cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Research Data Management

International Summer School in Uganda

Dr. Anja Perry

Oliver Watteler

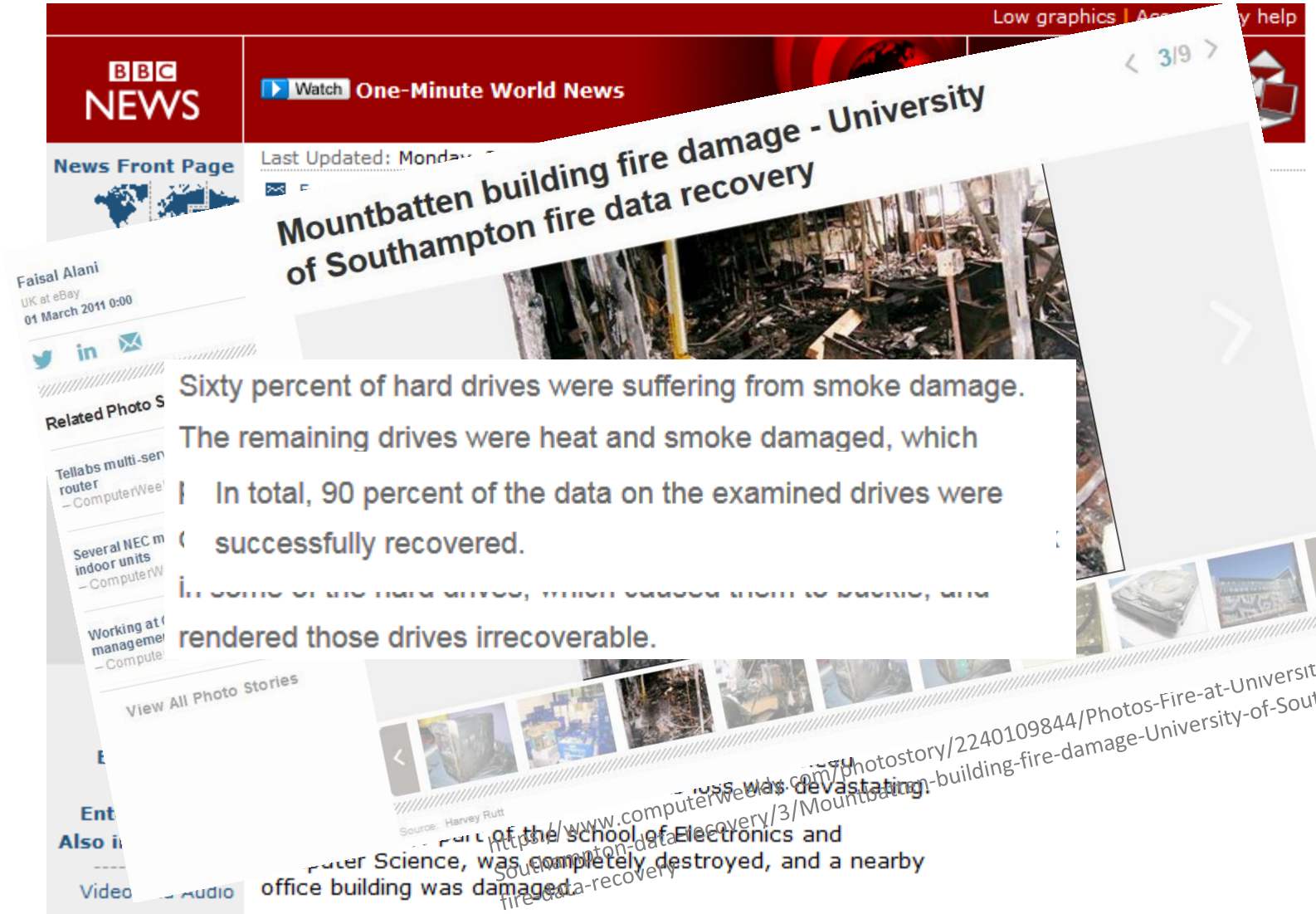
gesis Leibniz Institute for the Social Sciences

Image: pixabay (CC-0)

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Secure storage

Do you think your data are safe?



Protection against unauthorized access

- Personal data needs to be securely stored
- Using appropriate technical and organisational measures (Section 20(1), DPPA, Art. 5.1 GDPR)
- Consider state of the art
- „Integrity and confidentiality“

Technical and organisational measures

Organisational measures

- Separate personal data from research data
- Anonymize data
- Store personal data in countries with similar protection

Technical measures

- Control access to data
- Protect data with a secure password
- Encrypt data
- Destruct data that is no longer needed

Creating a strong password

- A strong password has
  - eight to fifteen characters or even more
  - a random distribution of characters
- Combine…
  - … upper case letters: A - Z
  - … lower case letters: a - z
  - … numerals: 0 - 9
  - … special characters: ! " # $ % & ' ( ) * + , - . / : *etc.*
- Use a ‘pass-sentence’ instead of a password

Passwords explained

https://www.youtube.com/watch?time_continue=209&v=jtFc6B5lmIM

Switch on English subtitles!

Data encryption

- **Maintains the security of data**
  - uses an algorithm to transform information
  - needs a “key” to decrypt
- **Use encryption**
  - to transfer data
  - to store data (back-ups)
  - on remote discs

  e.g. 7Zip, Gpg4win or Veracrypt

Bild: pixabay (CC-0)

Destructing sensitive / personal data

- **Temporary storage, according to**
  - informed consent
  - period of research project (reusability)
- **Destructing data**
  - Deleting data might probably not destruct them!
  - Shred physical files
  - Overwrite digital files (e.g. using *Eraser* or *BCWipe*)

    ⇒ if in doubt, seek to physically destroy the drive

Image: pixabay (CC-0)

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Data sharing

What is data sharing (for)?

- Make data available (for scholarly research) to other investigators
- Creates transparency
  - enables re-use of data in new research contexts
  - thereby contributes to the efficient use of public funds (i.e. tax money)

The spectrum of open data

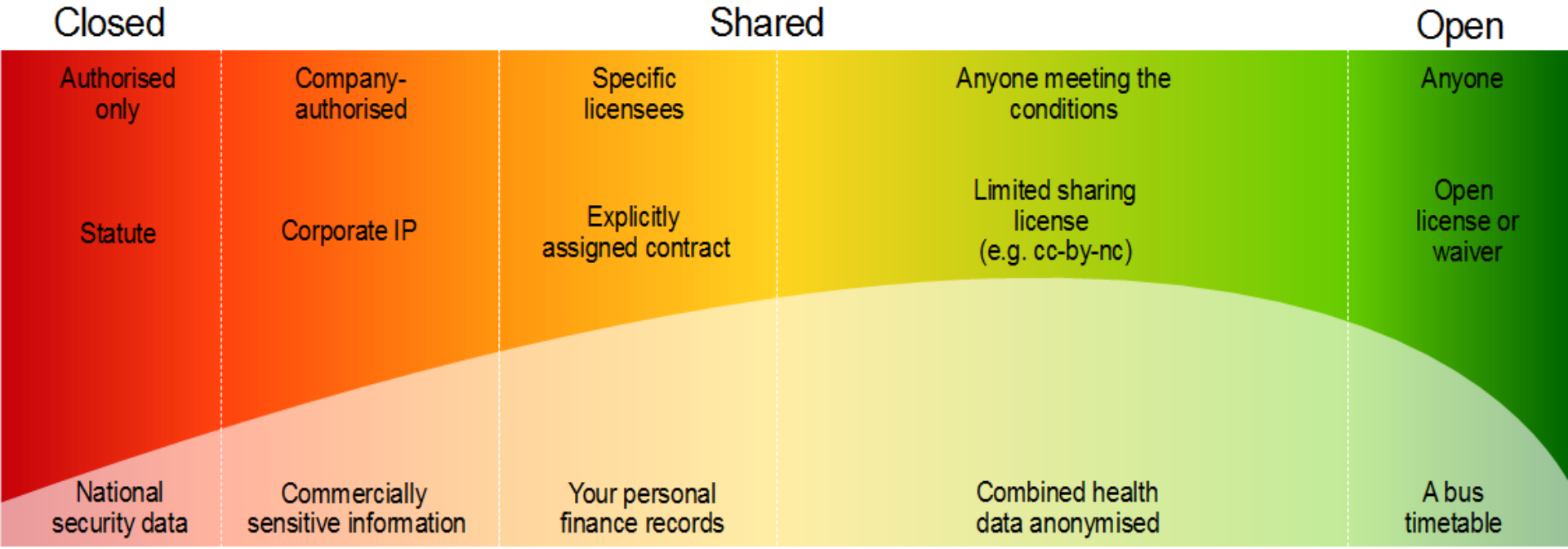
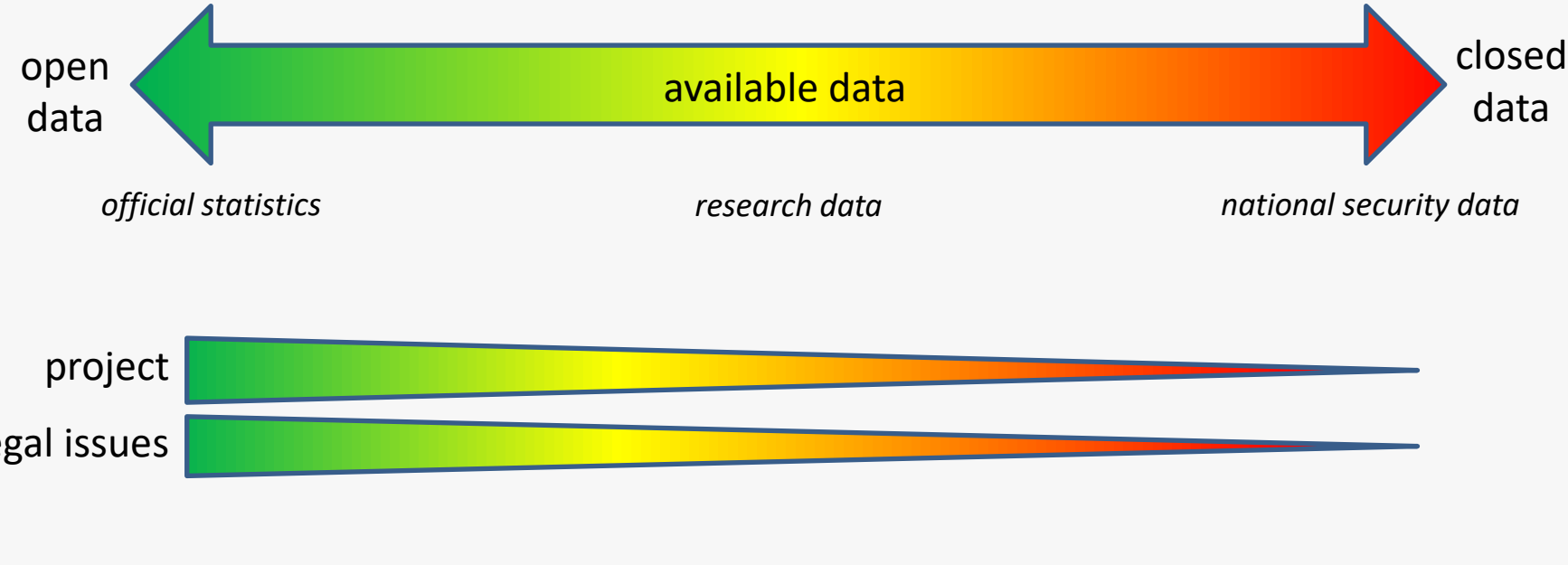


Image: "The Data Spectrum". Adapted from http://theodi.org/data-spectrum (Open Data Institute, cc-by)

Data Sharing and Accessibilty



⇒ as open as possible, as closed as necessary

Access and sharing during the research process

Private Research Domain

Who? Access for core research team

What? "Everything" is shared (working files, raw data…)

Icons: RRZE Erlangen (CC-BY-SA) https://github.com/RRZE-PP/rrze-icon-set

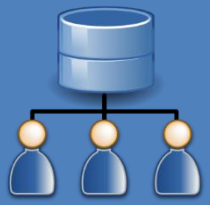
Access and sharing during the research process

Private Research Domain

Who? Access for core research team

What? "Everything" is shared (working files, raw data…)

Shared Research Domain



Who? Limited access for researchers outside core team (e.g. for peer review, replication)

What? "Stable" versions of data and documentation

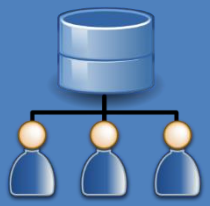
Access and sharing during the research process

Private Research Domain

Who? Access for core research team

What? "Everything" is shared (working files, raw data…)

Shared Research Domain



Who? Limited access for researchers outside core team (e.g. for peer review, replication)

What? "Stable" versions of data and documentation

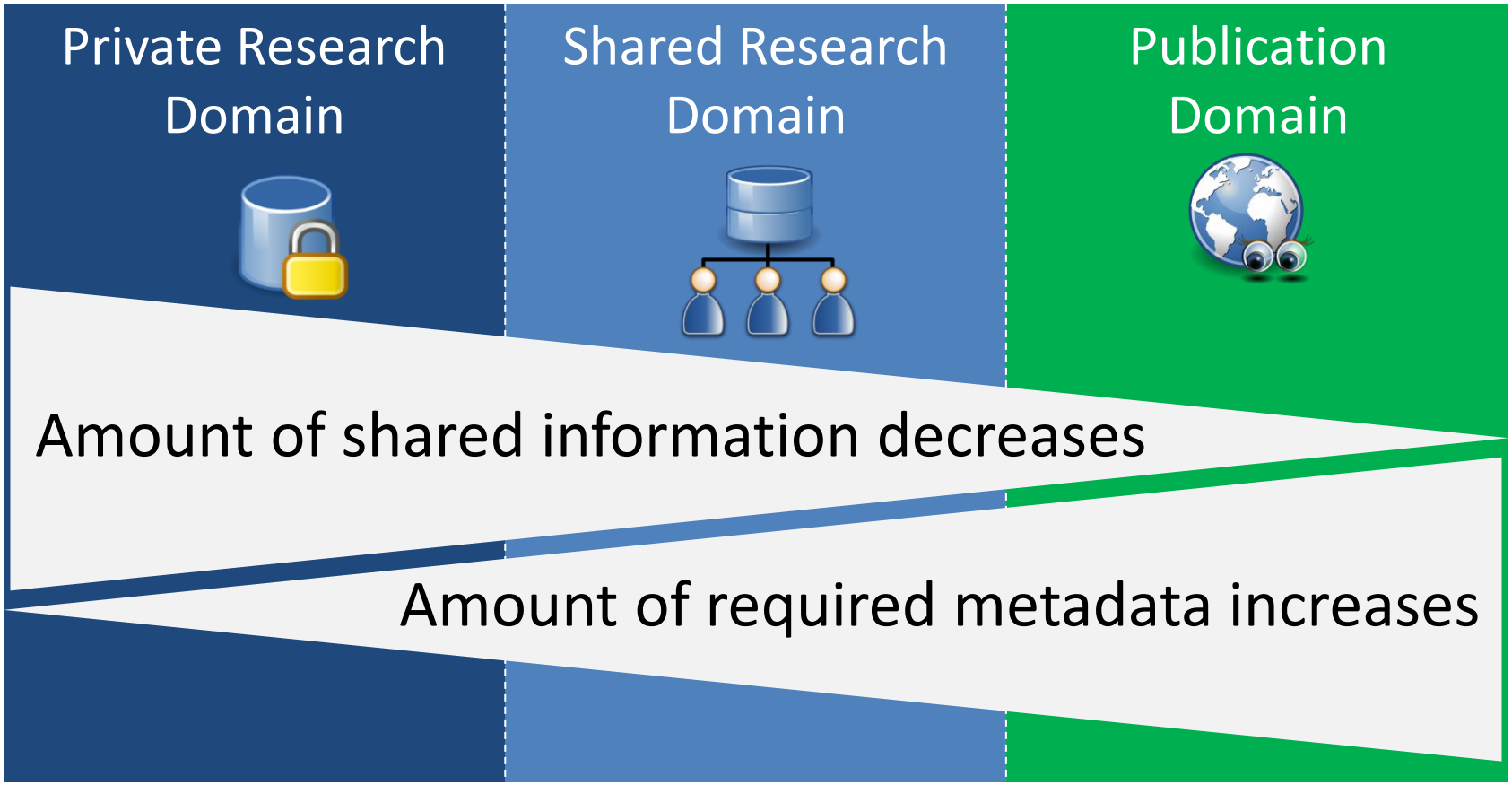
Publication Domain

Who? (Long-term) access for research community or public

What? "Final" versions of data and documentation

Icons: RRZE Erlangen (CC-BY-SA) https://github.com/RRZE-PP/rrze-icon-set

Access and sharing during the research process



Icons: RRZE Erlangen (CC-BY-SA) https://github.com/RRZE-PP/rrze-icon-set

Image: pixabay (CC-0)

cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Licenses

Intellectual Property Rights (IPR)

"Everyone has the right to the protection of the moral and material interests resulting from any scientific, literary or artistic production of which s/he is the author."

(United Nations Universal Declaration of Human Rights, Article 27)

What are Intellectual Property Rights?

- **IPR cover scientific work, as long as it**
  - is an intellectual creation
  - has an individual character
- **IPR do not cover**
  - (research) ideas
  - facts, e.g. single information in database

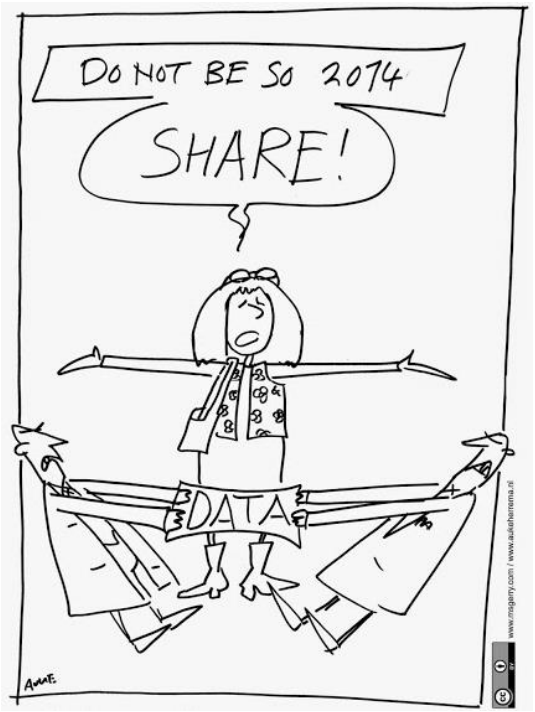


Image: Herrema, A. (2014): FOSTER Cartoon: Don't be so 2014 - Share. FOSTER-Project (EU-Funding 612425). Available at: https://www.fosteropenscience.eu/foster-taxonomy/research-data-management. (CC-by).

No Universal Intellectual Property Rights!

- **IPR are not universal but territorial rights**
  - ⇒ it matters
    - – where they are applied
    - – not where they originate
  - ⇒ vary from country to country
- **Basic regulations on IPR in EU by EU Directive 2001/29/EC**

Image: www.pixabay.com (CC-0).

No Intellectual Property Rights on Data?

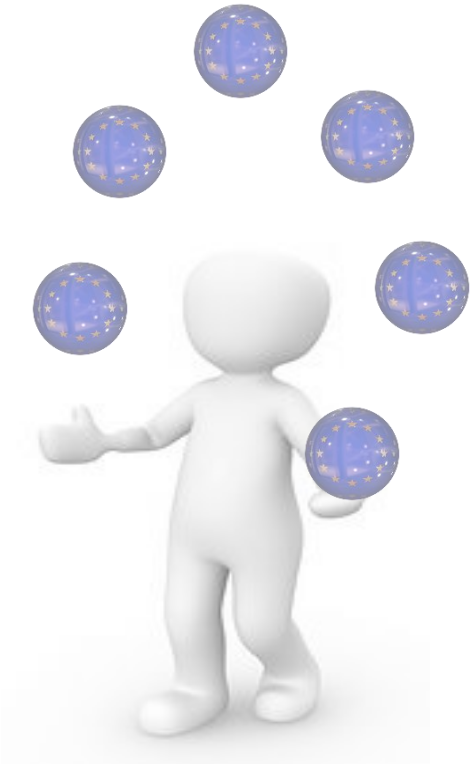
IPR do not cover facts!

i.e. single piece of information

⇒ no IPR on data?

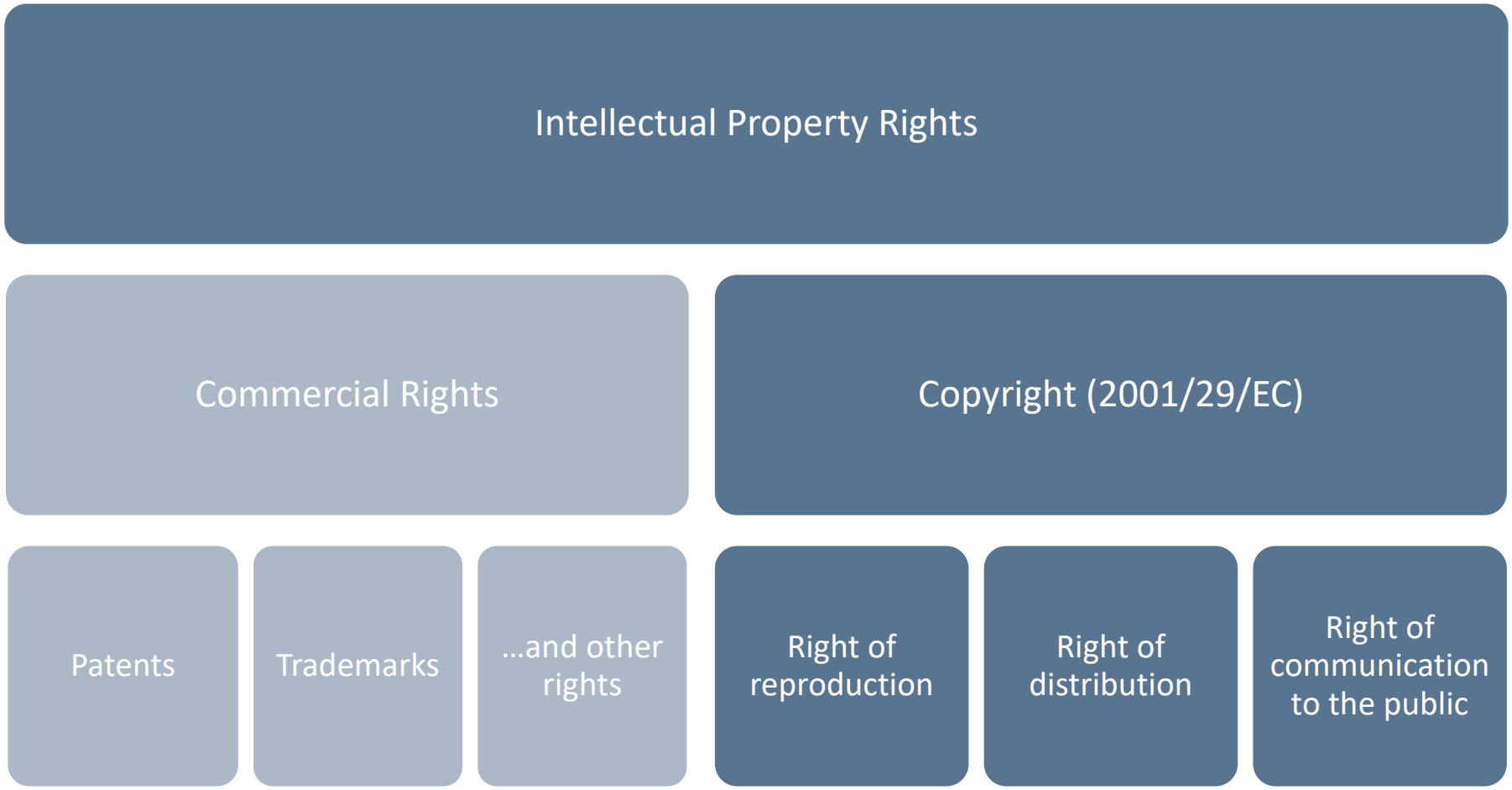
⇒ database = structured collection of information

⇒ IPR covers structure of collection but not the single facts included (EU Directive 2001/29/EC)



Images: www.pixabay.com (CC-0).

Intellectual Property Rights and Copyright



Copyright and the Right of Usage

- **Copyright**
  - assigns ownership
  - protects against unauthorized copies, derivation etc.
- **Right of usage**
  - permission to re-use the work of someone else

    ⇒ conditions of re-use (license)

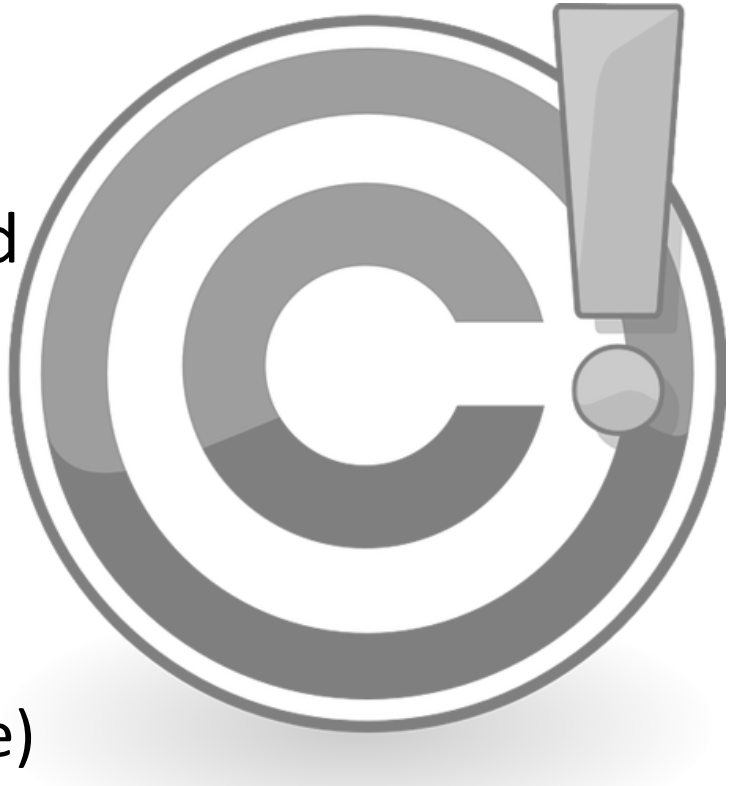


Image: www.pixabay.com (CC-0).

Re-Using Others' Work

- **Acknowledge the author, when re-using the work of someone else**
- **Many data are available with specific licenses, defining conditions of re-use**
  - clarify restrictions, i.e. what can and cannot be done with data
  - IPR can affect your ability to publish research outputs

Image: Herrema, A. (2014): FOSTER Cartoon: Open Data. FOSTER-Project (EU-Funding 612425). Available at: https://www.fosteropenscience.eu/foster-taxonomy/research-data-management. (CC-by).

Licensing

Licensing = give permission
to someone else
to do something

Licenses define

- conditions of access
- conditions of (re-)use

Image: www.pixabay.com (CC-0).

To License Scientific Work

Clarify ownership on

- data
  - might be you as the data creator
  - might be your employer or funder
    - ⇒ check your contract
    - ⇒ clarify with funders, research partners etc.
- methods, standards etc. employed

⇒ never assume that you are the owner, i.e. have IPR

How to get a License?

- **Your institution or funder may have a template license for research data**
- **Take care, (some) licenses**
  - are irrevocable
  - not suitable for personal data
- **A license example:** Creative Commons

Image: www.pixabay.com (CC-0).

Creative Commons (CC)

- **CC is a nonprofit organization**
  - provides a standardized way to give the public permission to use a work
- **CC licenses**
  - require attribution
  - can be a tailored mix of conditions for re-use

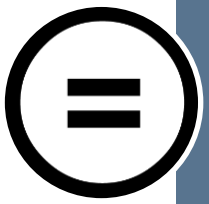
http://creativecommons.org/

Image: www.pixabay.com (CC-0).

Creative Commons Licenses

Attribution (by): copy, distribute, display, perform and derivate work as long as the author or licensor is cited in the manner specified

Non-commercial (nc): copy, distribute, display, perform and derivate work as long as the work is used only for non-commercial purposes

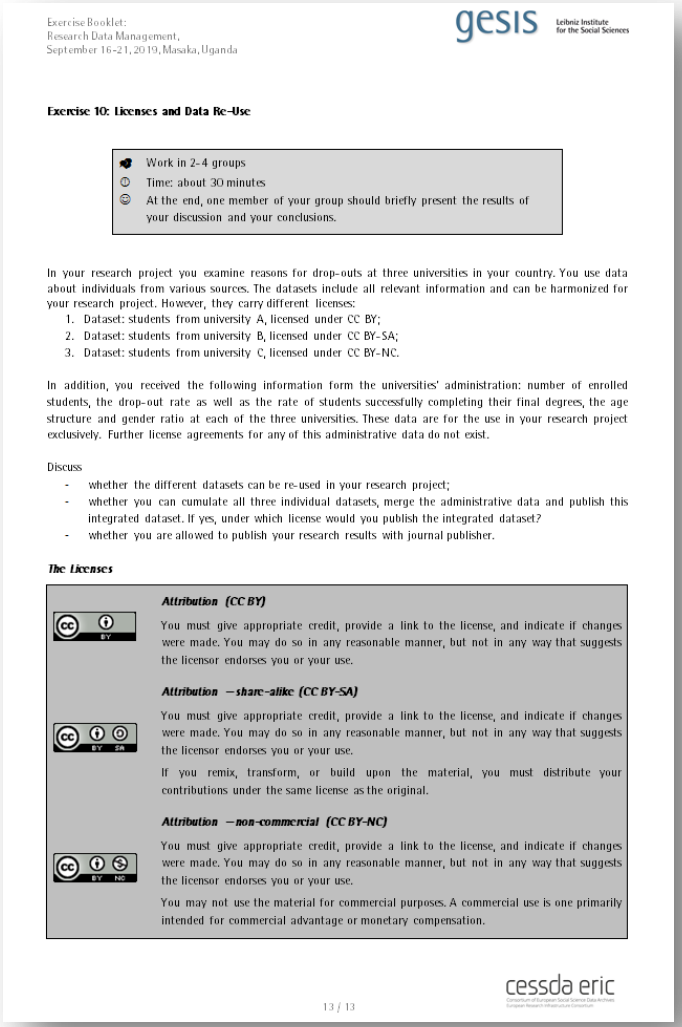


No Derivative Works (nd): copy, distribute, display, perform but no derivative works based on it

Share-alike (sa): copy, distribute, display, perform and derivate work as long as the derivative works is licensed identically to the license that governs the original work

Exercise: Licenses and Data Re-Use

- work in 5-6 groups
- time: about 30 minutes
- afterwards, we will commonly discuss your suggestions
- see Exercise-Booklet for details on Exercise 10



Exercise Booklet:
Research Data Management,
September 16-21, 2019, Masaka, Uganda

Exercise 10: Licenses and Data Re-Use

- Work in 2-4 groups
- Time: about 30 minutes
- At the end, one member of your group should briefly present the results of your discussion and your conclusions.

In your research project you examine reasons for drop-outs at three universities in your country. You use data about individuals from various sources. The datasets include all relevant information and can be harmonized for your research project. However, they carry different licenses:

1. Dataset: students from university A, licensed under CC BY;
2. Dataset: students from university B, licensed under CC BY-SA;
3. Dataset: students from university C, licensed under CC BY-NC.

In addition, you received the following information form the universities' administration: number of enrolled students, the drop-out rate as well as the rate of students successfully completing their final degrees, the age structure and gender ratio at each of the three universities. These data are for the use in your research project exclusively. Further license agreements for any of this administrative data do not exist.

Discuss

- whether the different datasets can be re-used in your research project;
- whether you can cumulate all three individual datasets, merge the administrative data and publish this integrated dataset. If yes, under which license would you publish the integrated dataset?
- whether you are allowed to publish your research results with journal publisher.

The Licenses

Attribution (CC BY)

You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Attribution – share-alike (CC BY-SA)

You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Attribution – non-commercial (CC BY-NC)

You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

You may not use the material for commercial purposes. A commercial use is one primarily intended for commercial advantage or monetary compensation.

13 / 13

Image: pixabay (CC-0)

A Final Note on Licensing

- **Licensing does usually not seek or assert ownership of data**

  ⇒ ownership remains with the original copyright holder

- **Archives**
  - manage the re-use via a license agreement
  - only hold data where the copyright holder has given permission for archiving and reuse

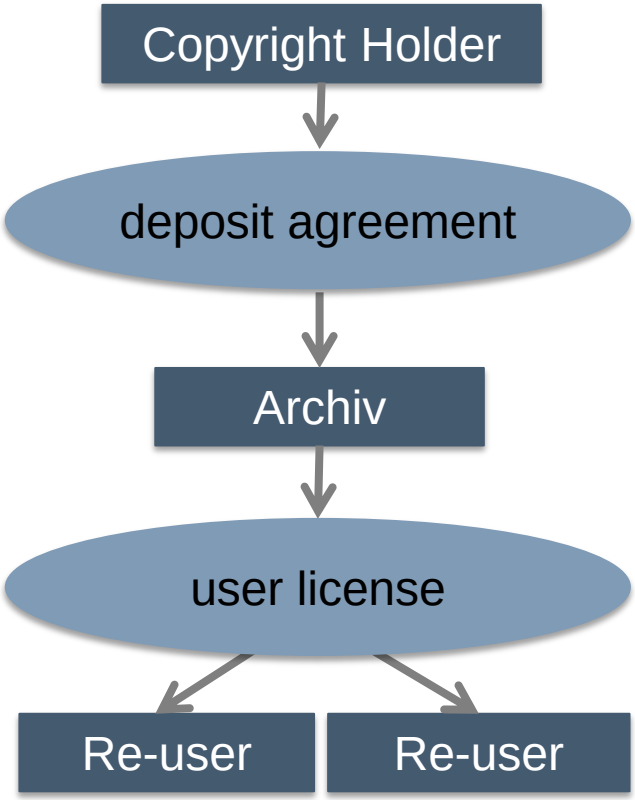


Image: pixabay (CC-0)

Adapt your DMP

Develop a back-up strategy

⇒ define clear and consistent guidelines

- what? all, something, only changed files
- where? at least in triplicates and different locations
- how long are different files (and versions) needed? (never destruct or overwrite original data while needed)
- how often are files backed-up?
- who: name researcher(s) and assign responsibilities

⇒ verify back-ups frequently (e.g. once a week), e.g. restoring the files (name researcher(s) and responsibilities)

Adapt your DMP

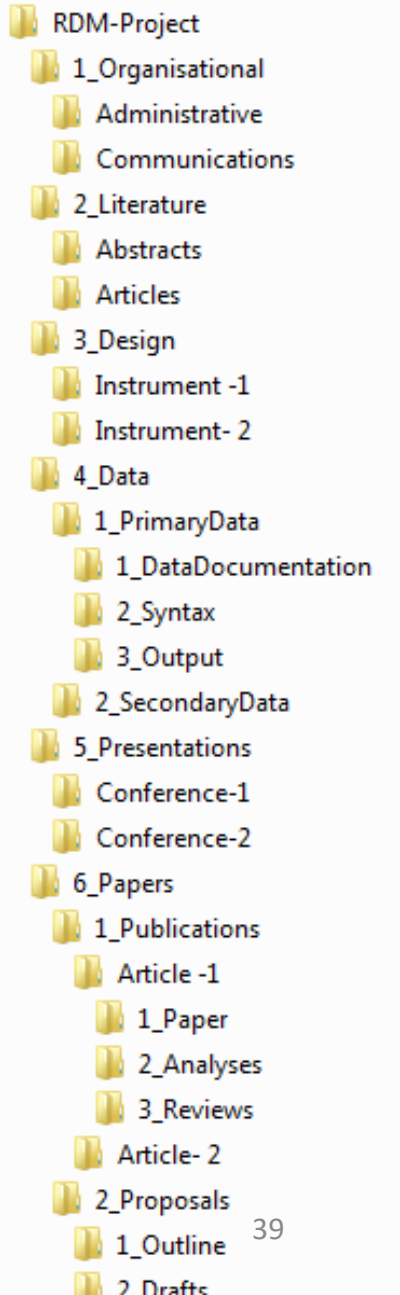
Develop guidelines to organize…

… folders

- define a consistent structure of folders
  ⇒ e.g. by topic

… and files

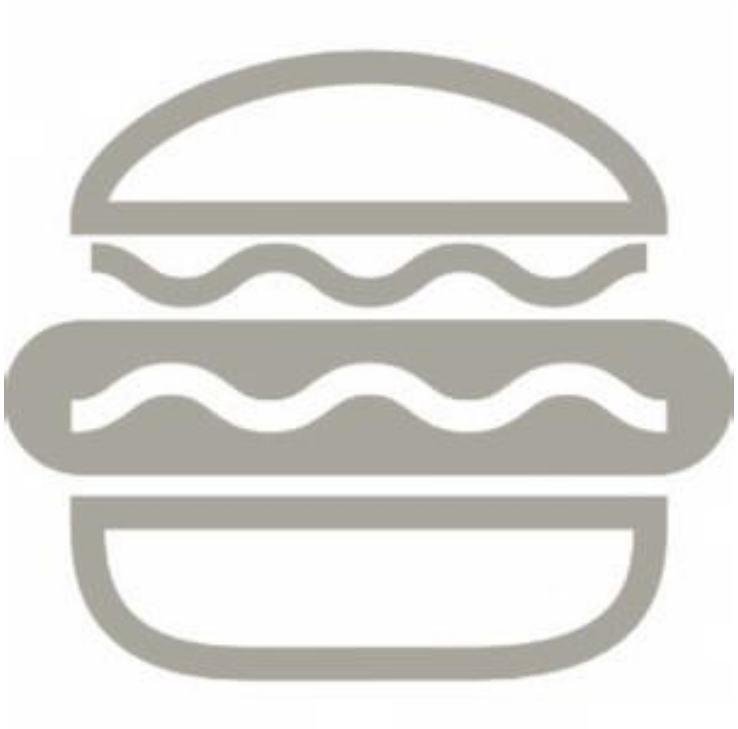
- to name files
  ⇒ e.g. [type_name_version]
- to version files
  ⇒ e.g. by the date and editor's acronyms
  data_RDMData_20150822sn



Adapt your DMP

- **Check consent form on data sharing, first**
- **Creative Commons (CC) Licenses:**
  - pros:
    - easy-to-use
    - standardized way to give the public permission
    - enable sharing and use of (creative) works
  - cons:
    - no controlled access and (non-commercial) use
    - do not cover “copylefts”
    - attribution stacking

⇒ CC-Licenses are not designed for (sensitive) data



cessda eric
Consortium of European Social Science Data Archives
European Research Infrastructure Consortium

Archiving

WHAT CAN DATA ARCHIVES DO FOR YOU?

We keep data safe!

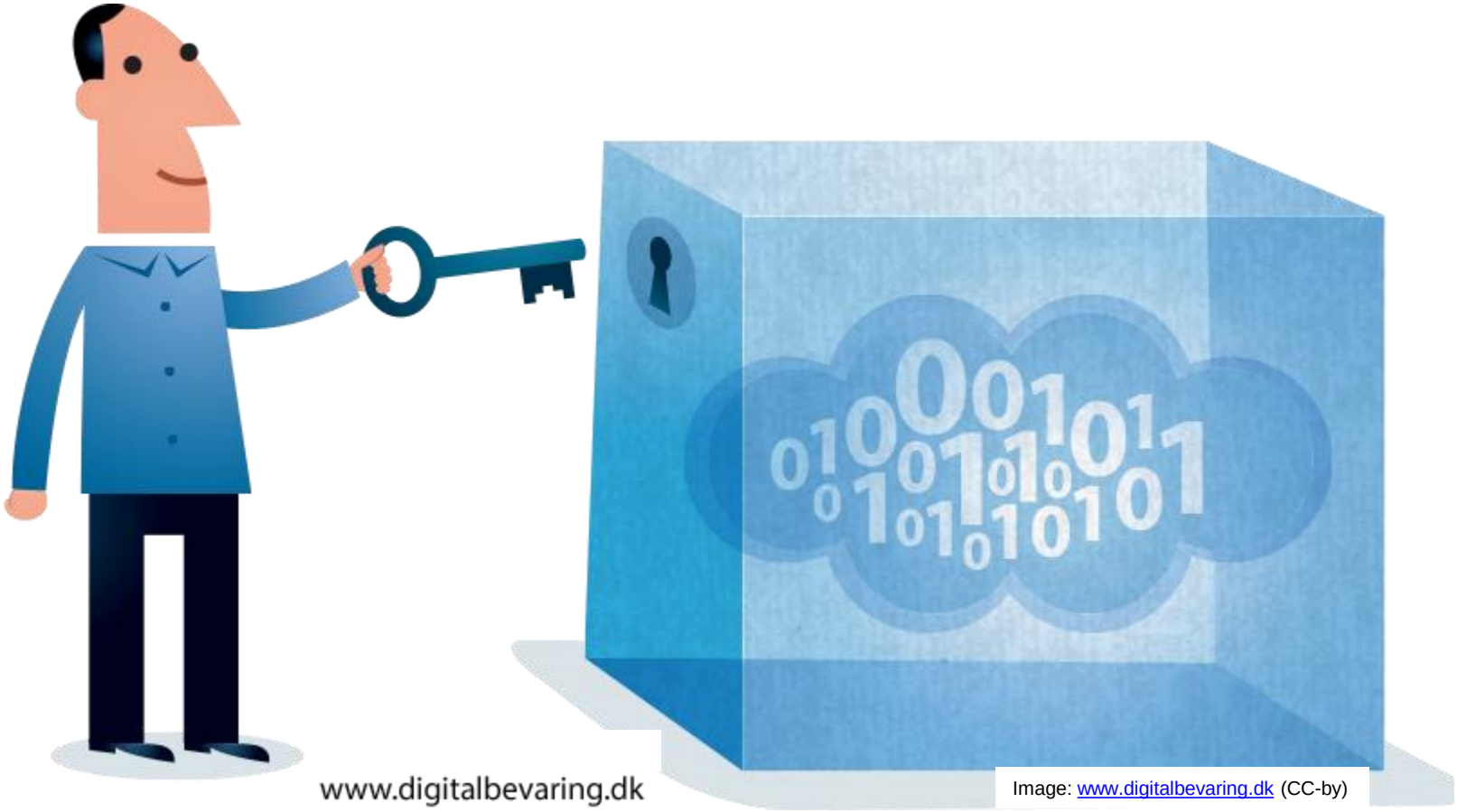


Image: www.digitalbevaring.dk (CC-by)

We keep data safe!

- Back-up and secure storage
- Long-term preservation
- Legal and ethical security
  - for data producers (IPR, licensing terms)
  - for participants in research (data protection)

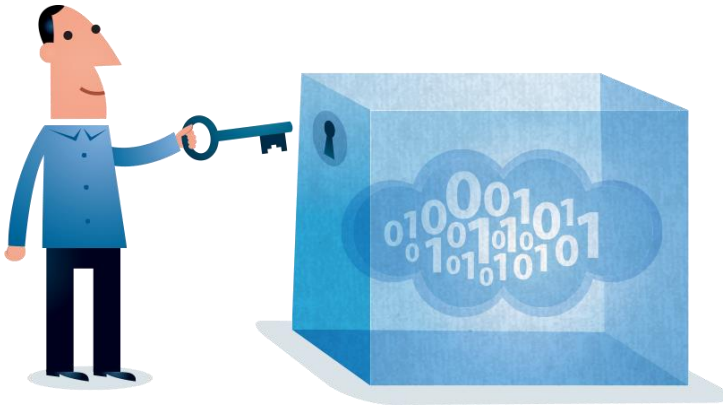
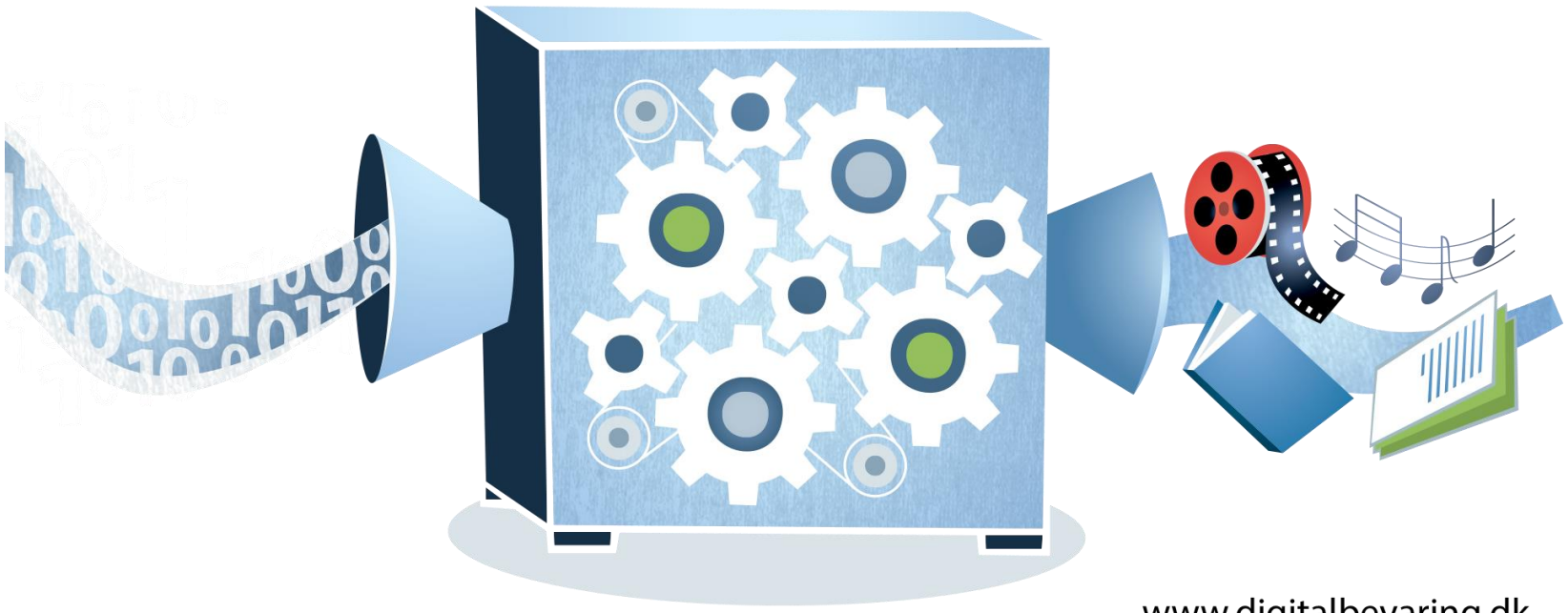


Image: www.digitalbevaring.dk (cc-by)

We make data accessible!



www.digitalbevaring.dk

Image: www.digitalbevaring.dk (CC-by)

We make data accessible!

- Accessible formats
- Increased visibility in the world wide web (e.g. through machine-readable metadata)
- Persistent identification
- High-quality documentation

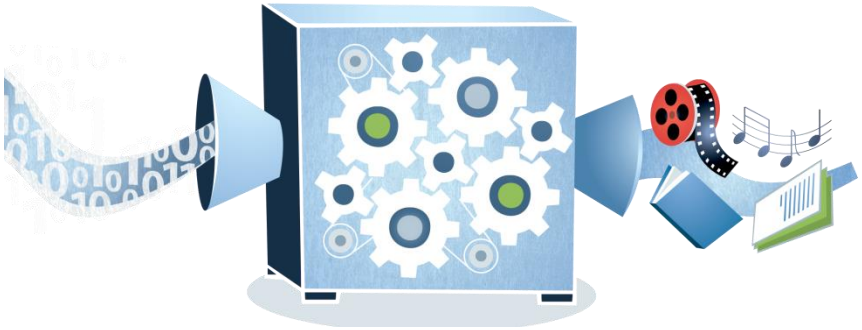


Image: www.digitalbevaring.dk (CC-by)

We offer support!

Image: www.digitalbevaring.dk (CC-by)

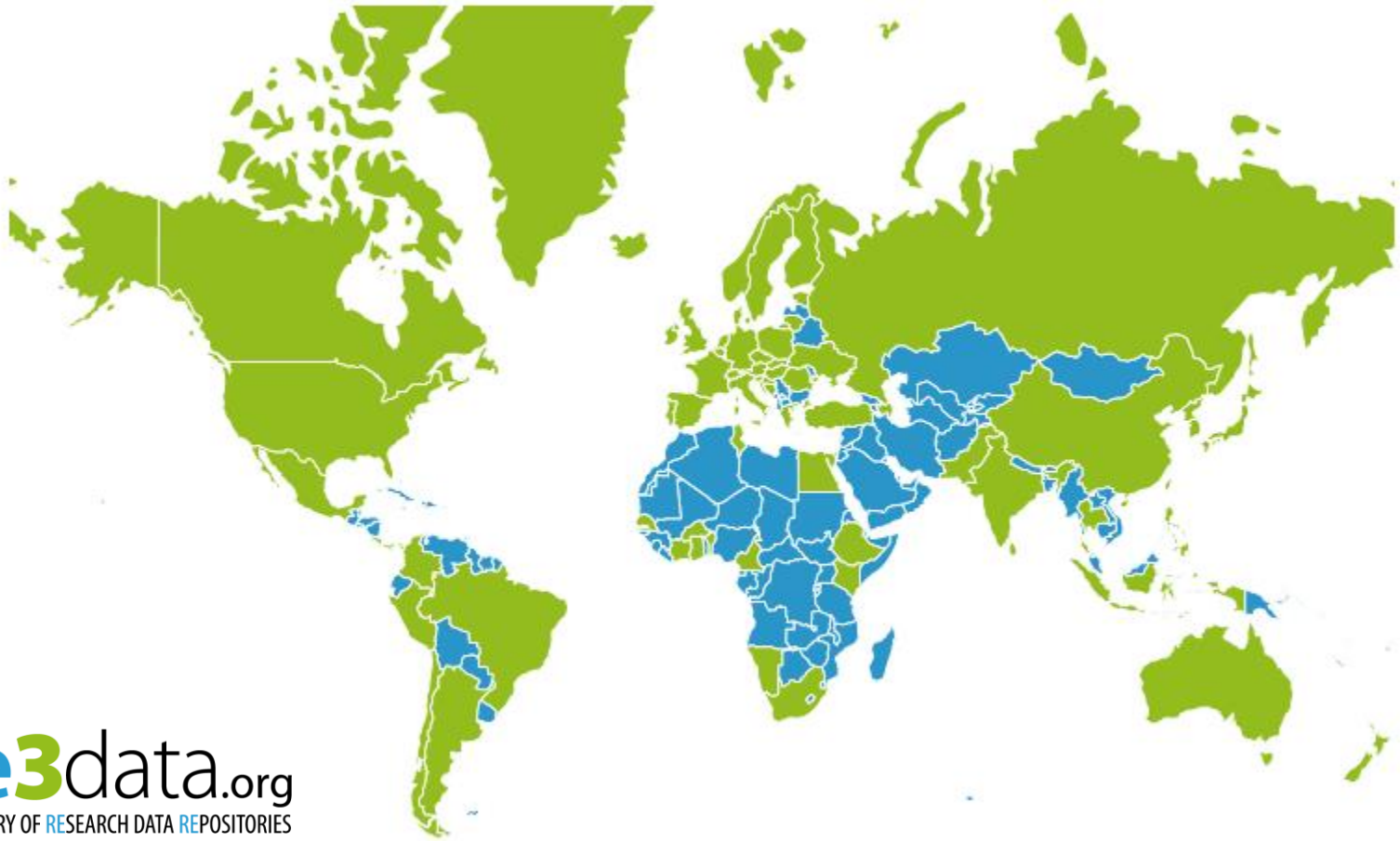
We offer support!

- We advise on
  - documentation
  - metadata
  - anonymization and data protection

Image: www.digitalbevaring.dk (CC-by)

DATA ARCHIVING IN AFRICA

Data archives worldwide



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

https://www.re3data.org/browse/by-country/

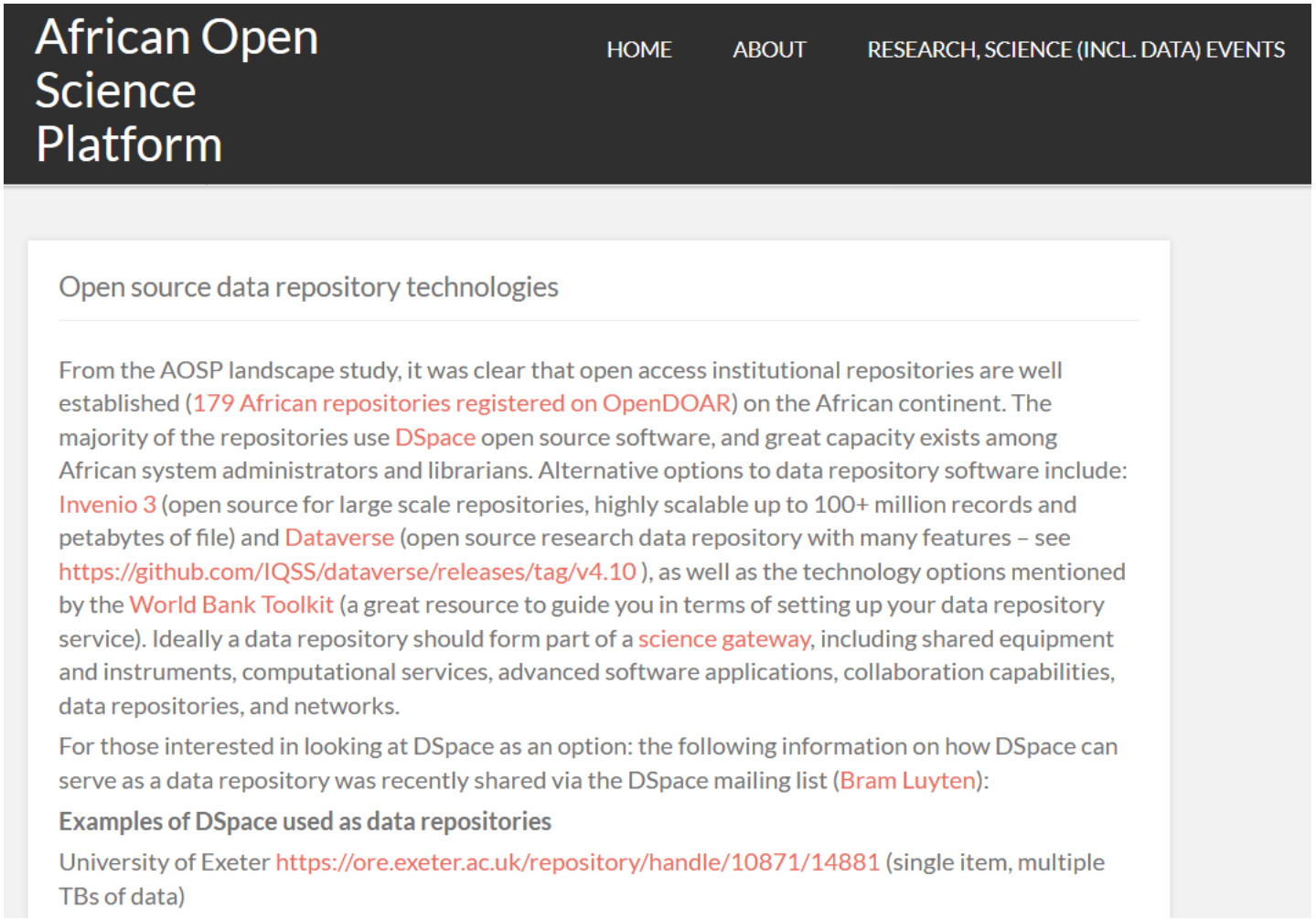
Data scattered globally after Ebola outbreak in West Africa

Only a selection of international responders is shown

But things are changing…

- Pan-African initiative
- System to find, deposit, manage, share and reuse data
- Pilot project is finished, operational launch of platform latest in 2020

http://africanopenscience.org.za/



http://africanopenscience.org.za/?p=1129

B			
BSUSpace	Type: Institutional	Software: DSpace	Country: Uganda
Busitema University Open Access Digital Repository	Type: Institutional	Software: DSpace	Country: Uganda
I			
IUIU Institutional Repository	Type: Institutional	Software: DSpace	Country: Uganda
K			
KIUIR	Type: Institutional	Software: DSpace	Country: Uganda
M			
Makerere University Business School Institutional Repository	Type: Institutional	Software: DSpace	Country: Uganda

https://v2.sherpa.ac.uk/view/repository_by_country/ug.html

GO FAIR
@GOFAIRofficial

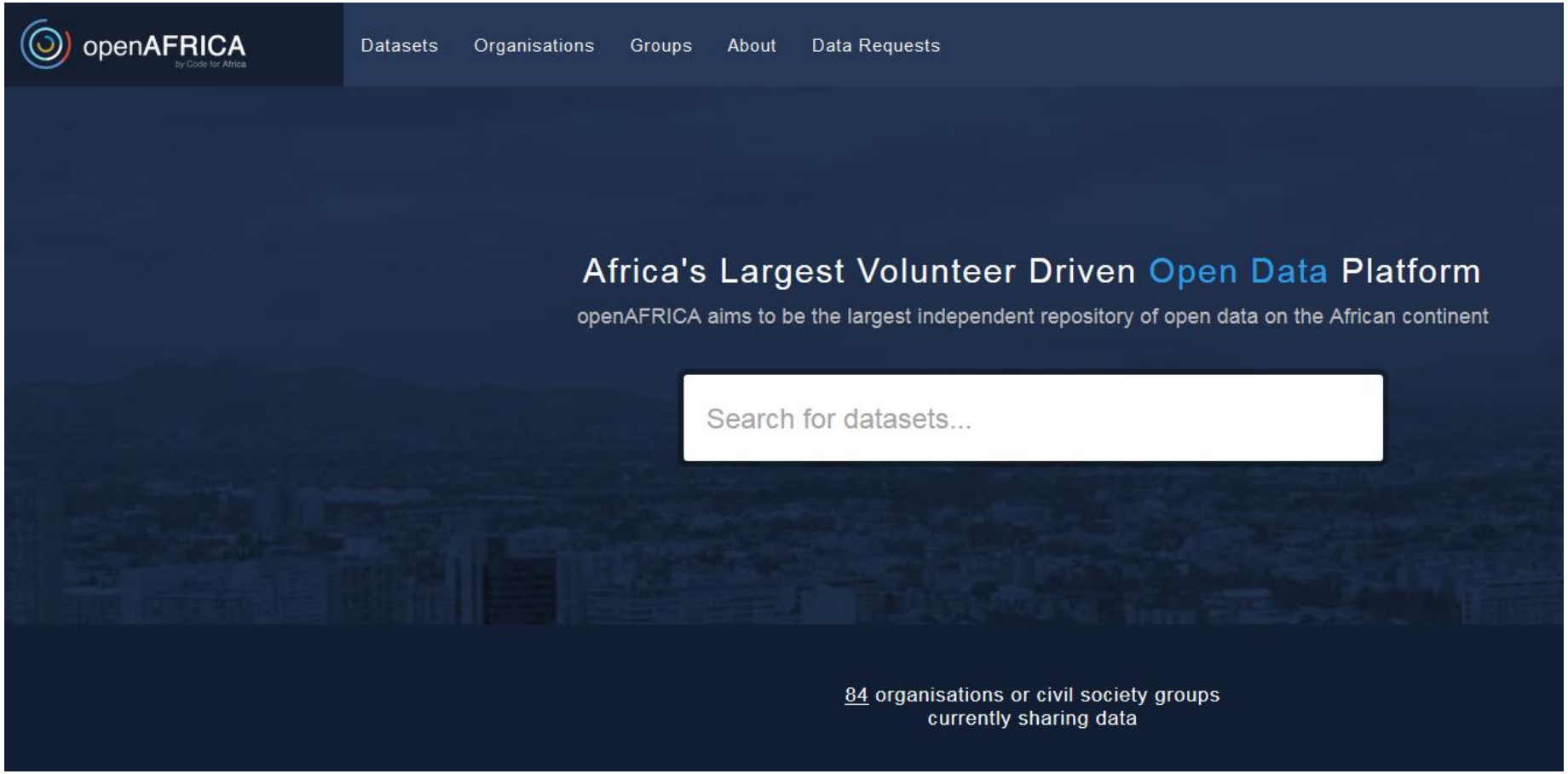
Yesterday, the IN-Africa was launched w/ participants from Kenya, Uganda, Ghana, Nigeria, Zimbabwe, Ethiopia, Senegal, Ivory Coast, South Africa, Tanzania, St Vincent & the Grenadines, Surinam, Indonesia, Kazakstan, UK, Netherlands. Learn more ➡ bit.ly /2L1F3lU #FAIRdata

Tweet übersetzen

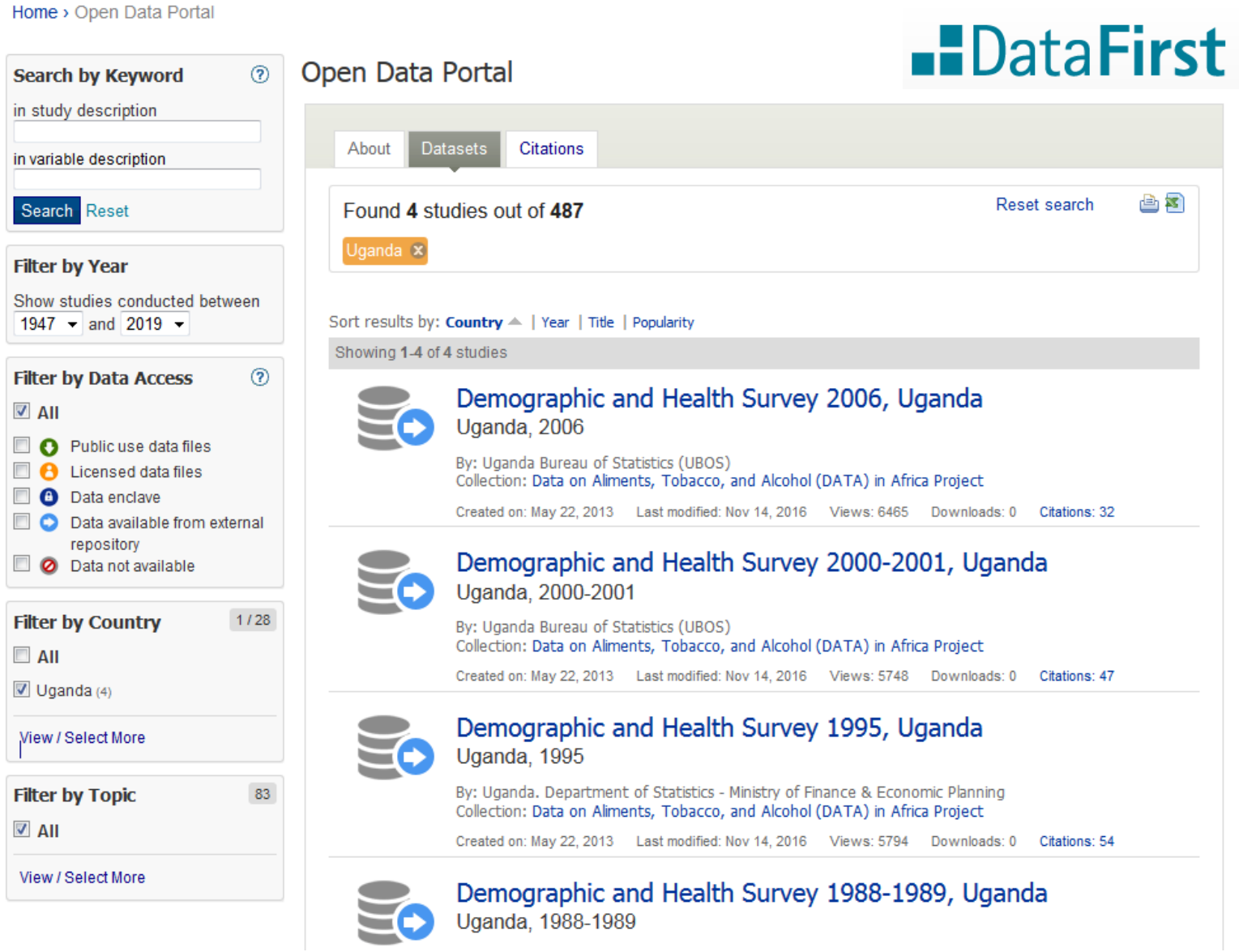
9:28 vorm. · 23. Aug. 2019 · TweetDeck

http://bit.ly/2L1F3lU

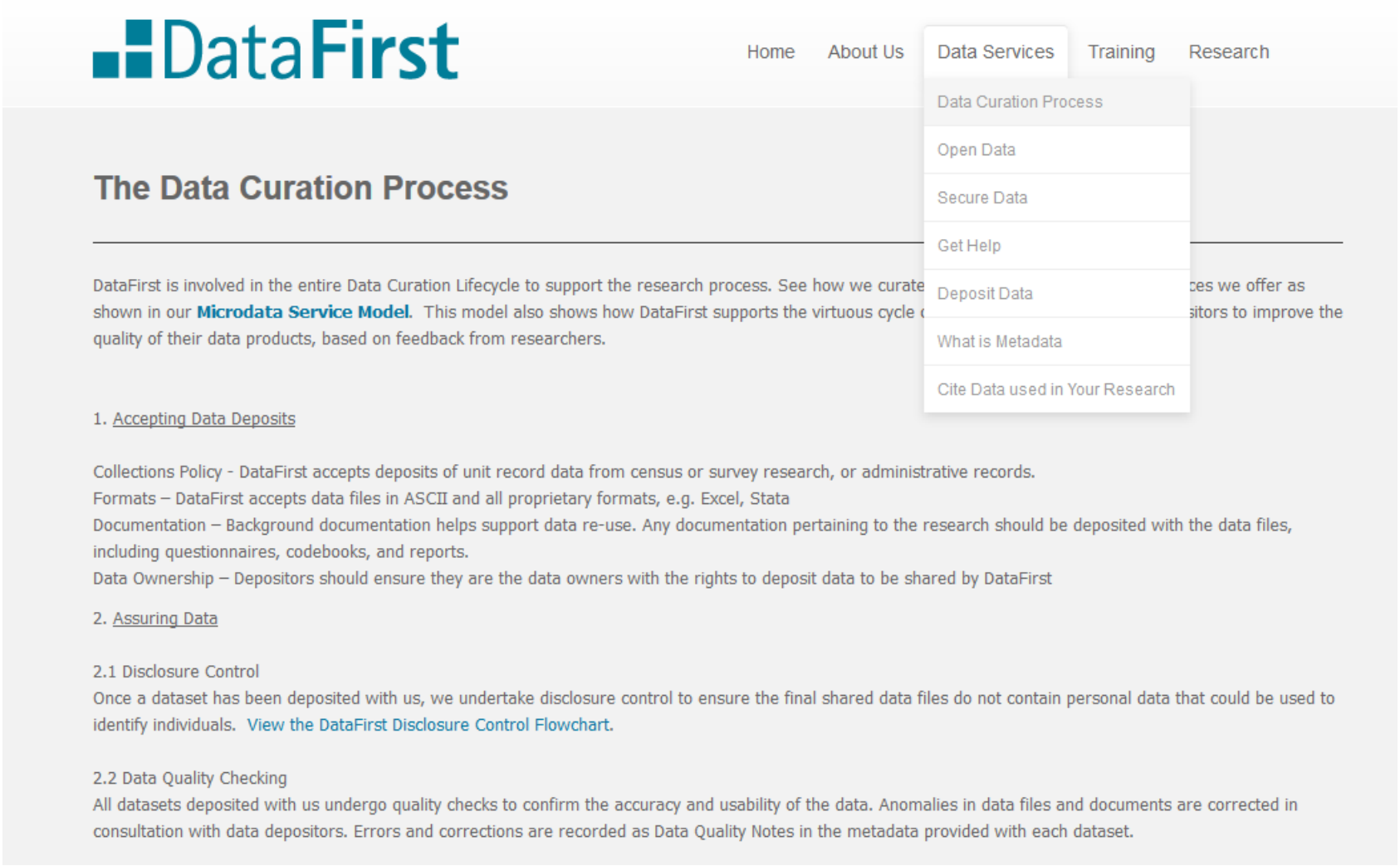
And there are alternatives…



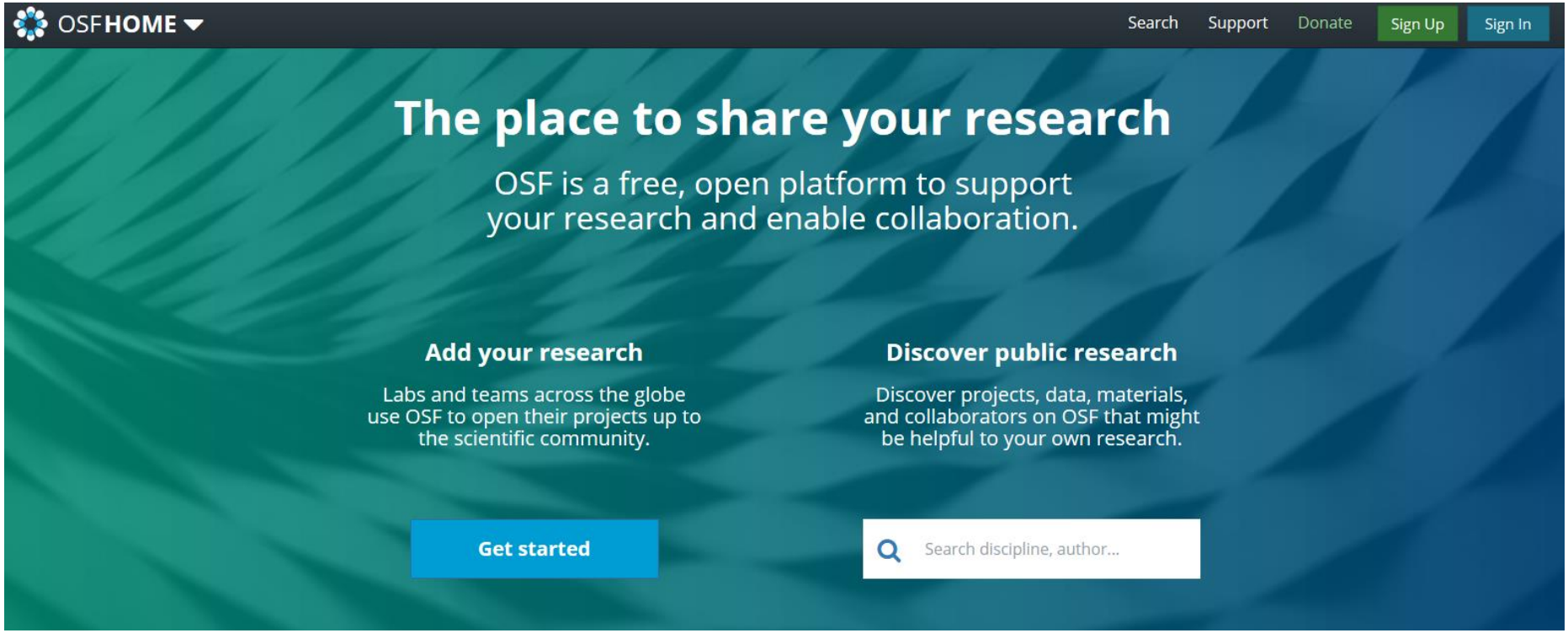
https://africaopendata.org/



https://www.datafirst.uct.ac.za/



https://www.datafirst.uct.ac.za/


OSF HOME
Search
Support
Donate
Sign Up
Sign In
The place to share your research
OSF is a free, open platform to support your research and enable collaboration.
Add your research
Labs and teams across the globe use OSF to open their projects up to the scientific community.
Get started
Discover public research
Discover projects, data, materials, and collaborators on OSF that might be helpful to your own research.
Search discipline, author...

https://osf.io/

Recent uploads

August 8, 2019 (vv3) Dataset Open Access

View

Analysis of correlation-based biomolecular networks from different omics data by fitting stochastic block models

Baum, Katharina; Rajapakse, Jagath C.; Azuaje, Francisco

Baum_et_al_2019_Supplementary_Figures.pdf: Supplementary Figures S1-S4. Legends are included under each figure. sbm-for-correlation-based-networks-master.zip: Archived source code of R and Python functions for the analyses and example workflow description at time of publication. Files are...

Uploaded on August 8, 2019

2 more version(s) exist for this record

August 6, 2019 (v3.1) Software Open Access

View

OpenScienceMOOC/Module-5-Open-Research-Software-and-Open-Source: 3.1

Jon Tennant; Julien Colomb; Lisa Matthias; Simon Worthington; Florian Kohrt; irrubio; Tania Allard; Philipp Zumstein; Daniel S. Katz; Alexander Morley; Tobias Steiner; Stephan Druskat; Zoran Pandovski; Arfon Smith; Gabriele Orlandi; Rutger Vos; José Raúl Canay Pazos; Paul Griffiths; Nithiya Streethran; Hollie Marshall; Luke W Johnston; Luis Camacho; Konrad Förstner; Heidi Seibold; EricDWilkey; Encarnación Martínez Álvarez; Brendan Palmer; Alessandro Sarretta; Alberto Marocchino; Abigail Cabunoc Mayes

Release now that the content is live on the open source platform, Eliademy. Please also see YouTube and Soundcloud for additional content.

Uploaded on August 6, 2019

3 more version(s) exist for this record

https://zenodo.org/

The Dataverse® Project

Open source research data repository software

Researchers

Enjoy full control over your data. Receive *web visibility, academic credit,* and *increased citation counts*. A personal dataverse is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. Want to set up your personal dataverse?

https://dataverse.org/

Conclusion: Sharing and Preservation

- Start thinking about this early
- Contact potential archives at the beginning of your research or
- Look for alternatives early if archives are not available
- Pay specific attention to: informed consent, legal restrictions, and where data is stored when you use free/generic services

SELF-ARCHIVING AT GESIS

gesis datorium

Features of datorium

- Data sharing even for smaller research projects
- Data producers describe their data based on metadata standards
- Easy upload of research data free of charge in a secure environment
- Licenses
- Researchers determine the conditions of data (re)use

Access categories

- Free Access (without Registration)
- Free Access (with Registration)
- Restricted Access: Users have to apply for permission to download the data by contacting the depositor
- Embargo: Data files can be published with an embargo date of one year maximum. The metadata though will be published in any case

In a nutshell

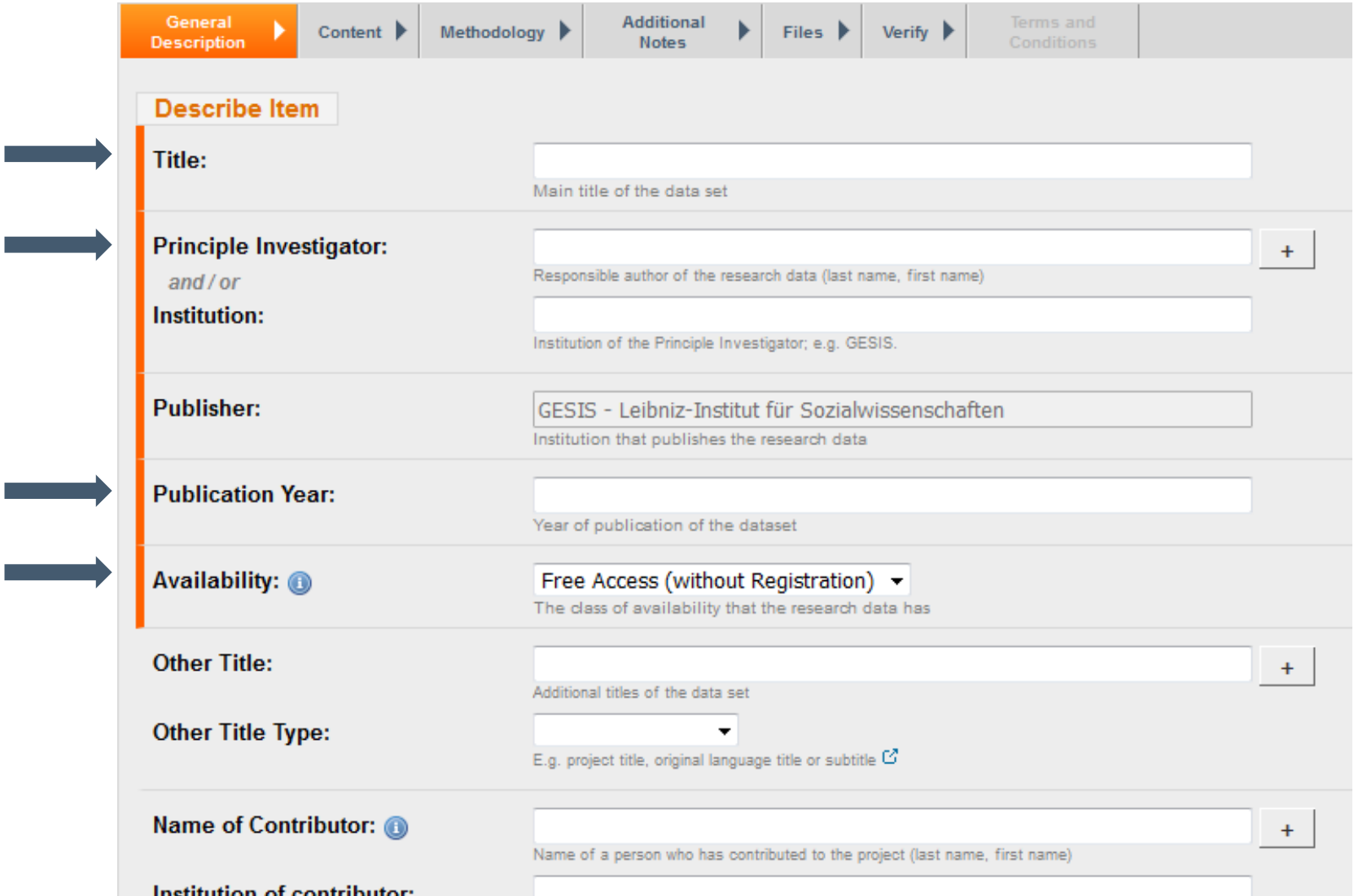
Documentation
&
Upload

Storage
&
Review

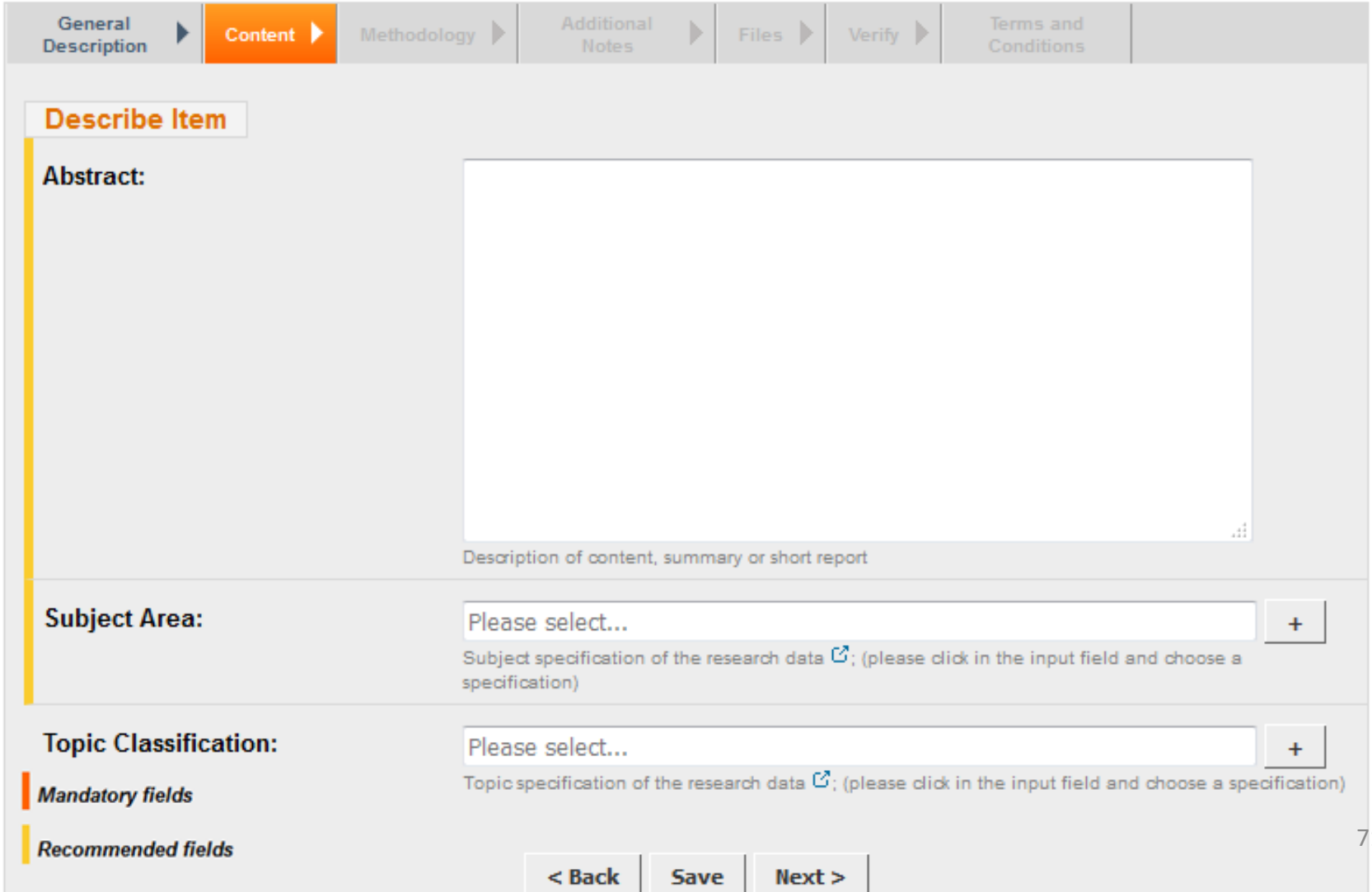
Publication

Search
&
Download

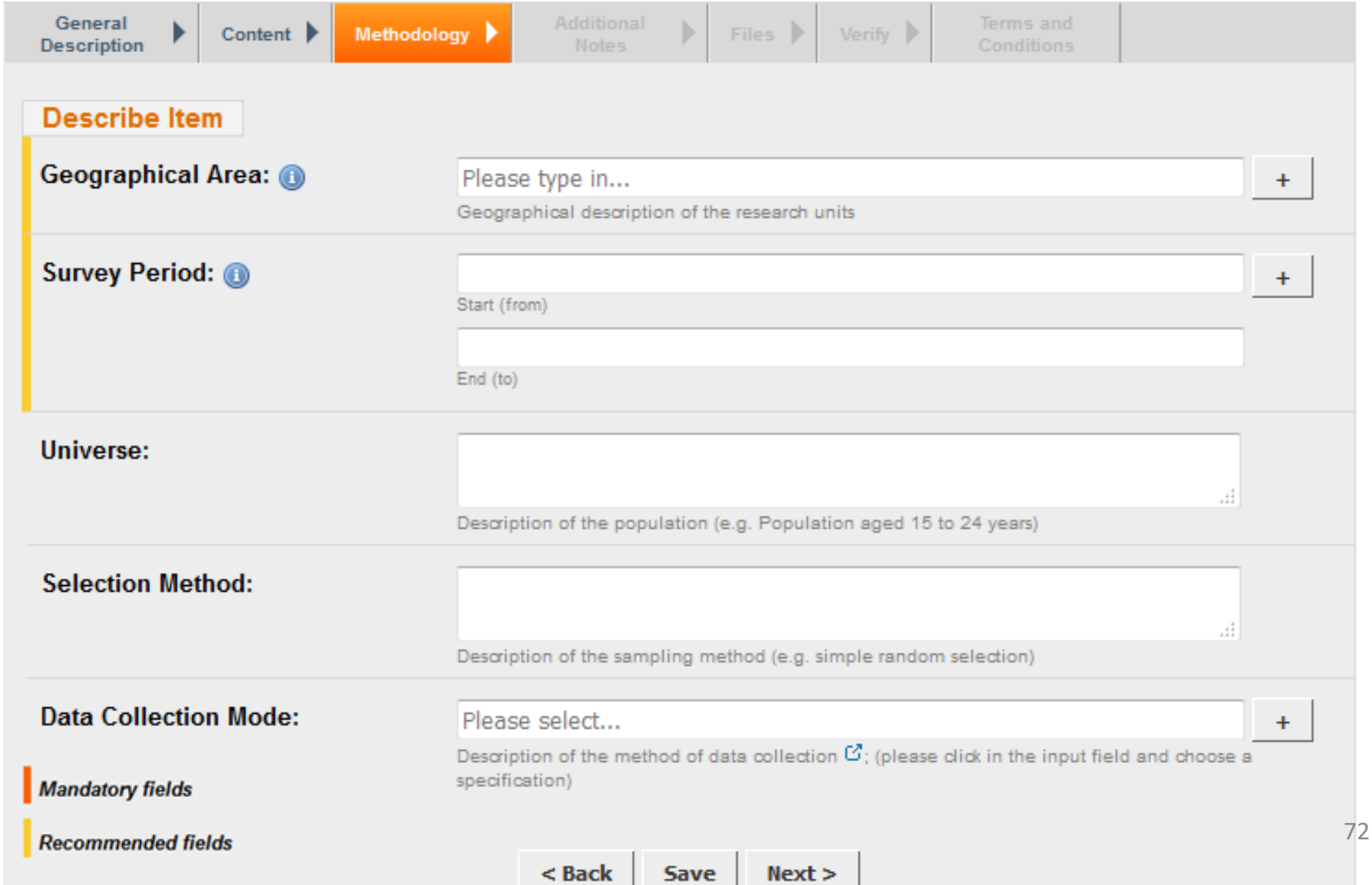
Only 4 mandatory fields.



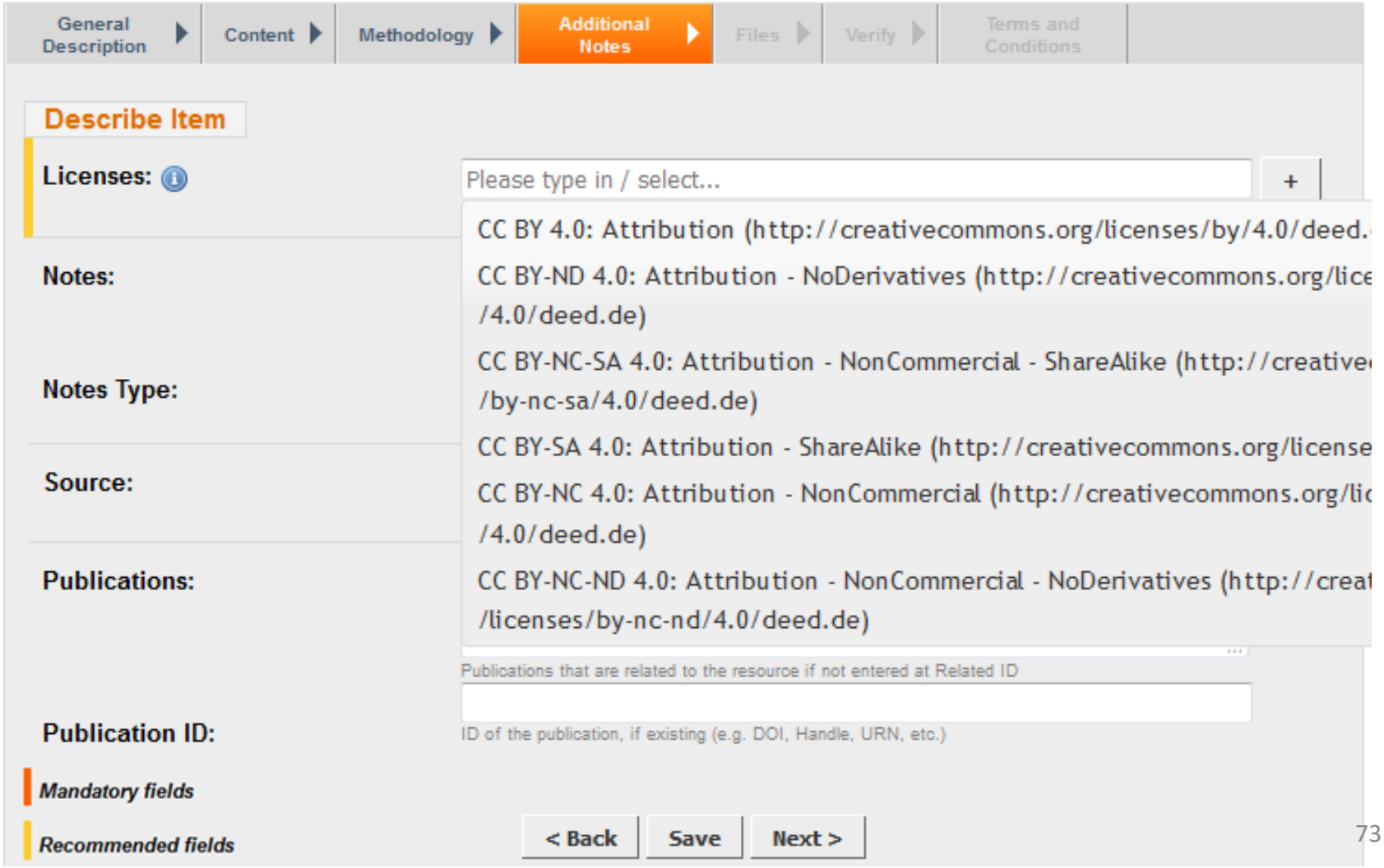
However, it makes sense for…



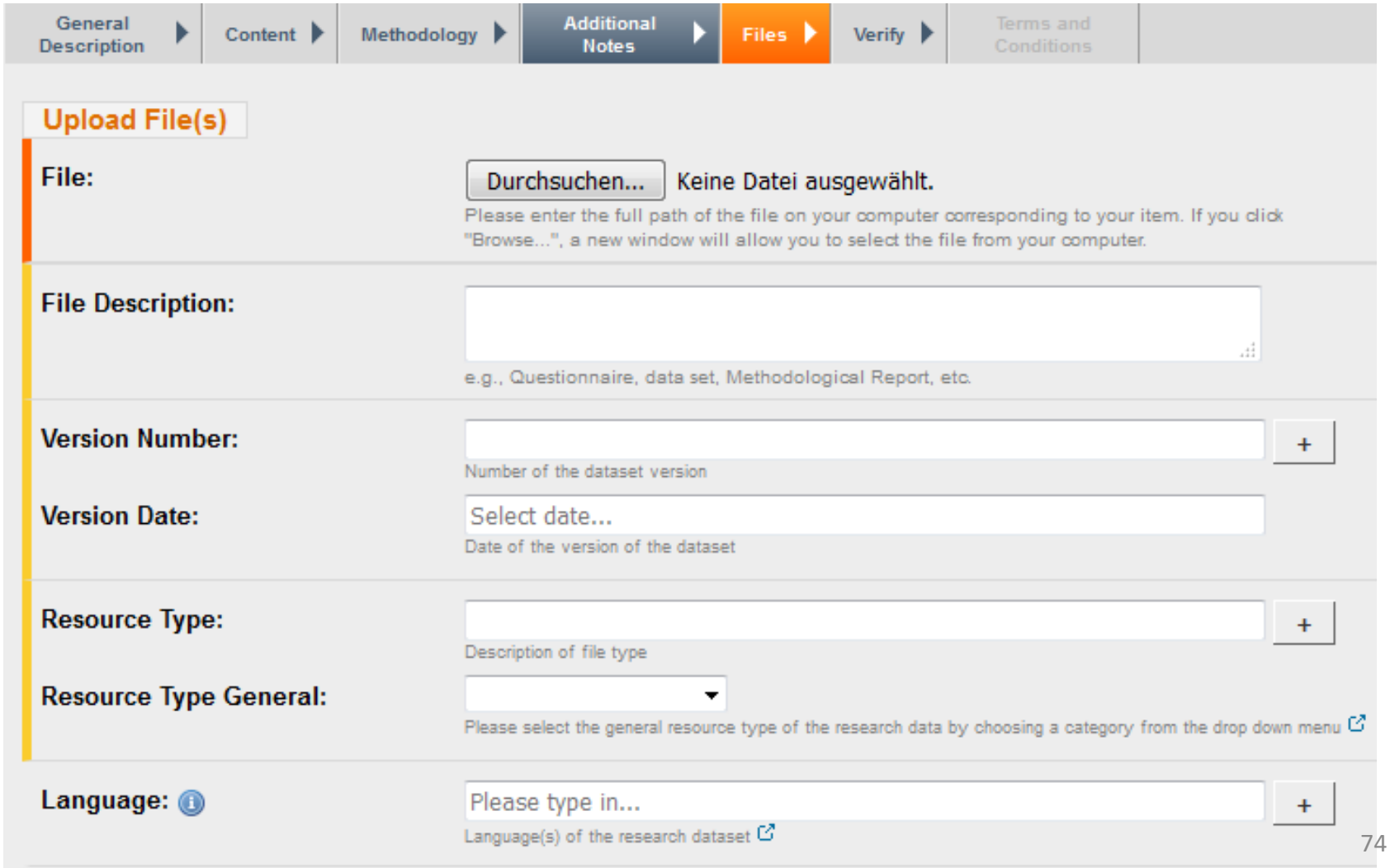
…a better understandig to…



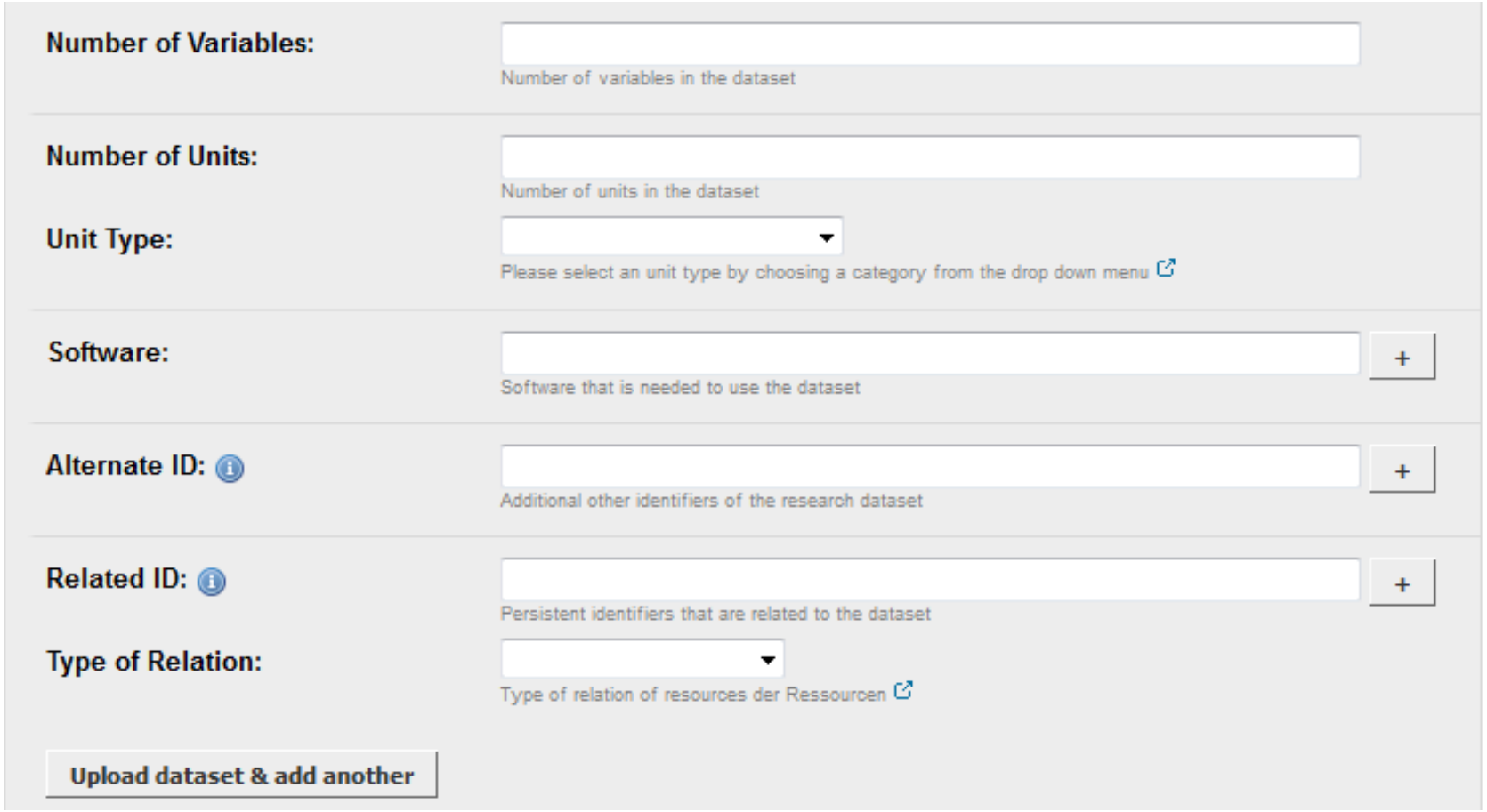
…add some more metadata.



It is also possible to describe…



...single files.



Search

Filter

- Publication Year
- Availability
- Subject Area
- Topic Classification
- Geographical Area
- Data Collection Method
- Replication Server

Mitglied der Leibniz-Gemeinschaft

Search

Go

Sort options

Results

Now showing items 1-10 of 53

1 2 3 4 ... 6 >

Daten: Feldexperimentelle Daten zum Einfluss der Wahl alter Sprachen auf die beruflichen Einstiegschancen

Primary Researcher / Institution:
Sawert, Tim; Universität Potsdam, Lehrstuhl Methoden der Empirischen Sozialforschung
Publication Year:
2016
Availability:
Eingeschränkter Zugang
Journal:
Zeitschrift für Soziologie

An Experimental Investigation into 'Pledge and Review' in Climate Negotiations

Primary Researcher / Institution:
Dannenberg, Astrid; University of Kassel, Germany
Publication Year:
2016
Availability:
Embargo (maximal 1 Jahr)

Survey on High-level Search Activities based on the Stratagem Level in Digital Libraries - Survey Data

Primary Researcher / Institution:
Carevic, Zeljko; GESIS - Leibniz-Institut für Sozialwissenschaften
Publication Year:

Record

General Description

Back

cite this data

Title	An Experimental Investigation into 'Pledge and Review' in Climate Negotiations
URI	http://dx.doi.org/10.7802/1282
Primary Researcher	Dannenberg, Astrid; University of Kassel, Germany
Publication Year	2016
Availability	Embargo (maximal 1 year)
Embargo (until)	2017-06-06
Contributor	Dannenberg, Astrid; University of Kassel; Researcher European Union (EU) Horizon 2020 program, action ERC-2014-STG, Project HUCO, grant number 636746; European Research Council; Funder Princeton Institute for International and Regional Studies research community on Communicating Uncertainty: Science, Institutions, and Ethics in the Politics of Global Climate Change; Princeton University; Sponsor Barrett,Scott; Columbia University School of International and Public Affairs & Earth Institute; Researcher Haita-Falah, Corina; University of Kassel; Kontaktperson (Contact Person)
Subject Area	Economics
Topic Classification	Natural Environment, Nature
Abstract	A novelty of the new Paris Agreement is the inclusion of a process for assessment and review of countries' nationally determined pledges and contributions. The intent is to reveal whether similar countries are making comparable pledges, whether the totality of such pledges will achieve the global goal, and whether, over the

Record

Files	Size	Format	Download Option
EBharmonized_QoL_data.dta	56.06Mb	application/x-stata	Please enter the purpose of your usage ▾ Download

File checksum: MD5:3cb9274766c25a59f098d13a5cb4ff10

Description :

File Description:
- Harmonized Eurobarometer Dataset on "Quality of Life" issues

Resource Type
- Survey data; Datensatz

Language:
- eng - (English)

Number of Variables:
- 484

Number of Units / Unit Type:
- 110612; Individuum

EBharmonized_QoL_data.sav	56.16Mb	application/x-spss-sav	Please enter the purpose of your usage ▾ Download

File checksum: MD5:f8dcb9aef9b6e9a5cfcbb2628c7335d3

Description :

File Description:
- Harmonized Eurobarometer Dataset on "Quality of Life" issues

Resource Type
- Survey data; Datensatz

Language:
- eng - (English)

Number of Variables:

Review process

Technical review:

check whether all delivered material is complete, correct and in a suitable technical condition (e.g. readable, virus free, etc)

Content review:

concerning plausibility, consistency, data weighting and data privacy

Final review

Benefits of datorium

- **Easy upload of research data free of charge in a secure environment**
- **All datasets receive a persistent identifier (DOI)**
- **Data accessible via**
  - datorium webpage
  - da|ra
  - DataCite
  - DBK
- **Well-balanced mixture of mandatory and optional metadata fields in its submission form**

Image: pixabay (CC-0)

Adequate Research Data Management

basic quality assurance	replicability	reusability
aims and goals of research data managment		
processing data for use within the research project	processing data for long-term storage and replication	processing data for sharing
minimal documentation, e.g. sampling, variable and codes	metadata to describe the entire research process	detailed documentation for reuse
data protection: informed consent for use of data within the project	data protection: informed consent on data storing and data anonymization	data protection: informed consent on data sharing and data anonymization
copyright on data and methods for use within the project	copyright on data and methods for long-term storage and replication	copyright on data and methods for sharing

Thank you for your attention!

anja.perry@gesis.org
oliver.watteler@gesis.org