

Module 3 – RDM Knowledge Blocks

- Persistent Identifiers
- Metadata standards
- Licenses
- Repositories
- Version Control Systems



Persistent Identifiers (PIDs) Optimised Research



Persistent Identifiers (PIDs)

- PID is a unique code



- There are various PIDs for various type of objects



- PID are persistent and linked to essential info (Metadata)



- PIDs helps to track objects itself



- PIDs link to other resources



Persistent Identifiers (PIDs)

- PIDs are created by PID Authorities Providers (DataCite, Crossref, ROR, ORCID, etc.)
- PIDs help in a number of things:
- PIDs make your research FAIR



*Dr. Diana Berger
Dr. Diana Smith*

*Dr. John Smith
Dr. J. Smith
Dr. John S.
PhD. John Smith*

PIDs examples

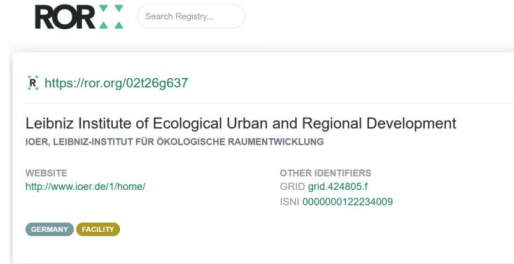
People



Name: **Stefano Della Chiesa**

<https://orcid.org/0000-0002-6693-2199>

Organization



<https://ror.org/02t26g637>

Research output



Research Data Management - Guidance

Della Chiesa, Stefano

Research Data Management (RDM) provides the strategies, guidelines, and measures to manage research outputs. Like a jigsaw, RDM requires multiple pieces of multidisciplinary knowledge and competencies. RDM Checklist guides the researchers to FAIRly manage the research outputs along the project lifecycle and beyond it. Research projects are inherently iterative in their try and error phases, and consequently, RDM is a continuous adapting process throughout the research project. Data Management Planning (DMP) is the operational tool to handle this complexity dynamically, in a structured way, ensuring compliance with policies and best practices. This document provides the RDM checklist and guidance for the Leibniz Institute of Ecological Urban and Regional Development (IOER).

<https://doi.org/10.5281/zenodo.6504928>

Controlled syntax

Host name Prefix Suffix

<https://doi.org/10.26084/ioerfdz-r10-urbgrn2018>

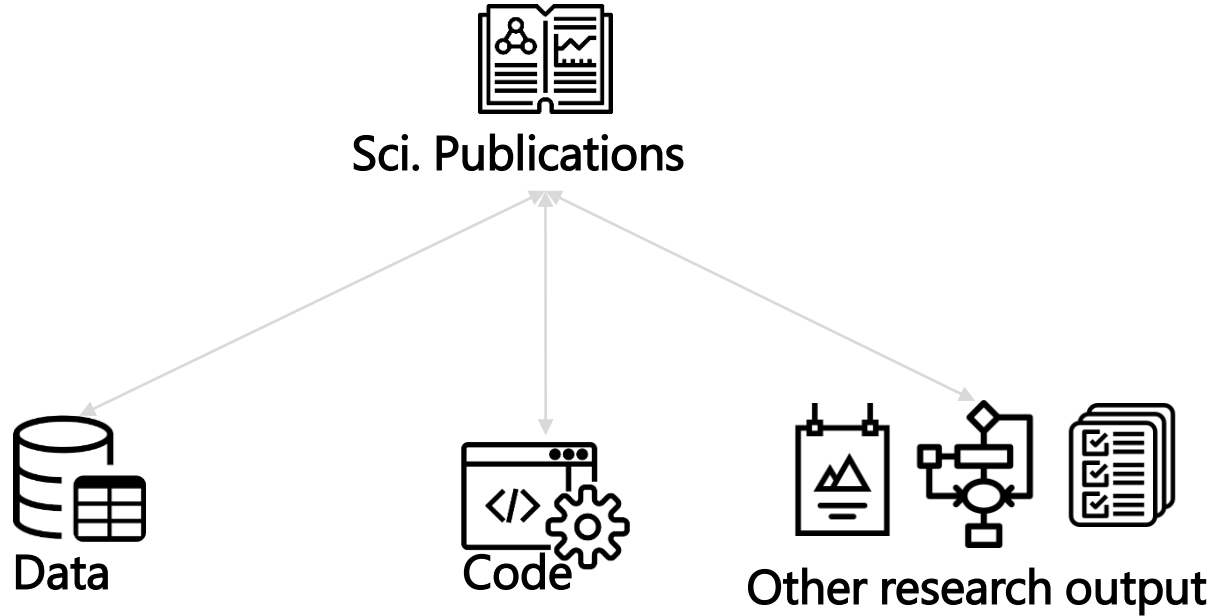
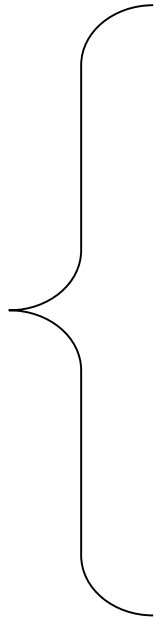


DRESDEN
concept

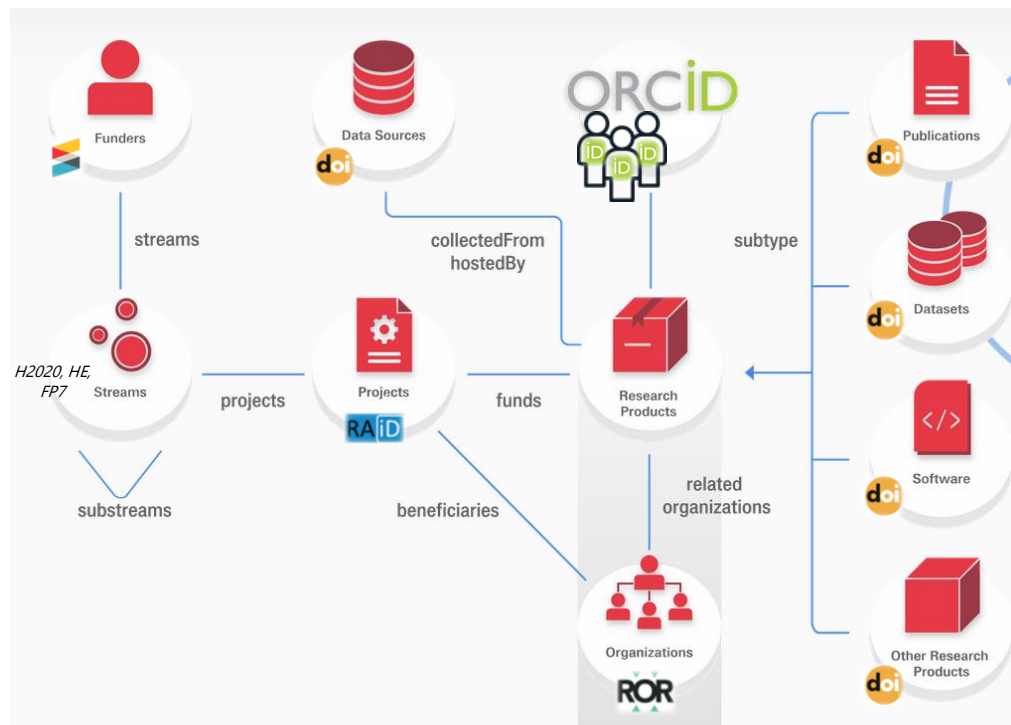








Leibniz Institute of
Ecological Urban and
Regional Development

DOI (Digital Object Identifier)



PID linked research



-  Crossref
-  Datacite
-  ORCID
-  Digital Object Identifiers (DOI)
-  Research Activity Identifier (RAiD)
-  Research Organization Registry (ROR)







Graphic adapted from: <https://graph.openaire.eu/>

Linked Research Outputs



The bigger picture

The PID-optimised research lifecycle

-  Crossref
-  Datacite
-  ORCID
-  Digital Object Identifiers (DOI)
-  Research Activity Identifier (RAID)
-  Research Organization Registry (ROR)

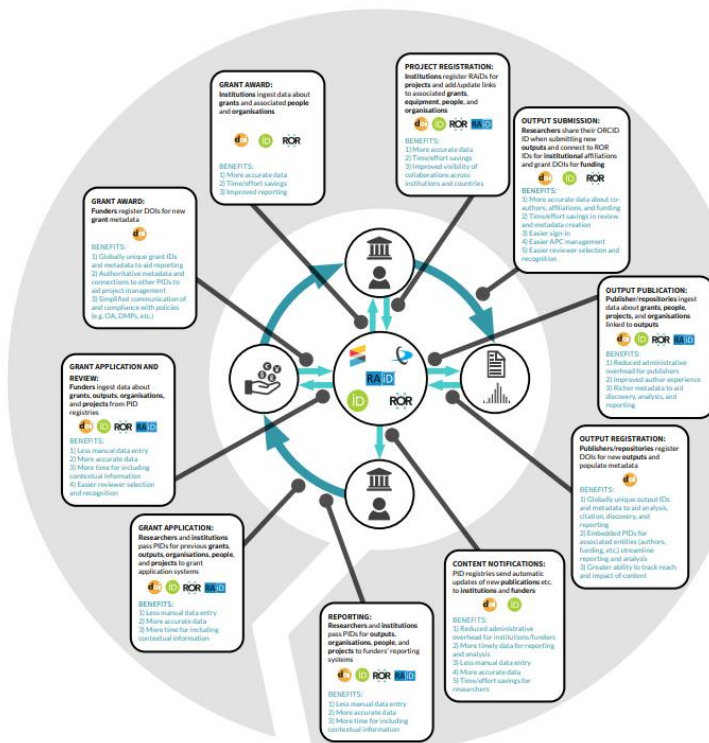


Example Funding dashboard: <https://monitor.openaire.eu/dashboard/cnr/funding>

Example Publication dashboard: <https://monitor.openaire.eu/dashboard/cnr/research-output/publications>

Example OA routes dashboard: <https://monitor.openaire.eu/dashboard/cnr/open-science/open-access/open-access-routes>

Example Impact dashboard: <https://monitor.openaire.eu/dashboard/cnr/impact->



<https://doi.org/10.5281/zenodo.4991733>

The PID Landscape

Heterogeneous PID Landscape & PID maturity

Mature PIDs

Emerging PIDs

Immature PIDs

Table 1. A Landscape Analysis of Research Entities, Identifying PID Types and Infrastructure Maturity (as of May 2018)

| Research Entity | PID Types Used | Maturity of PID Infrastructure |
|---|--|--------------------------------|
| Publication | DOI, accession number, handle, URN, Scopus EID, Web of Science UID, PMID, PMC, arXiv identifier, BibCode, ISSN, ISBN, PURL | mature |
| Citation | OCI (secondary aggregation of information) | emerging |
| Conference | DOI, accession number | emerging |
| Researcher (or scholar) | ORCID IDs, ISNI (also DAIs, VIAFs, arXivIDs, OpenIDs, ResearcherIDs, ScopusIDs) | mature |
| Organization | DOI, ISNI, GRID, Ringgold IDs, ROR IDs | emerging |
| Data | DOI, accession number, handle, PURL, URN, ARK | mature |
| Data repository | none | immature |
| Grants | DOI, PURL | emerging |
| Project | local identifier, accession number, RaID | emerging |
| Experiment | none | immature |
| Investigation | DOI, accession number | emerging |
| Analysis | GitHub gist | immature |
| Software | DOI, SHA-1 hash | emerging |
| Computer simulation | UUID | emerging |
| Software license | none | immature |
| Equipment | | |
| Instrument, device, sensor, platform, research facility | DOI, RRID, UID | emerging |
| Archival/storage facility | URI, DOI, UUID | emerging |
| Field station | none | immature |
| Sample | | |
| Geological or biological sample | accession number, RRID, DOI, IGSN | emerging |
| Cultural artifact | DOI, URN, accession number | emerging |
| Historical or mythical person | URI | emerging |
| Temporal period and historical place | ARK, URI, accession number | immature |
| Study Registration | | |
| Clinical trial; non-clinical registration | accession number, DOI | immature |
| Data management plan | DOI | immature |
| Workflow | URI, DOI | immature |
| Protocol | DOI | immature |

Adapted from Ferguson et al.[12]. ARK: Archival Resource Key, BibCode: Bibliographic Codes, DAI: Digital Author Identifier, DOI: Digital Object Identifier, ID: Identifier, IGSN: International Geo Sample Number, ISBN: International Standard Book Number, ISNI: International Standard Name Identifiers, ISSN: International Standard Serial Number, OCI: Open Citation Identifier, ORCID: Open Researcher and Contributor ID, PMID: PubMed ID, PURL: Persistent Uniform Resource Locators, RAID: Research Activity identifier, RRID: Research Resource ID, SHA-1: Secure Hash Algorithm 1, UUID: Universally Unique Identifiers, URI: Uniform Resource Identifier, URN: Uniform Resource Name, VIAF: Virtual International Authority Files

Connected Research: [The Potential of the PID Graph](#)

DOI minting process

Traditional
research outputs

Emerging
scientific outputs

| ISI Peer-review | DOI Minting Service | Landing page |
|---|---------------------|--------------|
| Publications (<i>Sci.Pub., Data Paper, Software Paper, Methods Paper</i>) | Journal | Journal |

| Other publications | DOI Minting Service | Landing page |
|------------------------|-------------------------------|---------------------------|
| Book, IOER collecitons | IOER Data Cite or Zenodo like | Zenodo, FDZ, QUCOSA, etc. |

| SPECIFIC doi Suffix | DOI Minting Service | Landing page |
|-----------------------|---------------------|--|
| Data | IOER Data Cite | Zenodo, FDZ, Other Repo. or Domain Specific Repository |
| Code | IOER Data Cite | Zenodo, Other Repo. (github for code versioninig) |
| Other research output | IOER Data Cite | Zenodo, FDZ, Other Repo. |

| NO SPECIFIC doi Suffix | DOI Minting Service | Landing page |
|------------------------|------------------------------------|--------------------------------------|
| Data | Zenodo, Domain Specific Repository | Zenodo, Domain Specific Repository |
| Code | Zenodo | Zenodo (github for code versioninig) |
| Other research output | Zenodo | Zenodo |

Metadata



No Metadata



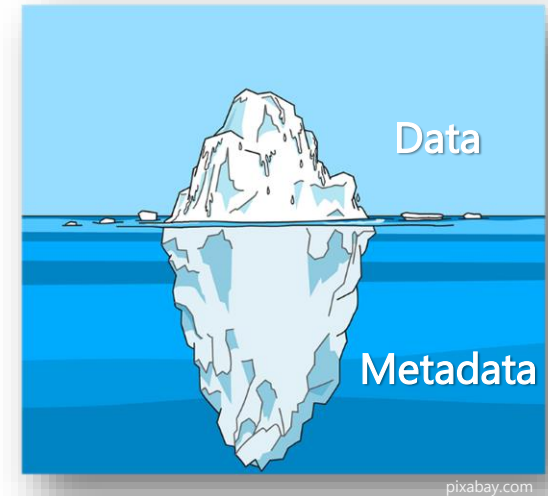
Metadata



hiclipart.com

Data and Metadata (MD)

- Research data includes **measurements, audio-visual information, texts, survey data, samples, questionnaires, algorithms, software, simulations, etc.**
- Metadata represents all the **information that add context and meaning** in a **structured** way to the (research) data.



Metadata Standards

- Metadata is fundamental as makes data findable and reusable.
- Metadata **standards** or **schemas** allow a structured description of your data by using specific MD elements.
- There are general purpose MD schemas.
- There are also disciplines with subject/domain specific metadata standards.



[List of Disciplinary Metadata](#)

[Use cases](#)

Metadata Standards

General purpose MD standard (e.g. Simple Dublin Core)

Table 1. The Fifteen Elements of “Simple Dublin Core”

| Identifier | Definition |
|-------------|--|
| Title | A name given to the resource. |
| Creator | An entity primarily responsible for making the content of the resource. |
| Subject | The topic of the content of the resource. |
| Description | An account of the content of the resource. |
| Publisher | An entity responsible for making the resource available. |
| Contributor | An entity responsible for making contributions to the content of the resource. |
| Date | A date associated with an event in the life cycle of the resource. |
| Type | The nature or genre of the content of the resource. |
| Format | The physical or digital manifestation of the resource. |
| Identifier | An unambiguous reference to the resource within a given context. |
| Source | A reference to a resource from which the present resource is derived. |
| Language | A language of the intellectual content of the resource. |
| Relation | A reference to a related resource. |
| Coverage | The extent or scope of the content of the resource. |
| Rights | Information about rights held in and over the resource. |



<https://www.dublincore.org/specifications/dublin-core/usageguide/2001-04-12/generic/#coverage>

Metadata Standards

General purpose MD standard (e.g. DataCite)

| Mandatory | Recommended | Optional |
|------------------|--------------------|-----------------|
| Identifier | Subject | Language |
| Creator | Contributor | Alternate ID |
| Title | Date | Size |
| Publisher | Resource Type | Format |
| Publication year | Related identifier | Version |
| | Description | Rights |
| | GeoLocation | |



https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

Metadata Standards



Disciplines MD standard (e.g. Data Documentation Initiative - DDI)

“For describing the data produced by surveys and observational methods in the social and economic sciences”.

A screenshot of the gesis website. The header includes the gesis logo (Leibniz Institute for the Social Sciences), a login button, a language selector for German, and an email icon. A search bar contains the text 'data'. Below the search bar are navigation links for Services, Research, and Institute. The main content area shows a '< Back' link and a search result titled 'Mobile Data Collection - Incentive Experiment' by Florian Keusch. The result includes the text 'GESIS Data Archive, Cologne. ZA6978 Data file Version 1.0.0, https://doi.org/10.4232/1.13247' and an abstract describing the study's goal to measure the influence of incentive schemes on smartphone data collection among German users.



Metadata Standards

Disciplines MD standard (e.g. ISO 19115)

“For describing for Geographical data”.



ISO 19115-1:2014
Geographic information – Metadata



Resource title: Corine Land Cover Change (CHA) 2012 - 2018, Version 2020_20u1

Resource abstract: CHA1218 is one of the Corine Land Cover (CLC) datasets produced within the frame the Copernicus Land Monitoring Service referring to land cover / land use status between the years 2012 and 2018. CLC service has a long-time heritage (formerly known as "CORINE Land Cover Programme"), coordinated by the European Environment Agency (EEA). It provides consistent and thematically detailed information on land cover and land cover changes across Europe.

CLC datasets are based on the classification of satellite images produced by the national teams of the participating countries - the EEA members and cooperating countries (EEA39). National CLC inventories are then further integrated into a seamless land cover map of Europe. The resulting European database relies on standard methodology and nomenclature with following base parameters: 44 classes in the hierarchical 3-level CLC nomenclature; minimum mapping unit (MMU) for status layers is 25 hectares; minimum width of linear elements is 100 metres. Change layers have higher resolution, i.e. minimum mapping unit (MMU) is 5 hectares for Land Cover Changes (LCC), and the minimum width of linear elements is 100 metres. The CLC service delivers important data sets supporting the implementation of key priority areas of the Environment Action Programmes of the European Union as e.g. protecting ecosystems, halting the loss of biological diversity, tracking the impacts of climate change, monitoring urban land take, assessing developments in agriculture or dealing with water resources directives. CLC belongs to the Pan-European component of the Copernicus Land Monitoring Service (<https://land.copernicus.eu/>), part of the European Copernicus Programme coordinated by the European Environment Agency, providing environmental information from a combination of air- and space-based observation systems and in-situ monitoring. Additional information about CLC product description including mapping guides can be found at https://land.copernicus.eu/user-corner/technical-library/clc2018technicalguidelines_final.pdf. CLC class descriptions can be found at <https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/>.

Resource type: Dataset

Resource Locator: <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012-2018>

Topic of category: Environment, Imagery/Base Maps/Earth Cover

Keyword: Copernicus Land Satellite Image Interpretation 2012 Corine Corine Land Cover Change 2012

Bounding Box:

West = -31.561261
East = 44.820775
North = 71.409109
South = 27.405827

Bounding Box:

West = -61.906047
East = -60.905616
North = 16.607552
South = 15.736333



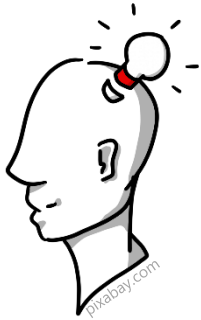
Leibniz Institute of Ecological Urban and Regional Development

IP, IPR & Licenses



IP & IPR

Intellectual Property (IP)



*"Creation of the mind"
(inventions, literary & artistic
work, designs, names)*

Intellectual Property Right (IPR)



"Legal tool that protect IP"

IP & IPR

Copyright is the common IPR
for research output

* Corresponding author.

E-mail address: stefano.dellachiesa@eurac.edu (S. Della Chiesa).



<https://doi.org/10.1016/j.geoderma.2019.02.010>

Received 19 June 2018; Received in revised form 1 February 2019; Accepted 4 February 2019
Available online 16 February 2019

0016-7061 © 2019 The Authors. Published by Elsevier B.V. This is an open access article under
(<http://creativecommons.org/licenses/by/4.0/>).

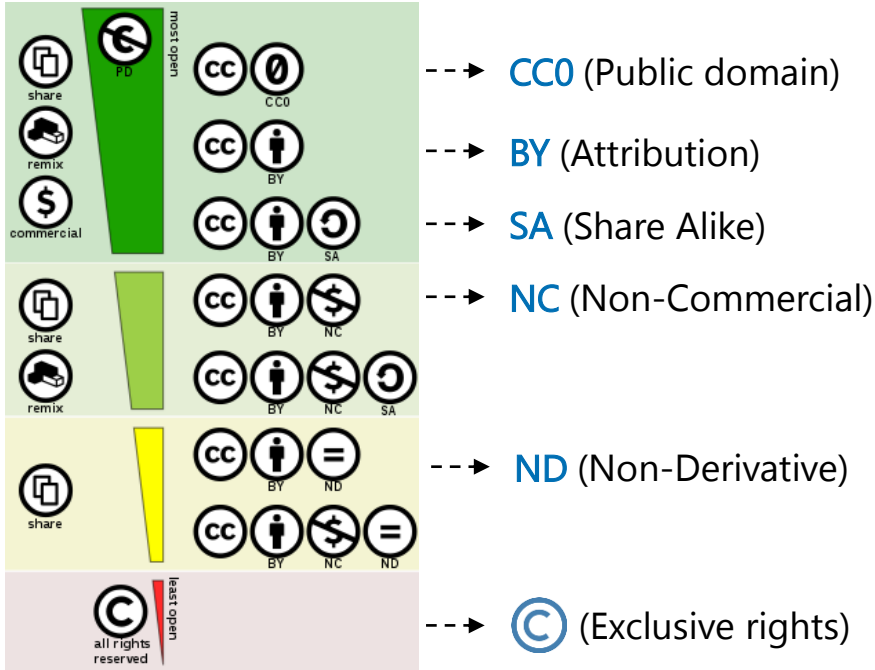
*Work protected from the moment is created
(No registration authority)*

Database IPR *sui generis database right*
Note: RAW data is generally not
protected by IPR.

Copyright protect you by default.
License provides the use conditions.

External resources: [The EU IP-Helpdesk](#)

Research Output Licenses



CREATIVE COMMONS LICENSES

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

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

Shaddim; [original CC license symbols by Creative Commons](#), CC BY 4.0

Creative Commons licenses, [by Foter](#) CC-BY-SA

License: Inbound vs Outbound



Compatibility Checker

| | PUBLIC DOMAIN | PUBLIC DOMAIN | BY | BY SA | BY NC | BY ND | BY NC SA | BY NC ND |
|---------------|---------------|---------------|----|-------|-------|-------|----------|----------|
| PUBLIC DOMAIN | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| PUBLIC DOMAIN | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| BY | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ |
| BY SA | ✓ | ✓ | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ |
| BY NC | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ |
| BY ND | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |
| BY NC SA | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ |
| BY NC ND | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ |

https://wiki.creativecommons.org/License_Compatibility

Repositories for Research Output

Where is my data?



The needle in a haystack

Data repository

A data repository is an archiving space for researchers to deposit research data linked with their research.

“A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future”.



www.coretrustseal.org

Repositories

General-purpose Repositories



The
Dataverse[®]
Project



Domain/Subject/Discipline Specific Repositories

gesis

Leibniz Institute
for the Social Sciences



Institutional Repositories



OpARA

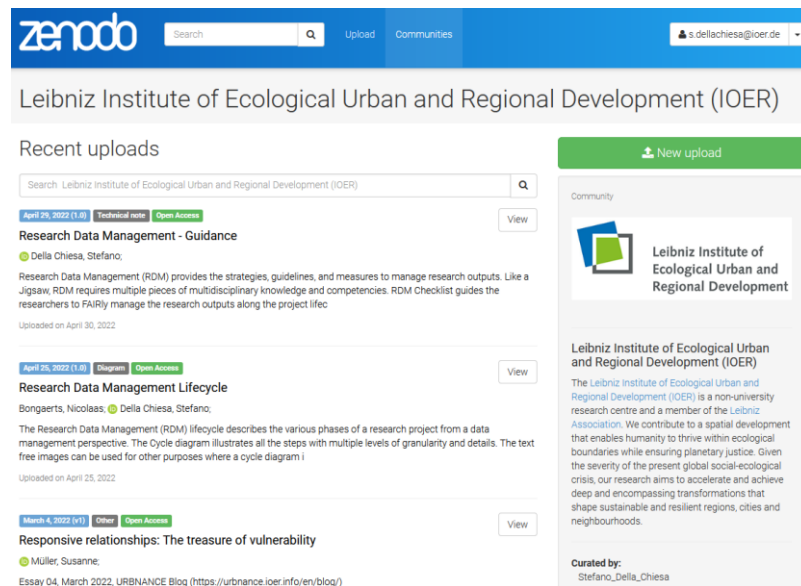


FIZ Karlsruhe – Leibniz Institute for Information Infrastructure

Zenodo Community

- Create your own research output collection
- Establish your own repository
- Central hub linking research outputs

Zenodo IOER Community



The screenshot displays the Zenodo interface for the Leibniz Institute of Ecological Urban and Regional Development (IOER) community. The top navigation bar includes the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. The user profile 's.dellachiesa@ioer.de' is visible in the top right.

The main heading is 'Leibniz Institute of Ecological Urban and Regional Development (IOER)'. Below this, the 'Recent uploads' section lists three items:

- Research Data Management - Guidance** (April 29, 2022, 1.00) by Della Chiesa, Stefano. Description: Research Data Management (RDM) provides the strategies, guidelines, and measures to manage research outputs. Like a Jigsaw, RDM requires multiple pieces of multidisciplinary knowledge and competencies. RDM Checklist guides the researchers to FAIRly manage the research outputs along the project life cycle. Uploaded on April 30, 2022.
- Research Data Management Lifecycle** (April 25, 2022, 1.00) by Bongaerts, Nicolaas and Della Chiesa, Stefano. Description: The Research Data Management (RDM) lifecycle describes the various phases of a research project from a data management perspective. The Cycle diagram illustrates all the steps with multiple levels of granularity and details. The text free images can be used for other purposes where a cycle diagram is. Uploaded on April 25, 2022.
- Responsive relationships: The treasure of vulnerability** (March 4, 2022, v1) by Müller, Susanne. Description: Essay 04, March 2022, URBANANCE Blog (<https://urbanance.ioer.info/en/blog/>)

On the right side, the community profile for 'Leibniz Institute of Ecological Urban and Regional Development (IOER)' is shown, including a logo and a description: 'The Leibniz Institute of Ecological Urban and Regional Development (IOER) is a non-university research centre and a member of the Leibniz Association. We contribute to a spatial development that enables humanity to thrive within ecological boundaries while ensuring planetary justice. Given the severity of the present global social-ecological crisis, our research aims to accelerate and achieve deep and encompassing transformations that shape sustainable and resilient regions, cities and neighbourhoods.'

Registry of research data repositories

Browser resources in the global registry of research data repositories

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Filter

[Reset all](#)

Subjects

[Humanities and Social Sciences \(1\)](#)

Natural Sciences (2)

[Geosciences \(including Geography\) \(2\)](#)

[Atmospheric Science and Oceanography \(1\)](#)

[Atmospheric Science \(1\)](#)

[Oceanography \(1\)](#)

[Geology and Palaeontology \(1\)](#)

[Water Research \(1\)](#)

[Hydrogeology, Hydrology, Limnology, Urban Water Management, Water Chemistry, Integrated Water Resources Management \(1\)](#)

Content Types

[Archived data \(1\)](#)

Databases (2)

[Images \(2\)](#)

[Plain text \(2\)](#)

[Scientific and statistical data formats \(2\)](#)

[Standard office documents \(2\)](#)

[Structured graphics \(1\)](#)

[Structured text \(1\)](#)

Countries

Germany (2)

[International \(1\)](#)

API

Certificates

[CoreTrustSeal \(2\)](#)



Browse resources by subject



Standards

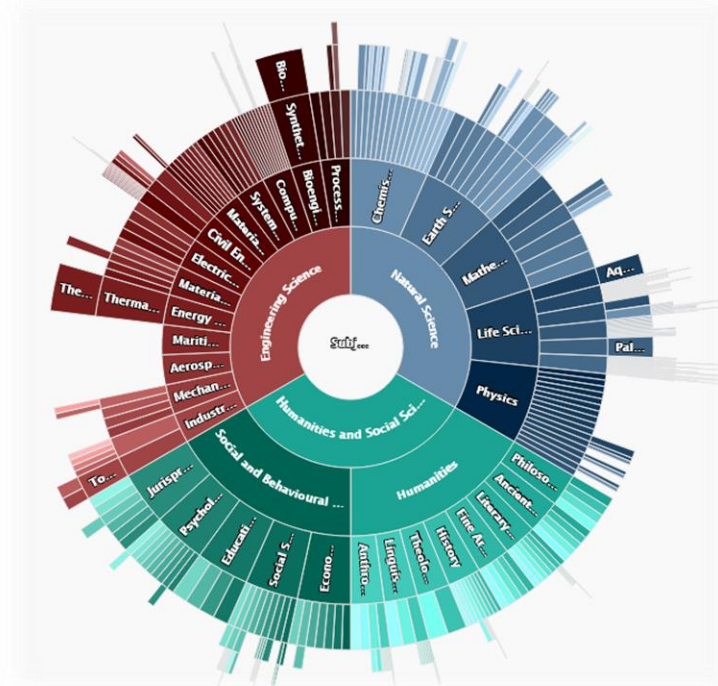


Database/Repositories



Policies

freicons.io/



<https://fairsharing.org/browse/subject>

Data organization

Naming conventions



Naming convention

No way!

Outline.docx

My.conf~abstrc#qwertz> <

Presentation 1 .ppt

figure.jpg

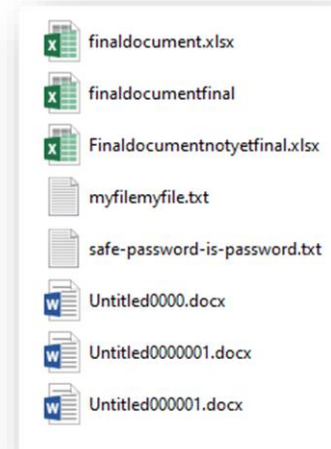
Better!

2022-05-20_RDM_Seminar_Outline_v01.docx

2022-05-22_Abstract_Nexus_Conference.docx

01_DMP_Introduction.pptx

Fig01_ScatterPlot_QuercusPetrea_Height_vs_Age.jpg



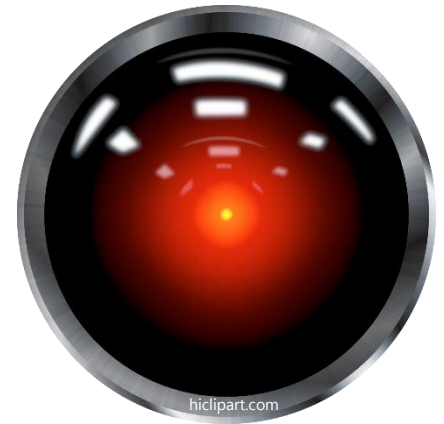
Someone else's folder?

Naming convention

Machine readable

(is not their fault if you are human)

- **No spaces, or special characters** `{[" § $%&?*+ #,.;` ^ < > °]}`
programs might not be able to read the files
- **Use delimiters "_" or "-" to frame keywords and metadata**
(makes search easier, listing and extract info from file names)
- **File ordering, left pad numbers with zeros** (01_Draft_*
02_Draft_* and not 1_Draft_*, 10_Draft_*)



Naming convention

Human readable

(pepl mght undrst ths bt isnt tht practcl)

- Use the **standard ISO 8601** strings for dates (*2022-05-20* or *20220520*)
- **Left pad numbers with zeros** for logical ordering (*01_Introduction.txt* *02_StudyArea*) or use the dates (*20220520_Kickoff_Presentation.pptx*)
- Use **Capital letters** *20220520_MeetingNotes_ProjectKickoff*)
- **File versioning** (*01_Introduction_v01.txt*)
- **Encode other Metadata** (*Author/Revision John Smith*
01_Introduction_v01_Rev_JoS.txt)



Think you search a file one hour before a deadline

Naming convention

Document your Naming Convention

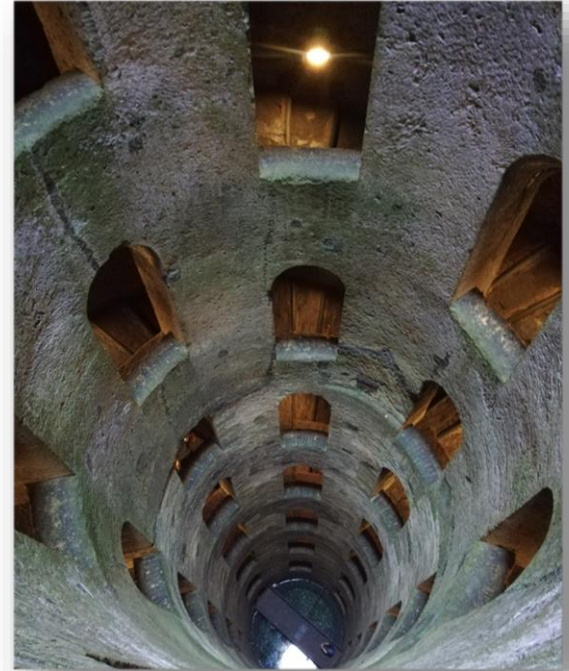
- Some naming best practices are project independent
- Some other are specific
- Think about what better fits your needs/project
- Attach your documentation as appendix to the DMP

[Additional resources](#)



Data organization

Storage, Backup & Recovery Strategy



Storage, Backup & Recovery Strategy

Organizations → 3-2-1 Backup rule



Primary
data



3 copies



2 types
Storage Media



1 Off-Site



Research Data Management

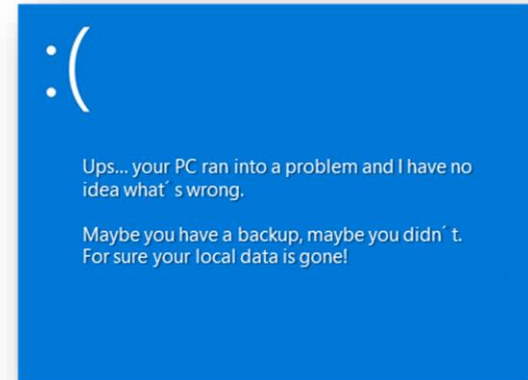
Storage, Backup & Recovery Strategy



<https://doi.org/10.5281/zenodo.5792643>

Cloud and collaboration

- Never save you data on your local PC!
- Your organization can provide you cloud storage for collaboration, with versioning and backup (FILR/[MS SharePoint](#))



Blue screen of death

[RDM horror stories Episode 1 - Lost Data](#)

Version Control Systems

Where is my FAIR code?

```
17 string sInput;
18 int iLength, iN;
19 double dblTemp;
20 bool again = true;
21
22 while (again) {
23     iN = -1;
24     again = false;
25     getline(cin, sInput);
26     system("cls");
27     stringstream(sInput) >> dblTemp;
28     iLength = sInput.length();
29     if (iLength < 4) {
30         again = true;
31         continue;
32     } else if (sInput[iLength - 3] != '.') {
33         again = true;
34         continue;
35     } while (++iN < iLength) {
36         if (isdigit(sInput[iN])) {
37             continue;
38         } else if (iN == (iLength - 3)) {
39             continue;
40         }
41     }
42 }
```

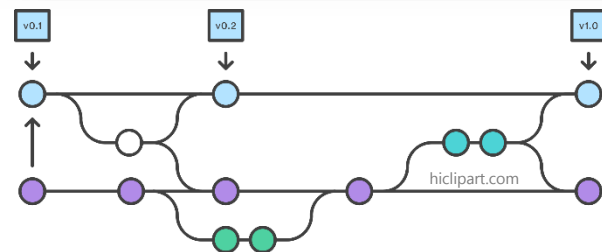
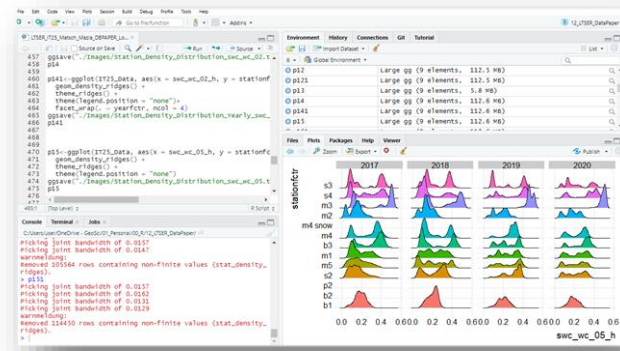
pixabay.com

Nowadays many journals require FAIR algorithms that reproduce results described in a scientific publication.

Management of algorithm and software

Data processing, model building are commonly executed in a scripting environment using various high-level programming languages (e.g. [R](#), [Python](#), [MATLAB](#))

Researcher that work with code and scripts use version control systems to manage changes and collaborate. [GitHub](#) and [GitLab](#) are web-based hosting platform to manage and publish code.



Github-Zenodo Integration

- Easy to publish on Zenodo your Github repository
- Ensure reproducibility and long-term archiving
- Zenodo version management by issuing a new DOI each time you create a new GitHub release

[How to link Github and Zenodo](#)

Data Browser Matsch | Mazia

DOI 10.5281/zenodo.6513553 license Apache-2.0 test passing

Introduction

Data Browser Matsch | Mazia is a user-friendly web-based application to visualize and download micrometeorological and biophysical time series of the [Long-Term Socio-Ecological Research site Matsch | Mazia in South Tyrol, Italy](#). It is designed both for the general public and researchers. The Data Browser Matsch | Mazia drop-down menus allow the user to query the InfluxDB database in the backend by selecting the measurements, time range, land use and elevation. Interactive Grafana dashboards show dynamic graphs of the time series.

- [IT25 LT\(S\)ER Eurac Website](#)
- [DEIMS ID](#)
- [LTER Italy Val Mazia](#)
- [Homepage Public Dashboards](#)
- [Software Description Paper](#)

Coding best practices

Make sure you can understand and reuse your code (*six months later*)

[Checkout some simple best practices](#)

When your code is working...



...but it is a mess