# Research Data Management

## NI4O-Europe Toolbox and good practices at EOSC

Elli Papadopoulou
Athena Research Center

elli.p@athenarc.gr

# Go to menti.com

# code: 3709 4826

# Open and FAIR
# Research Data Management

# About research data

- 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

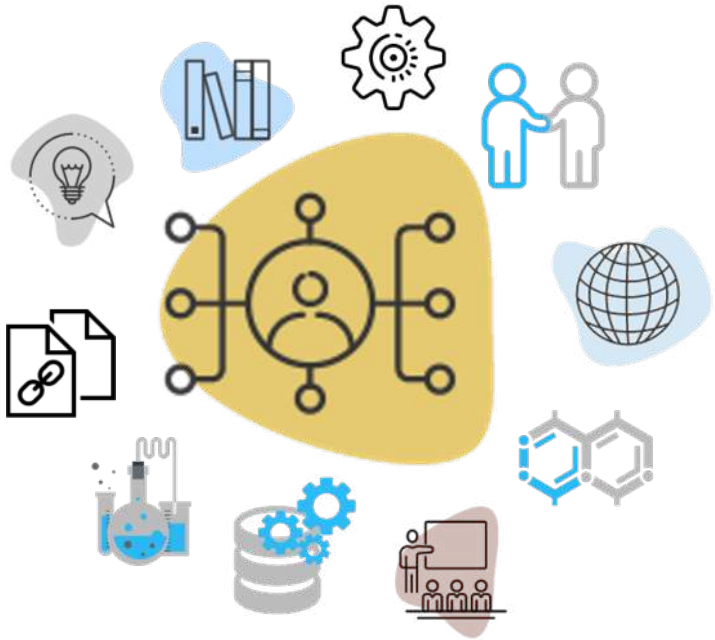
# Who is involved

## 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

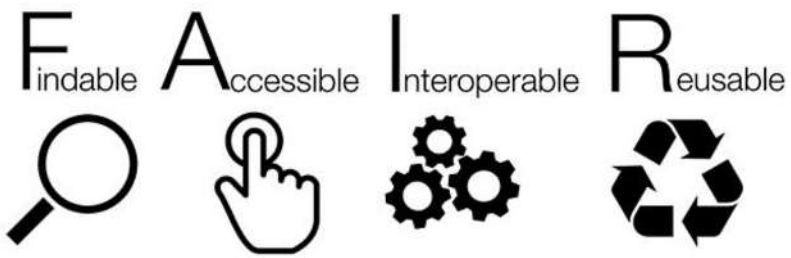
- Monitor research
- Better control of funds
- Research excellence
- Innovation

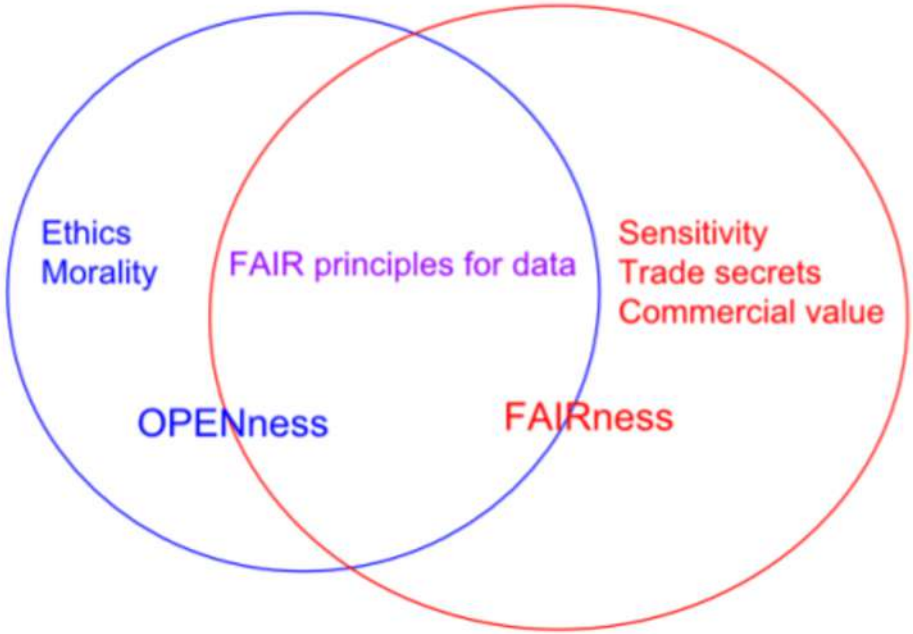
## Service Providers

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

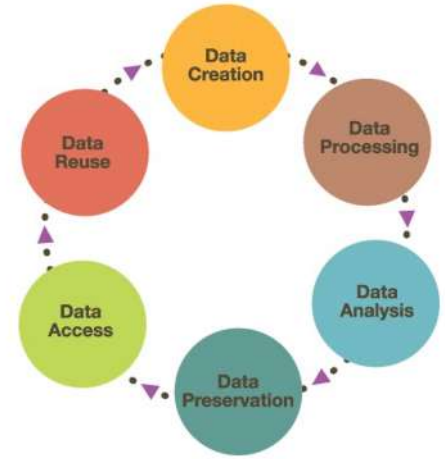
# Open and FAIR

| 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 |

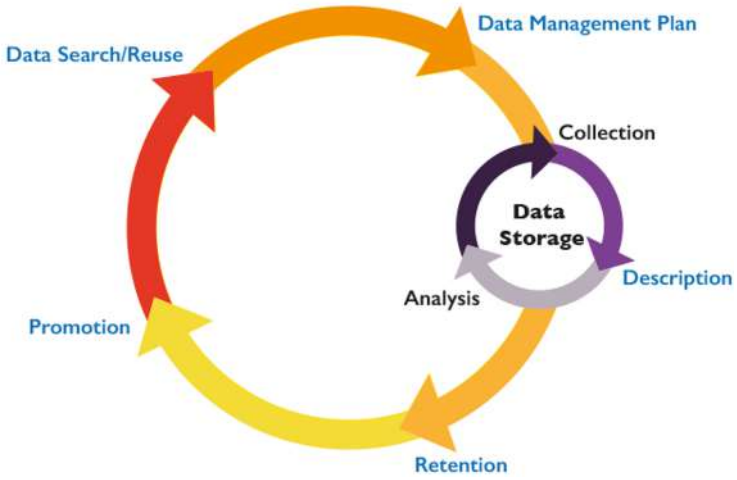


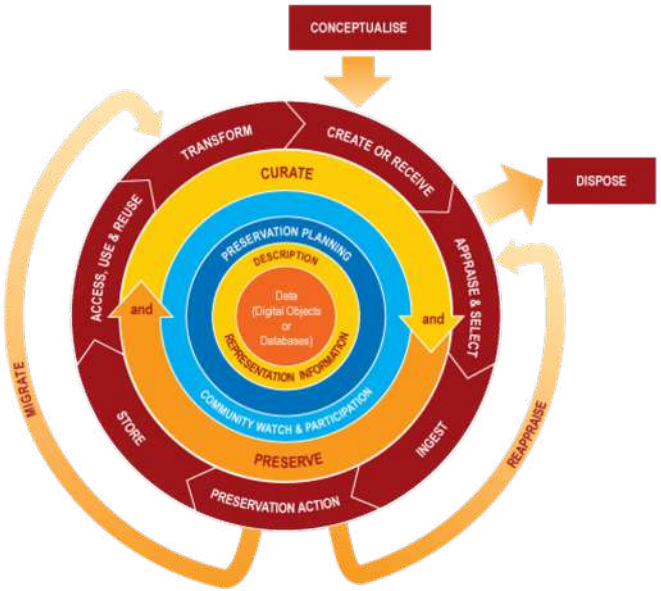



# RDM lifecycles

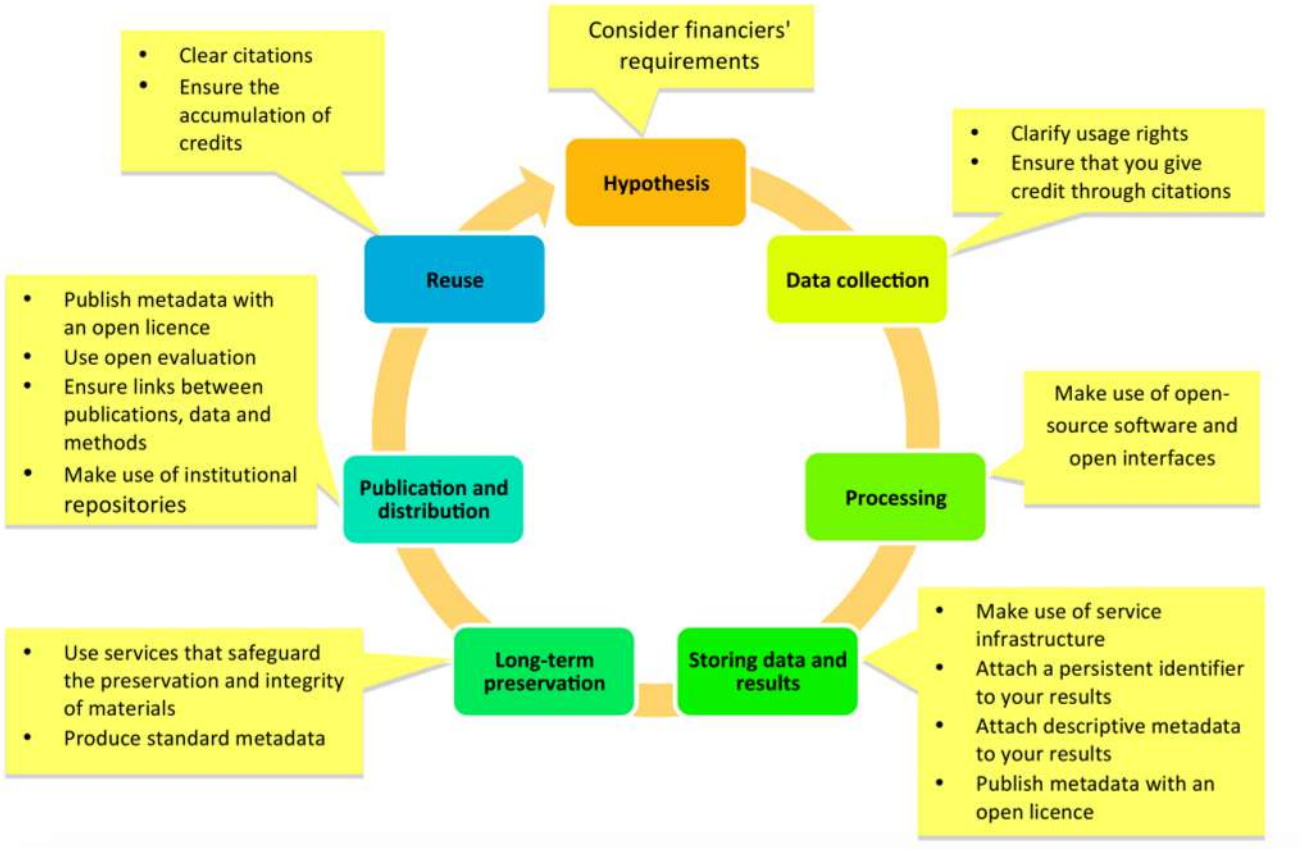






# Open Science Practices




(Open Science and Research Initiative, 2014)

# Plan

# Plan – Costing RDM

***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

# Plan – Costs for management and curation


**Estimating costs RDM tool**

| DMP PHASE | ACTIVITY | COMMENTS AND SUGGESTIONS | COSTS |
|---|---|---|---|
| *Preparing* | **Make a Data Management Plan** | Make a DMP before you start creating data; make decisions about managing your data. You can find the template for H2020 DMPs here.<br><br>Check if there is a department within your organization to support data management planning. | 2 hrs to 2 days, depending on the complexity of your project |
| *1. Data Collection* | **Acquiring External datasets**<br><br>Do you plan to use existing data, and is the data available at a commercial partner? | Your library may be able to help you acquire a license to a crucial database<br><br>In research data repositories, data can be available at no or low costs | Example: A faculty licence on a database for macro-economic analysis: €18.000/y |

Activity: Go to menti and answer the questions based on the following resource

https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs

# Plan – Costs for publishing and IP application


- Browse from the list
- Contribute

IP Cost Tool Services

- 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

# IP agreements

## IP Toolkit for Universities and PRIs: IP Commercialization and Knowledge Transfer

The IP Toolkit is designed to help universities and PRIs with knowledge/technology transfer and IP commercialization. It provides university managers, knowledge transfer officers and researchers with a baseline with which to develop their own entrepreneurial approach to IP management.

- Model agreements DOC
- Academic intellectual asset map DOC
- Hypothetical case studies DOC


- Material transfer agreements
- Provision of material
- Possession of material
- Safety
- Use of material
- New IP
- Publications
- …

**Material Transfer Agreement – Academic**

**IN THIS AGREEMENT,** effective as of the day of , [year]

*2 , a *3 , located at *4 ("the Owner")

AND

* , a , located at *5 ("the Recipient")

AGREE AS FOLLOWS:

BACKGROUND:

A. The Owner owns or has rights to the Material.

B. The Recipient has asked the Owner to provide a sample of the Material to the Recipient.

1. **MEANINGS**

# Plan – DMPs 1/3

## What is a DMP?

Deliverable and “living” document

documents processes undertaken throughout data management lifecycle, including costs

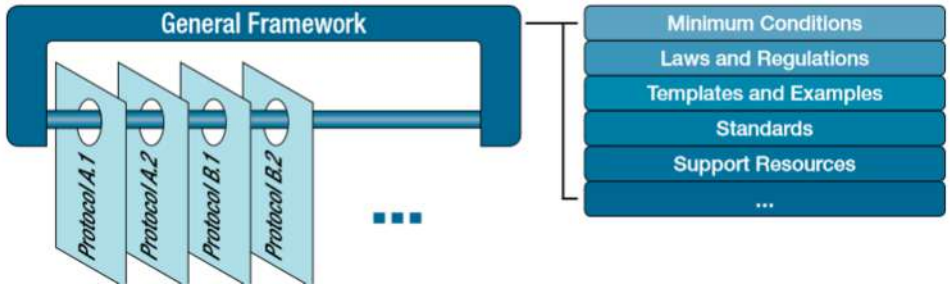
## What is not a DMP?

Research assessment method

# Plan – DMPs 2/3

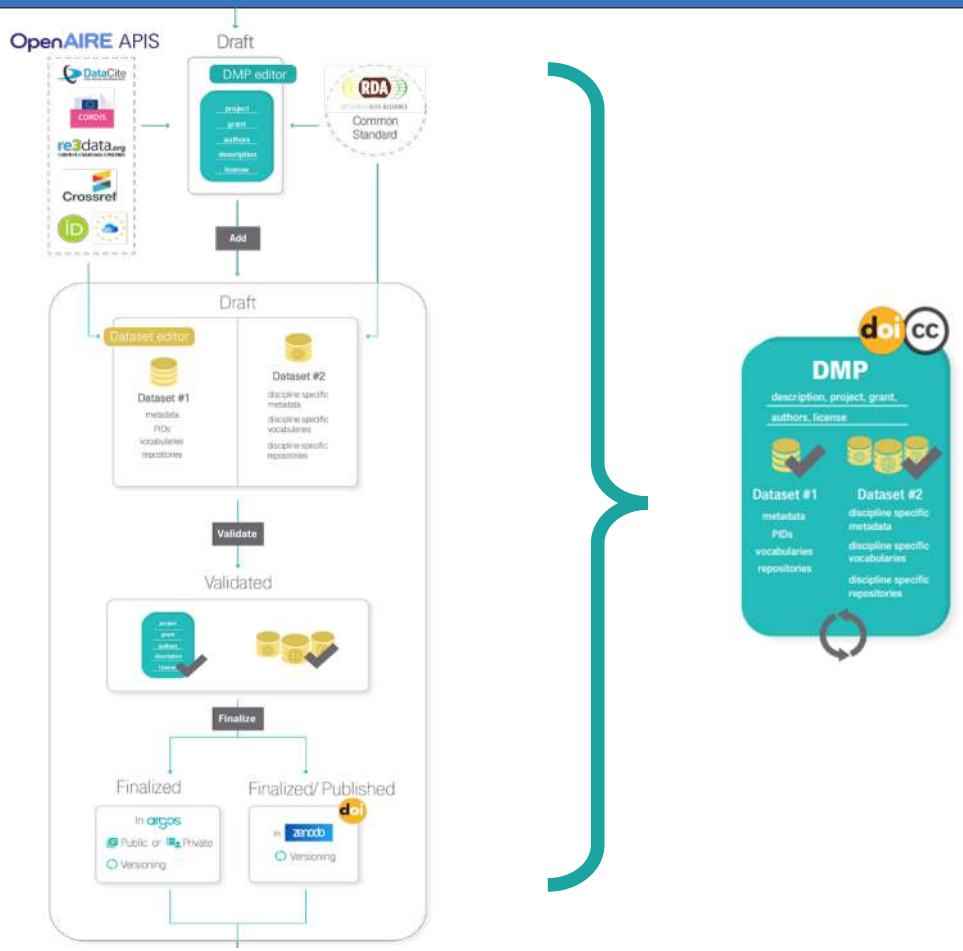
- Depends on the funder/institution requirements
- Differences in research communities
  - Formats, standards, documentation etc

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




Activity: Go to Menti and answer the questions

# Plan DMPs 3/3







- ARGOS is an open source, configurable and extensible tool for planning Research Data Management (RDM) activities according to Open Access & FAIR data policies.
- Website: https://argos.openaire.eu/

# 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** | • State the purpose of the data collection/generation<br>• Explain the relation to the objectives of the project<br>• Specify the types and formats of data generated/collected<br>• Specify if existing data is being re-used (if any)<br>• Specify the origin of the data<br>• State the expected size of the data (if known)<br>• Outline the data utility: to whom will it be useful |
| **2. FAIR Data**<br>2.1. Making data findable, including provisions for metadata | • Outline the discoverability of data (metadata provision)<br>• 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?<br>• Outline naming conventions used<br>• Outline the approach towards search keyword<br>• Outline the approach for clear versioning<br>• 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 | • Specify which data will be made openly available? If some data is kept closed provide rationale for doing so<br>• Specify how the data will be made available<br>• 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)?<br>• Specify where the data and associated metadata, documentation and code are deposited<br>• Specify how access will be provided in case there are any restrictions |
| 2.3. Making data interoperable | • Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.<br>• 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) | • Specify how the data will be licenced to permit the widest reuse possible<br>• Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed<br>• 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<br>• Describe data quality assurance processes<br>• Specify the length of time for which the data will remain re-usable |
| **3. Allocation of resources** | • Estimate the costs for making your data FAIR. Describe how you intend to cover these costs<br>• Clearly identify responsibilities for data management in your project<br>• Describe costs and potential value of long term preservation |
| **4. Data security** | • Address data recovery as well as secure storage and transfer of sensitive data |
| **5. Ethical aspects** | • 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** | • Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any) |

# Create / Collect

# Find / Collect


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


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

# Create / Collect

- **Metadata -> Standards**
  - For discovery (minimum)
  - For interoperability (rich)
  - General or Domain specific
  - For different outputs, eg instruments
- **Include PIDs**

**Recommended Minimum Metadata Elements**

The following are recommended as a minimum set of metadata elements. It is important t
may choose to use more elements based on the needs of your project.

**Title/Name** – Name given to the resource.

**Description** – A description of the resource and its spatial, temporal or subject coverage.

**Format** – File format, physical medium, dimensions of the resource, or hardware and soft

**Metadata** – Description of the metadata to be provided along with the generated data and
found.

**Identifier** – A unique identification assigned to the resource.

**Rights Holder** – The entities or persons who hold the rights to the data.

**Rights** – Information about the rights held in and over the resource.

**Contact Information** – Identity of, and means to communicate with persons or entities as

| 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<br>contributor type |
| Funding Reference (MA) | oaire:fundingReference | funderIdentifier type |
| Alternate Identifier (R) | datacite:alternateIdentifier | alternateIdentifier type |
| Related Identifier (R) | datacite:relatedIdentifier | relatedIdentifier type<br>relation type<br>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/

# Why metadata

http://opengeospatial.github.io/e-learning/metadata/text/main.html

## Attributes

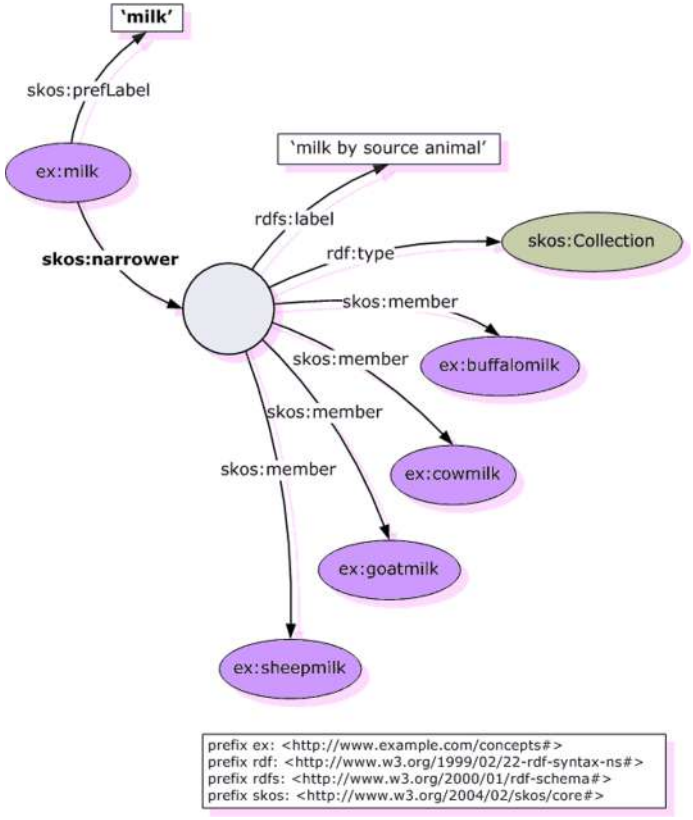
| Type | Value |
|---|---|
| External Id | SAMN15743948 |
| INSDC center name | CSIR-Institute of Genomics and Integrative Biology |
| INSDC first public | 2020-08-10T00:00:00Z |
| INSDC last update | 2020-08-10T13:01:21.373Z |
| INSDC secondary accession | SRS7175641 |
| INSDC status | live |
| NCBI submission model | Pathogen.cl |
| NCBI submission package | Pathogen.cl.1.0 |
| SRA accession | SRS7175641 |
| organism | Severe acute respiratory syndrome coronavirus 2 |
| replicate | Biological Replicate 759 |
| strain | SARS-CoV-2 |
| title | Negative Control 7 |

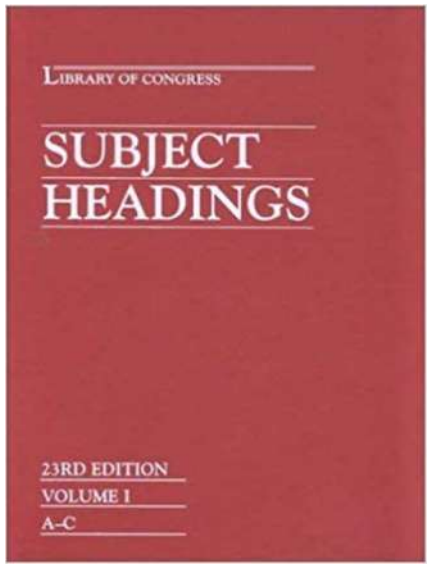
https://www.ebi.ac.uk/biosamples/samples/SAMN15743948

# Controlled Vocabularies

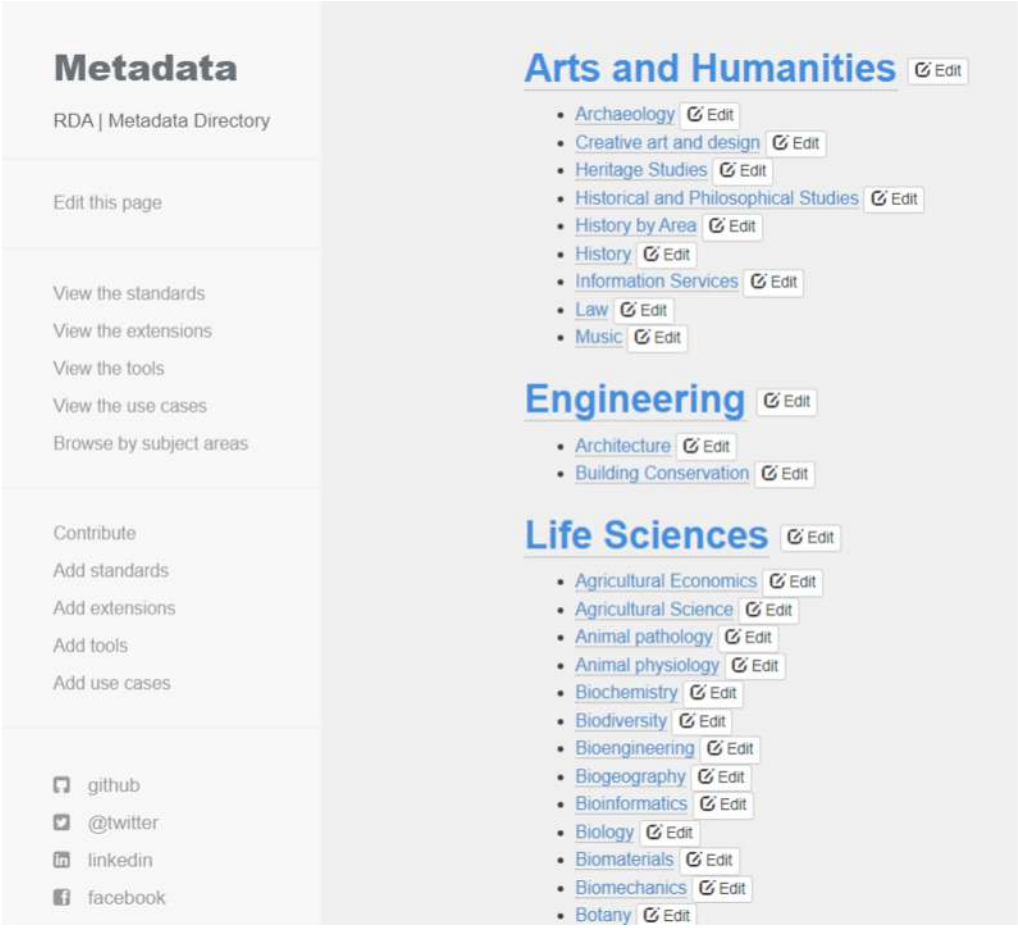
- Ontologies
- Taxonomies
- Thesauri
- Subject Headings

schema.org



# Metadata




Activity: Go to the RDA Metadata Directory. Does a standard for your domain exist?

https://rdamsc.bath.ac.uk/

# Process

# Process

***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: Match the left to the right

# Structured vs Unstructured

- Variable -> Column
- Observation -> Row

| | treatmenta | treatmentb |
|---|---|---|
| John Smith | — | 2 |
| Jane Doe | 16 | 11 |
| Mary Johnson | 3 | 1 |

| | John Smith | Jane Doe | Mary Johnson |
|---|---|---|---|
| treatmenta | — | 16 | 3 |
| treatmentb | 2 | 11 | 1 |

| name | trt | result |
|---|---|---|
| John Smith | a | — |
| Jane Doe | a | 16 |
| Mary Johnson | a | 3 |
| John Smith | b | 2 |
| Jane Doe | b | 11 |
| Mary Johnson | b | 1 |

# 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

# Handling sensitive data

- **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)

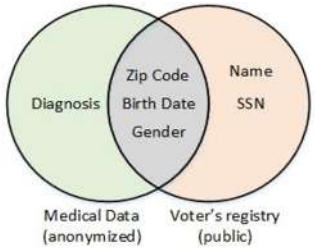
# Data processing - anonymisation

- 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

amnesia

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




Activity: Download Amnesia and perform an anonymization to your dataset

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

# Analyze

# Analyze

***Start producing outputs and prepare for sharing***

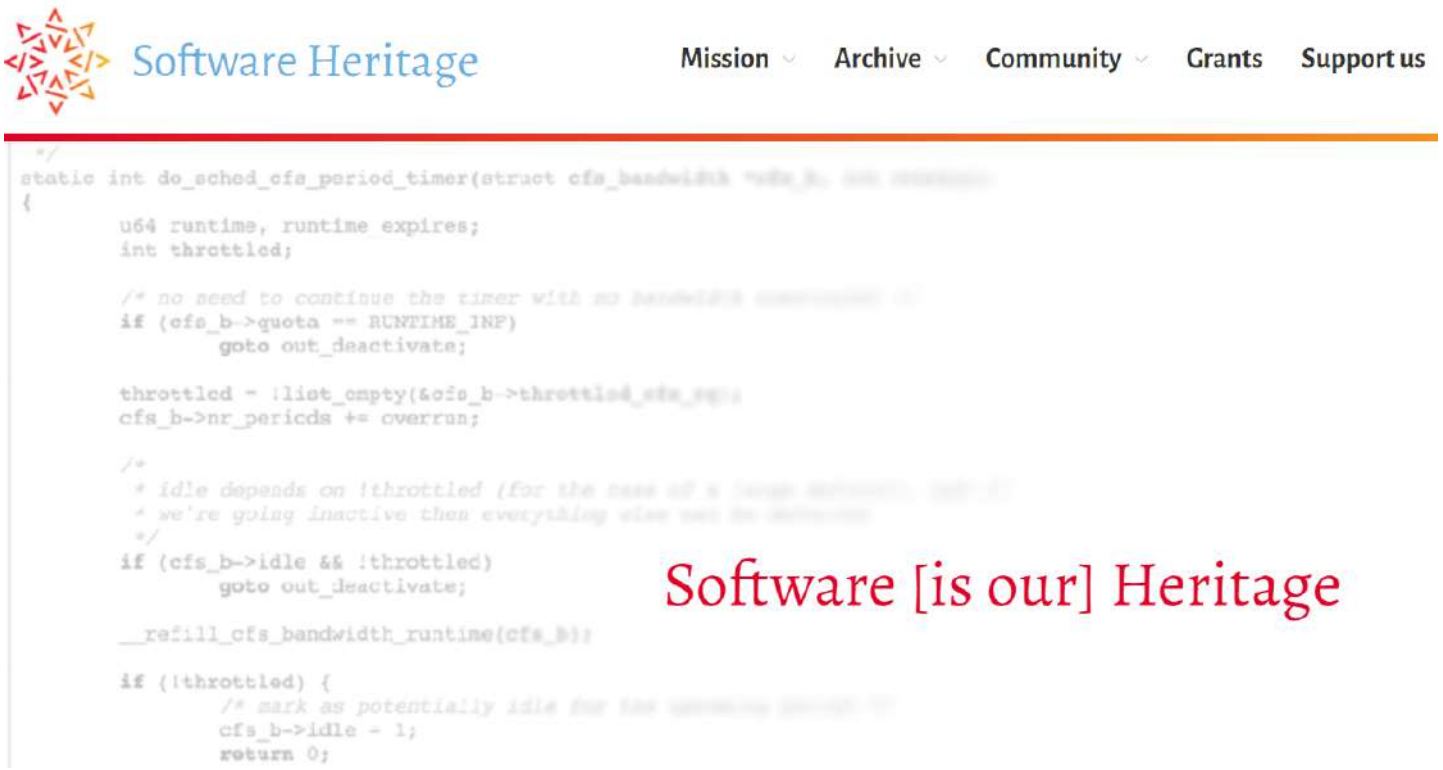
- **Methods**
  - Lab notebooks, end-to-end code/scripts for statistics, etc
- **Software**
  - R, MatLab, Python, etc

# Data analysis





# Archive software




0. RUN

1. STUDY

3. SHARE

4. IMPROVE

https://www.softwareheritage.org/

# Preserve

# Preserve

## *Store data in the short and/or long term*

- **What to preserve?**

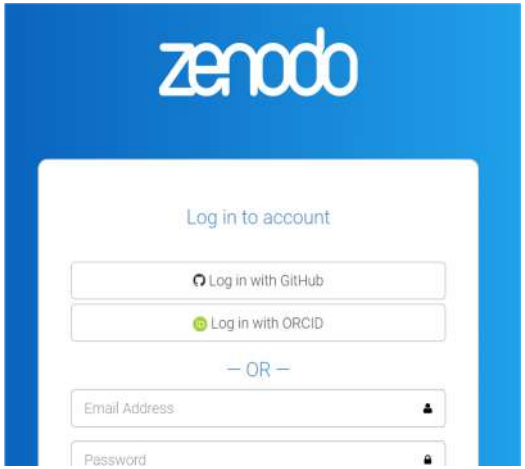
  data underlying publications, recreation purposes and value, legal issues, etc

- **Move from proprietary formats**
  - Prefer open, lossless formats
    - Check repository policy to see for suggested options

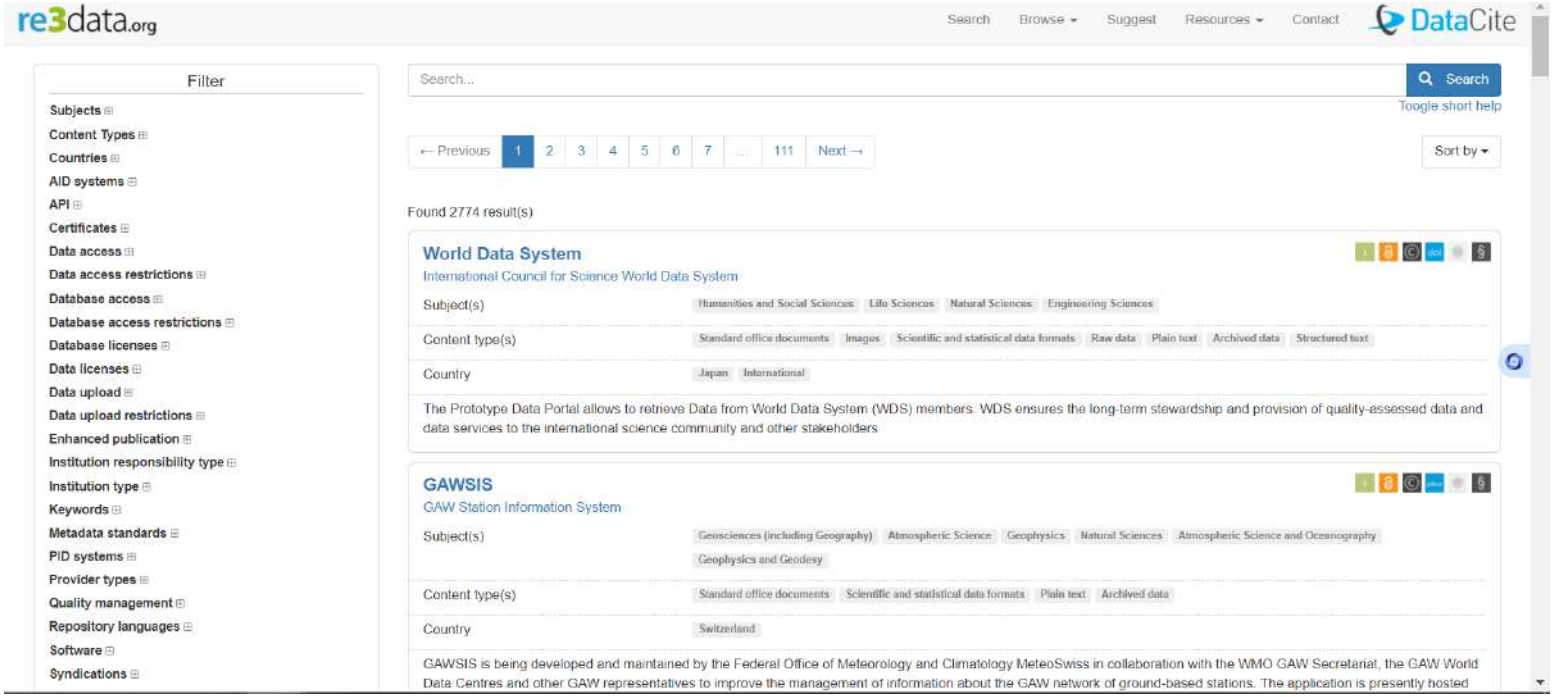
Activity: Guess the format of data (proprietary vs non-proprietary)

| Format | Open | Proprietary |
|---|---|---|
| Xml | | |
| Psd | | |
| Doc | | |
| Jpg | | |
| mp3 | | |
| Rar | | |

# Archive / Deposit



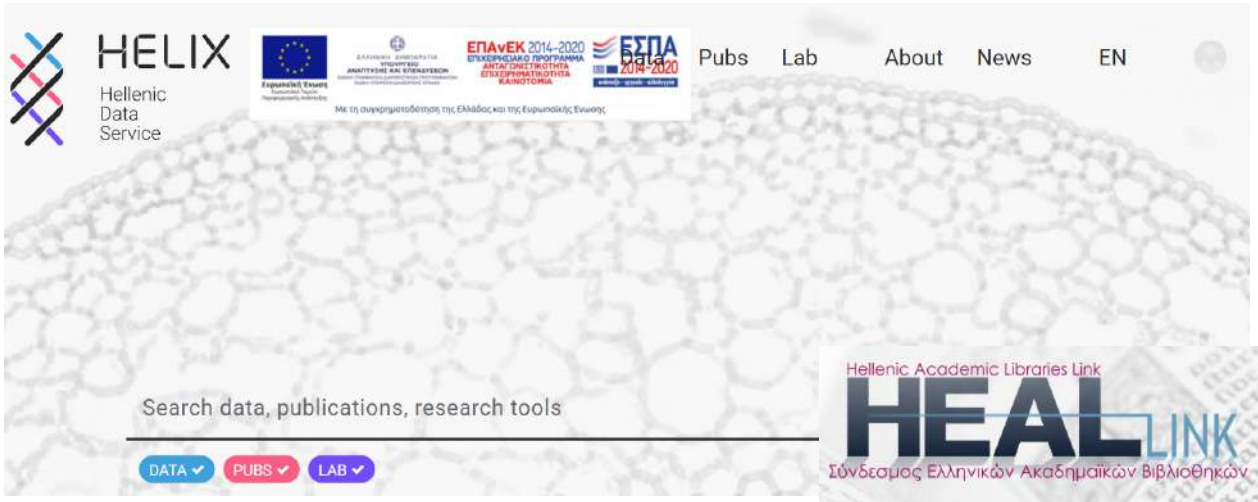

https://zenodo.org/




https://www.re3data.org/

Activity: Go to re3data, find a repository for your domain and add your answer on menti

# HELIX / HARDMIN

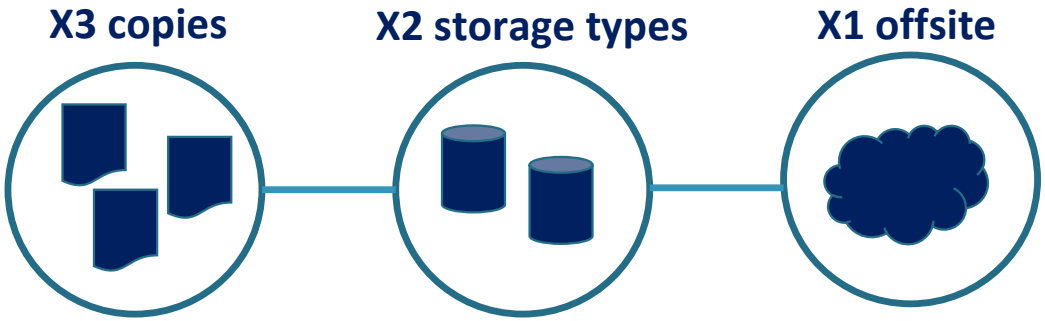



https://hellenicdataservice.gr/main/


https://hardmin.heal-link.gr

# Store / Preserve

- **Risk-assessment / Back-ups**
  - Retention
  - Frequency of back-up
  - Storage and methods




Activity: Go to menti and answer the question

# Persistent identifiers

## Digital Objects



ARK

PURL

## Researchers & Organisations



## Other activities

# PID systems

| | | | |
|---|---|---|---|
| http://dx.doi.org/ | 10.5284 | / | 1000389 |
| | | | |
| resolver service | prefix<br>(assigning body) | | suffix<br>(resource) |

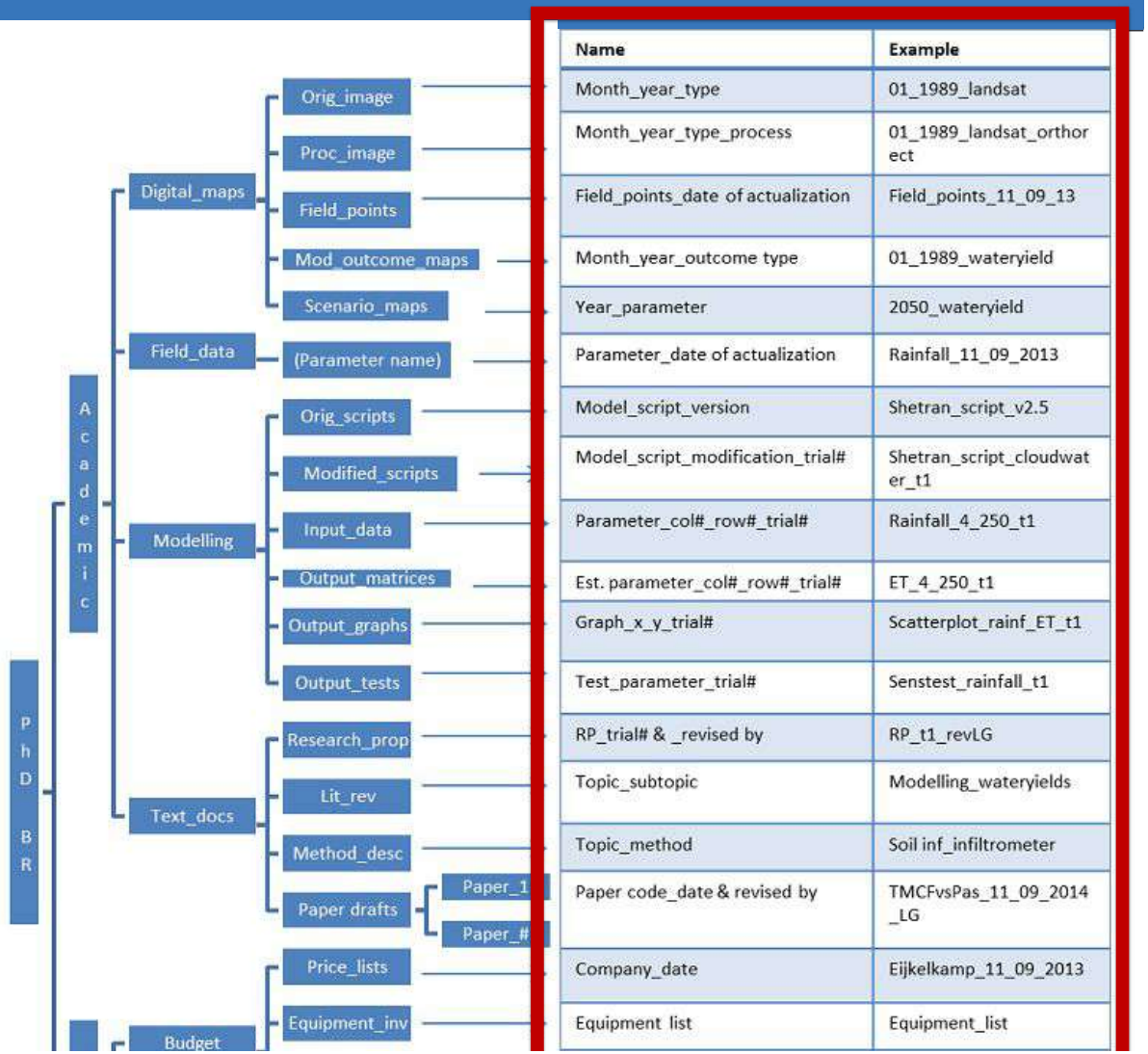
- 1. Integrate services
- 2. Provide PIDs for data and digital objects

# Share

# Share

- **Naming conventions**, so that data are understandable by others
  - e.g. letters, characters, abbreviations

- **Not over 25 characters**
- **~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { }**
- **Not spaces between words**
- **Continuous numbering (01-99 ; 0001, 0002…1001, etc)**
- **Date: YYYYMMDD**
- **Authors**




| Name | Example |
|---|---|
| Month_year_type | 01_1989_landsat |
| Month_year_type_process | 01_1989_landsat_orthorect |
| Field_points_date of actualization | Field_points_11_09_13 |
| Month_year_outcome type | 01_1989_wateryield |
| Year_parameter | 2050_wateryield |
| Parameter_date of actualization | Rainfall_11_09_2013 |
| Model_script_version | Shetran_script_v2.5 |
| Model_script_modification_trial# | Shetran_script_cloudwater_t1 |
| Parameter_col#_row#_trial# | Rainfall_4_250_t1 |
| Est. parameter_col#_row#_trial# | ET_4_250_t1 |
| Graph_x_y_trial# | Scatterplot_rainf_ET_t1 |
| Test_parameter_trial# | Senstest_rainfall_t1 |
| RP_trial# & _revised by | RP_t1_revLG |
| Topic_subtopic | Modelling_wateryields |
| Topic_method | Soil inf_infiltrometer |
| Paper code_date & revised by | TMCFvsPas_11_09_2014_LG |
| Company_date | Eijkelkamp_11_09_2013 |
| Equipment list | Equipment_list |

# Share

- **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

# Access

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

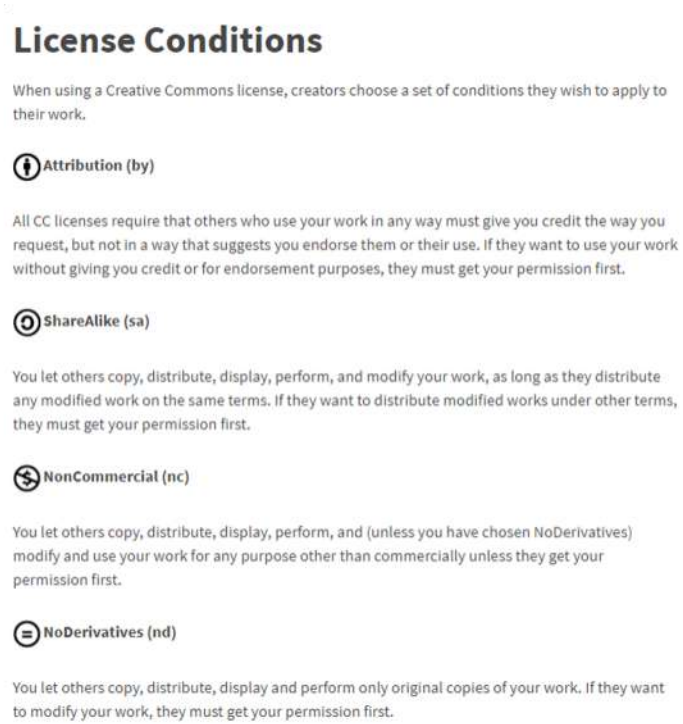
| | Open data | Safeguarded data | Controlled data |
|---|---|---|---|
| Security requirement | Suitable for fully anonymised data or data with agreement to publish personal details | Partially anonymised data or data with agreement to publish personal details, and where owner wishes to track usage | 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 |

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

# Reuse

# Reuse

- **Licenses**
  - Conditions
  - Types
- **Citations**
  - Specify required data citation
  - Open citations



## 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.

# Licenses


- Wizard-based form to guide you through the required fields
- Two distinct workflows for derivative works
  - Resource driven
  - License driven
- Available for registered and guest users
- Custom report generation with all clearance information
  - Personal & ethics information included if available
- Clearance history for registered users

↗ https://lct.ni4os.eu/

Wiki: https://wiki.ni4os.eu/index.php/License_Clearance_Tool_-_Description_and_Documentation

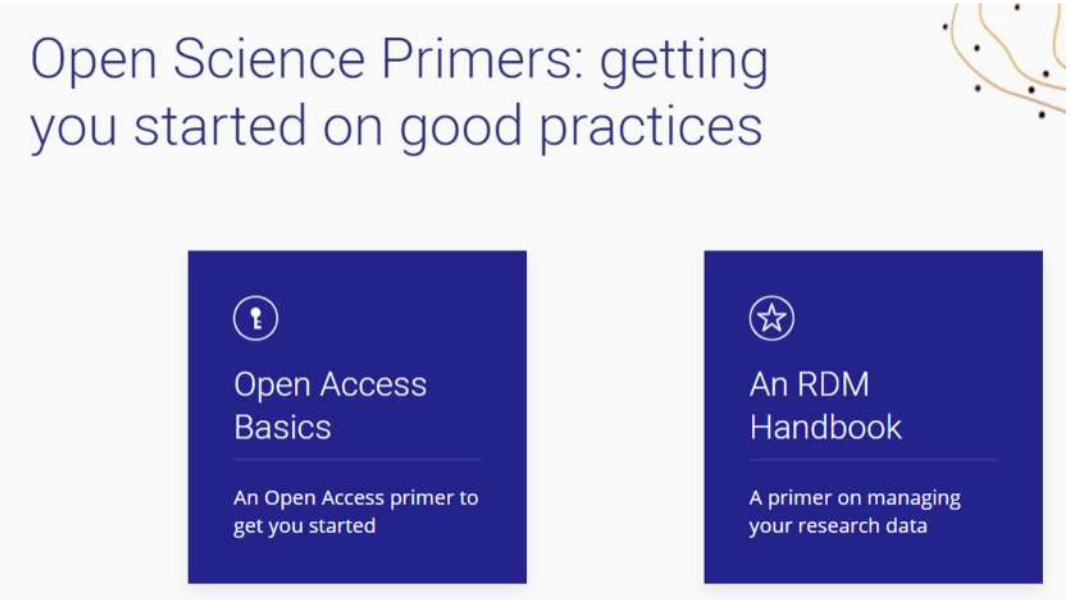
Source: https://github.com/ni4os-europe/license-clearance-application

License: EUPL (European Union Public License 1.2+)

# Training and resources

# Learn and get informed 1/5

OpenAIRE



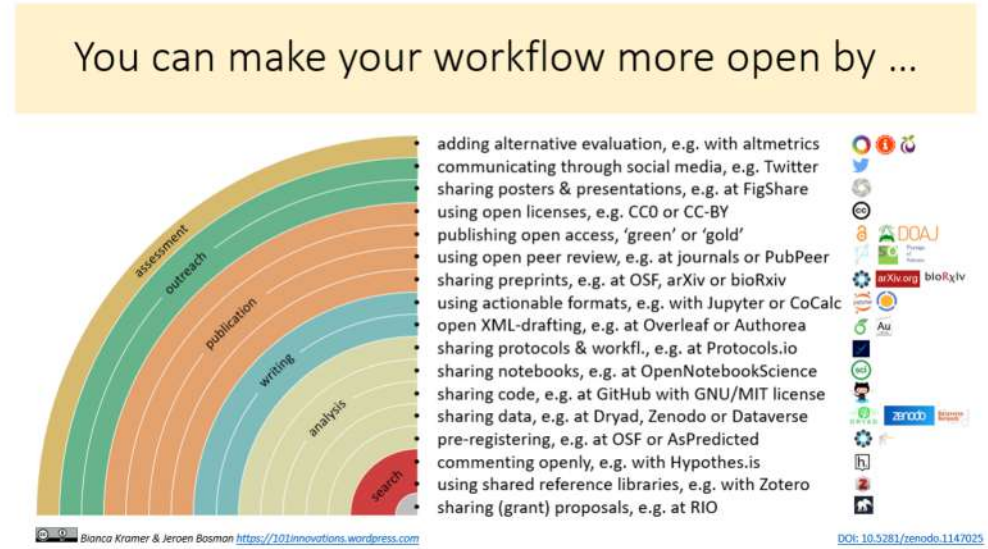

https://www.openaire.eu/os-primers


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

# Learn and get informed 2/5

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

https://www.fosteropenscience.eu/




- Best practices for open workflows
- Indicative tools

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

# Learn and get informed 3/5


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

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


- Review
- Assess
- Disseminate

https://www.openuphub.eu/

# Learn and get informed 4/5



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

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

# Learn and get informed 5/5

THE CARPENTRIES




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

https://carpentries.org/

# Assess data FAIRness

Activity: Go to OpenAIRE, find a dataset and assess it FAIRness with F-UJI

FAIR Aware

FAIR questions

**FINDABLE**

1. Are you aware that a data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository? ○ Yes ○ No

2. Are you aware that when you deposit a data(set) in a data repository, you will need to provide discovery metadata in order to make the data(set) findable, understandable and reusable to others? ○ Yes ○ No

3. Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans? ○ Yes ○ No

**ACCESSIBLE**

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which the data(set) can be reused? ○ Yes ○ No

5. Are you aware that metadata should remain available over time, even if the data(set) is no longer accessible? ○ Yes ○ No

**INTEROPERABLE**

https://fairaware.dans.knaw.nl/

F-UJI Automated FAIR Data Assessment Tool

Assessment Results:

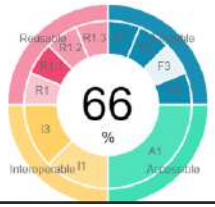
Evaluated Resource:

Carex nigra (L.) Reichard (BR0000011325505)

Save | {JSON} | New

| | |
|---|---|
| **FAIR level:** | moderate |
| **Resource PID/URL:** | 10.5281/zenodo.2767016 |
| **DataCite support:** | enabled |
| **Metric Version:** | metrics_v0.4 |
| **Metric Specification:** | https://doi.org/10.5281/zenodo.4081213 |
| **Software version:** | v1.3.8 |
| **Download assessment results:** | {JSON} |
| **Save and share assessment results:** | |

Summary:




| | Score earned: | Fair level: |
|---|---|---|
| **Findable:** | 6 of 7 | moderate |
| **Accessible:** | 2 of 3 | moderate |
| **Interoperable:** | 3 of 4 | moderate |
| **Reusable:** | 5 of 10 | moderate |

https://www.f-uji.net/index.php?action=home

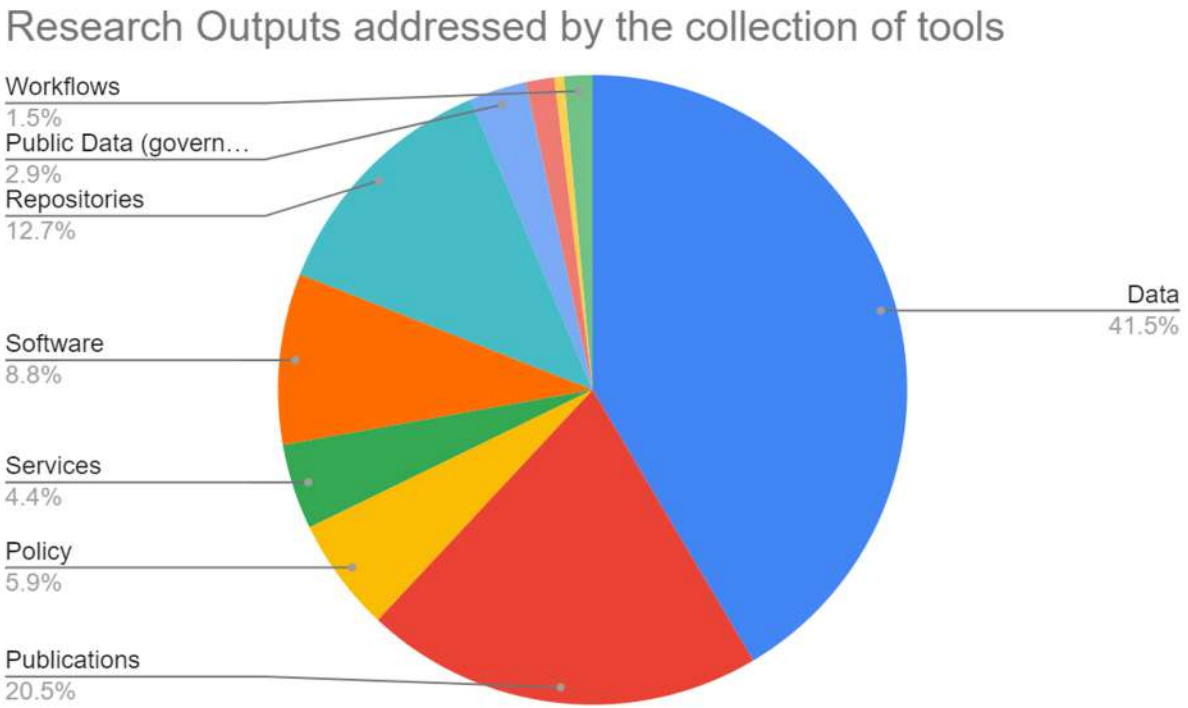
# Wrap-Up

# D4.3 Mapping of legal, procedural, technical tools

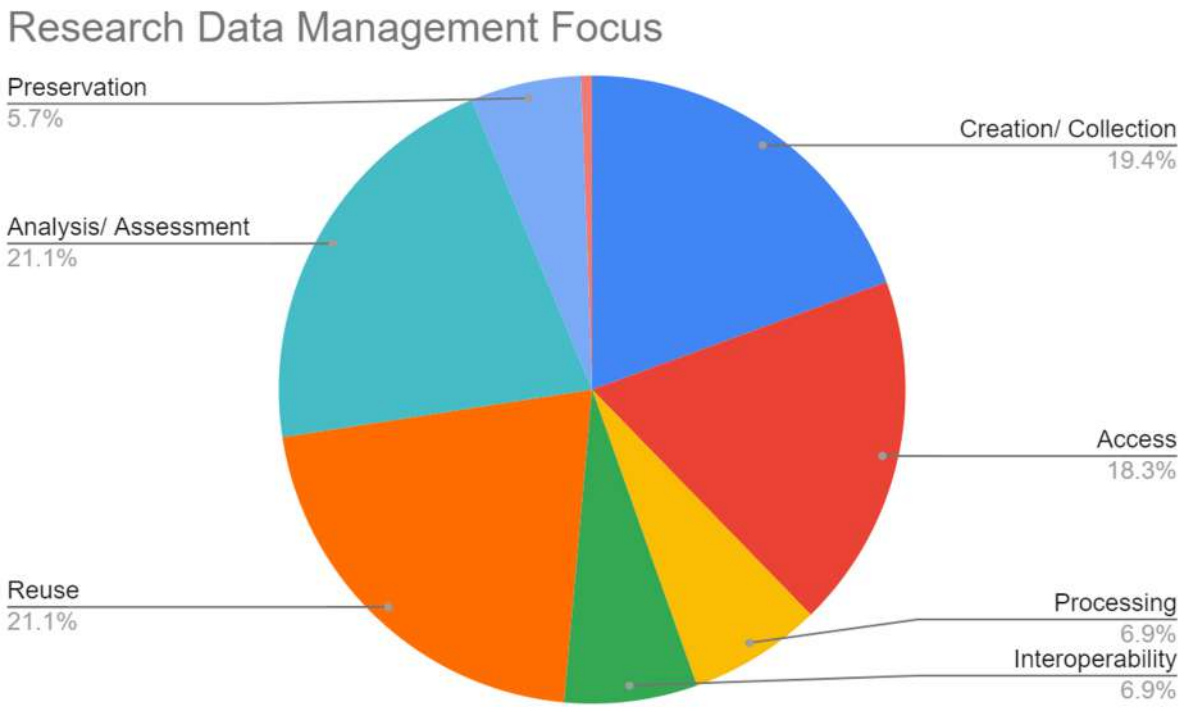
| A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Name | Description | Published by | Type (guidelines, tool, model) | Use (certification, decision making, | Users (Developer or End-users: e.g. RPOs = Research | Tags | Research output (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 unders 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 Proccessing 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/ryregier/APCDOI |

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

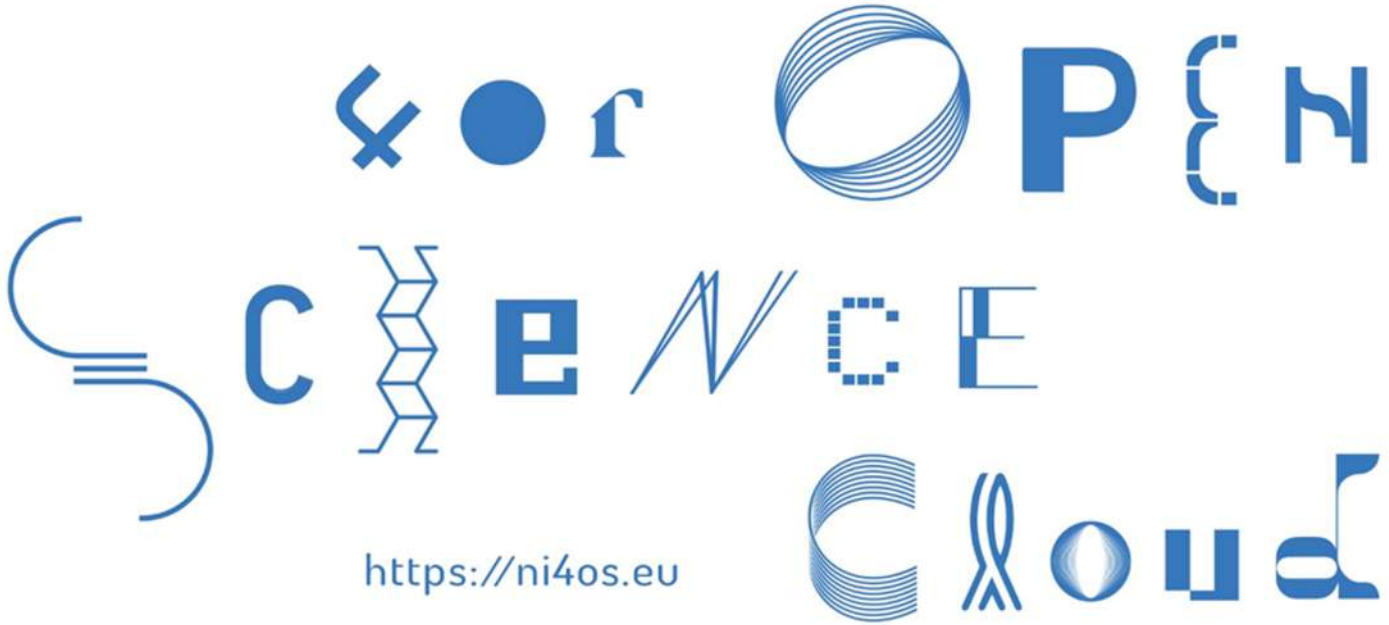
# D4.3 Results


Research Outputs addressed by the collection of tools
Workflows
1.5%
Public Data (govern…
2.9%
Repositories
12.7%
Software
8.8%
Services
4.4%
Policy
5.9%
Publications
20.5%
Data
41.5%

# D4.3 Results


Research Data Management Focus
Preservation
5.7%
Creation/ Collection
19.4%
Analysis/ Assessment
21.1%
Access
18.3%
Reuse
21.1%
Processing
6.9%
Interoperability
6.9%

# Thanks!




@NI4OS_eu @NI4OS