

# DATA PUBLICATION 101

# OUTLINE

- What is data publishing?
- Why data sharing?
- Data sharing policy
- Data sharing using repositories
  - What is a data repository?
  - Examples
  - How repositories lead to FAIR and/or open data (metadata, persistent identifiers, standards)
  - How to select a data repository?
  - How to use a data repository
- Data re-use, data citation, licenses
- Biblio data register (refer to demos)

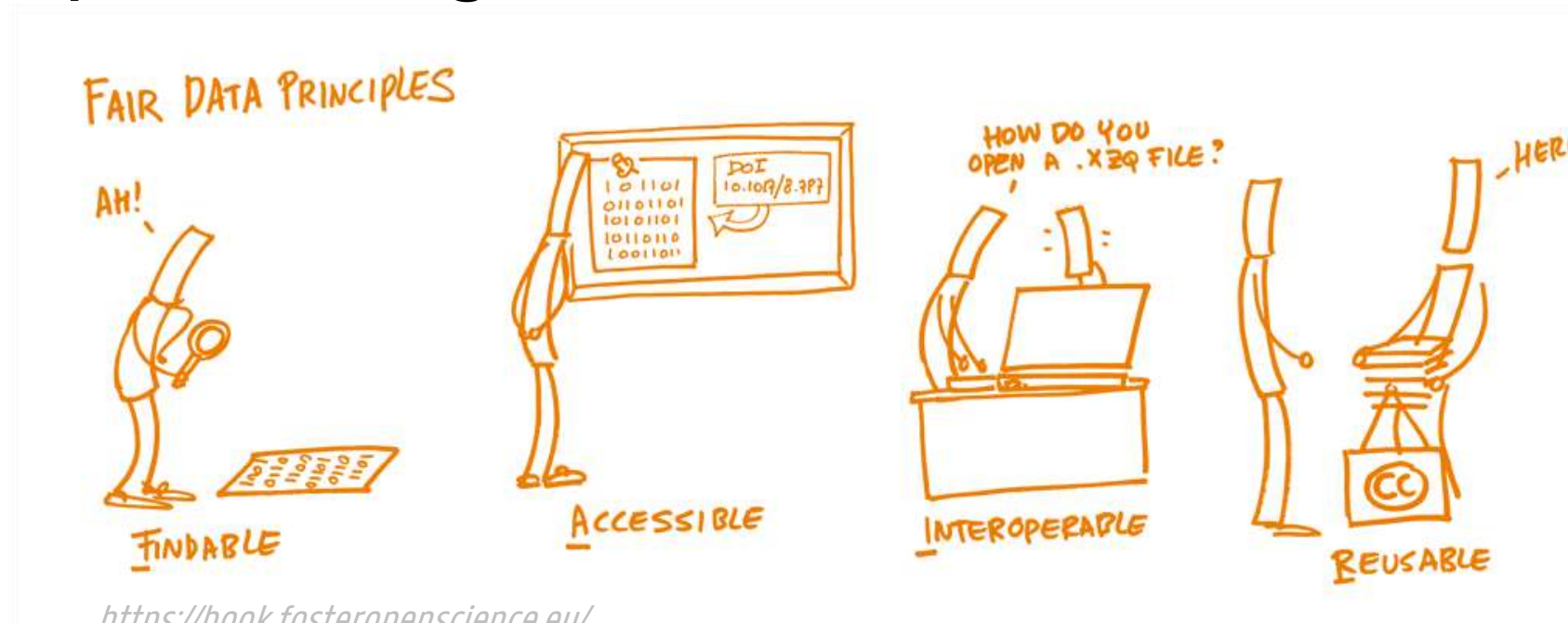
# PUBLISHING DATA

- Sharing data from
  - **completed** (parts of) research projects
  - **outside** your project or research team.
- Not
  - exchanging data with collaborators while your research is active.
  - Share data with your promotor



# RATIONALE FOR DEVELOPING FAIR PRINCIPLES

- Urgent need to improve the infrastructure supporting the reuse of research data
- Specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals
- The FAIR principles are a guideline to enhance the data reusability



# WHY SHARE DATA?

## THE BENEFITS





### FOR RESEARCHERS:

-  More visibility and citations.
-  Opportunities for collaboration.
-  Career recognition.
-  Decrease of non-compliance risks (legal, ethical, institutional and funders' policies).

### FOR SCIENCE:

-  Facilitates data finding and re-use.
-  Enables new research and new insights on the data.
-  Valuable data is protected.
-  Supports research integrity and reproducibility.

### FOR SOCIETY:

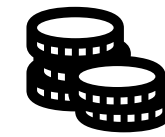
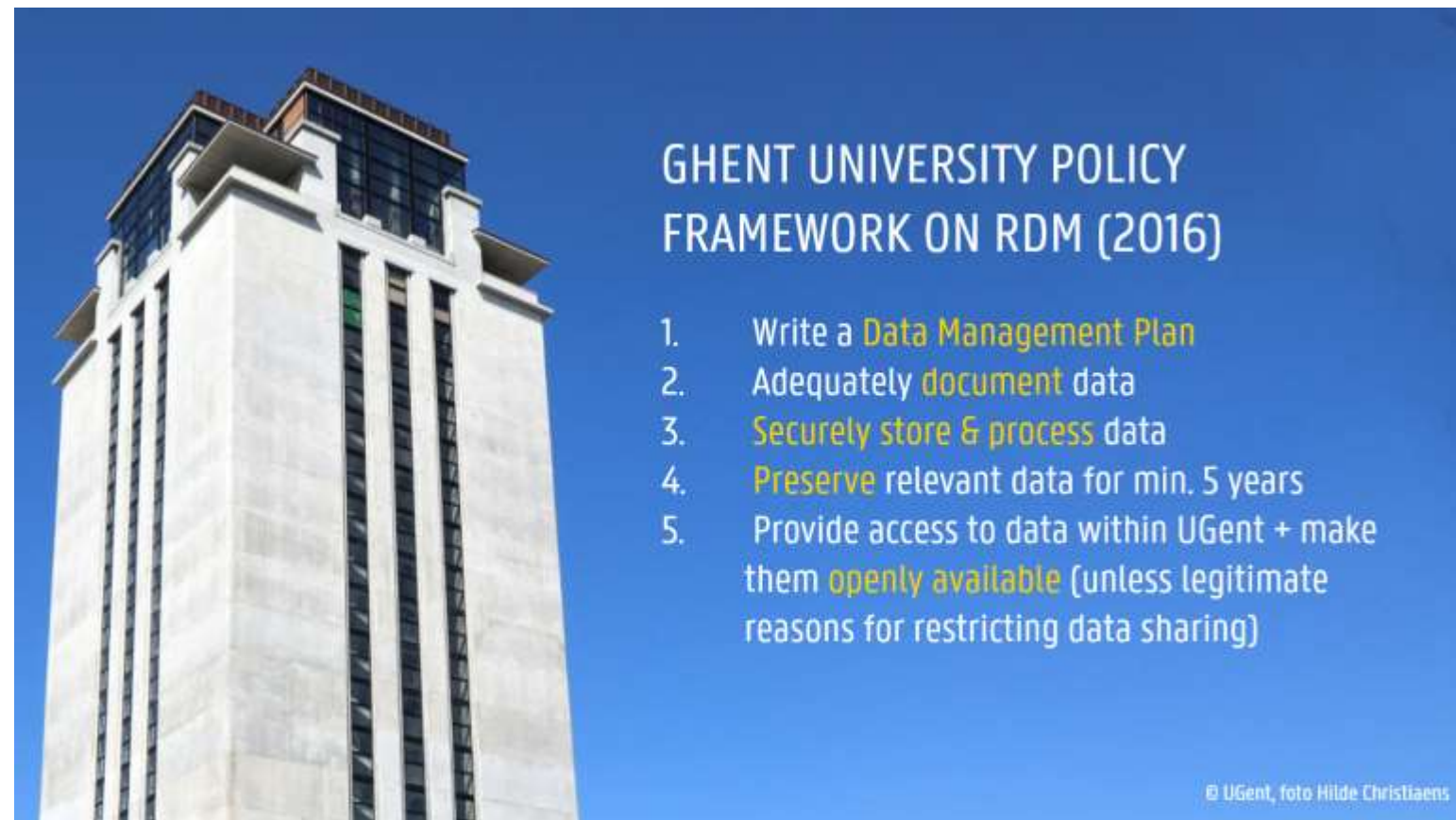
-  Efficient use of public resources.
-  Better quality research can lead to better decision-making.
-  Opportunities for citizen science.
-  Increased transparency and trust in science.



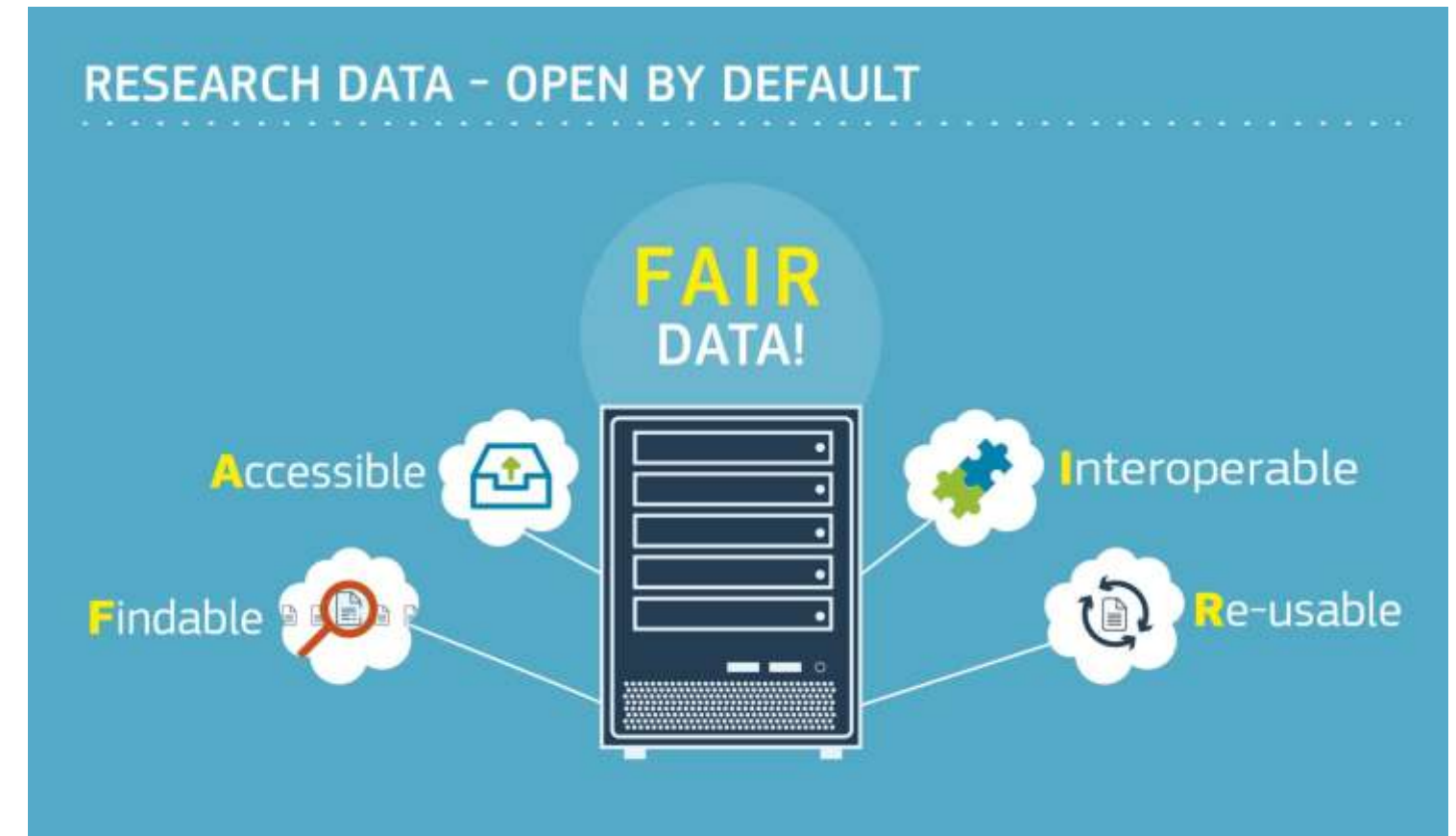
# POLICIES AND REQUIREMENTS ON DATA SHARING



## Institutional policies

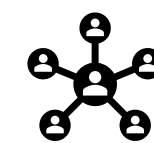


## HEU Open Science requirements



## Journals & publishers

***"PLOS journals require authors to make all data necessary to replicate their study's findings publicly available without restriction at the time of publication. When specific legal or ethical restrictions prohibit public sharing of a data set, authors must indicate how others may obtain access to the data."*** [PLOS ONE Data availability](#)



## ALLEA European Code of Conduct for Research Integrity



- Access as open as possible, as closed as necessary
- Data as legitimate and citable products of research

- RDM = key part of FWO policy for all support channels
  - “Sound data management is required for all research supported by the FWO.”
- Specific expectations:
  1. preserve data for a **minimum of 5 years** after end of research
  2. submit a **Data Management Plan**
  3. encourages researchers to ensure that their data management complies as much as possible with the **'FAIR' principles**
  4. advises researchers to apply the principle **'as open as possible, as closed as necessary'**
- Reporting
  - the FWO obliges to **register all relevant output** via the institutional ‘CRIS’ systems



### 3. Valuation of diverse research outputs

- **Research applications**
  - Adapted application form: Space for narrative, CV with a broad set of activities, 5 main publications and/or achievements, other scientific output and impact
  - Adapted scoring grid for evaluation
- **Scientific reporting**
  - Specific question: 'Are one or more referable datasets linked to this publication or to the project?'
  - Questions about metadata and accessibility

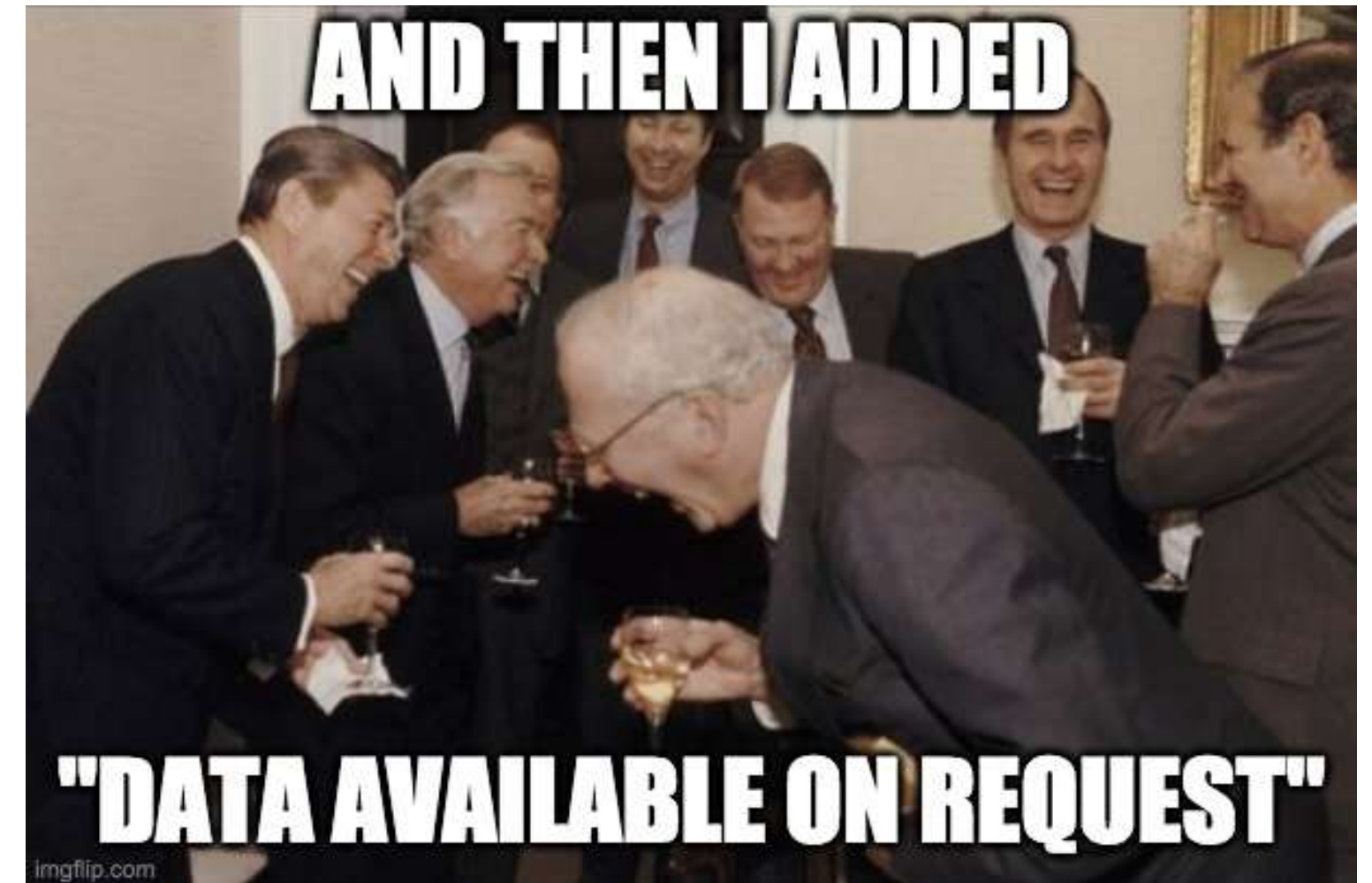
# WAYS OF SHARING DATA

# DATA UPON REQUEST

Often unavailable

Recent research has found that as little as **6,8%** of data was readily made available upon request

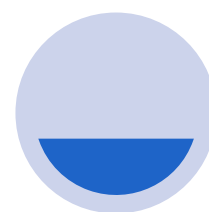
- Only 14% responded



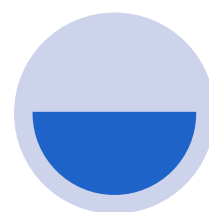
[https://figshare.com/articles/presentation/Handson\\_fMRI\\_code\\_and\\_data\\_sharing/12911165/1](https://figshare.com/articles/presentation/Handson_fMRI_code_and_data_sharing/12911165/1)



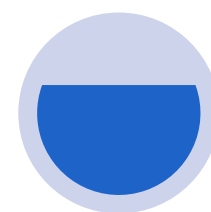
# WAYS OF SHARING DATA



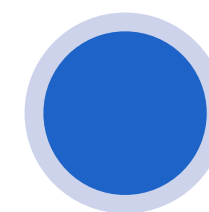
Upon request



Personal or  
project website



Article suppl.  
material



Data repository  
or archive



# DATA LINKED TO ARTICLE PUBLICATION

## Data availability

---

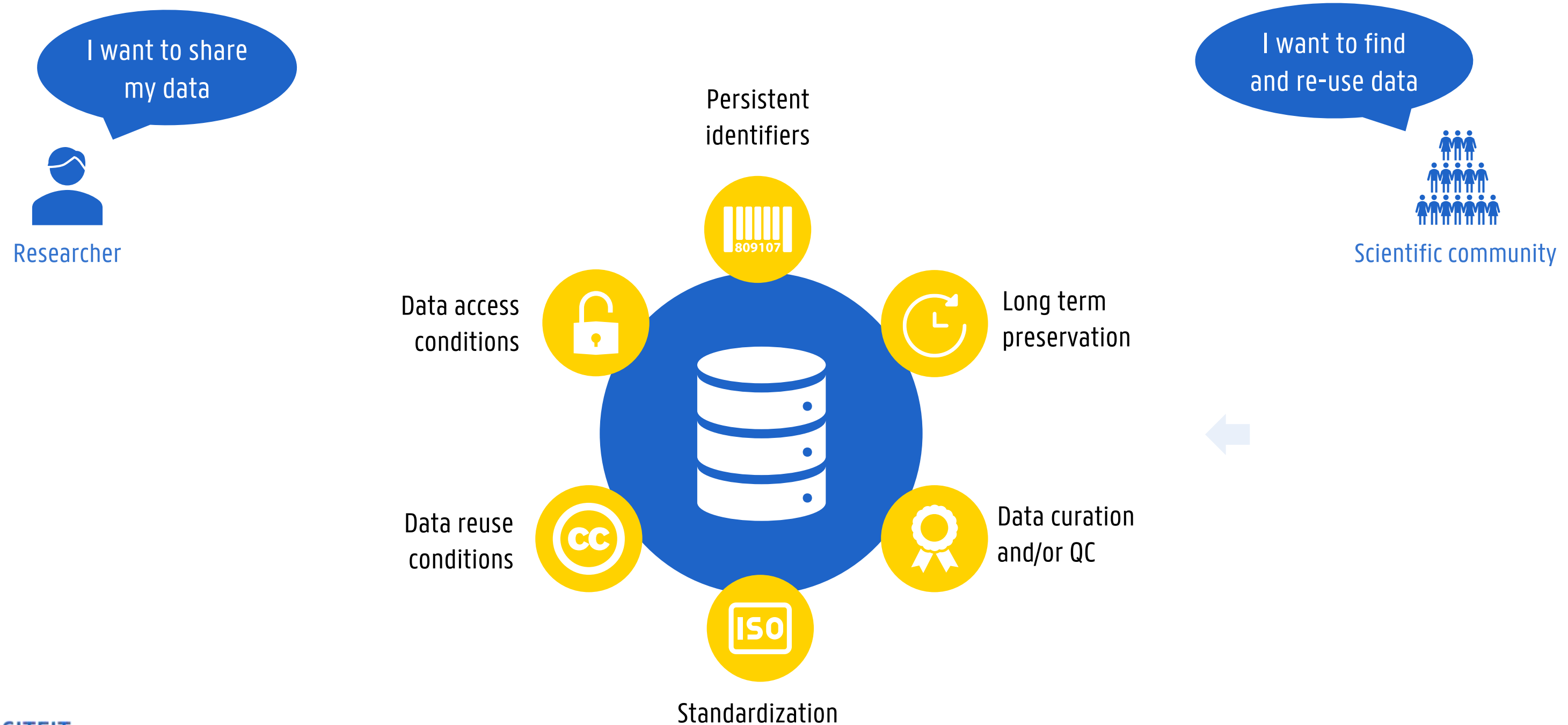
Raw data of scRNA-seq (and output files from CellRanger) and spatial transcriptomics (Visium and stereo-seq) are deposited at the GEO under accession number [GSE207592](#). Processed count matrices and cell annotations (scRNA-seq and Visium) are available online (<https://marinelab.sites.vib.be/en>). Molecular cartography data are available at Zenodo (<https://zenodo.org/record/6856193#.Ytj-fnZBz-g>). All other data supporting the findings of this study are available on reasonable request. [Source data](#) are provided with this paper.

# FAIR DATA IN REPOSITORIES



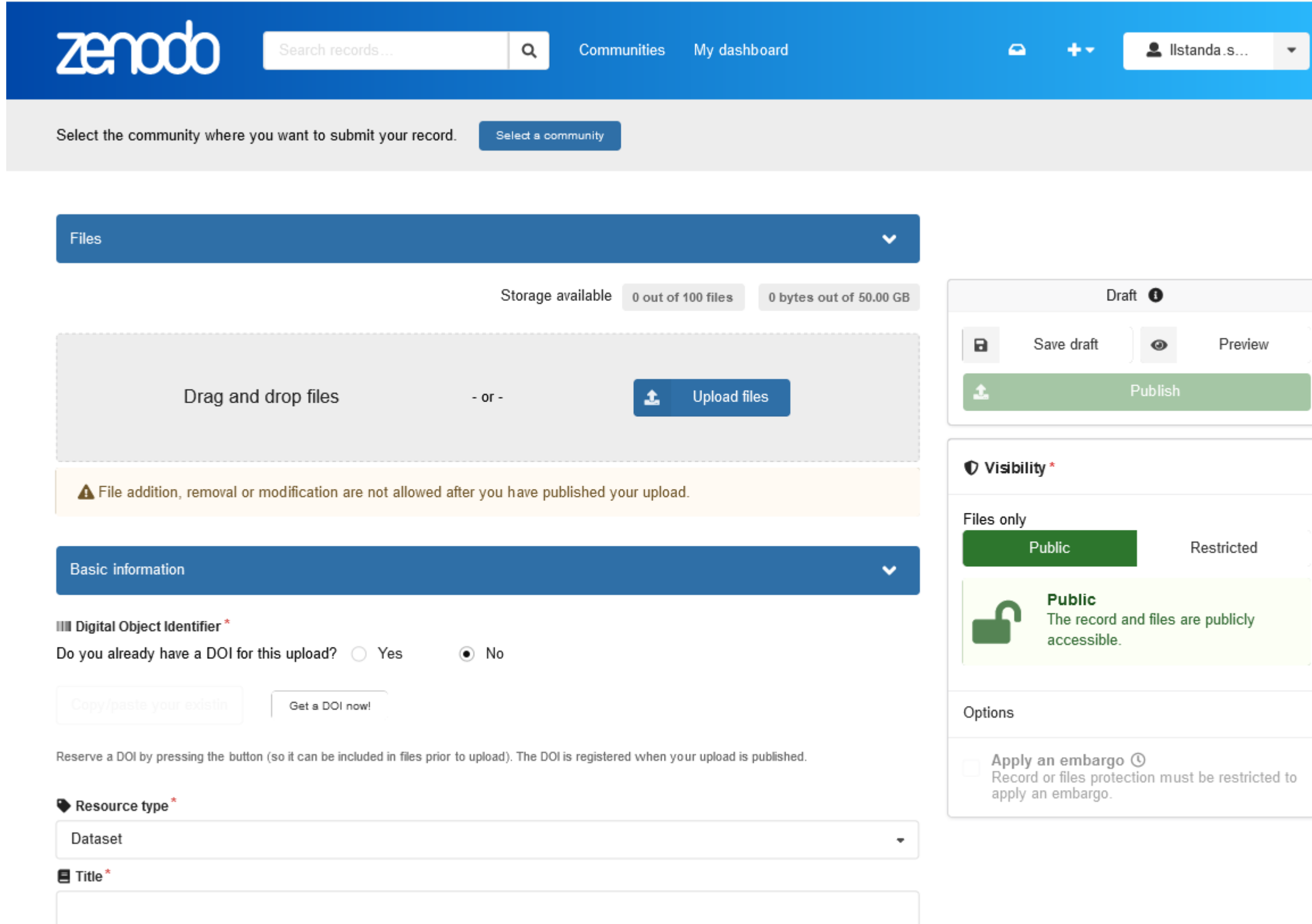
# WHAT IS A DATA REPOSITORY?

A database infrastructure that collects, manages and stores data for preservation and sharing



# EXPLORING DATA REPOSITORIES

- Some repositories have a test environment: [sandbox.zenodo.org/](https://sandbox.zenodo.org/) (vs. [zenodo.org/](https://zenodo.org/))



The screenshot shows the Zenodo upload interface. At the top is a blue header with the Zenodo logo, a search bar, and navigation links for 'Communities' and 'My dashboard'. Below the header is a grey bar with the text 'Select the community where you want to submit your record.' and a 'Select a community' button. The main content area is divided into two columns. The left column has a 'Files' tab, a storage status bar showing '0 out of 100 files' and '0 bytes out of 50.00 GB', a large dashed box for 'Drag and drop files' with an 'Upload files' button, and a warning message: 'File addition, removal or modification are not allowed after you have published your upload.' Below this is a 'Basic information' tab. Under 'Basic information', there is a section for 'Digital Object Identifier' with a question 'Do you already have a DOI for this upload?' and radio buttons for 'Yes' and 'No'. There are buttons for 'Copy/paste your existing DOI' and 'Get a DOI now!'. A note states: 'Reserve a DOI by pressing the button (so it can be included in files prior to upload). The DOI is registered when your upload is published.' Below this is a 'Resource type' dropdown menu set to 'Dataset' and a 'Title' input field. The right column contains a 'Draft' section with 'Save draft', 'Preview', and 'Publish' buttons. Below this is a 'Visibility' section with 'Files only' and 'Public' (selected) and 'Restricted' options. A green box indicates 'Public: The record and files are publicly accessible.' At the bottom is an 'Options' section with a checkbox for 'Apply an embargo' and a note: 'Record or files protection must be restricted to apply an embargo.'

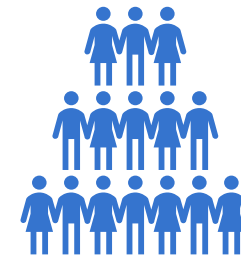
# WHAT IS A DATA REPOSITORY?

A database infrastructure that collects, manages and stores data for preservation and sharing



DATA  
FINDABILITY

FAIR 



RESEARCH COMMUNITY  
SUPPORT



PERSISTENT & UNIQUE  
IDENTIFIER

FAIR 



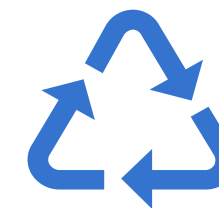
LONG TERM  
DATA PRESERVATION

FAIR 



DATA ACCESS CONDITIONS

FAIR 



DATA RE-USE CONDITIONS

FAIR 

# DATA REPOSITORIES & FAIR DATA

- Key infrastructures that provide functionalities to make your data FAIR.
- Shared responsibilities between repositories and researchers in making data FAIR

FAIR principle	In practice	Repository	Researcher
FINDABLE	Assign a persistent identifier (PID)	●	
	Data are described with <u>rich</u> metadata	●	●
	Metadata include the PID, are indexed & searchable online	●	
ACCESSIBLE	(Meta)data retrievable via PID using a standardised communications protocol (e.g. html)	●	
	Authentication and authorisation procedure	●	
INTEROPERABLE	Metadata standards	●	
	Controlled vocabularies	●	●
	References to other (meta)data	●	●
REUSABLE	(Machine readable) license / terms of use	●	●
	Detailed provenance documentation		●
	Domain standards	●	●

# DATA REPOSITORIES & FAIR DATA

- Key infrastructures that provide functionalities to make your data FAIR.
- Shared responsibilities between repositories and researchers in making data FAIR


FAIR principle	In practice	Repository	Researcher
FINDABLE	Assign a persistent identifier (PID)	●	
	Data are described with rich metadata	●	●
	Metadata include the PID, are indexed & searchable online	●	
ACCESSIBLE	(Meta)data retrievable via PID using a standardised communications protocol (e.g. html)	●	
	Authentication and authorisation procedure	●	
INTEROPERABLE	Metadata standards	●	
	Controlled vocabularies	●	●
	References to other (meta)data	●	●
REUSABLE	(Machine readable) license / terms of use	●	●
	Detailed provenance documentation		●
	Domain standards	●	●

Persistent Identifier

# PERSISTENT IDENTIFIERS

	Identifier	Globally unique identifier (guid, uuid)	Persistent identifier (PID)
What	a name that identifies a unique object gene identifier ENSG00000139618 book identifier ISBN 817525766-0	ensures that there are no two identical identifiers that point to different digital objects 123e4567-e89b-12d3-a456-426614174000	ensures that the digital object will remain findable over time and reduces the risk of broken links
How	can be (arbitrarily) assigned by anyone	by using a mechanism (e.g. using a registry service) for assigning identifiers	is determined by the commitment of the organisation that assigns and manages the identifier

## Examples of globally unique, persistent identifiers





UNIVERSITEIT  
GENT



Identifier:

Organisation:



Scope:



Digital objects



Researchers



Research organisations



# ORCID FOR IDENTIFICATION OF RESEARCHERS



Connection to contributions and affiliations

The screenshot shows the ORCID website interface. At the top, there's a header with the ORCID logo and navigation links. Below, a search results table is displayed, showing 40 of 65156 results. The table has columns for ORCID ID, First name, Family name, Other names, and Affiliations. The results list several entries for 'Juan Garcia' with their respective ORCID IDs and affiliations.

ORCID ID	First name	Family name	Other names	Affiliations
<a href="https://orcid.org/0000-0003-2957-9469">https://orcid.org/0000-0003-2957-9469</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0002-8047-6724">https://orcid.org/0000-0002-8047-6724</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0002-8913-427X">https://orcid.org/0000-0002-8913-427X</a>	Juan	Garcia		Instituto Tecnológico de Colima
<a href="https://orcid.org/0000-0002-1237-9009">https://orcid.org/0000-0002-1237-9009</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0001-7847-1193">https://orcid.org/0000-0001-7847-1193</a>	JUAN	GARCIA	Carey	Universidad de Guadalajara
<a href="https://orcid.org/0000-0003-4688-0077">https://orcid.org/0000-0003-4688-0077</a>	Juan	garcia		
<a href="https://orcid.org/0000-0001-6539-9524">https://orcid.org/0000-0001-6539-9524</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0001-9237-8729">https://orcid.org/0000-0001-9237-8729</a>	Juan	Garcia		GAD Municipal del Cantón Cuenca, Universidad de Cuenca, Universidad Católica de Cuenca
<a href="https://orcid.org/0000-0001-5525-3527">https://orcid.org/0000-0001-5525-3527</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0002-5326-9498">https://orcid.org/0000-0002-5326-9498</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0001-8876-2170">https://orcid.org/0000-0001-8876-2170</a>	Juan	Garcia		
<a href="https://orcid.org/0000-0002-4850-1491">https://orcid.org/0000-0002-4850-1491</a>	Juan	Garcia	Juan García Rodríguez	Universidad Complutense de Madrid
<a href="https://orcid.org/0000-0002-4582-4703">https://orcid.org/0000-0002-4582-4703</a>	Juan	Garcia		Complejo Hospitalario de Navarra
<a href="https://orcid.org/0000-0003-4516-4669">https://orcid.org/0000-0003-4516-4669</a>	Juan	Garcia		

Unambiguous identity



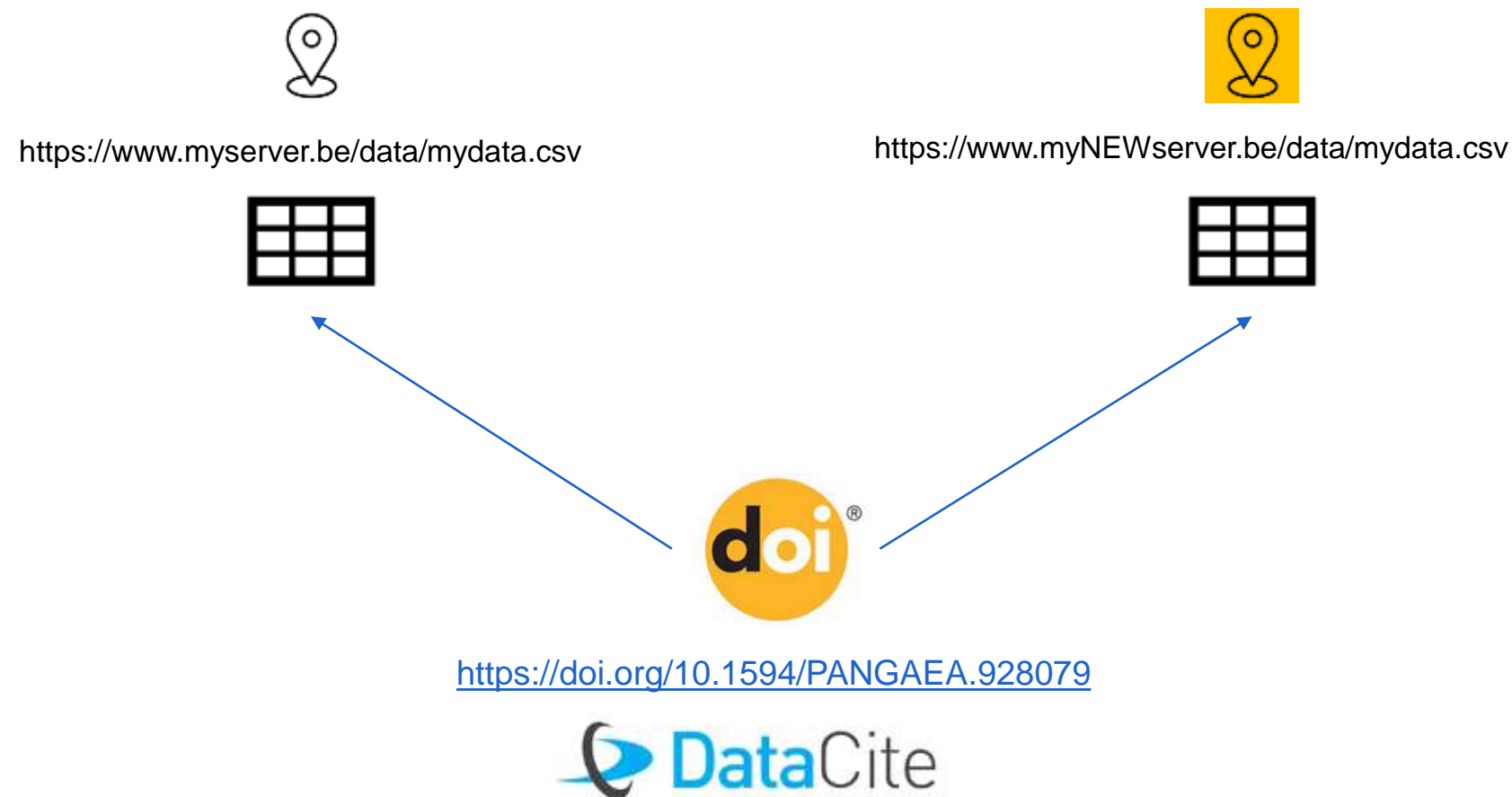
Commonly used and required

# PID: PERSISTENT IDENTIFIER

**Globally unique:** an unambiguous reference for humans or machines, an identifier associated with only one resource at any time

**Persistent:** stable and resolvable identifier for the long term

e.g. DOI: Digital Object Identifier



# DATA REPOSITORIES & FAIR DATA

- Key infrastructures that provide functionalities to make your data FAIR.
- Shared responsibilities between repositories and researchers in making data FAIR

FAIR principle	In practice	Repository	Researcher
FINDABLE	Assign a persistent identifier (PID)	●	
	Data are described with <u>rich</u> metadata	●	●
	Metadata include the PID, are indexed & searchable online	●	
ACCESSIBLE	(Meta)data retrievable via PID using a standardised communications protocol (e.g. html)	●	
	Authentication and authorisation procedure	●	
INTEROPERABLE	Metadata standards	●	
	Controlled vocabularies	●	●
	References to other (meta)data	●	●
REUSABLE	(Machine readable) license / terms of use	●	●
	Detailed provenance documentation		●
	Domain standards	●	●

Metadata



WHAT



WHEN



WHERE



HOW



WHO



WHY

## DOCUMENTATION

Any contextual information that allows data to be understandable and reusable.

The dataset...  
[Illegible text represented by yellow bars]

## METADATA

Information about data, recorded in a highly structured way, using a set of defined information fields or elements to be machine readable.

Title	[Illegible text represented by yellow bars]
Creator	[Illegible text represented by yellow bars]
Abstract	[Illegible text represented by yellow bars]
Keywords	[Illegible text represented by yellow bars]
Temporal scope	[Illegible text represented by yellow bars]
File type	[Illegible text represented by yellow bars]
License	[Illegible text represented by yellow bars]
Related publication	[Illegible text represented by yellow bars]

# METADATA IS DATA...

- About data
- A (very) structured type of documentation: records information in sets of pre-defined fields or elements.
- Meant to be read & exchanged by machines
- To facilitate (online) data discovery
- To quickly judge usefulness of data

## Toetsenbordindeling ^

De toetsenbordindeling  
verschilt per land. Het bepaalt... >

- ☒ Qwerty (Nederland) (328)
- ☐ Azerty (Belgie) (41)
- ☐ Azerty (Frankrijk) (16)

## Verbinding ^

Op welke manier wil jij dat jouw  
toetsenbord verbinding maak... >

- ☒ Draadloos (328)
- ☐ Bedraad (761)

## Wat vind je belangrijk? ^



Te veel keuze?

Beantwoord 5 korte vragen en vind jouw artikel

1 Top 10

★ Best beoordeeld

♥ Meest bewaard

328 resultaten

Sorteer



6 maanden  
gratis Norton

Logitech

**Logitech LGT-K360 Draadloos Keyboard Kantoor Usb 2.0 Us  
International Zwart**

Voorzien van een ISO Qwerty indeling (korte shift links, hoge enter)

Radio(RF) | QWERTY toetsenbord lay-out

★★★★★ (99)

# METADATA IN DATA REPOSITORIES


[Zenodo](#) data upload form

- Repositories: infrastructure to allow us to share & find data
- Provide **services to create & manage metadata**
- Typically, by filling a form with
  - Pre-defined fields
  - Lists of controlled terms to choose from (e.g. standard licenses)
- Metadata adheres to a specific metadata standard (interoperability)

## New upload


**Instructions:** (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.


[Files](#) >


[Communities](#)  recommended >


Upload type


required ▼


 Publication ☒


 Poster ☐


 Presentation ☐


 Dataset ☐


 Image ☐

 Video/Audio ☐

 Software ☐

 Lesson ☐

 Physical object ☐

 Other ☐


Publication type

Journal article ▼


Publication type.

Basic information


required ▼

 Digital Object Identifier


Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.




 Publication date \*

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.


 Title \*


Required.














 Authors \*

 ORCID (e.g.: 0000-0002-1825-0097)  

Optional.

 Add another author

 Description \*

  **B** *I* **S**  $x_2$   $x^2$           *I<sub>x</sub>*  $\Sigma$   $\Omega$   Source 



# DATA REPOSITORIES & FAIR DATA















- Key infrastructures that provide functionalities to make your data FAIR.
- Shared responsibilities between repositories and researchers in making data FAIR

FAIR principle	In practice	Repository	Researcher
FINDABLE	Assign a persistent identifier (PID)	●	
	Data are described with <u>rich</u> metadata	●	●
	Metadata include the PID, are indexed & searchable online	●	
ACCESSIBLE	(Meta)data retrievable via PID using a standardised communications protocol (e.g. html)	●	
	Authentication and authorisation procedure	●	
INTEROPERABLE	Metadata standards	●	
	Controlled vocabularies	●	●
	References to other (meta)data	●	●
REUSABLE	(Machine readable) license / terms of use	●	●
	Detailed provenance documentation		●
	Domain standards	●	●

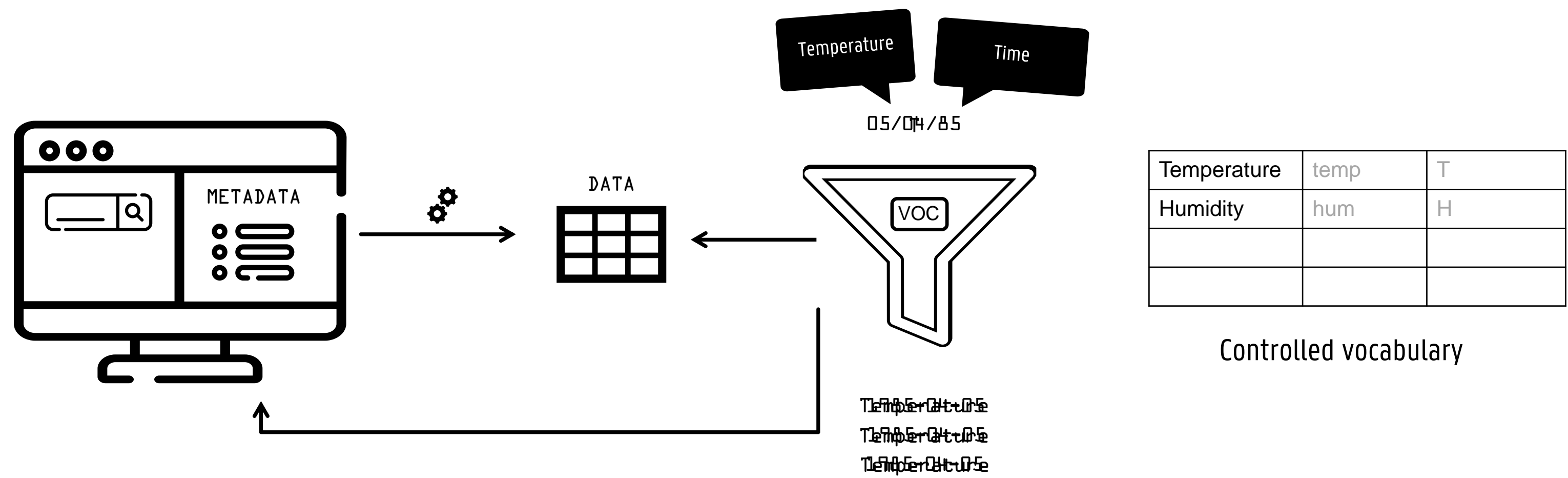
Vocabularies

# VOCABULARIES AND STANDARDS

## STANDARDIZED METADATA

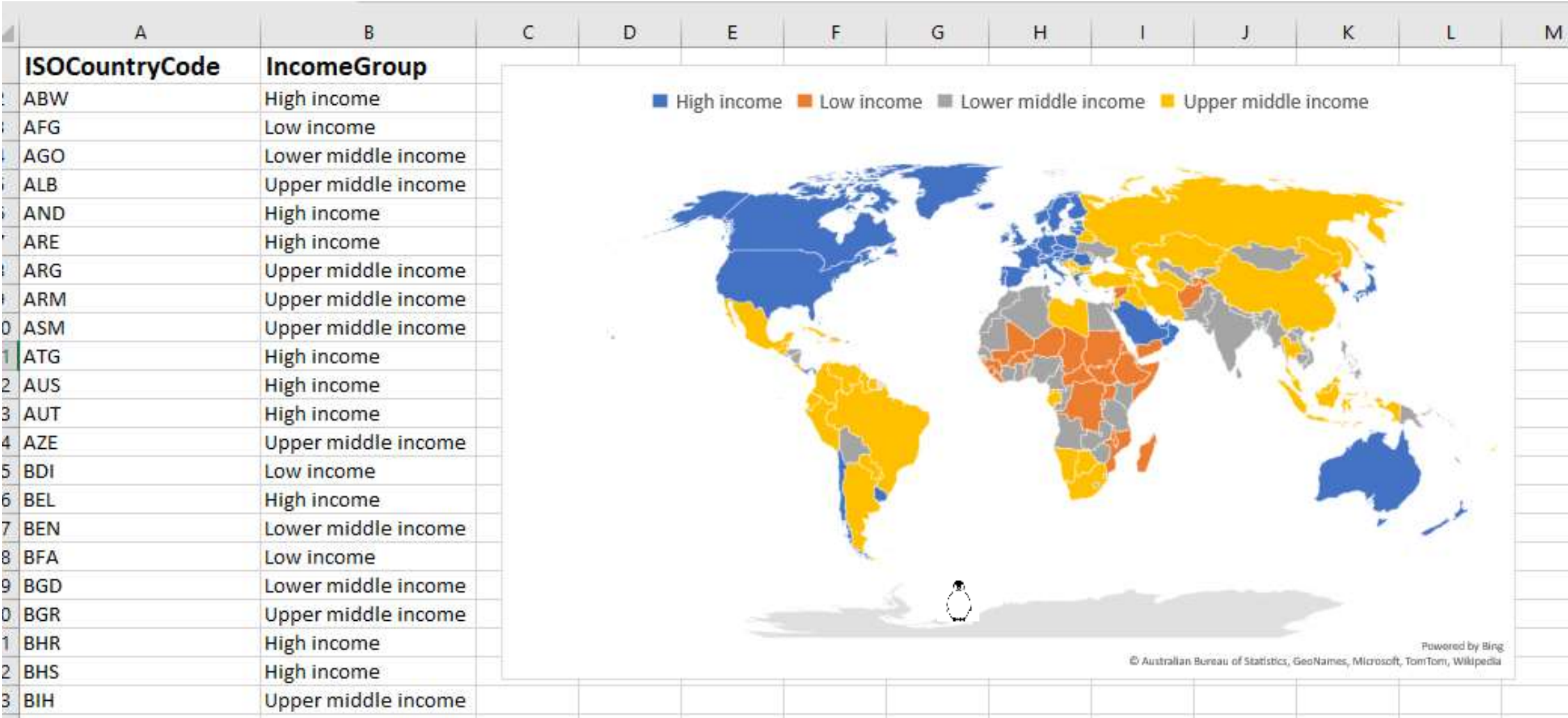
	Metadata fields or elements	Metadata values Vocabularies, ontologies, ISO standards
	Title	
	Creator	ORCID
	Description	
	Subject	
	Date	ISO-...
	Format	
	Rights	
	Related identifier	

# CONTROLLED VOCABULARIES TO STANDARDIZE TERMINOLOGY



# STANDARD VOCABULARIES & ONTOLOGIES

- Enable machine readability (use might be required in some domain repositories)
- Day-to-day example:



MS Excel recognizes [ISO Country codes](#) (or other place names in geonames.org) to automatically create maps

# FINDING REPOSITORIES

# WHICH TYPES OF REPOSITORIES EXIST?

## Types of repositories

### GENERIC REPOSITORY



### DOMAIN-SPECIFIC REPOSITORY



### INSTITUTIONAL REPOSITORY

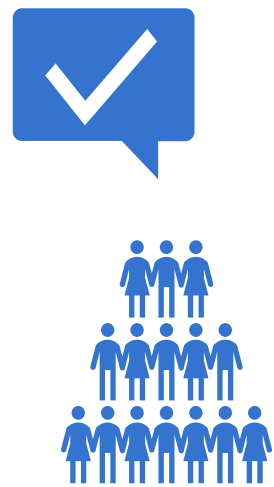




# HOW TO SELECT A DATA REPOSITORY?

## Recommended repositories

RESEARCH DOMAIN'S  
CONSENSUS



RESEARCHERS'  
BEST PRACTICES



JOURNAL  
RECOMMENDATIONS



e.g. [data repositories](#)  
recommended by Nature

DATA REPOSITORY  
REGISTRIES

**re3data.org**  
REGISTRY OF RESEARCH DATA REPOSITORIES

**FAIR**sharing

[re3data.org demo](#)

# HOW TO SELECT A DATA REPOSITORY?

- Does your publisher or funder recommend a repository?

For  
example:

## Data Repository Guidance

*Scientific Data* mandates the release of datasets accompanying our Data Descriptors,

### View data repositories

- **Biological sciences:** Nucleic acid sequence; Protein sequence; Molecular & supramolecular structure; Neuroscience; Omics; Taxonomy & species diversity; Mathematical & modelling resources; Cytometry and Immunology; Imaging; Organism-focused resources
- **Health sciences**
- **Chemistry and Chemical biology**
- **Earth, Environmental and Space sciences:** Broad scope Earth & environmental sciences; Astronomy & planetary sciences; Biogeochemistry and Geochemistry; Climate sciences; Ecology; Geomagnetism & Palaeomagnetism; Ocean sciences; Solid Earth sciences
- **Physics**
- **Materials science**
- **Social sciences**
- **Generalist repositories**

# HOW TO SELECT A DATA REPOSITORY?

- Does your publisher or funder recommend a repository?



Horizon Europe

For  
example:

**Deposit data in a trusted repository** and provide **open access** to it ('as open as possible, as closed as necessary')

## ELIXIR Deposition Database list

Deposition Database	Data type	International collaboration framework <sup>1</sup>
<a href="#">ArrayExpress</a>	Functional genomics data. Stores data from high-throughput functional genomics experiments.	
<a href="#">BioModels</a>	Computational models of biological processes.	
<a href="#">BioSamples</a>	BioSamples stores and supplies descriptions and metadata about biological samples used in research and development by academia and industry.	NCBI BioSamples database
<a href="#">BioStudies</a>	Descriptions of biological studies, links to data from these studies in other databases, as well as data that do not fit in the structured	

# GENERIC VS. DOMAIN-SPECIFIC REPOSITORIES

zenodo

covid-19

Upload

Communities

All versions

Found 1318 results.

< 1 2 3 4 5 6 7 8 9 >

Access Right

☒ Open (41069)

☐ Closed (8886)

☐ Restricted (107)

☐ Embargoed (23)

File Type

☐ Pdf (21859)

☐ Png (7726)

☐ Html (5956)

☐ Jpg (3216)

☐ Zip (1604)

☐ Hdf5 (624)

☐ Xml (224)

April 20, 2020 (Version 1.0)

Dataset

Open Access

COVID-19 CT Lung and Infection Segmentation Dataset

Ma Jun;

Ge Cheng;

Wang Yixin;

An Xingle;

Gao Jiantao;

Yu Ziqi;

Zhang Mingqin;

Liu >

Shucheng;

Wei Hao;

Mei Sen;

Yang Xiaoyu;

Nie Ziwei;

Li Chen;

Tian Lu;

Zhu Yuntao;

Zhu Qi

This dataset contains 20 labeled COVID-19 CT scans. Left lung, right lung, and infections are labeled by two experienced radiologist. To promote the studies of annotation-efficient deep learning methods, we set up 1 tasks based on

Uploaded on April 20, 2020

April 24, 2020 (v2)

Dataset

Open Access

Linked COVID-19 Data: Ontology

Florian Thiery;

Linked COVID-19 Data Ontology maintained at <https://github.com/Research-Squirrel-Engineers/COVID-19> 10.5281/zenodo.3757283 10.5281/zenodo.3757279

Uploaded on April 24, 2020

COVID-19 datasets in Zenodo



About ▾ Partners Related resources Bulk downloads Submit data

Viral Sequences

Host Sequences

Expression

Proteins

Biochemistry

Imaging

Literature

Viral sequences

Raw and assembled sequence and analysis of SARS-CoV-2 and other coronaviruses

Search

Search

Examples: ACE2, Severe acute respiratory syndrome 2...

Data types

All (737,152)

Sequences (72,683)

Reference sequences (2)

Raw reads (336,995)

Sequenced samples (314,481)

Studies (277)

Genes (22)

Browser (1)

Variants (12,691)

Organisms

☐ Severe acute respiratory syndrome coronavirus 2 (72,652)

☐ Canine respiratory coronavirus (23)

☐ Human coronavirus OC43 (8)

Showing 15 of 72,683 in Viral sequences > Sequences

powered by ENA

Edit table view

Download

Statistics

Phylogeny

<input type="checkbox"/>	Accession	Collection date	Country	Host	Strain	Isolate	Location
<input type="checkbox"/>	MN908947 <div>i</div>	Dec, 2019	China	Homo sapiens		Wuhan-Hu-1	
<input type="checkbox"/>	LR991698 <div>B.1.1.7</div> <div>i</div>	Sep 21, 2020	United Kingdom	Homo sapiens			
<input type="checkbox"/>	MT612306	May 23, 2020	Australia	Homo sapiens		SARS-CoV-2/human/AUS/VIC1725/2020	
<input type="checkbox"/>	MW185086	Aug 5, 2020	Australia	Homo sapiens		SA	2/1

COVID-19 data portal





# GENERIC VS. DOMAIN REPOSITORIES

- Domain-specific repositories
  - Visibility: where your peers will likely look for data
  - Use of discipline standards & vocabularies
    - Search filters, visualization tools tailored to domain
    - Interoperability: easier to integrate multiple datasets
  - Dataset curation by repository advisors often part of submission process
- Generic repositories:
  - Provide rich metadata to make it more findable & reusable
- Generic AND domain:
  - DOI, link to publication(s), dataset(s) or other outputs
  - Support minimum degree of FAIRness





# DOMAIN SPECIFIC: LIFE SCIENCES



[Tips on how to search](#)

- ABOUT
- SUBMISSION
- BROWSE
- ACCESS
- DOWNLOAD
- METADATA
- Helpdesk
- Log in

## Study

### WTCCC case-control study for Coronary Artery Disease

Study ID	Alternative Stable ID	Type
EGAS000000000003		GWAS

#### Study Description

WTCCC genome-wide case-control association study for Bipolar disorder (CAD) using the 1958 British Birth Cohort and the UK National Blood Service collections as controls.

#### Study Datasets 3 datasets.

Click on a Dataset ID in the table below to learn more, and to find out who to contact about access to these data





zenodo

caretta caretta

Upload Communities

All versions

Access Right

- ☐ Open (27)
- ☐ Closed (14)

File Type

- ☐ Pdf (18)
- ☐ Png (4)
- ☐ Xlsx (3)
- ☐ 7z (1)
- ☐ Html (1)
- ☐ Xml (1)
- ☐ Zip (1)

Keywords

- ☐ Biodiversity (22)
- ☐ Taxonomy (22)
- ☐ Animalia (15)

*"Caretta caretta" in Zenodo*

Occurrences

Search all fields

Simple Advanced

Occurrence status

License

Scientific name

☒ Caretta caretta (Linnaeus, 1758)

Basis of record

Location

Administrative areas (gadm.org)

Coordinate uncertainty in meters

Year

No filter

1000 2021

Month

Dataset

Country or area

Search filters tailored to domain & data type -> facilitated by domain metadata standard

*"Caretta caretta" in GBIF*

SEARCH OCCURRENCES | 14,390 WITH COORDINATES

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Map showing distribution of *Caretta caretta* occurrences across the Mediterranean and surrounding regions.

SEARCH OCCURRENCES | 16,744 RESULTS

Default Custom

CURRENT FILTER

Scientific name Caretta caretta (Linnaeus, 1758)

OCCURRENCES PER MONTH

OCCURRENCES PER YEAR

OCCURRENCES PER LICENSE

OCCURRENCES PER BASIS OF RECORD

Observation Machine observation Human observation Material sample Living specimen Unknown


CC BY-NC 4.0 CC0 1.0 CC BY 4.0

Data visualization -> facilitated by data standardization




## Jupyter notebooks

**STEP 1** → SHE DESCRIBES THE EXPERIMENTS AS A

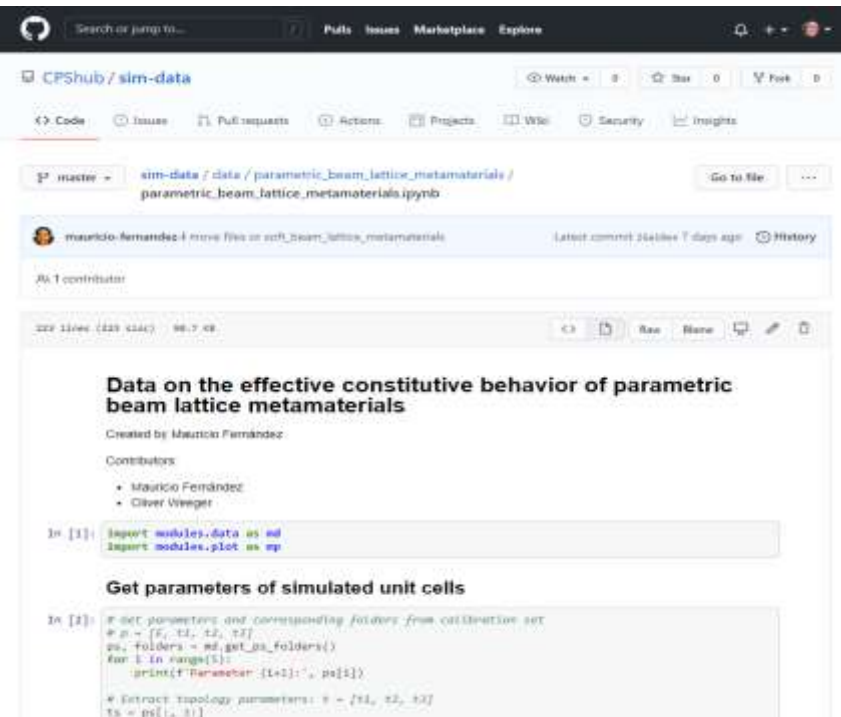
 **NOTEBOOK, MIXING:**

PROSE  
CODE &  
VISUALIZATION

AND RESOURCES:  
SOURCE CODE,  
DATA,  
MEDIA...



<https://opendreamkit.org/2017/11/02/use-case-publishing-reproducible-notebooks/>



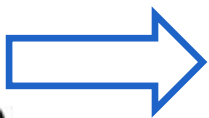
<https://github.com/CPSHub/sim-data>



CODE OCEAN  
<https://codeocean.com>



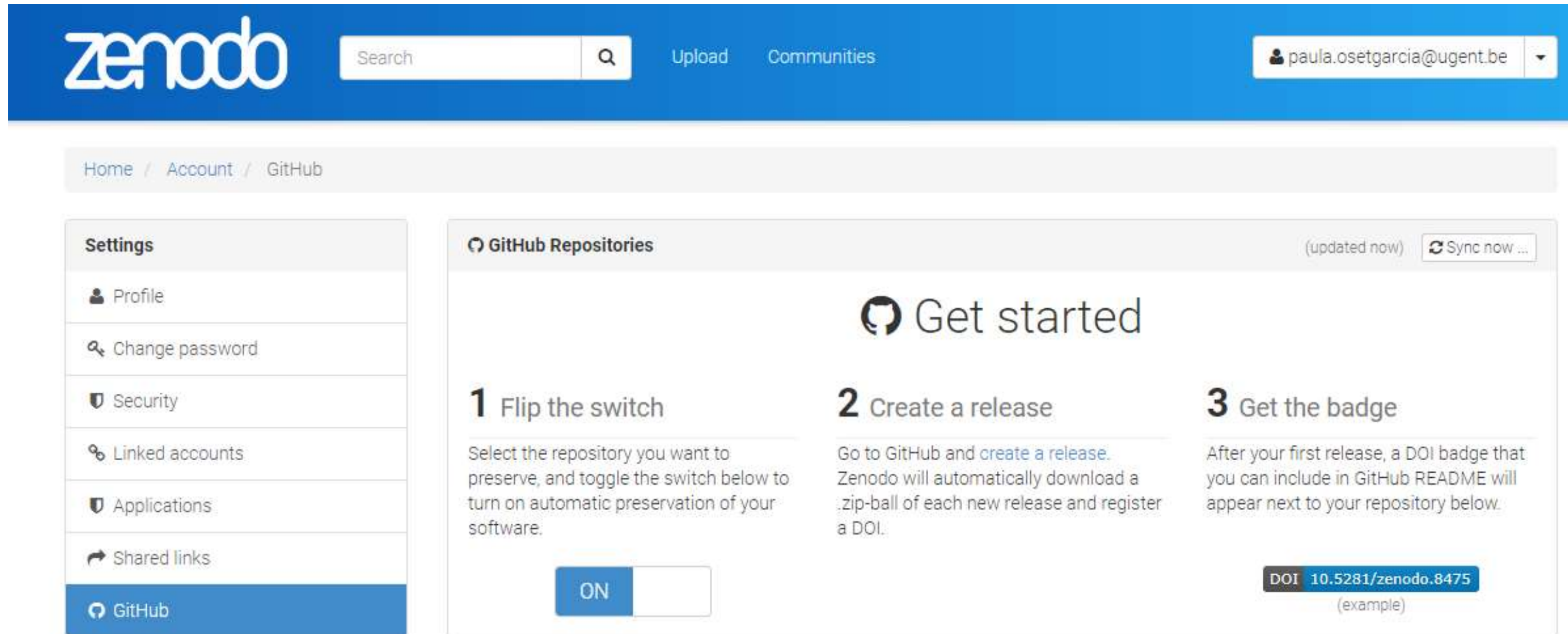
## Publish git repositories



<https://guides.github.com/activities/citable-code/>

# SHARING CODE AND MAKING CODE CITABLE

- GitHub – Zenodo integration: <https://guides.github.com/activities/citable-code/>



The screenshot shows the Zenodo website interface. At the top is a blue header with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. A user profile for 'paula.osetgarcia@ugent.be' is visible in the top right. Below the header is a breadcrumb trail: 'Home / Account / GitHub'. The main content area is divided into two sections. On the left is a 'Settings' sidebar with options: Profile, Change password, Security, Linked accounts, Applications, Shared links, and GitHub (which is highlighted in blue). The right section is titled 'GitHub Repositories' and includes a '(updated now)' status and a 'Sync now ...' button. Below this is a 'Get started' guide with three steps: 1. Flip the switch (with a toggle switch currently set to 'ON'), 2. Create a release (with instructions to go to GitHub and create a release), and 3. Get the badge (with instructions on how to use the DOI badge). An example DOI badge is shown: 'DOI 10.5281/zenodo.8475 (example)'.

# THE JOURNEY TO FAIR DATA



# PREPARING DATA

## DATA DEPOSITION CHECKLIST (GENERIC)

- Gather all relevant (meta)data needed for reproducibility:
  - experimental method details, raw data files, organized data tables, scripts and code, data visualizations, and statistical output
- Migrate data to recommended, reusable formats
- Organize files in a logical schema
- Provide a separate readme file



Jørgen Stamp -  
Digitalbevaring.dk



# DATA LINKED TO ARTICLE PUBLICATION

<https://doi.org/10.1098/rspb.2022.2456>

- **Neutral processes underlying the macro eco-evolutionary dynamics of mixed-ploidy systems**
- [Felipe Kauai](#), [Frederik Mortier](#), [Silvija Milosavljevic](#), [Yves Van de Peer](#) and [Dries Bonte](#)
- All relevant data are within the manuscript and its electronic supplementary material. The code developed for running numerical simulations is available on a GitHub repository at <https://github.com/KauaiFe/Polyploidy-V1.0>. Readers will also find the code documentation within the electronic supplementary material
- Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.6456230>

# DATA LINKED TO ARTICLE PUBLICATION

<https://doi.org/10.1111/mec.16865>

- Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an island-colonizing bird
- [Claudia A. Martin](#), [Eleanor C. Sheppard](#), [Juan Carlos Illera](#), [Alexander Suh](#), [Krystyna Nadachowska-Brzyska](#), [Lewis G. Spurgin](#), [David S. Richardson](#)
- The genomic data supporting this study and code used to perform the data analyses within this article are openly available in the Dryad Digital Repository: <https://doi.org/10.5061/dryad.ksn02v75k> (Martin et al., [2022](#)).

Supporting Information ^	
Filename	Description
<a href="#">mec16865-sup-0001-AppendixS1.pdf</a> PDF document, 1.1 MB	<b>Appendix S1</b>

# DATA LINKED TO ARTICLE PUBLICATION

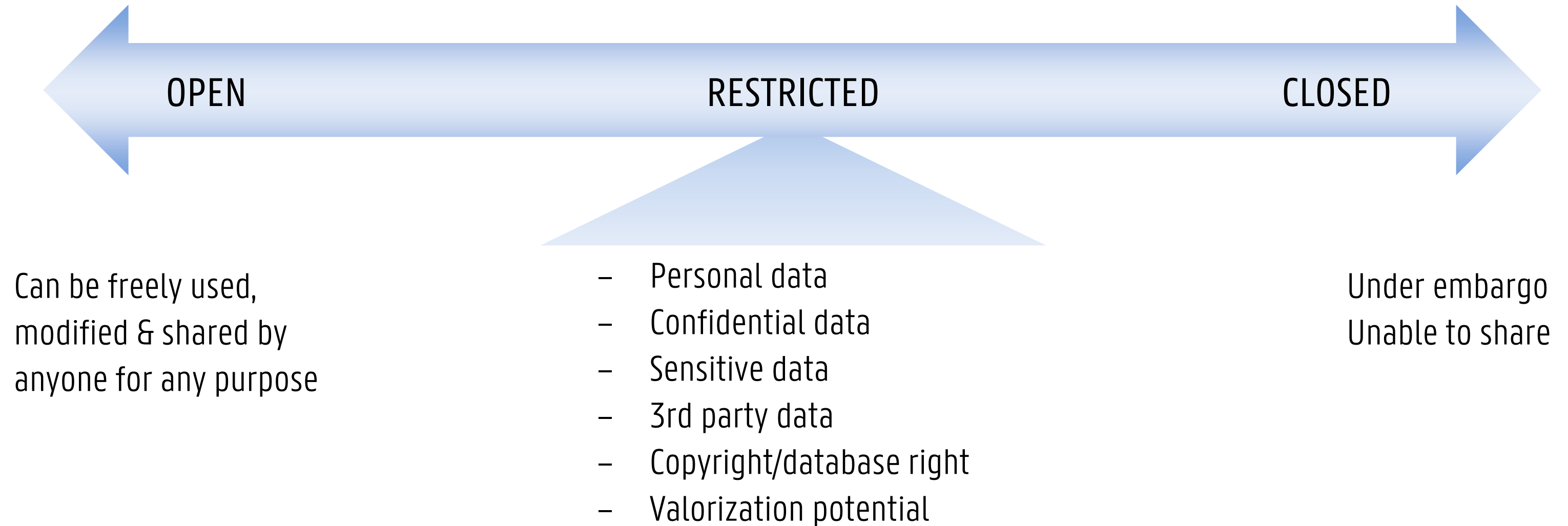
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9627755/>

- **At age 9, the methylome of assisted reproductive technology children that underwent embryo culture in different media is not significantly different on a genome-wide scale**
- Rebekka M Koeck, Florence Busato, Jorg Tost, Heleen Zandstra, Sylvie Remy, Sabine Langie, Marij Gielen, Ron van Golde, John C M Dumoulin, Han Brunner, Masoud Zamani Esteki, and Aafke P A van Montfoort
- The raw and processed array data from this study have been uploaded to the Gene Expression Omnibus (GEO) database and are available under the accession number GSE196432 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE196432> ). The raw array data from the FLEHS study are available within the GEO database under the accession number GSE110128.

# ALWAYS OPEN SHARING?

# WHAT ABOUT SHARING OF SENSITIVE OR CONFIDENTIAL DATA?

“As open as possible, as closed as necessary”



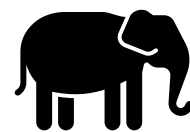
# RESTRICTIONS ON DATA SHARING



Personal data



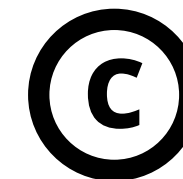
Confidential data



Sensitive data



Third-party data



Copyright



Valorization



# TAKE-HOME MESSAGES

- Publishing data can provide personal (career) advantages, and it benefits science and society at large
- Sharing data is a growing requirement from research institutes, funders and journals
- The recommended way to share data is using data repositories
  - Data repositories provide essential functionalities to make your data FAIR
- Check repository requirements to prepare for deposit during research
- Putting your data in a repository  $\neq$  making data open

# Ghent University Data Stewards

RESEARCH DEPARTMENT - UNIVERSITY LIBRARY



[rdm.support@ugent.be](mailto:rdm.support@ugent.be)



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



[ugent.be/en/research/datamanagement](https://ugent.be/en/research/datamanagement)