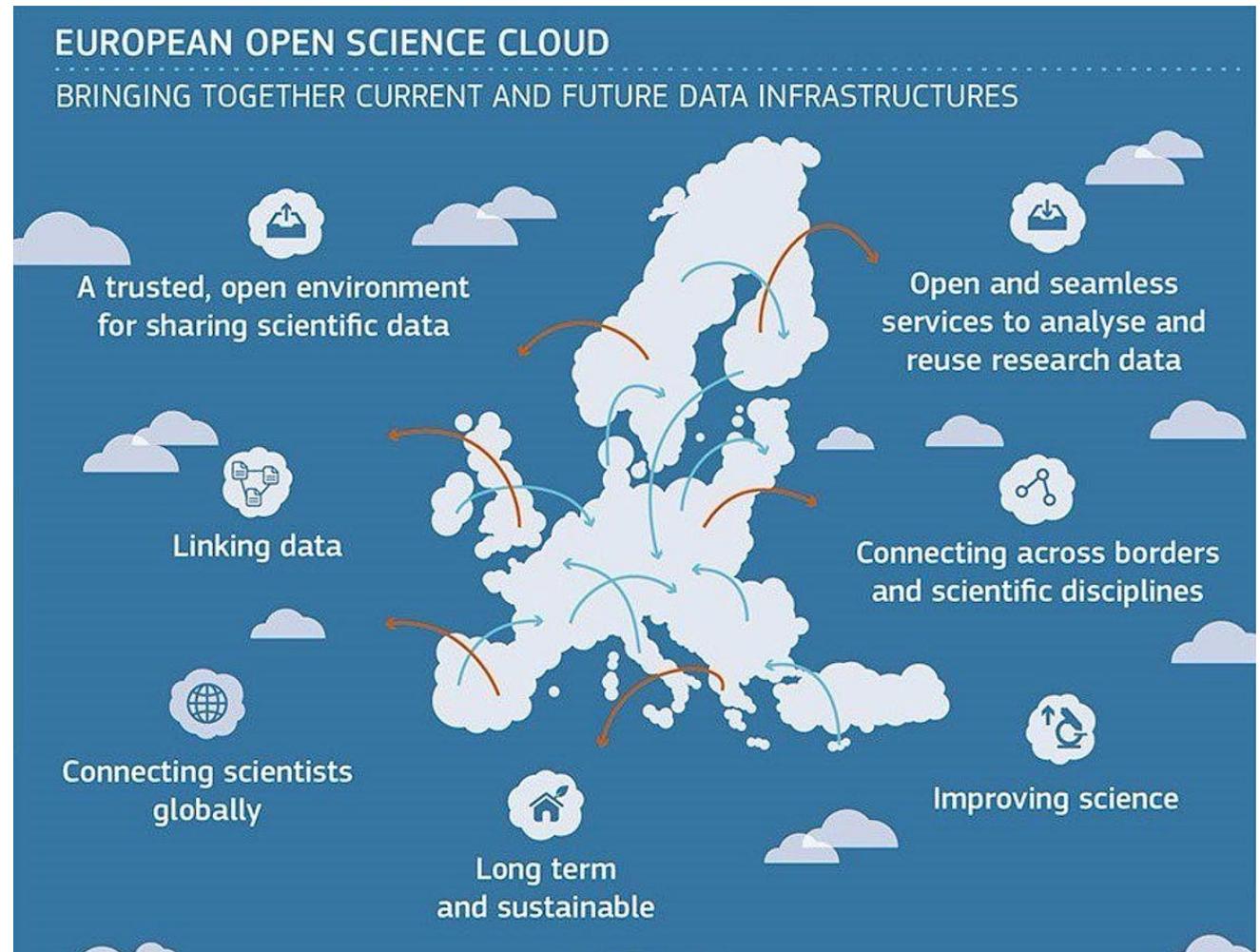
Policies and Directives



Papadopoulou, Elli, & Tsiavos, Prodromos. (2019, January 7). D3.1: Policy Landscape Review (Version 4.2). Zenodo. <http://doi.org/10.5281/zenodo.3387562>



- An **ethical, open, secure** and **cost-effective** environment to support Open Science practices and research communication in the EU



EOSC portal: <https://www.eosc-portal.eu/>

Why Open Science and EOSC

Research

- Reach wider audience
- Re-use research outputs
- Validate research
- Prevent information and data loss

Economy

- Stimulate innovation
- Strengthen regional and national markets
- New job openings

Researchers

- Promote integrity
- Increase use of citations and get more credits
- Rewards in the EOSC

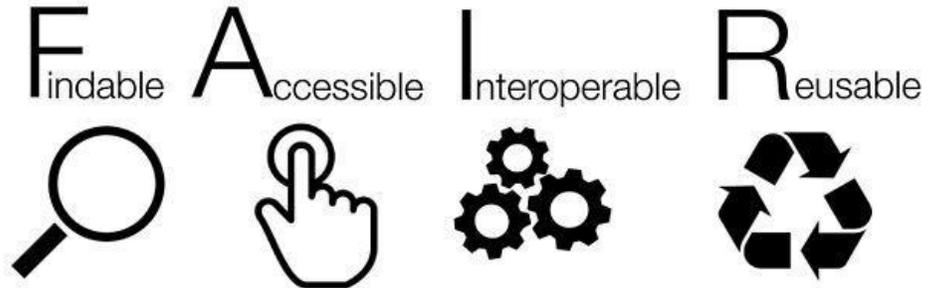
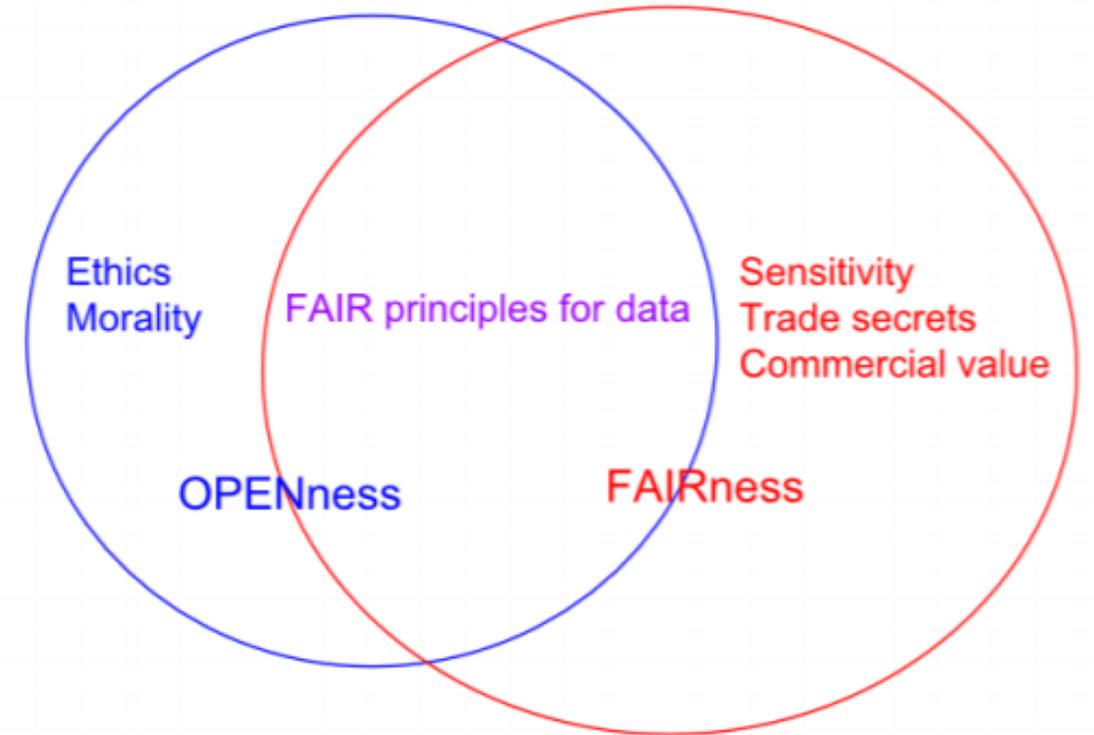
Society

- Transparency
- Build trust
- Collaboration on national and EU level
- Citizen Science

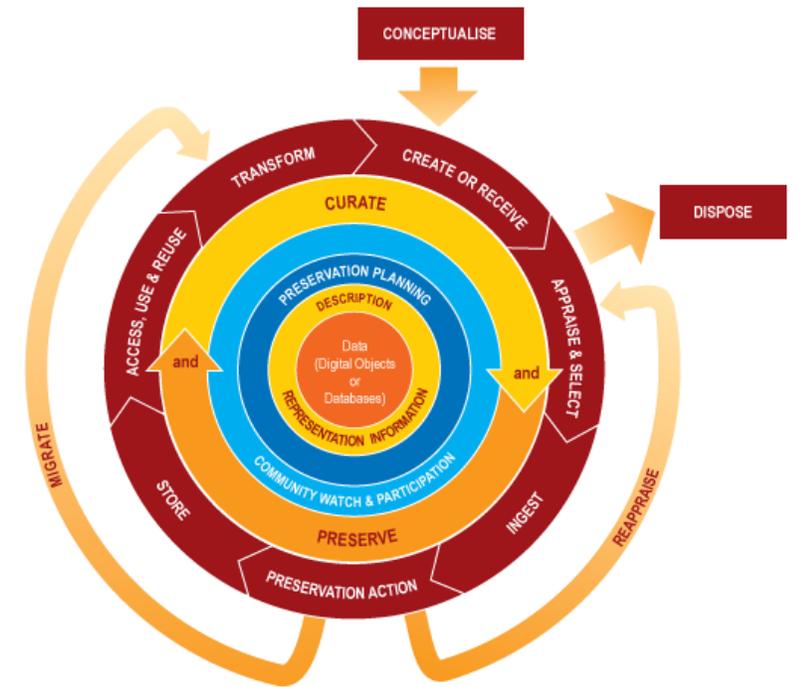
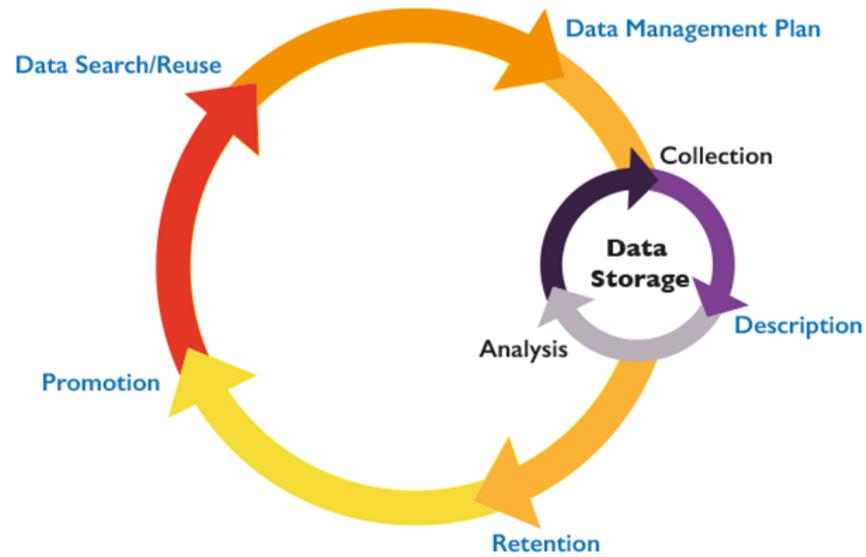
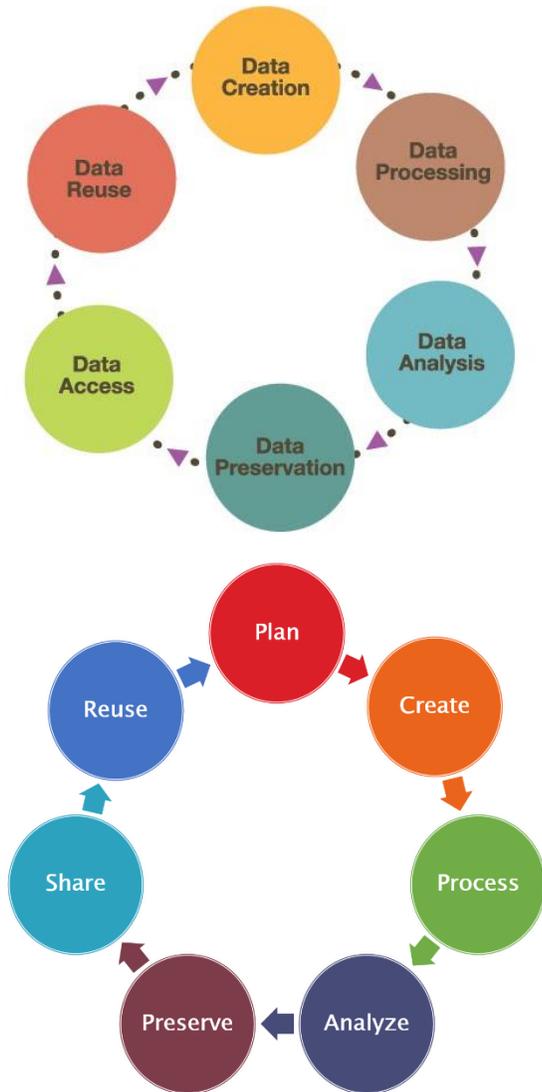
Open and FAIR RDM

Open and FAIR

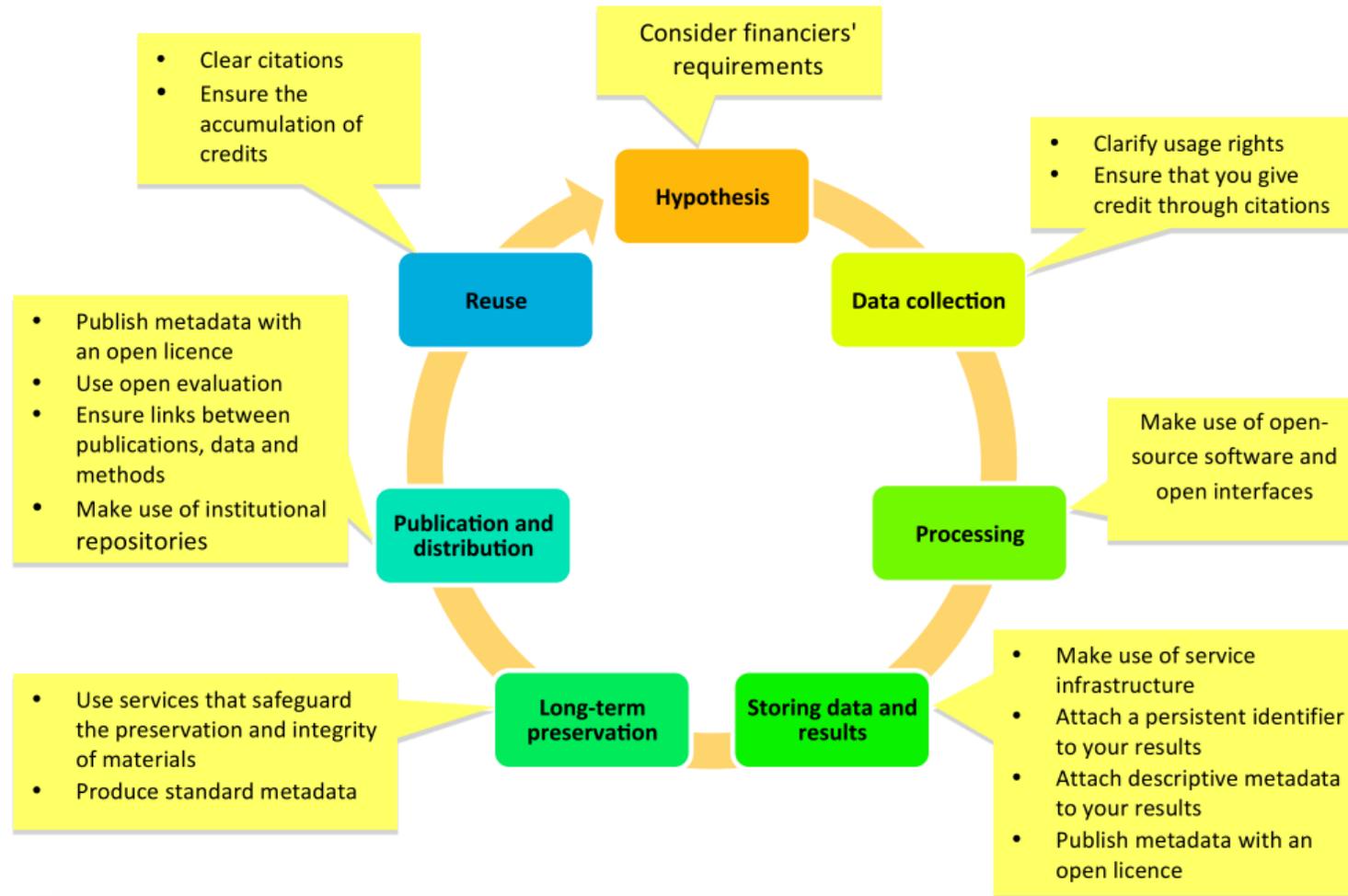
ACCESS	READER RIGHTS	REUSE RIGHTS	COPYRIGHTS	AUTHOR POSTING RIGHTS	AUTOMATIC POSTING	MACHINE READABILITY	ACCESS
OPEN ACCESS	Free readership rights to all articles immediately upon publication	Generous reuse & remixing rights (e.g., CC BY license)	Author holds copyright with no restrictions	Author may post any version to any repository or website with no delay	Journals make copies of all articles automatically available in trusted third-party repositories (e.g., PubMed Central, OpenAire, institutional) immediately upon publication	Article full text, metadata, supporting data (including format and semantic markup) & citations may be accessed via API, with instructions publicly posted	OPEN ACCESS
	Free readership rights to all articles after an embargo of no more than 6 months	Reuse, remixing, & further building upon the work subject to certain restrictions & conditions (e.g., CC BY-NC & CC BY-SA licenses)	Author retains/publisher grants broad rights, including author reuse (e.g., of figures in presentations/teaching, creation of derivatives) and authorization rights (for others to use)	Author may post some version (determined by publisher) to any repository or website with no delay	Journals make copies of all articles automatically available in trusted third-party repositories (e.g., PubMed Central, OpenAire, institutional) within 6 months	Article full text, metadata, & citations may be accessed via API, with instructions publicly posted	
	Free readership rights to all articles after an embargo greater than 6 months	Reuse (no remixing or further building upon the work) subject to certain restrictions and conditions (e.g., CC BY-ND license)	—	Author may post some version (determined by publisher) to any repository or website with some delay (determined by the publisher)	Journals make copies of all articles automatically available in trusted third-party repositories (e.g., PubMed Central, OpenAire, institutional) within 12 months	Article full text, metadata, & citations may be crawled without special permission or registration, with instructions publicly posted	
	Free and immediate readership rights to some, but not all, articles (including "hybrid" models)	Some reuse rights beyond fair use for some, but not all, articles (including "hybrid models")	Author retains/publisher grants limited rights for author reuse (e.g., of figures in presentations/teaching, creation of derivatives)	Author may post some version (determined by publisher) to certain repositories or websites, with or without delays	Journals make copies of some, but not all, articles automatically available in trusted third-party repositories (e.g., PubMed Central, OpenAire, institutional) within 12 months	Article full text, metadata, & citations may be crawled with permission, with instructions publicly posted	
CLOSED ACCESS	Subscription, membership, pay-per-view, or other fees required to read all articles	No reuse rights beyond fair use/dealing or other limitations or exceptions to copyright (All Rights Reserved)	Publisher holds copyright, with no author reuse beyond fair use	Author may not deposit any versions to any repositories or websites at any time	No automatic posting in third-party repositories	No full text articles available for crawling	CLOSED ACCESS



RDM lifecycles



Open Science Practices



(Open Science and Research Initiative, 2014)

Zooming in – The basics

□ What is research data?

what has been used or generated (including software) during research process and support/validate its findings

□ Why manage research data?

- ✓ Data are understandable, re-usable and reproducible
 - ✓ Avoid data loss
 - ✓ Get credit
 - ✓ Avoid fraudulent/ bad science
-

Plan data management of research activities following research data lifecycle steps

- Costing RDM
 - Preparing (DMP)
 - **Data collection**, eg database, formatting, transcription, etc
 - **Data documentation**, eg data description, metadata
 - **Data storage and back-up**
 - **Data access and security**, eg TTP, encryption
 - **Data sharing & reuse**, eg anonymization, copyright, cleaning, digitization
 - **Overall**, eg roles & responsibilities

Activity: Plan your research and estimate your costs

What is a DMP?

Deliverable and “living” document

documents processes undertaken throughout data management lifecycle, including costs



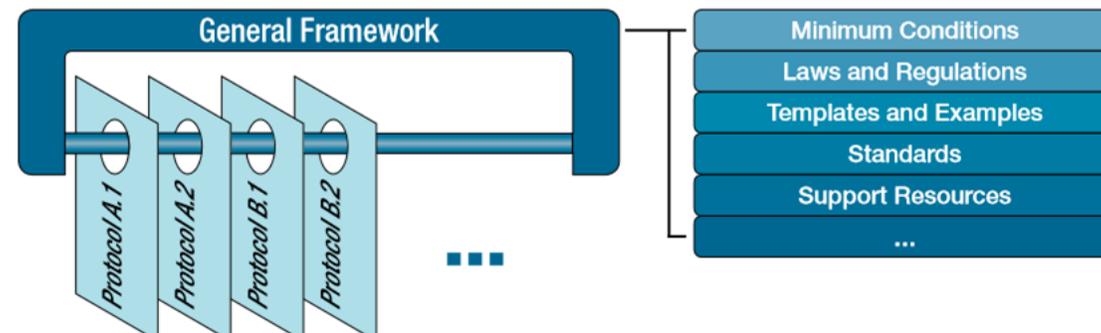
What is not a DMP?

Research assessment method

Plan – DMPs 2/2

- ❑ Depends on the funder/institution requirements
- ❑ Differences in research communities
 - ❑ Formats, standards, documentation etc

-> Minimum requirements: Science Europe – **DDPs** (Domain Data Protocols)



Activity: Search for a DDP for your domain. If none exists, create your own

The case of H2020 DMP template

SUMMARY TABLE 1

FAIR Data Management at a glance: issues to cover in your Horizon 2020 DMP

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined above.

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none">• State the purpose of the data collection/generation• Explain the relation to the objectives of the project• Specify the types and formats of data generated/collected• Specify if existing data is being re-used (if any)• Specify the origin of the data• State the expected size of the data (if known)• Outline the data utility: to whom will it be useful
2. FAIR Data 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none">• Outline the discoverability of data (metadata provision)• Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?• Outline naming conventions used• Outline the approach towards search keyword• Outline the approach for clear versioning• Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

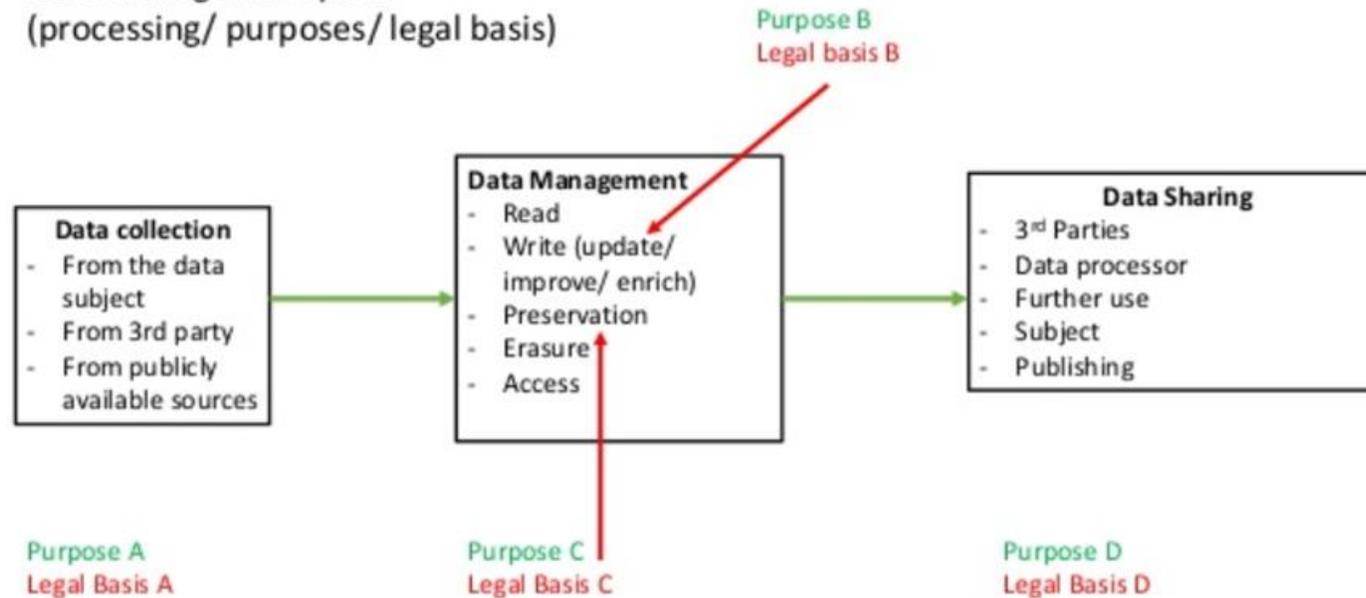
The case of H2020 DMP template

2.2 Making data openly accessible	<ul style="list-style-type: none">• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so• Specify how the data will be made available• Specify 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)?• Specify where the data and associated metadata, documentation and code are deposited• Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	<ul style="list-style-type: none">• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.• Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none">• Specify how the data will be licenced to permit the widest reuse possible• Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed• Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why• Describe data quality assurance processes• Specify the length of time for which the data will remain re-usable
3. Allocation of resources	<ul style="list-style-type: none">• Estimate the costs for making your data FAIR. Describe how you intend to cover these costs• Clearly identify responsibilities for data management in your project• Describe costs and potential value of long term preservation
4. Data security	<ul style="list-style-type: none">• Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	<ul style="list-style-type: none">• To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	<ul style="list-style-type: none">• Refer to other national/funder/sectorial/departmental procedures (if any)

Activity: Go to ARGOS and create a DMP

Plan - GDPR and DMP

Data management plan
(processing/ purposes/ legal basis)



- Follow the data
- Different types of data processing may have different purposes and legal bases
- Always stay within the legal basis

- ✓ Mostly forms of **public interest** (needs to be specifically documented per institution and research project)
- ✓ **Contract** (tender)
- ✓ **Consent** (specific research)
- ✓ Could **change** from collection, to retaining to sharing. There always needs to be one covering the purpose of processing.

❑ Metadata -> Standards

- ❑ For discovery (minimum)
- ❑ For interoperability (rich)

❑ General or Domain specific

- ❑ For different outputs, eg instruments

❑ Include PIDs and citations in metadata

Term Name: contributor	
URI:	http://purl.org/dc/elements/1.1/contributor
Label:	Contributor
Definition:	An entity responsible for making contributions to the resource.
Comment:	Examples of a Contributor include a person, an organization, or
Term Name: coverage	
URI:	http://purl.org/dc/elements/1.1/coverage
Label:	Coverage
Definition:	The spatial or temporal topic of the resource, the spatial applic
Comment:	Spatial topic and spatial applicability may be a named place or administrative entity or a geographic place to which the resource places or time periods can be used in preference to numeric id
References:	[TGN] http://www.getty.edu/research/tools/vocabulary/tgn/ind
Term Name: creator	
URI:	http://purl.org/dc/elements/1.1/creator
Label:	Creator
Definition:	An entity primarily responsible for making the resource.
Comment:	Examples of a Creator include a person, an organization, or a se
Term Name: date	
URI:	http://purl.org/dc/elements/1.1/date
Label:	Date
Definition:	A point or period of time associated with an event in the lifecyc
Comment:	Date may be used to express temporal information at any level
References:	[W3CDTF] http://www.w3.org/TR/NOTE-datetime
Term Name: description	

<https://www.dublincore.org/specifications/dublin-core/dces/>

OpenAIRE-Field	Metadata Element	Refinement by Vocabulary
Title (M)	datacite:title	title type
Creator (M)	datacite:creator	name type
Contributor (MA)	datacite:contributor	name type contributor type
Funding Reference (MA)	oaire:fundingReference	funderIdentifier type
Alternate Identifier (R)	datacite:alternateIdentifier	alternateIdentifier type
Related Identifier (R)	datacite:relatedIdentifier	relatedIdentifier type relation type resourceType general
Embargo Period Date (MA)	datacite:date	date type
Language (MA)	dc:language	IETF BCP 47, ISO 639-3
Publisher (MA)	dc:publisher	
Publication Date (M)	datacite:date	date type
Resource Type (M)	oaire:resourceType	COAR Resource Type Vocabulary
Description (MA)	dc:description	
Format (R)	dc:format	
Resource Identifier (M)	datacite:identifier	identifier type
Access Rights (M)	datacite:rights	COAR Access Right Vocabulary
Source (R)	dc:source	
Subject (MA)	datacite:subject	
License Condition (R)	oaire:licenseCondition	
Coverage (R)	dc:coverage	
Size (O)	datacite:size	
Geo Location (O)	datacite:geoLocation	
Resource Version (R)	oaire:version	COAR Version Vocabulary
File Location (MA)	oaire:file	
Citation Title (R)	oaire:citationTitle	
Citation Volume (R)	oaire:citationVolume	
Citation Issue (R)	oaire:citationIssue	
Citation Start Page (R)	oaire:citationStartPage	
Citation End Page (R)	oaire:citationEndPage	
Citation Edition (R)	oaire:citationEdition	

<https://openaire-guidelines-for-literature-repository-managers.readthedocs.io/en/v4.0.0/>

The operational phase during which raw data is being manipulated to result to meaningful information

- ❑ Handling/curation of data
- ❑ Between data collection and data preservation
- ❑ Involves processes such as: ingestion, aggregation, analysis, classification, metadata enrichment, organisation, validation, storing, etc.
- ❑ There might be re-processing of data (e.g. data migration)
- ❑ Data disposal

Process – Clean and tidy

- ❑ Representation inconsistencies
- ❑ Numerical inconsistencies
- ❑ Misspelling/ typos
- ❑ Abbreviated content and inconsistencies
- ❑ Duplicates
- ❑ Other data

M, male, m., fem., F,
Female

05/05/2020 OR 5th May 2020 OR May 5 2020
VS 05-05-2020 “%d-%m-%Y”

SUM; Notes

“hello world” -> “hello world
munchen, Munich, Muenchen

Mixed scales, eg million vs age

Activity: Find the wrong and fix it OR
Match the left to the right

Types of sensitive data

- ❑ **Personal data** (and metadata)
- ❑ **Confidential data** (trade secrets, investigations,...)
- ❑ **Security data** (passwords, financial information, national safety, military,...)
- ❑ **Data protected by Intellectual Property Rights (IPR)**
- ❑ **Location Data/GPS/mobile phones**
- ❑ **Endangered (plant or animal) species**, where their survival is dependent on the protection of their location data (biodiversity community)
- ❑ **Combination** of different datasets could lead to sensitive data?
- ❑ racial or ethnic origin
- ❑ political opinions
- ❑ religious or philosophical beliefs
- ❑ trade union membership
- ❑ genetic data, biometric data
- ❑ physical or mental health
- ❑ sex life or sexual orientation
- ❑ criminal offences

Best practices

- ❑ **Access controls**

passwords, firewall (viruses, hacking)

- ❑ **Anonymisation**

removing or aggregating variables or reducing the precision or detailed textual meaning of a variable

- ❑ **Encryption**

encoded digital information

- ❑ **Share in a secure place**

no cloud drives

- ❑ **Store in an isolated machine**

server not connected to Internet

- ❑ **Secure disposal**

no data recovery is possible (uninstall)

Start producing outputs and prepare for sharing

□ **Methods**

- Lab notebooks, end-to-end code/scripts for statistics, etc

□ **Software**

- R, MatLab, Python, etc

Store data in the short and/or long term

❑ What to preserve?

data underlying publications, recreation purposes and value, legal issues, etc

❑ Move from proprietary formats

- ❑ Prefer open, lossless formats (e.g. rtf, xml, tif, wav VS doc, jpg, mp3)
 - ❑ Check repository policy to see for suggested options

Activity: guess the format of data
(proprietary vs non-proprietary)

❑ Archiving: Certified repositories/ FAIR enabled repositories

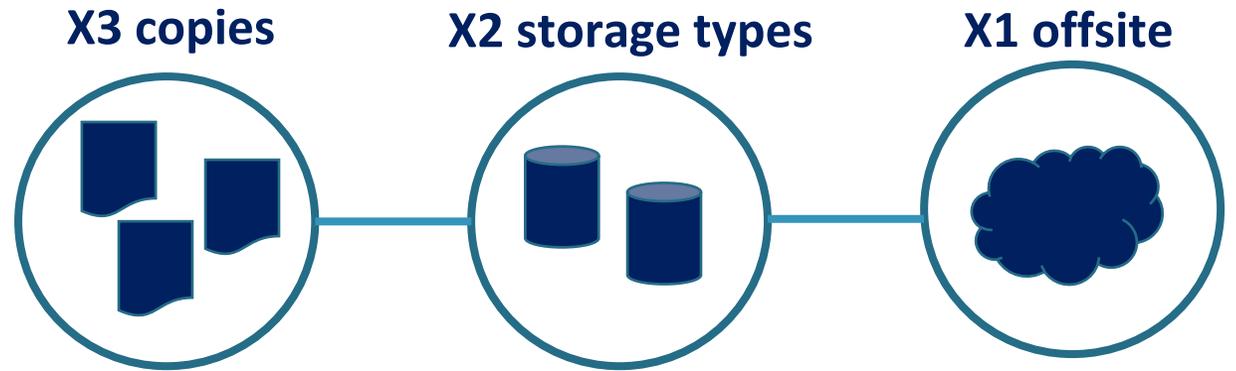
- ❑ Thematic
- ❑ Institutional

Activity: select a repository for your data

Preserve

❑ Risk-assessment/ Back-ups

- ❑ Retention
- ❑ Frequency of back-up
- ❑ Storage and methods



Activity: What is the difference between back-up and preservation?

❑ PIDs

- ❑ Findability
- ❑ Actionability -> accessible by web browser

- ❑ **Naming conventions**, so that data are understandable by others
 - ❑ e.g. letters, characters, abbreviations
- ❑ **Means of sharing**
 - ❑ Commercial cloud, e.g. Google Drive
 - ❑ Cloud infrastructure for research, e.g. B2SHARE
 - ❑ Ftp server
 - ❑ USB Drives
- ❑ **Create links**
 - ❑ Link research outputs

Access

- ❑ Immediate; metadata only
 - ❑ Check embargoes
- ❑ Restrictions (Copyright, IPR etc)
- ❑ Access controls

	Open data	Safeguarded data	Controlled data
Security requirement	Suitable for fully anonymised data or data with agreement to publish personal details	Partially anonymised data or data with agreement to publish personal details, and where owner wishes to track usage	Too detailed, confidential or sensitive to be downloaded
Level of access	Accessible without user registration	Accessible to authenticated users	Accessible to authenticated users, using secure remote access or secure onsite room
Legal conditions	Under open licence, either Open Government Licence (OGL) for Crown Copyright data or Creative Commons for other data	Requiring an End User Licence and, where appropriate, special conditions agreed to, or data owner approval	Requires user accreditation and registration through training and approval by a data access committee

Activity: Which category does your dataset fall under?

<https://ukdataservice.ac.uk/deposit-data/how-to/regular-depositors/negotiate>

□ Licenses

- Conditions
- Types

□ Citations

- Specify required data citation
- Open citations

□ Peer-reviews

□ Teaching & learning

License Conditions

When using a Creative Commons license, creators choose a set of conditions they wish to apply to their work.

Attribution (by)

All CC licenses require that others who use your work in any way must give you credit the way you request, but not in a way that suggests you endorse them or their use. If they want to use your work without giving you credit or for endorsement purposes, they must get your permission first.

ShareAlike (sa)

You let others copy, distribute, display, perform, and modify your work, as long as they distribute any modified work on the same terms. If they want to distribute modified works under other terms, they must get your permission first.

NonCommercial (nc)

You let others copy, distribute, display, perform, and (unless you have chosen NoDerivatives) modify and use your work for any purpose other than commercially unless they get your permission first.

NoDerivatives (nd)

You let others copy, distribute, display and perform only original copies of your work. If they want to modify your work, they must get your permission first.



Activity: Go to License Selector and choose a license for a dataset you are currently working

Go to [menti.com](https://www.menti.com)

code: 488464



Stakeholders and RDM

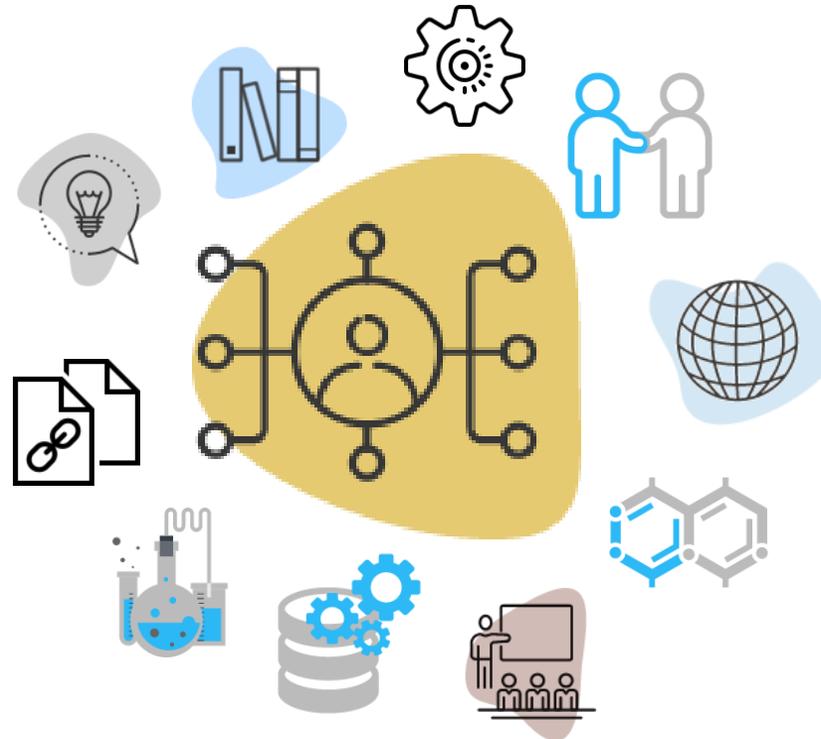
Who is involved

Researchers

- Quality data
- Follow best practices
- Comply with RDM policies
- Credits

Research Performing Organisations

- Research Excellence
- Scholarly Communication
- Monitor research
- Support research conduct



Research Funding Organisations

- Monitor research
- Better control of funds
- Research excellence
- Innovation

Service Providers

- Apply standards and best practices
- New tools
- Data-intensive activities

Research Performing Organisations

Research Performing Organisations

- Policy

- Data Services & Tools

- Support

RECOMMENDATIONS

**COMMISSION RECOMMENDATION (EU) 2018/790
of 25 April 2018
on access to and preservation of scientific information**

THE EUROPEAN COMMISSION,

Having regard to the Treaty on the Functioning of the European Union, and in particular Article 292 thereof,

Whereas:

- (1) The European Commission adopted in July 2012 a scientific information package, consisting of the communication 'Towards better access to scientific information: Boosting the benefits of public investments in research' ⁽¹⁾, and of Commission Recommendation 2012/417/EU ⁽²⁾. Recommendation 2012/417/EU states that the Commission will review the progress made across the Union to assess whether further action is needed to achieve the objectives laid down.

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018H0790&from=EN>

- ❑ **Open access to scientific publications from 2020**
 - ❑ IPR retention for self-archiving and re-use, immediate access, etc
- ❑ **Management of research data**
 - ❑ DMPs, FAIR, licenses, funding, etc
- ❑ **Preservation and re-use**
 - ❑ Policies and platforms, curation, interlinking
- ❑ **Infrastructures**
 - ❑ Certifications, access to resources and services (inc. EOSC), indicators, etc
- ❑ **Skills and competences**
 - ❑ At the higher education and training system
- ❑ **Incentives and rewards**
 - ❑ New generation metrics



Project acronym: RECODE
Project title: Policy RECommendations for Open access to research Data in Europe
Grant number: 321463
Programme: Seventh Framework Programme for Science in Society
Objective: SiS-2012.1.3.3-1: Scientific data: open access, dissemination, preservation and use
Contract type: Co-ordination and Support Action
Start date of project: 01 February 2013
Duration: 24 months
Website: www.recodeproject.eu

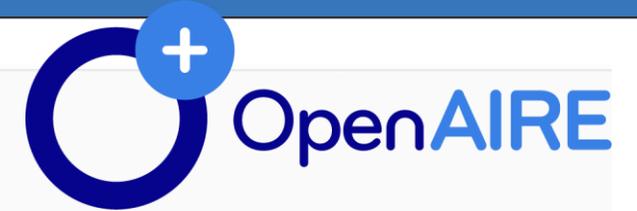


<http://www.pasteur4oa.eu/resources>



**LEARN Toolkit of Best Practice
for Research Data Management**

<http://learn-rdm.eu/wp-content/uploads/RDMToolkit.pdf>



Toolkit for Policy Makers on Open Science and Open Access

The OpenAIRE Toolkit for Policy Makers has been designed to assist the design and adoption of Open Science policies aligned with EU developments in the field. It therefore targets stakeholders at national, institutional or funder level with a key role in the adoption of Open Science / Open Access policies (university rectors, directors of research centres, directors of funding agencies, rectors' summit, ministries etc.).

The Toolkit aims to assist NOADs in promoting OS/OA policies in their country and enabling them to become the national hub on OS/OA by enhancing their expertise.

Yet, the Toolkit can also be used by other stakeholders seeking to learn, adopt or align their OS/OA policies and this is why all related material and resources are public.

Open Science Policy Checklist for
Research Funding Organisations

Open Science Policy Checklist for
Research Performing Organisations

Model Policy on Open Science for
Research Funding Organisations

Model Policy on Open Science for
Research Performing Organisations

<https://www.openaire.eu/toolkit-for-policy-makers-on-open-science-and-open-access>



https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf



http://www.scienceeurope.org/media/nsxdyvqn/se_guidance_document_rdmeps.pdf

□ Core Requirements for DMPs

- Data Description and collection
- Documentation and data quality
- Storage and backup
- Legal, ethics, CoC
- Data sharing and long-term preservation
- Data management responsibilities

Building data services - audits

DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment

<http://www.repositoryaudit.eu/>



Four steps to effective data management

Home

Background

Methodology

Users

Findings

Documents

Launch

Contact Us

Methodology

The Data Asset Framework recommends that audits of research data assets proceed as a four step process:

In the planning stage the purpose and scope of the audit is defined. Preliminary research is conducted and meetings scheduled to optimise time spent with the organisation's staff. The purpose of the second stage, identifying research data, is to establish what data assets exist and classify them according to the purposes of the survey. The classification step determines the scope of further activities, as only selected aspects will be assessed in greater detail. The information collected in Stage 3 will assist auditors to identify weaknesses in data policy and current data creation and curation procedures. This will provide the basis of recommendations in the final stage. The knowledge gained from the assessment survey will enable organisations to plan improvements in data management.

The current version of the [methodology](#) is available to download. An

- 1 PLANNING THE AUDIT
- 2 IDENTIFYING AND CLASSIFYING ASSETS
- 3 ASSESSING MANAGEMENT OF DATA ASSETS
- 4 REPORTING AND RECOMMENDATIONS

<https://www.data-audit.eu/>

Building data services - curation



Home About **Understand costs** Compare costs Read more Vendor services Help Sign Up Sign In

Understand costs

CCEX can't predict your costs for you, but we can help get you started on costing your curation activities. Learn about the **basic concepts** used to assess curation costs. Follow our tips on finding a suitable **cost model** to describe your costs. Identify the **cost drivers** that may impact on the investment decisions you make. Prepare for the future using our advice on **planning to sustain** your curation services and digital assets over time.



Basic cost concepts

What are the basics I need to know to get started with costing curation activities?



Cost models

Are there existing cost models that can help me describe my organisation's activities?



Cost drivers

How can I be sure that my costs are justifiable?



Sustainability planning

What do I need to consider to sustain my organisation's investment in curation?

<https://www.curationexchange.org/understand-your-costs>

Basic concepts

- Organisational context, Stakeholders, Services/activities and Resources

Cost models

- Spending and efficiencies, Benefit and value, assets and use cases, etc

Cost drivers

- Financial, customer-related, learning and growth, operational efficiency

Sustainability planning

- Business model (customer, activity, financial perspective, and the value proposition)

□ Organize research data events

□ Training methods and courses



[Homepage - Open Field Guides - Engaging Researchers with Data Management: The Cookbook](#)

Engaging Researchers with Data Management: The Cookbook

Connie Clare, Maria Cruz, Elli Papadopoulou, James Savage, Marta Teperek, Yan Wang, Iza Witkowska, and Joanne Yeomans

- <https://orcid.org/0000-0002-4369-196X> <https://orcid.org/0000-0001-9111-182X>
- <https://orcid.org/0000-0002-0893-8509> <https://orcid.org/0000-0002-4737-5673>
- <https://orcid.org/0000-0001-8520-5598> <https://orcid.org/0000-0002-6317-7546>
- <https://orcid.org/0000-0002-7914-3957> <https://orcid.org/0000-0002-0738-7661>

Paperback	ISBN: 978-1-78374-797-9	£15.95	Add to Cart
Hardback	ISBN: 978-1-78374-798-6	£25.95	Add to Cart
PDF	ISBN: 978-1-78374-799-3	£0.00	Download
epub	ISBN: 978-1-78374-800-6	£5.99	Add to Cart
mobi	ISBN: 978-1-78374-801-3	£5.99	Add to Cart
XML	ISBN: 978-1-78374-802-0	£0.00	Download



[READ THE PDF](#) [READ THE HTML](#)

<https://www.openbookpublishers.com/product/1080>

The Open Science Training Handbook

Readme	1.1
Introduction	1.2
Open Science Basics	1.3
Open Concepts and Principles	1.3.1
Open Research Data and Materials	1.3.2
Open Research Software and Open Source	1.3.3
Reproducible Research and Data Analysis	1.3.4
Open Access to Published Research Results	1.3.5
Open Licensing and File Formats	1.3.6
Collaborative Platforms	1.3.7
Open Peer Review, Metrics and Evaluation	1.3.8
Open Science Policies	1.3.9
Citizen Science	1.3.10
Open Educational Resources	1.3.11
Open Advocacy	1.3.12
On Learning and Training	1.4
Organizational Aspects	1.5
Examples and Practical Guidance	1.6
Glossary	1.7
References	1.8
About the Authors & Facilitators	1.9
Languages	1.10

<https://book.fosteropenscience.eu/en/book.pdf>

Researchers

- ❑ **Awareness and training**
- ❑ **Plan data management activities**
 - ❑ Estimate Costs
 - ❑ Publish in OA
 - ❑ Curation of data
 - ❑ IP costs
 - ❑ IP
 - ❑ Licenses
 - ❑ Metadata and standards
 - ❑ PIDs
 - ❑ Protocols
 - ❑ Data processing
 - ❑ Data Analysis



Open Science Primers: getting you started on good practices



Open Access Basics

An Open Access primer to get you started



An RDM Handbook

A primer on managing your research data

- ❑ Essential information and tutorials on basic concepts
- ❑ Supporting material
- ❑ Support through NOADs and the Helpdesk

<https://www.openaire.eu/os-primers>

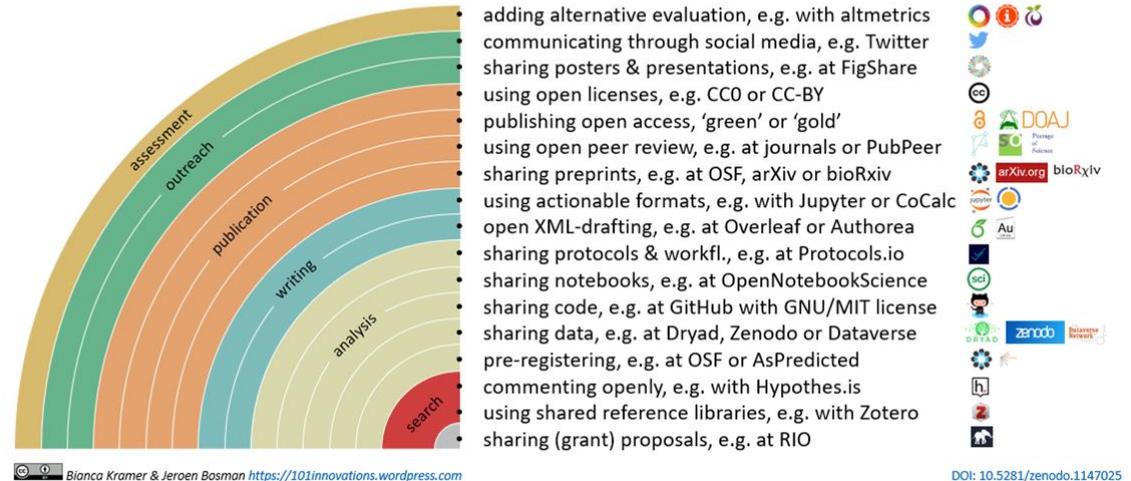
Learn and get informed 2/4



- ❑ Modules about Open Science subjects
- ❑ Self-paced e-learning option (badges)
- ❑ Anyone can add content

<https://www.fosteropenscience.eu/>

You can make your workflow more open by ...



- ❑ Best practices for open workflows
- ❑ Indicative tools

https://zenodo.org/record/1147025#.XrF_IKgzY2w

Learn and get informed 3/4

Top 10 FAIR Data & Software Things

about github repository download/cite license #top10fair

The Top 10 FAIR Data & Software Things are brief guides (stand alone, self paced training materials), called "Things", that can be used by the research community to understand how they can make their research (data and software) more FAIR (Findable, Accessible, Interoperable and Reusable). Each discipline/topic has its own specific list:

Nanotechnology

Astronomy

Linked Open Data

Imaging

Music

- FAIR in disciplines
 - Basic concepts
 - Best practices
 - Activities

<https://librarycarpentry.org/Top-10-FAIR/>

OpenUP HUB



- Review
- Assess
- Disseminate

<https://www.openuphub.eu/>

□ E-learning

- Under development – 2 courses open for enrollment
- 10 modules
 - Rationale
 - Learning Outcomes
 - Resources
 - Tools
 - Research Articles and Reports
 - Key Posts
 - Other

**OPEN
SCIENCE
MOOC**
FREE | OPEN | LEARNING

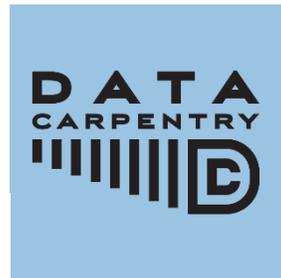
<https://opensciencemooc.eu/>



The screenshot shows the top navigation bar of the EPO e-learning centre. On the left is the EPO logo with the text 'Europäisches Patentamt', 'European Patent Office', and 'Office européen des brevets'. To its right is the text 'e-learning centre'. Further right is a language selector set to 'English (en)'. Below the language selector is a dropdown menu for 'All courses' and a search box with a magnifying glass icon. A dark grey bar below the navigation bar contains the text 'You are currently using guest access (Log in)'. The main content area features a banner with a gold textured background on the left containing the text 'Lectures on PCT at the EPO' and a white background on the right containing the text 'How to get a European patent EURO-PCT'.

- ❑ Seminars on Patent Cooperation Treaty
- ❑ Recent developments
- ❑ How to fill an international application

<https://e-courses.epo.org/course/view.php?id=178>



- ❑ Essential skills for Data Management and Software development
- ❑ 3 streams: Software, Data, Library
- ❑ Open courses
- ❑ Pedagogy
- ❑ Instructors

<https://carpentries.org/>

Plan – Costs for management and curation

THE COST OF DATA MANAGEMENT

HOW TO CALCULATE COSTS?

HOW TO USE THIS COSTING TOOL?

ESTIMATING COSTS RDM TOOL

ACTIVITY	COMMENTS AND SUGGESTIONS	✓	COST
Data description <ul style="list-style-type: none"> Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc.? Are labels consistent? Do textual data like interview transcripts need description of context, e.g. included as a heading page? 	<ul style="list-style-type: none"> if data description is carried out as part of data creation, data input or data transcription – low or no additional cost if needed to be added afterwards – higher cost codebooks for datasets can often be easily exported from software packages 		
Data cleaning <ul style="list-style-type: none"> Do quantitative data need to be cleaned, checked or verified before sharing, e.g. check validity of codes used, check for anomalous values? Will data match documentation, e.g. same number of variables, cases, records, files? Does textual information in data need to be spell-checked? 	<ul style="list-style-type: none"> if carried out as part of data entry and preparation before data analysis – low or no additional cost if needed afterwards – higher cost 		
Documentation <ul style="list-style-type: none"> Do you have documentation for the data that describes the context and methodology of how data were gathered, created, processed and quality controlled? 	<ul style="list-style-type: none"> often essential contextual and methods documentation will be written up in publications and reports if all data creation steps are well documented and documentation is kept well organised during research – low or no additional cost if documentation to be written or compiled specifically afterwards – higher cost 		
Metadata <ul style="list-style-type: none"> Do structured metadata need to be created when data are shared via a data centre or archive, e.g. completing a deposit form for the UK Data Archive? 	<ul style="list-style-type: none"> completing a UK Data Archive deposit form may take one to two hours other data centres will have their own metadata forms 		
Formatting and organising <ul style="list-style-type: none"> Are your data files, spreadsheets, interview transcripts, records etc. all in a uniform format or style? Are files, records and items in the 	<ul style="list-style-type: none"> if planned beforehand by developing templates and data entry forms for individual data files (transcripts, spreadsheets, databases) and by constructing clear file structures – low or no additional cost 		

Estimating costs RDM tool

DMP PHASE	ACTIVITY	COMMENTS AND SUGGESTIONS	COSTS
<i>Preparing</i>	Make a Data Management Plan	<p>Make a DMP before you start creating data; make decisions about managing your data. You can find the template for H2020 DMPs here.</p> <p>Check if there is a department within your organization to support data management planning.</p>	2 hrs to 2 days, depending on the complexity of your project
<i>1. Data Collection</i>	Acquiring External datasets <p>Do you plan to use existing data, and is the data available at a commercial partner?</p>	<p>Your library may be able to help you acquire a license to a crucial database</p> <p>In research data repositories, data can be available at no or low costs</p>	<i>Example:</i> A faculty licence on a database for macro-economic analysis: €18.000/y
<i>1. Data Collection</i>	Formatting and organising <p>Are your data files, spreadsheets, measurements, interview transcripts, records etc. all in a uniform format or style?</p>	<p>If planned beforehand by developing templates and data entry forms for individual data files (transcripts, spreadsheets, databases) and by constructing clear file structures – low or no additional cost</p> <p>If needed afterwards – higher cost</p>	Per project organize style, format, names can be done by a student assistant at level 1* salary or data manager at level 2* salary

<https://ukdataservice.ac.uk/media/622368/costingtool.pdf>

Plan – Costs for publishing and IP application

OPEN APC

ABOUT OLAP SERVER GITHUB INTACT

The Open APC initiative releases data sets on fees paid for Open Access journal articles by universities and research institutions under an Open Database License.

Select a region from the map or use the list below to view APC data for individual institutions.

You may also browse the following aggregated collections:

- The [OpenAPC](#) data set is an aggregation of all institutional APC data.
- The [Transformative Agreements](#) dataset contains articles published under transformative deals such as the Springer



- Browse from the list
- Contribute



- Select a type of protection
 - Copyright, design, patent, trade mark, utility
 - Application, registration, annuity, examination

<https://www.intact-project.org/openapc/>

<http://www.latinamerica-ipr-helpdesk.eu/node/26>



- ❑ First DMP tool
- ❑ Big collection of DMP templates and guidance
- ❑ Fee to use in institution

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



- ❑ Smart DMPs
- ❑ Literacy on FAIR principles and FAIR assessment
- ❑ Mostly for Life Science

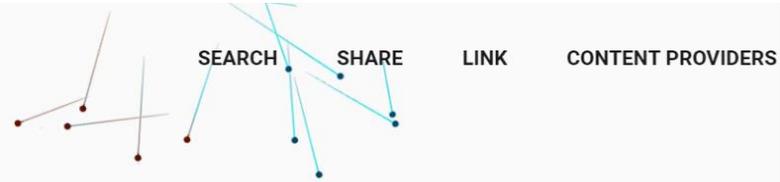
<https://ds-wizard.org/>



- ❑ Machine-actionable
- ❑ FAIR DMPs
- ❑ Publish DMPs
- ❑ Collaborative writing

<https://argos.openaire.eu/>

Find/ Collect



Search in OpenAIRE for scholarly works



32mi publications, 2mi research data, 3mi other research products, 105K software from 16K content providers and 18 funders linked together for an integrated discovery of research outcomes



Explore



Share



Link



Browse the service: <https://explore.openaire.eu/>



BIP! Finder

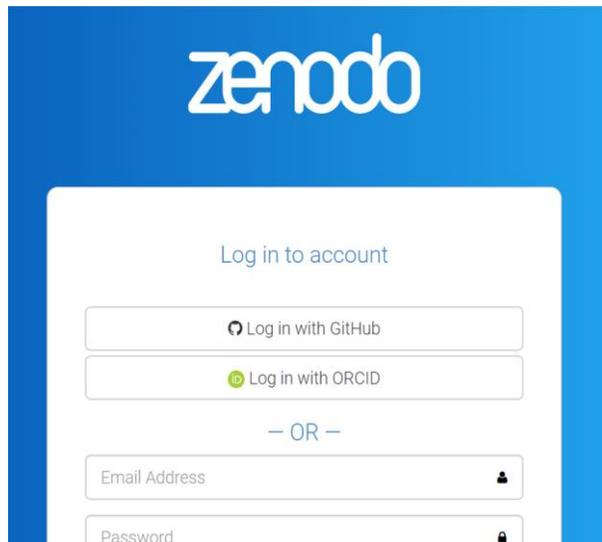
Enter keywords to retrieve articles...

Find!

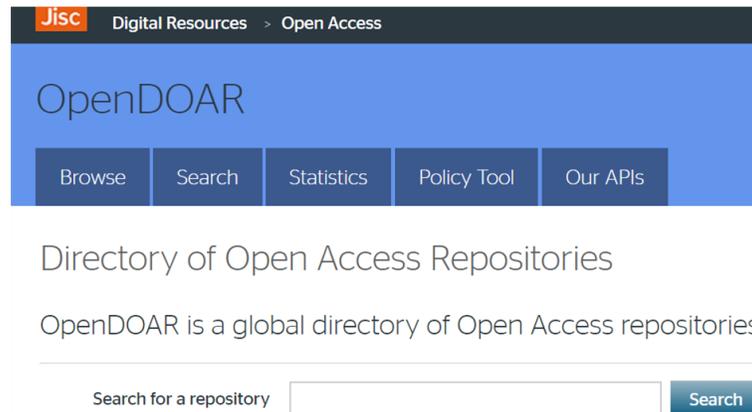
Order by: Popularity Influence Year Combine with keyword relevance: Yes No

<https://bip.imsi.athenarc.gr/>

Find/ Collect & Deposit



<https://zenodo.org/>

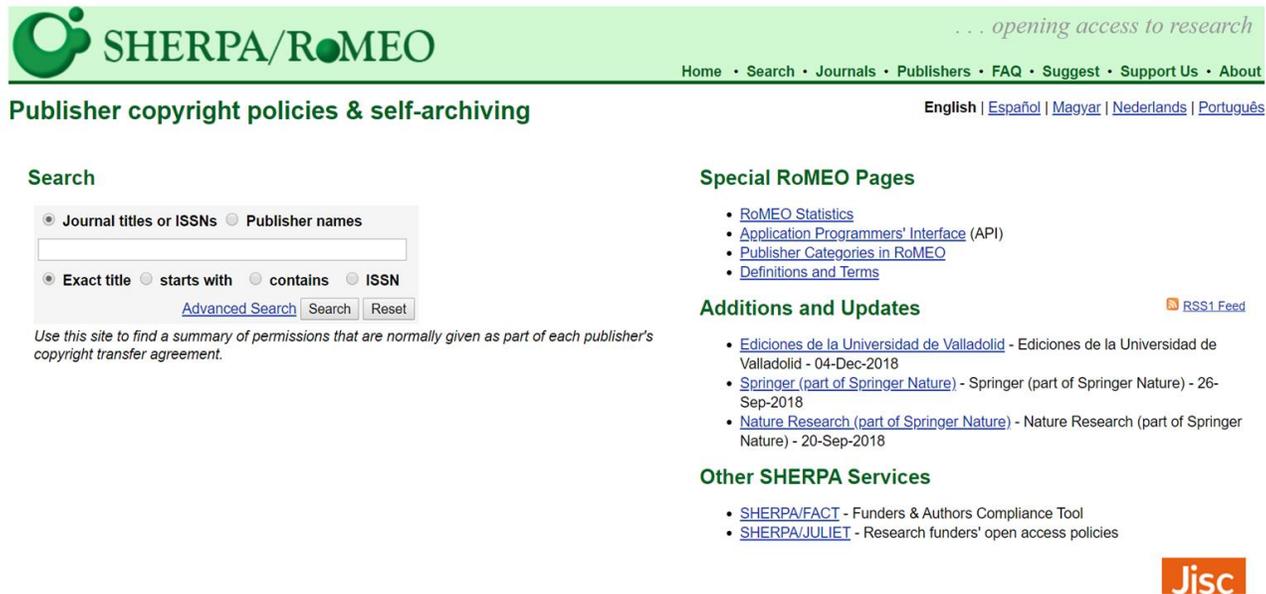


<https://v2.sherpa.ac.uk/opendoar/>



<https://www.re3data.org/>

Publish in OA – articles policies



The screenshot shows the SHERPA/RoMEO website. At the top left is the SHERPA/RoMEO logo. To its right is the tagline "... opening access to research". Below the logo is a navigation menu with links: Home, Search, Journals, Publishers, FAQ, Suggest, Support Us, and About. Below the navigation menu is the page title "Publisher copyright policies & self-archiving" and a language selection menu with options: English, Español, Magyar, Nederlands, and Português. The main content area is divided into three columns. The left column is titled "Search" and contains a search form with radio buttons for "Journal titles or ISSNs" and "Publisher names", a search input field, and radio buttons for "Exact title", "starts with", "contains", and "ISSN". Below the search form are buttons for "Advanced Search", "Search", and "Reset". Below the search form is a note: "Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement." The middle column is titled "Special RoMEO Pages" and contains a list of links: "RoMEO Statistics", "Application Programmers' Interface (API)", "Publisher Categories in RoMEO", and "Definitions and Terms". Below this is a section titled "Additions and Updates" with an "RSS1 Feed" icon and a list of links: "Ediciones de la Universidad de Valladolid - Ediciones de la Universidad de Valladolid - 04-Dec-2018", "Springer (part of Springer Nature) - Springer (part of Springer Nature) - 26-Sep-2018", and "Nature Research (part of Springer Nature) - Nature Research (part of Springer Nature) - 20-Sep-2018". Below this is a section titled "Other SHERPA Services" with a list of links: "SHERPA/FACT - Funders & Authors Compliance Tool" and "SHERPA/JULIET - Research funders' open access policies". The right column is empty. At the bottom right of the page is the Jisc logo.

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

<https://v2.sherpa.ac.uk/romeo/>

Publish in OA – data policies

A project by Center for Open Science

TOP GUIDELINES
TRANSPARENCY AND OPENNESS PROMOTION

TOP Standards ? ^

- Data Citation
- Data Transparency
- Analysis Code Transparency
- Materials Transparency
- Design and Analysis Guidelines
- Study Preregistration
- Analysis Plan Preregistration
- Replication
- Registered Reports

Search Journal Titles

Journal ↑	Total	Data Citation	Data Transparency	Analysis Code Transparency	Materials Transparency	Design and Analysis Guidelines
Attention, Perception, and Psychophysics  Springer Nature	14	2	1	1	1	1
Behavioral Development  American Psychological Association	1	0	0	0	0	0
EMBO Journal 	5	1	1	1	1	1
AEA Papers & Proceedings  American Economics Association	7	0	3	3	0	0
AEJ: Applied Economics  American Economics Association	7	0	3	3	0	0

Policies



FAIRsharing policies: A catalogue of data preservation, management and sharing policies from international funding agencies, regulators and journals.

[Contribute by adding a policy](#)

[Any problems? Please tell us!](#)

Search Policies

Search

Reset

Advanced

Metadata

Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

 github

 @twitter

 linkedin

 facebook

Arts and Humanities

- [Archaeology](#) 
- [Creative art and design](#) 
- [Heritage Studies](#) 
- [Historical and Philosophical Studies](#) 
- [History by Area](#) 
- [History](#) 
- [Information Services](#) 
- [Law](#) 
- [Music](#) 

Engineering

- [Architecture](#) 
- [Building Conservation](#) 

Life Sciences

- [Agricultural Economics](#) 
- [Agricultural Science](#) 
- [Animal pathology](#) 
- [Animal physiology](#) 
- [Biochemistry](#) 
- [Biodiversity](#) 
- [Bioengineering](#) 
- [Biogeography](#) 
- [Bioinformatics](#) 
- [Biology](#) 
- [Biomaterials](#) 
- [Biomechanics](#) 
- [Botany](#) 

Standards

FAIRsharing.org standards, databases, policies

Search all of FAIRsharing Standards Databases Policies Collections Add/Claim Content Stats Log in or Reg

Search Standards Search Reset Advanced

Help us curate and collate the databases and standards useful for COVID-19 research. Our COVID-19 Collection is growing, but not yet complete. Join our community curathon and tell us what's missing. Email contact@fairsharing.org to join!

Showing records 1 - 50 of 1401.

View as Table View as Grid

Sort by Name

Recommended Records

Recommended

Associated Publication?

No Publication Has Publication

Claimed?

No Maintainer Has Maintainer

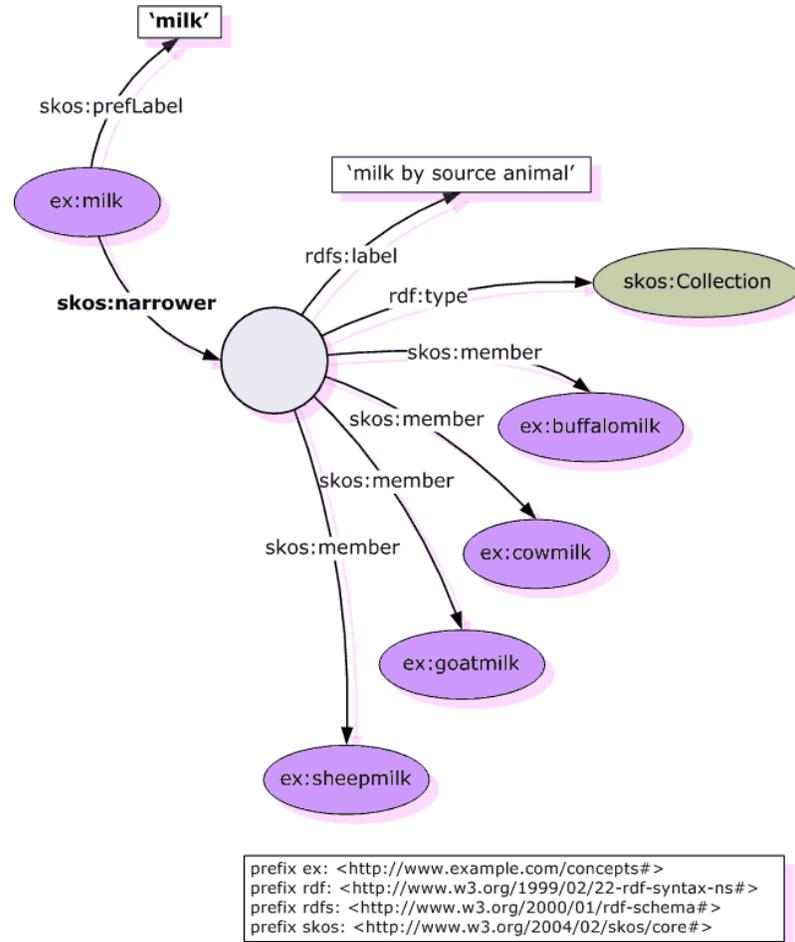
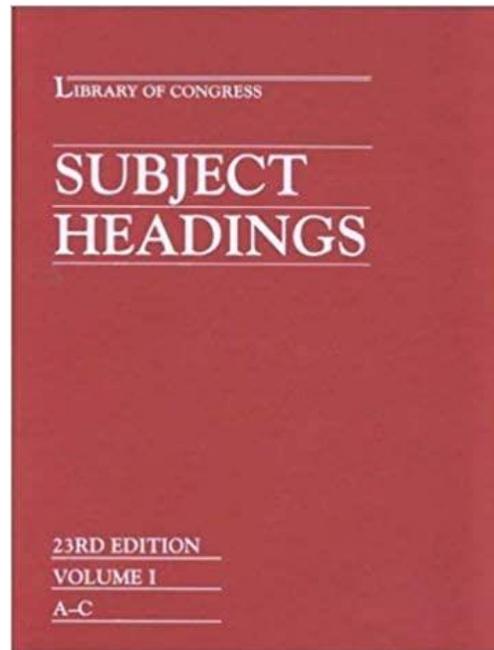
Record Status

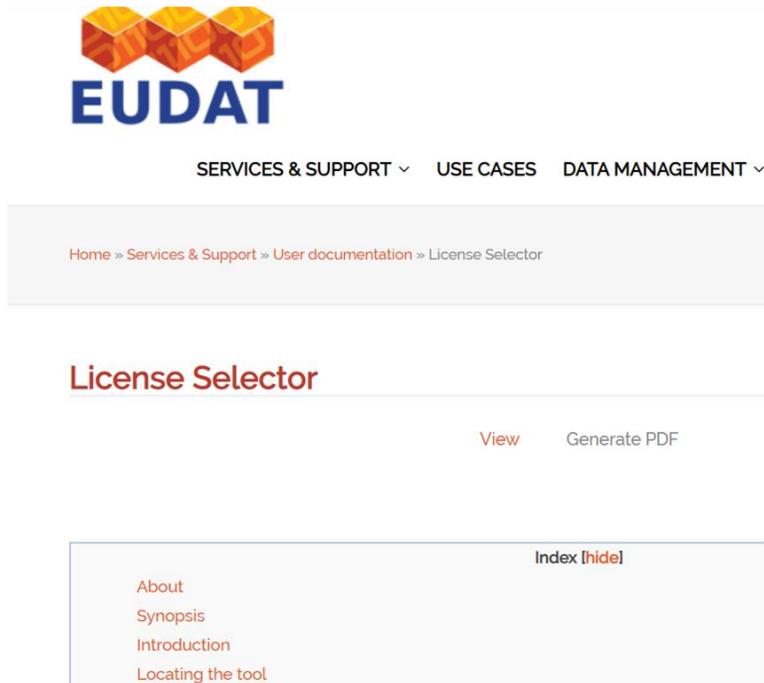
Uncertain Deprecated In development Ready

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
ABA Adult Mouse Brain	ABA	Standard	Neuroscience	Brain Gene Expression Brain Imaging	None	None	NeuroMorpho Org	None	None	None	R
Access to Biological Collection Data	ABCD	Standard	Biochemistry Biology Life Science	None	None	None	GBIF ALA IPT - GBIF Australia Repository GBIF Spain IPT - GBIF Spain Repository Canadensys IPT - GBIF Canadensys Repository SIB Colombia IPT - GBIF Colombia Repository Plus 1 more...	ABCCDNA ABCDEFG GOBN	None	TOWG Biodiversity Information Standards	R
Access to Biological Collection Databases Extended for Geosciences	ABCDEFG	Standard	Earth Science Geology Paleontology Soil Science	None	None	None	GeoCase Data Portal	ABCD XML	None	None	R

Vocabularies

- Ontologies
- Taxonomies
- Thesauri





The screenshot shows the EUDAT website's License Selector page. At the top is the EUDAT logo, followed by navigation links for SERVICES & SUPPORT, USE CASES, and DATA MANAGEMENT. A breadcrumb trail reads: Home » Services & Support » User documentation » License Selector. The main heading is "License Selector" with "View" and "Generate PDF" buttons. A sidebar on the left contains a table of contents with links for About, Synopsis, Introduction, and Locating the tool, and an "Index [hide]" link.

Choose an open source license

An open source license protects contributors and users. Businesses and savvy developers won't touch a project without this protection.

Which of the following best describes your situation?



I need to work in a community.

Use the **license preferred by the community** you're contributing to or depending on. Your project will fit right in.

If you have a dependency that doesn't have a license, ask its maintainers to **add a license**.



I want it simple and permissive.

The **MIT License** is short and to the point. It lets people do almost anything they want with your project, like making and distributing closed source versions.

Babel, **.NET Core**, and **Rails** use the MIT License.



I care about sharing improvements.

The **GNU GPLv3** also lets people do almost anything they want with your project, *except* distributing closed source versions.

Ansible, **Bash**, and **GIMP** use the GNU GPLv3.

Publications

Data

Software

Combined

Joinup Licensing Assistant (JLA)

Select licence terms below

Can	Must	Cannot	Compatible	Law	Support
Use/reproduce	Incl. Copyright	Hold liable	None N/A	EU/MS law	Strong Community
Distribute	Royalty free	Use trademark	Permissive	US law	Governments/EU
Modify/merge	State changes	Commerce	GPL	Licensor's law	OSI approved
Sublicense	Disclose source	Modify	Other copyleft	Other law	FSF Free/Libre
Commercial use	Copyleft/Share a.	Ethical clauses	Linking freedom	Not fixed/local	
Use patents	Lesser copyleft	Pub sector only	Multilingual	Venue fixed	
Place warranty	SaaS/network	Sublicense	For data		
	Include licence		For software		
	Rename modifs.				

or enter [licence SPDX identifier](#)

RESET

1. Select type of License



Content

Licenses for content

✓ (Selected)



Software

Licenses for software



Services

Licenses for services

2. Select the 2 licenses that you want to combine

-- select 1st license --

-- select 2nd license --

3. Compatibility results

You must first select the 2 licenses that you want to see the compatibility of

Persistent identifiers

Digital Objects



ARK

PURL

Researchers & Organisations



Other activities



Data processing – cleaning data

Facet / Filter Undo / Redo 13 Extract... Apply...

227 rows Extensions: Freebase

Show as: rows records Show: 5 10 25 50 rows < first < previous 1 - 10 next > last >

Filter:

1. Remove / rows
2. Text transform on 57 cells in column Column: `grel:value.replace(" ", "")`
3. Create new column Is Winner based on column Column by filling 227 rows with `grel:not(value.startsWith("****"))`
4. Create new column Year based on column Column by filling 57 rows with `grel:value[1,5]`
5. Text transform on 57 cells in column Column: `grel:value.substring(6)`
6. Text transform on 170 cells in column Column: `grel:value.substring(2)`
7. Fill down 170 cells in column Year
8. Split 227 cell(s) in column Column into several columns by separator
9. Split 227 cell(s) in column Column 2 into several columns by separator
10. Rename column Column 1 to actress
11. Rename column Column 2 1 to film
12. Rename column Column 2 2 to character

All	actress	film	character	Year	Is Winner
☆	1. [[Meena Kumar]] (winner)	"[[Baju Bawra (1952 film) Baju Bawra]]"	Gauri	1954	true
☆	2. [[Meena Kumar]] (winner)	"[[Parineeta (1953 film) Parineeta]]"	Lolita	1955	true
☆	3. [[Kamini Kaushal]] (winner)	"[[Bira] Bahu]]"	Bira] Chakravorty	1956	true
☆	4. [[Geeta Bai]]	"[[Vachan]]"	Kamla	1956	false
☆	5. [[Meena Kumar]]	"[[Azaad (1955 film) Azaad]]"	Shobha	1956	false
☆	6. [[Nutan]] (winner)	"[[Seema (1955 film) Seema]]"	Gauri	1957	true
☆	7. [[Nargis]] (winner)	"[[Mother India]]"	Radha	1958	true
☆	8. [[Vyjayanthimala]] (winner)	"[[Madhumati]]"	Madhumati	1959	true
☆	9. [[Meena Kumar]]	"[[Sahara (1958 film) Sahara]]"	Leela	1959	false
☆	10. [[Vyjayanthimala]]	"[[Sadhna]]"	Champabai	1959	false



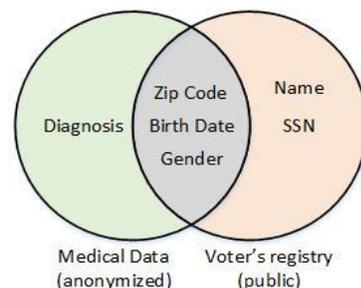
Data processing - anonymisation

- Micro data often reveal important private information, e.g., medical condition of a person
 - Individuals are afraid to provide their data
 - Companies are afraid to share data with experts
 - GDPR makes a strict protection scheme obligatory



<https://amnesia.openaire.eu/>

- The key idea in anonymization is that identifying information is removed from the published data, so no sensitive information can be attributed to a person – not even after data linking



- The aim of anonymization methods is to allow sharing such data, without compromising the privacy of the users.

Data analysis



Data Sharing



Assessing FAIRness

How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- A persistent identifier is assigned to your data
- There are rich metadata, describing your data
- The metadata are online in a searchable resource e.g. a catalogue or data repository
- The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- Following the persistent ID will take you to the data or associated metadata
- The protocol by which data can be retrieved follows recognised standards e.g. http
- The access procedure includes authentication and authorisation steps, if necessary
- Metadata are accessible, wherever possible, even if the data aren't

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- Data is provided in commonly understood and preferably open formats
- The metadata provided follows relevant standards
- Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- Qualified references and links are provided to other related data

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- The data are accurate and well described with many relevant attributes
- The data have a clear and accessible data usage license
- It is clear how, why and by whom the data have been created and processed
- The data and metadata meet relevant domain standards

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir



SATIFYD

Self-Assessment Tool to Improve the FAIRness of Your Dataset

Welcome to SATIFYD: the DANS Self-Assessment Tool to Improve the FAIRness of Your Dataset. This tool will show you how FAIR (Findable, Accessible, Interoperable, Reusable) your dataset is and will provide you with tips to score (even) higher on FAIRness. Ideally, you use this tool prior to the deposit in EASY.

The 12 questions touch upon the FAIR data principles but do not strictly follow them. While answering the questions, the score per letter will be displayed underneath each letter. The more 'blue' the letters get, the more FAIR your dataset is. An overall score is provided at the end of the page.

Some questions are posed more than once (e.g. on metadata and data standards or usage licences), because the topics are relevant in more than one letter.

Want to know more? Please click [here](#)

If you have any questions, please let us know by sending an e-mail

FINDABLE

1. Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data?

2. Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset?

- Controlled vocabularies
- Taxonomies (thesauri)
- Ontologies
- There are no standards for my discipline

3. Did you provide rich and detailed additional documentation?

- Readme file
- Versioning
- Provenance

ACCESSIBLE

4. Is the metadata publicly accessible even if the data is no longer available?

5. Does your dataset contain personal data?

6. Which of the usage licenses provided by EASY did you choose in order to comply with the access rights attached to the data?

INTEROPERABLE

<https://zenodo.org/record/1065991#.XqN-XGgzY2w>

<https://satifyd.dans.knaw.nl/>

Assessing FAIRness

FAIR Evaluation Services

Resources and guidelines to assess the FAIRness of digital resources.



Import MI Tests

Import Maturity Indicators Tests as YAML [smartAPI](#) interface annotation

Get started



Create collections

Assemble Maturity Indicators Tests into community centered collections

Get started



Evaluate resources

Evaluate resources FAIRness against Collections of Maturity Indicator Tests

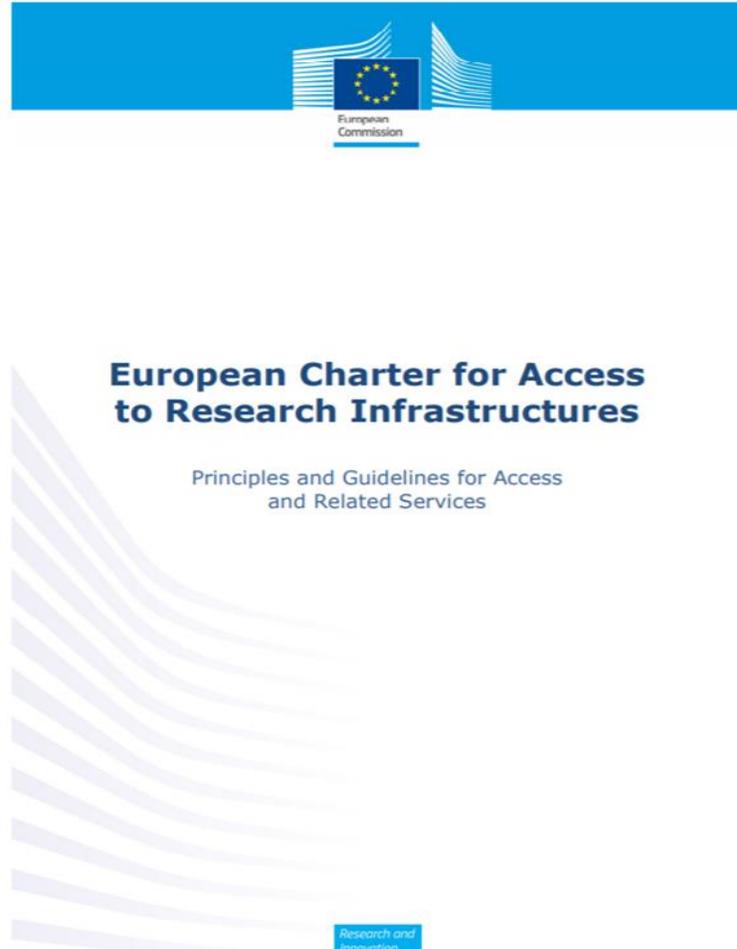
Get started

<https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/#/>

Service Providers

Overview

- Policies
- Building data services
- FAIR-aligned services
- Repositories



https://ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf

- Access policy
- Access mode
 - Excellence driven
 - Market driven
 - Wide Access
- Access restrictions and processes
- Quality Assurance
- Research Data Management Plan
- Education and Training

Part I: The Plan S Principles

"With effect from 2021*, all scholarly publications on the results from research funded by public or private grants provided by national, regional and international research councils and funding bodies, must be published in Open Access Journals, on Open Access Platforms, or made immediately available through Open Access Repositories without embargo."

*For funders agreeing after January 2020 to implement Plan S in their policies, the start date will be one year from that agreement

In addition:

- 01** Authors or their institutions retain copyright to their publications. All publications must be published under an open license, preferably the Creative Commons Attribution license (CC BY), in order to fulfil the requirements defined by the [Berlin Declaration](#).
- 02** The Funders will develop robust criteria and requirements for the services that high-quality Open Access journals, Open Access platforms, and Open Access repositories must provide.
- 03** In cases where high-quality Open Access journals or platforms do not yet exist, the Funders will, in a coordinated way, provide incentives to establish and support them when appropriate; support will also be provided for Open Access infrastructures where necessary.
- 04** Where applicable, Open Access publication fees are covered by the Funders or research institutions, not by individual researchers; it is acknowledged that all researchers should be able to publish their work Open Access.
- 05** The Funders support the diversity of business models for Open Access journals and platforms. When Open Access publication fees are applied, they must be commensurate with the publication services delivered and the structure of such fees must be transparent to inform the market and funders potential standardisation and capping of payments of fees.
- 06** The Funders encourage governments, universities, research organisations, libraries, academies, and learned societies to align their strategies, policies, and practices, notably to ensure transparency.
- 07** The above principles shall apply to all types of scholarly publications, but it is understood that the timeline to achieve Open Access for monographs and book chapters will be longer and requires a separate and due process.
- 08** The Funders do not support the 'hybrid' model of publishing. However, as a transitional pathway towards full Open Access within a clearly defined timeframe, and only as part of transformative arrangements, Funders may contribute to financially supporting such arrangements.
- 09** The Funders will monitor compliance and sanction non-compliant beneficiaries/grantees.
- 10** The Funders commit that when assessing research outputs during funding decisions they will value the intrinsic merit of the work and not consider the publication channel, its impact factor (or other journal metrics), or the publisher.

<https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/>

Practical Policy Template

1. Contextual metadata extraction
2. Data access control
3. Data backup
4. Data format control
5. Data retention
6. Disposition
7. Integrity (including replication)
8. Notification
9. Restricted searching
10. Storage cost reports
11. Use agreements



Building data services

DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment



Four steps to effective data management

Home **Methodology**

Background The Data Asset Framework recommends that audits of research data assets proceed as a four step process:

Methodology In the planning stage the purpose and scope of the audit is defined. Preliminary research is conducted and meetings scheduled to optimise time spent with the organisation's staff. The purpose of the second stage, identifying research data, is to establish what data assets exist and classify them according to the purposes of the survey. The classification step determines the scope of further activities, as only selected aspects will be assessed in greater detail. The information collected in Stage 3 will assist auditors to identify weaknesses in data policy and current data creation and curation procedures. This will provide the basis of recommendations in the final stage. The knowledge gained from the assessment survey will enable organisations to plan improvements in data management.

Users

Findings

Documents

Launch

Contact Us The current version of the [methodology](#) is available to download. An

The diagram illustrates a four-step process for effective data management, shown as a staircase with an upward-pointing arrow. The steps are:

- 1 PLANNING THE AUDIT
- 2 IDENTIFYING AND CLASSIFYING ASSETS
- 3 ASSESSING MANAGEMENT OF DATA ASSETS
- 4 REPORTING AND RECOMMENDATIONS

CESSDA-CDM

Introduction

Background of CESSDA

Background of the model

Model Components

Glossary

Versions

Part 1 - CRA1: Organisational Infrastructure

Part 2 - CRA2: Digital Object

Management

Part 3 - CRA3: Technical Infrastructure

FAIR-aligned services

B2FIND / Findable				
FAIR: Findable	reduce	respect	enable	comment
F1. (meta)data are assigned a <u>globally unique and eternally persistent identifier</u> .		F1		B2FIND relies on the harvested repositories to attach a PID to its records and expose that as part of the metadata. If the provided metadata contains a PID B2FIND represents this PID and uses it to link the harvested metadata to the original data object in the repository.
F2. data are described with <u>rich metadata</u> .			F2	B2FIND relies on the provided metadata by the harvested repositories. B2FIND does not further enrich metadata. enhances F2 with citation metadata
F3. (meta)data are <u>registered or indexed in a searchable resource</u> .			F3	By being on B2FIND the (meta)data is registered and indexed and can be found through search.
F4. metadata <u>specify</u> the data identifier.			F4	In the current data infrastructure landscape B2FIND takes the position of a metadata registry and indexer. It extends

<https://zenodo.org/record/3688762#.XrE4FagzY2y>

Persistent identifiers

Digital Objects



ARK

PURL

Researchers & Organisations



Other activities



Assessing FAIRness of data

FAIR Data Maturity Model: specification and guidelines

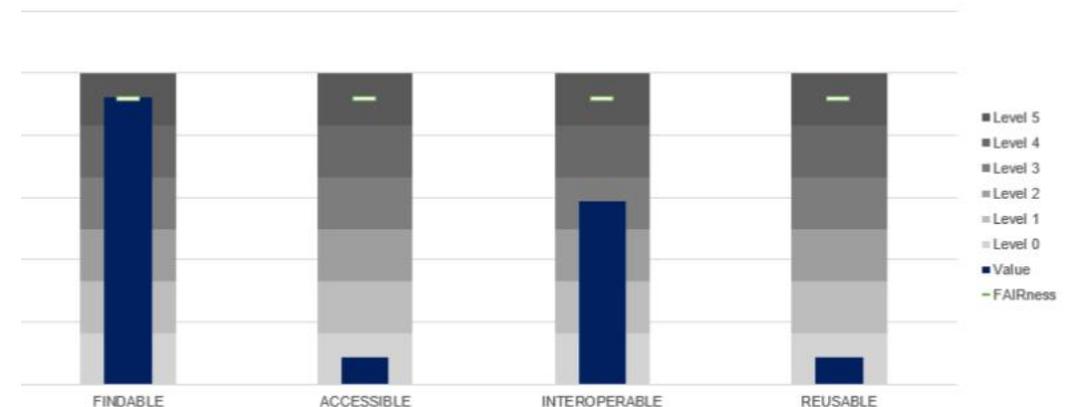
[Home](#) » [Data Management](#) » FAIR Data Maturity Model: Specification And Guidelines



Priority	Principle				Grand Total
	Findable	Accessible	Interoperable	Reusable	
Essential	7	8	0	5	20
Important	0	3	7	4	14
Useful	0	1	5	1	7
Grand Total	7	12	12	10	41

What level describes your digital object?

Level 0	Not FAIR
Level 1	FAIR essential criteria only
Level 2	FAIR essential criteria + 50 % of important criteria
Level 3	FAIR essential criteria + 100% of important criteria
Level 4	FAIR essential criteria + 100% of important criteria + 50% of useful criteria
Level 5	FAIR essential criteria + 100% of important criteria + 100% of useful criteria



Repositories - Overview

- ❑ Registries
 - ❑ OpenDOAR
 - ❑ Re3data
 - ❑ Policy
 - ❑ Metadata / Interoperability
 - ❑ Licenses
 - ❑ Repository
 - ❑ Metadata
 - ❑ Content
 - ❑ FAIR Metrics
 - ❑ Certification
-

Registries

Jisc Digital Resources > Open Access

OpenDOAR

Browse Search Statistics Policy Tool Our APIs Suggest Admin

Directory of Open Access Repositories

OpenDOAR is a global directory of Open Access repositories and their policies.

Search for a repository Search

OpenDOAR is the quality-assured global directory of academic open access repositories. It enables the identification, browsing and search for repositories, based on a range of features, such as location, software or type of material held [more...](#)

re3data.org Search Browse Suggest Resources Contact DataCite

Search... Search

← Previous 1 2 3 4 5 6 7 ... 98 Next → Sort by ▾

Found 2428 result(s)

CancerData.org

Sharing data for cancer research

Subject(s) Basic Biological and Medical Research Medicine Biology Life Sciences

Content type(s) Standard office documents Databases Images Structured graphics Scientific and statistical data formats

Raw data Plain text Archived data other

Country Netherlands

The CancerData site is an effort of the Medical Informatics and Knowledge Engineering team (MIKE for short) of Maastricht Clinic, Maastricht, The Netherlands. Our activities in the field of medical image analysis and data modelling are visible in a number of projects we are running. CancerData is offering several datasets. They are grouped in collections and can be public or private. You can search for public datasets in the NBIA (National Biomedical Imaging Archive) image archives without logging in.

Repository Policies

Metadata Policy (for information describing items in the repository)

- Anyone may access the metadata free of charge.
- The metadata may be re-used in any medium without prior permission for not-for-profit purposes provided the OAI Identifier or a link to the original metadata record are given.
- The metadata must not be re-used in any medium for commercial purposes without formal permission.

Data Policy (for full-text and other full data items)

- Anyone may access full items free of charge.
- Single copies of full items can be reproduced, and displayed or performed in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge.
- Full items must not be sold commercially in any format or medium without formal permission of the copyright holders.

Content Policy (for document types and datasets)

- This is an institutional or departmental repository OR Multi-institution subject-based repository [list subjects]:
- The repository holds all types of items OR The repository only permits the following item types [list]
- All items are individually tagged with their peer-review status and publication status.

Submission Policy (concerning depositors, quality and copyright)

- Items may only be deposited by accredited members of the organisation, or their delegated agents.
- Authors may only submit their own work for archiving.
- The administrator only vets items for the exclusion of spam
- The validity and authenticity of the content of submissions is the sole responsibility of the depositor.
- Items may not be deposited until any publishers' or funders' embargo period has expired.
- Any copyright violations are entirely the responsibility of the authors/depositors.
- If the repository receives proof of copyright violation, the relevant item will be removed immediately.

Preservation Policy

- Items will be retained indefinitely.
- The repository will try to ensure continued readability and accessibility.
- The repository regularly backs up its files according to current best practice.
- Items may not normally be removed from the repository.
- Acceptable reasons for withdrawal include:
 - Proven copyright violation or plagiarism
 - Legal requirements and proven violations
 - National Security
 - Falsified research
- Withdrawn items are not deleted per se, but are removed from public view.
- Withdrawn items' identifiers/URLs are retained indefinitely.
- URLs will continue to point to 'tombstone' citations, to avoid broken links and to retain item histories.

Repositories Interoperability

- ❑ Literature
- ❑ Data
- ❑ CRIS
- ❑ (Software)



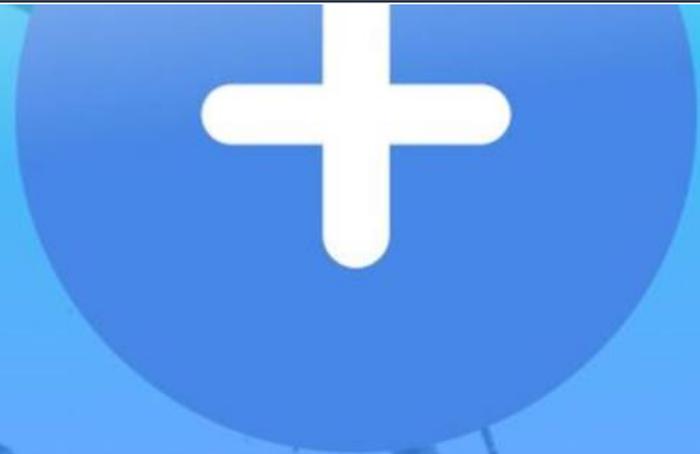
Current Guidelines

- [OpenAIRE Guidelines for Literature Repositories](#)
- [OpenAIRE Guidelines for Data Archives](#)
- [OpenAIRE Guidelines for CRIS Managers](#)
- [Draft OpenAIRE Guidelines for Software Repository Managers](#)
- [Draft OpenAIRE Guidelines for Other Research Products](#)

The guidelines specifically provide guidance on how to specify:

- Access right
- Funding information
- Related publications, datasets, software etc..

Validator



Openaire Validator

OpenAIRE Validator functionalities are now available through the [provide.openaire.eu](https://www.openaire.eu) dashboard

[TRY NOW](#)

<https://www.openaire.eu/validator/>

Certification

- ❑ R0. Context
- ❑ R1. The repository has an explicit mission to provide access to and preserve data in its domain.
- ❑ R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.
- ❑ R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.
- ❑ R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.



CoreTrustSeal Requirements v02.00-2020-2022 (doi:10.5281/zenodo.3638211)

2. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

Response

Guidance:

Repositories must have an appropriate rights model covering data access and use, communicate about them with users, and monitor compliance. This Requirement relates to the access regulations and applicable licenses set by the data repository itself, as well as any codes of conduct that are generally accepted in the relevant sector for the exchange and proper use of knowledge and information. Evidence should demonstrate that the repository has sufficient controls in place according to the access criteria of their data holdings, as well as evidence that any relevant licenses or processes are well managed.

For this Requirement, please describe:

- License agreements in use.
- Conditions of use (Intellectual Property Rights, distribution, intended use, protection of sensitive data, etc.).
- Documentation on measures in the case of noncompliance with conditions of access and use.

Note that if all data holdings are completely public and without conditions imposed on users—such as attribution requirements or agreement to make secondary analysis openly available—then it can simply be stated.

The ethical and privacy provisions that impact on licenses are dealt with in R4 (Confidentiality/Ethics). Assurance that deposit licenses provide sufficient rights for the repository to maintain, preserve, and offer access to data should be covered under R10 (Preservation Plan).

<https://www.coretrustseal.org/why-certification/requirements/>

- ❑ R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.
- ❑ R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).
- ❑ R7. The repository guarantees the integrity and authenticity of the data.
- ❑ R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.



CoreTrustSeal Requirements v02.00-2020-2022 (doi:10.5281/zenodo.3638211)

Digital Object Management

7. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

Response

Guidance:

The repository should provide evidence to show that it operates a data and metadata management system suitable for ensuring integrity and authenticity during the processes of ingest, archival storage, and data access. This Requirement covers the entire data lifecycle within the repository.

To protect the integrity of data and metadata, any intentional changes to data and metadata should be documented, including the rationale and originator of the change. Measures should be in place to ensure that unintentional or unauthorized changes can be detected and correct versions of data and metadata recovered.

Authenticity covers the degree of reliability of the original deposited data and its provenance, including the relationship between the original data and that disseminated, and whether or not existing relationships between datasets and/or metadata are maintained.

For this Requirement, responses on data integrity should include evidence related to the following:

- Description of checks to verify that a digital object has not been altered or corrupted (i.e., fixity checks) from deposit to use.
- Documentation of the completeness of the data and metadata.
- Details of how all changes to the data and metadata are logged.
- Description of version control strategy.
- Usage of appropriate international standards and conventions (which should be specified).

Evidence of authenticity management should relate to the following questions:

- Does the repository have a strategy for data changes? Are data producers made aware of this strategy?
- Does the repository maintain provenance data and related audit trails?
- Does the repository maintain links to metadata and to other datasets? If so, how?
- Does the repository compare the essential properties of different versions of the same file? How?
- Does the repository check the identities of depositors?

- ❑ R9. The repository applies documented processes and procedures in managing archival storage of the data.
- ❑ R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
- ❑ R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality related evaluations.
- ❑ R12. Archiving takes place according to defined workflows from ingest to dissemination.



CoreTrustSeal Requirements v02.00-2020-2022 (doi:10.5281/zenodo.3638211)

10. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

Response

Guidance:

The repository, data depositors, and Designated Community need to understand the level of responsibility undertaken for each deposited item in the repository. The repository must have the rights to undertake these responsibilities. Procedures must be documented and their completion assured.

For this Requirement, responses should include evidence related to the following questions:

- Does the repository have a documented approach to preservation?
- Is the level of responsibility for the preservation of each item understood? How is this defined?
- Are plans related to future migrations or similar measures to address the threat of obsolescence in place?
- Does the contract between depositor and repository provide for all actions necessary to meet the responsibilities?
- Is the transfer of custody and responsibility handover clear to the depositor and repository?
- Does the repository have the rights to copy, transform, and store the items, as well as provide access to them?
- Are actions relevant to preservation specified in documentation, including custody transfer, submission information standards, and archival information standards?
- Are there measures to ensure these actions are taken?

Rights concerning data access and use, and the monitoring of their compliance should be covered under R2 (Licenses).

Certification

- ❑ R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
- ❑ R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.
- ❑ R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
- ❑ R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.



CoreTrustSeal Requirements v02.00-2020-2022 (doi:10.5281/zenodo.3638211)

16. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

Response

Guidance:

The repository should analyze potential threats, assess risks, and create a consistent security system. It should describe damage scenarios based on malicious actions, human error, or technical failure that pose a threat to the repository and its data, products, services, and users. It should measure the likelihood and impact of such scenarios, decide which risk levels are acceptable, and determine which measures should be taken to counter the threats to the repository and its Designated Community. This should be an ongoing process.

For this Requirement, please describe:

- Your IT security system, employees with roles related to security (e.g., security officers), and any risk analysis tools (e.g., DRAMBORA¹) you use.
- What levels of security are required, and how these are supported.
- Any authentication and authorization procedures employed to securely manage access to systems in use (e.g., Shibboleth, OpenAthens).

The storage processes and technical infrastructure that utilize these security measures should be covered in R9 (Documented storage procedures) and R15 (Technical Infrastructure), respectively.

Wrap-Up

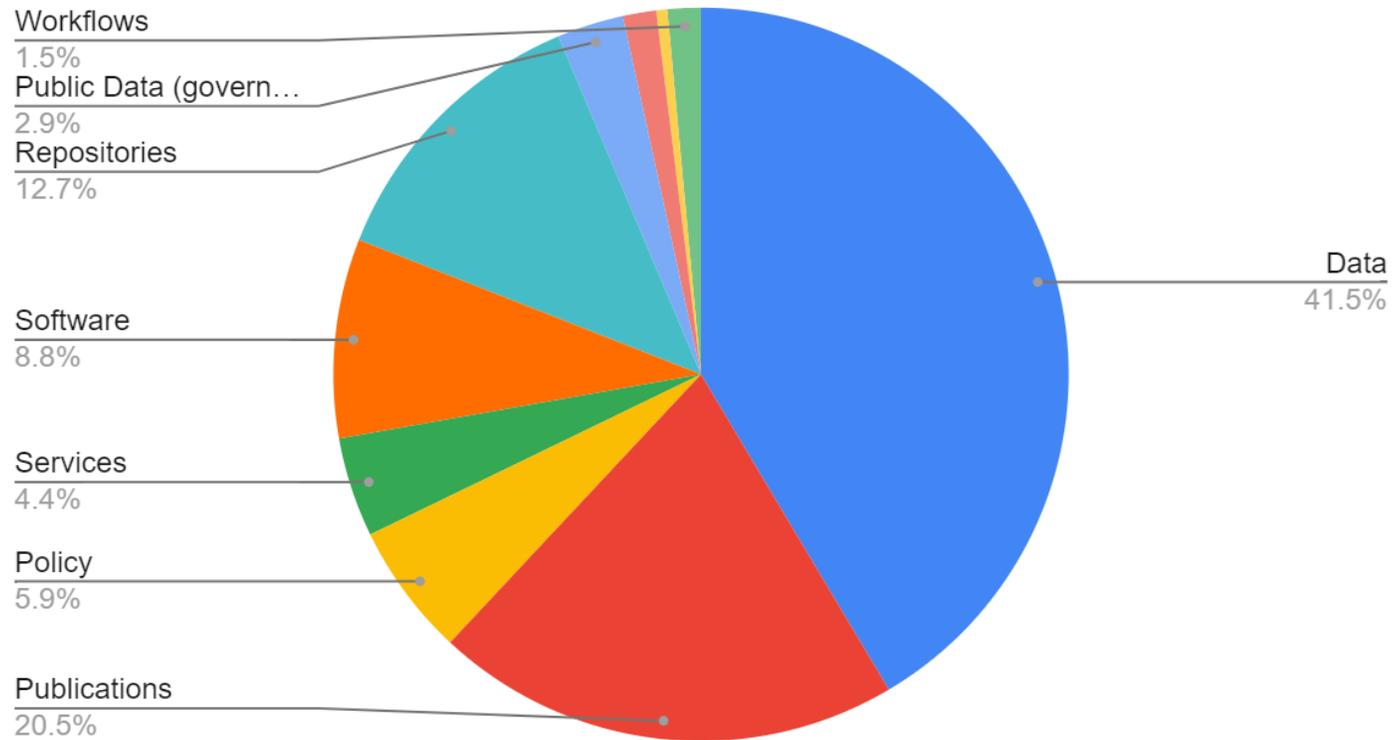
D4.3 Mapping of legal, procedural, technical tools

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Name	Description	Published by	Type (guidelines, tool, model)	Use (certification, decision making)	Users (Developer or End-users: e.g. RPOs = Research)	Tags	Research output (publications, data, software, services, workflows, articles)	Focus (e.g. RDM, FAIR)	License	Development stage - just for tools (integrated tool, primary)	Development stage (concept / pilot /)	Geographic area of usage (if the tool is designed for /)	Link
A collection of Citizen Science Guidelines and publications	The collection pays particular attention to six components of Responsible Research and Innovation (RRI) - Governance, Science Education, Ethics, Open Access, Gender and Public Engagement. A reference to these six components of RRI is made in order to explore how they are relevant to and addressed by citizen science. For this reason, the RRI components have been adopted as key criteria to sort the review presented in this collection.	European Citizen Science Association	Guidelines	Support; Decision making	CS practitioners; researchers; decision makers	Citizen science; EU; Open access;	Publications	Citizen science	Each set of guidelines is under its own license	-	Operational	EU	https://ecsa.citizen-science.net/blog/collection-citizen-science-guidelines-and-publications
A design framework and exemplar metrics for FAIRness.	This paper proposes a general framework and early stage indicators to measure the FAIRness of data.	Mark D. Wilkinson; Susanna-Assunta Sansone; Erik Schultes; Peter Doorn; Luiz Olavo Bonino da Silva Santos; Michel Dumontier	Guidelines & Policies	Support	RPOs; RFOs; Service Providers	Research data; FAIR data; Metrics	Data	FAIR	CC-BY 4.0 International license	-	Concept	International	https://www.biorxiv.org/content/10.1101/225490v3
ADA-M Automatable Discovery and Access Matrix	The Automatable Discovery and Access Matrix (ADA-M) provides a standardized way to unambiguously represent the conditions related to data discovery and access. By adopting ADA-M, data custodians can generally describe what their data are (the Header section), who can access them (the Permissions section), terms related to their use (the Terms section), and special conditions (the Meta-Conditions). By doing so, data custodians can	GA4GH - Global Alliance for Genomics & Health IRDIRC - International Rare Disease Research Consortium	Model	Support	RPOs; Service providers	OA policies; Open research data; Data sharing; Policy standardisation	Data	Discovery; Data sharing; Metadata	Open	Primary tool	Pilot	International	https://github.com/ga4gh/ADA-M
Amnesia	Amnesia is a data anonymization tool, that allows to remove identifying information from data. Amnesia not only removes direct identifiers like names, SSNs etc but also transforms secondary identifiers like birth date and zip code so that individuals cannot be identified in the data.	OpenAIRE	Tool	Support	Researchers; RPOs	research data; personal data; management; anonymisation	Data	Data processing; Anonymisation	C	Integrated tool	Operational	EU; international	https://amnesia.openaire.eu/
APCDOI	This is a python program for which you can enter a list of DOIs in a csv file and it will make use of the Unpaywall API and a JSON file of Article Processing Charges (APCs) assembled to find out what journal articles are Gold or Hybrid open access and how much the APC for each of these articles are.	GitHub	Tool	Decision making	Developers; Researchers; RPOs	Research data; Journals	Publications	Costing RDM	Open	Integrated tool	Pilot	International	https://github.com/rvregier/APCDOI

<https://drive.google.com/file/d/1Sk8YmcUbxoMDelcyEKyXGPxDG61zEmDc/view>

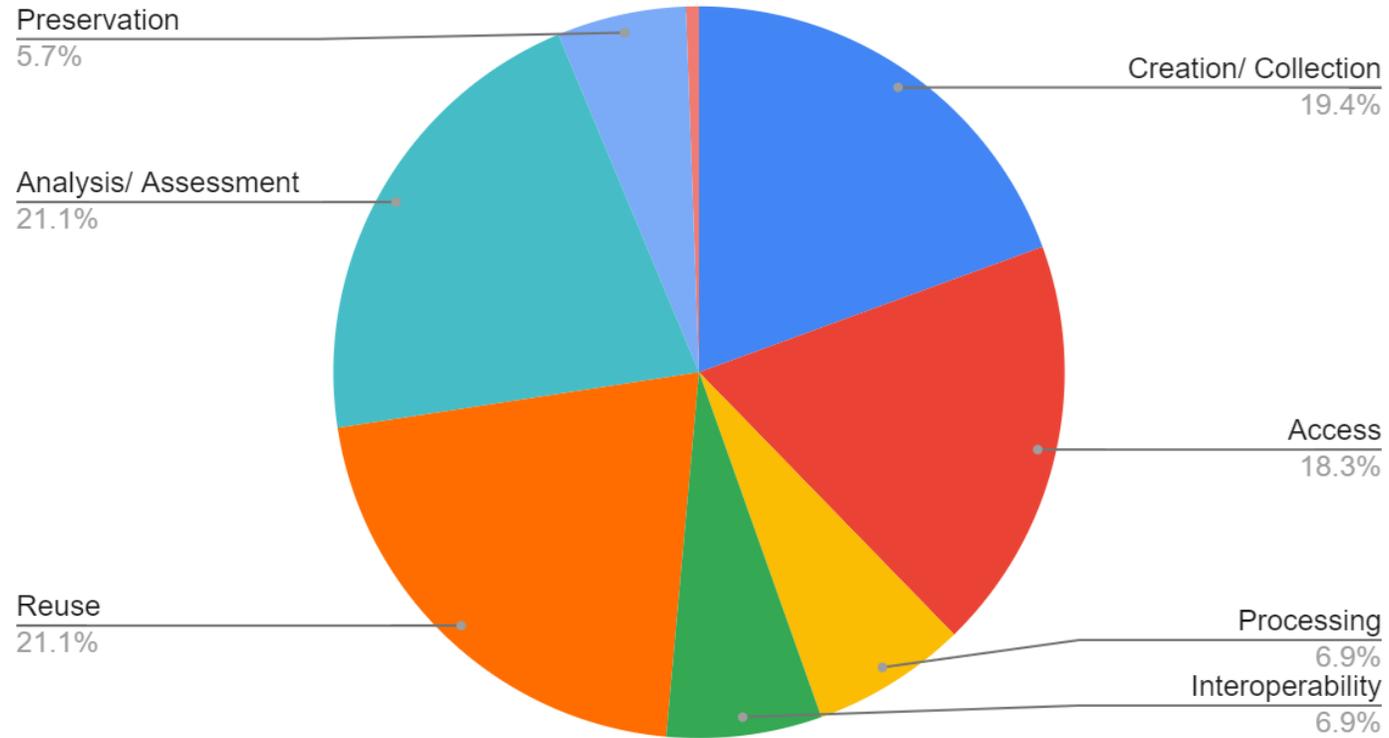
D4.3 Results

Research Outputs addressed by the collection of tools

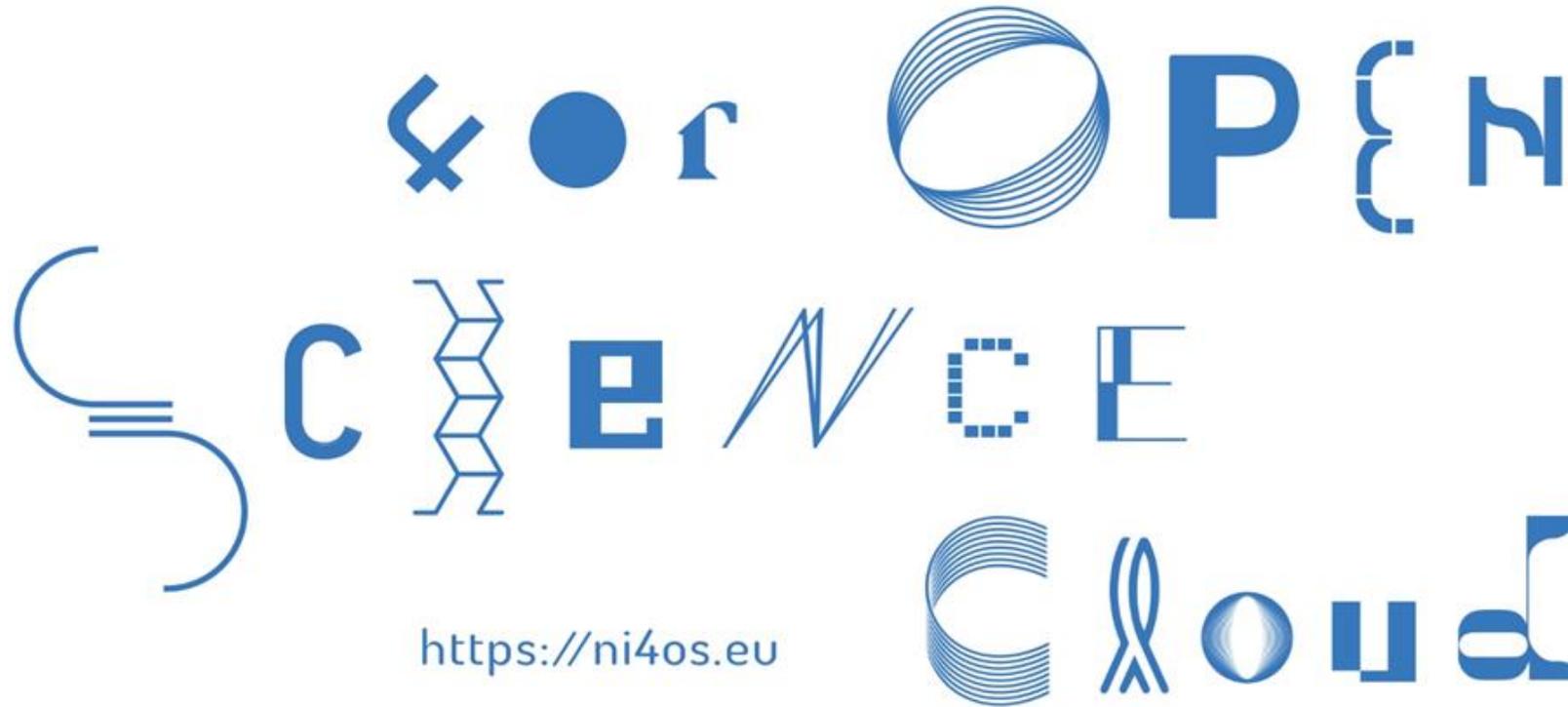


D4.3 Results

Research Data Management Focus



Thanks!



 [@NI4OS_eu](https://twitter.com/NI4OS_eu)

 [@NI4OS](https://www.facebook.com/NI4OS)