Data Management (Plan)

MoSiG M2
SMEE

Céline Coutrix — October 24 2024

DEFINITION OF RESEARCH DATA

What is research data

?

What is research data

- **Facts**
- **Observations**
- **Images**
- **Computer program results**
- **Recordings**
- **Measurements or experiences**

on which is based an

- **Argument**
- **Theory**
- **Test**
- **Hypothesis**
- **Other research output**

- Numerical
- Descriptive
- Visual
- Tactile

- Raw
- Cleaned
- Processed

Any format or media

« What Is Research Data », 2017. https://www.ands.org.au/__data/assets/pdf_file/0006/731823/Whatis-research-data.pdf

What is research data

In a data management plan, we do not include :

- Preliminary analyses and draft of scientific papers
- Programs of future work
- Peer reviews
- Personal communications with colleagues
- Physical objects
- Scientific publications
- Training materials
- Administrative data

➔ Some of them must be kept permanently

see: Référentiel de gestion des archives de la recherche (FR)

PERSONAL DATA

?

PERSONAL DATA

- **Directly identify the person**
  E.g., last name, first name, address, picture, voice recording, etc.
- ***In*directly identify the person**
  E.g., telephone number, cross-referencing information such as the son of the research director, the latter living in Grenoble, etc.

≠ **Irreversibly anonymized data**
No longer allow for the re-identification of a person

SENSITIVE PERSONAL DATA

?

SENSITIVE PERSONAL DATA

- Presumed race or ethnicity
- Political opinions
- Philosophical or religious beliefs
- Trade union membership
- Sexual orientation
- Health-related data
- Biometric data that can identify a person
- Genetic data

(+ Special status for *social security number* and data about *law violation and conviction*)

SENSITIVE PERSONAL DATA

Processing sensitive personal data is forbidden

Except, for example,

- After consent of the person
- Data made manifestly public by the person
- Important public interest
- Safeguarding human life

What is a data management plan?

Data Management plan

= written document describing

1. **Research data** you expect to **acquire** or **generate** ***during*** the course of a research project
2. The mechanisms you will use ***during*** the course of the project
   - To **manage** the data
   - To **describe** the data
   - To **analyze** the data
   - To **store** the data
3. The mechanisms you will use ***at the end*** of the project
   - To **share** the data
   - To **preserve** the data

MOTIVATION
FOR A DATA MANAGEMENT PLAN

Data Management

Aims

FAIR data = **Findable**, **Accessible**, **Interoperable** and **Reusable** data

Does not necessarily mean opening up all your research data

➔ Follow the principle "***as open as possible, as closed as necessary***”

Data Management

Aims

Conduct **time-efficient research**

Allow your research to be **reproduced or reused** (by others or yourself)

Help **find and understand** if data collected by others can be reused for your research purpose

Conduct research that is **safe** for the people participating

Comply with **funder** mandates

Data Management

Aims

Conduct **time-efficient research**

Allow your research to be **reproduced or reused** (by others or yourself)

Help **find and understand** if data collected by others can be reused for your research purpose

Conduct research that is **safe** for the people participating

Comply with **funder** mandates

Increase your visibility and impact

Researcher's Incentives for Data Management

Time efficiency

Avoid problems that would otherwise appear later on

Researcher’s Incentives for Data Management

Impact

Correlation between sharing data and citation rate of scientific papers

by how much?

69% or 25% or 9% (depending on domain and citation prediction model)

Piwowar, Day and Frisma, “Sharing detailed research data is associated with increased citation rate”, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308
Piwowar and Vision, “Data reuse and the open data citation advantage”, http://dx.doi.org/10.7717/peerj.175
Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, et Barbara McGillivray. 2020. « The citation advantage of linking publications to research data ». PLOS ONE 15 (4): e0230416. https://doi.org/10.1371/journal.pone.0230416.

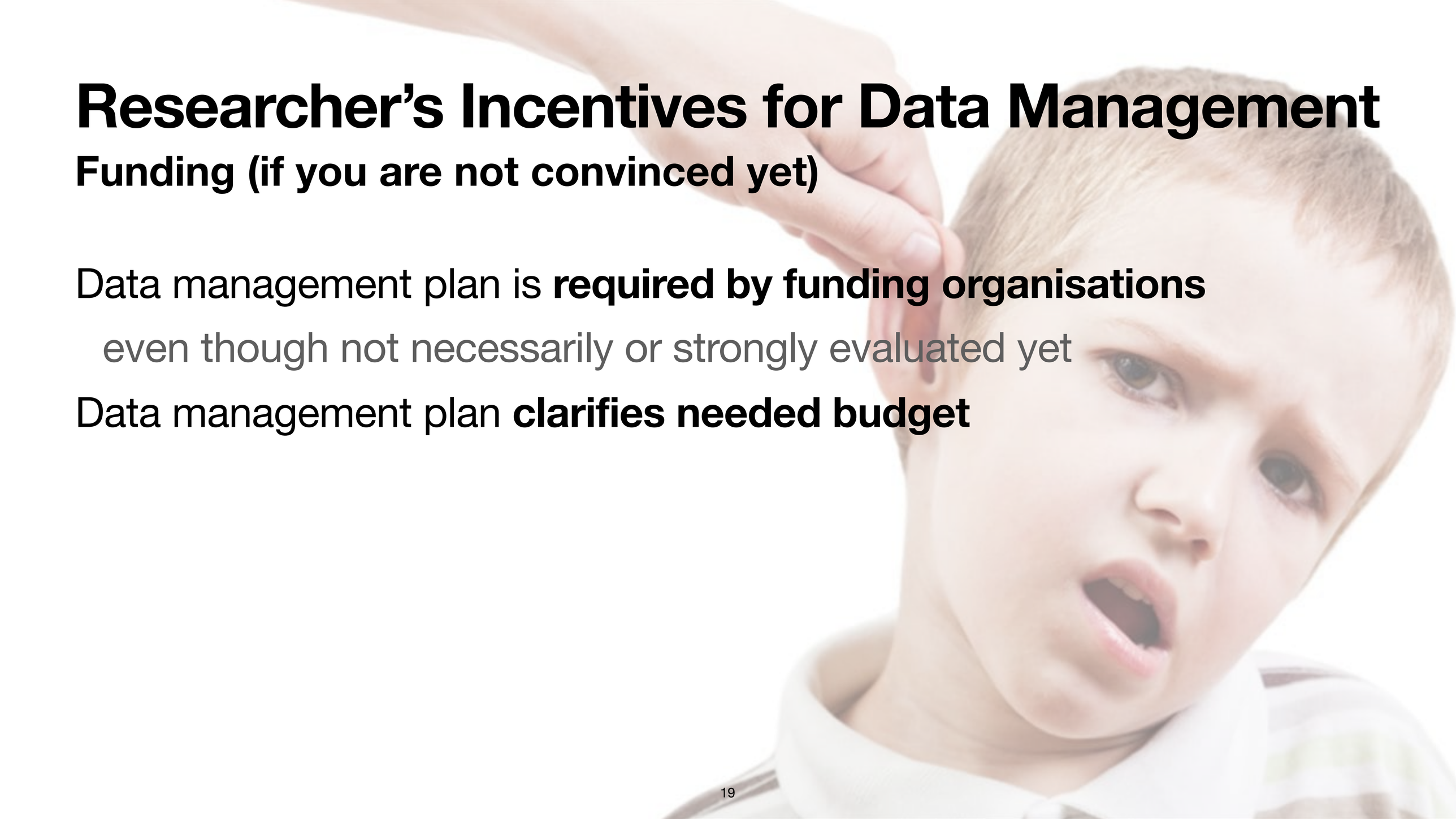
Researcher’s Incentives for Data Management

Funding (if you are not convinced yet)

Data management plan is **required by funding organisations**

even though not necessarily or strongly evaluated yet

Data management plan **clarifies needed budget**




2021
Optional
Mandatory
European funding


2019
Optional
Mandatory
French funding (ANR)

How to
handle,
organise,
document,
and **store**
your data ?

technical

organisational

legal

ethical

sustainability

aspects to take into account

Data Management

Data management should be **thought through**, **structured**, and **documented**

1. what will be necessary for **using/collecting** data

research project

2. Maintain the **integrity** of the data (e.g., avoid data loss)
Provide **access** to relevant people at the appropriate time

3. Detailed and structured **documentation** to share your data in the long-term

Data Management Plan is meant to be updated

To plan and conduct a research project,
e.g., a master internship or a PhD,
what are **relevant aspects of data management**
that you can consider from the very beginning?

in groups of 2-3
with Post-it notes
on Miro

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

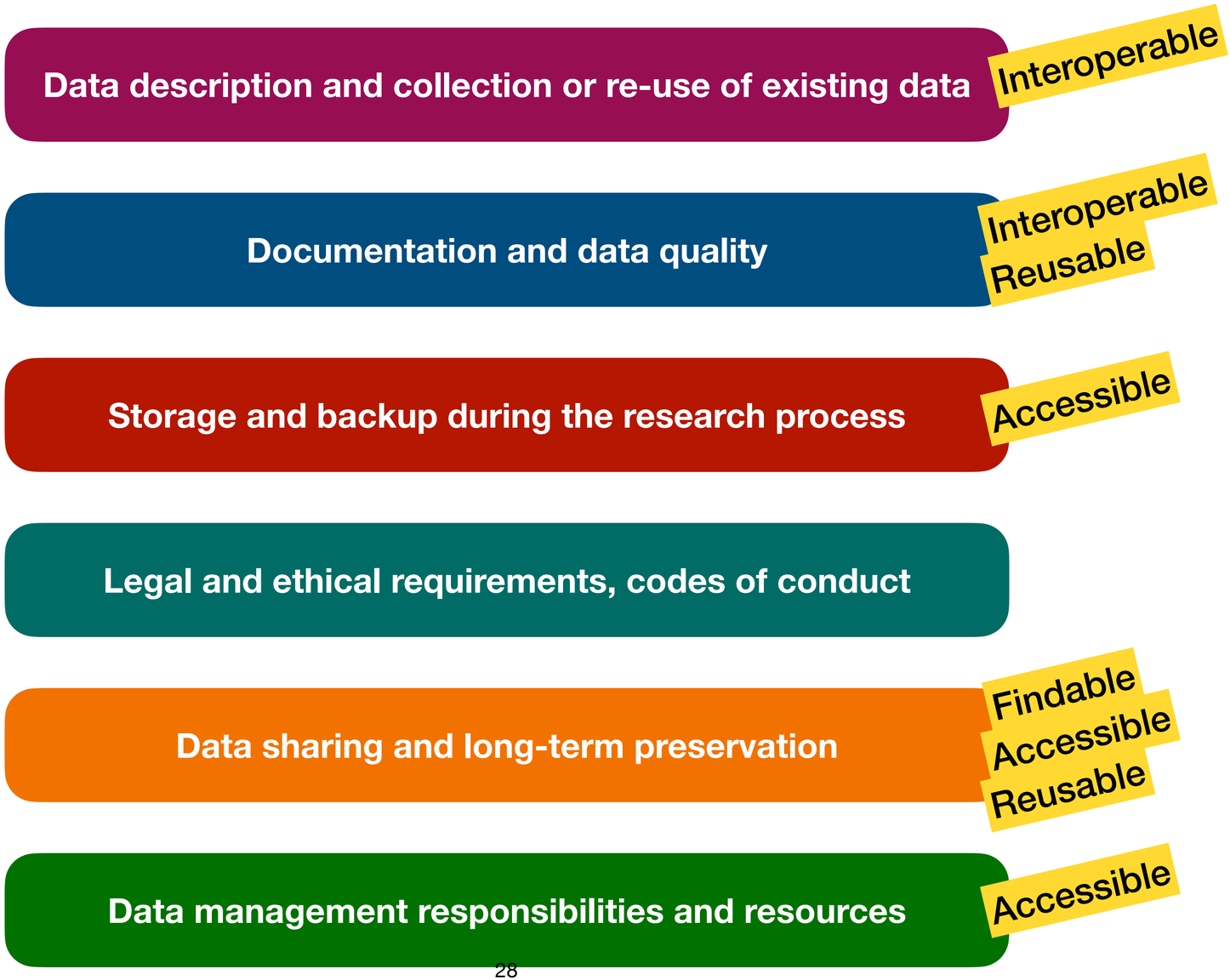
Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS



Writing a data management plan

- By hand
- Using tools like argos (EN) or DMP OPIDoR (FR)

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

Data description and collection or re-use of existing data

- Describe **data to be re-used**: your **own** or **third-party**
  e.g., Open Street Map cartographic data, UK Biobank, data shared on Zenodo
  - Describe **source**, **licenses**, **conditions of use**, **price**

    ➔ Looking for datasets: DataCite, OpenAIRE, etc.
- Describe **new data** to be collected or produced
  - Briefly state the **reasons why not re-using** any existing data
  - Explain which **methodologies** or software will be used **to collect or produce the new data**
- Explain how data provenance will be documented

Data description and collection or re-use of existing data

Detail the **kind** of data, e.g.,

- Numeric (databases, spreadsheets)
- Textual (documents)
- Image
- Audio
- Video
- Mixed media

Detail the data **format**: the way in which the data is encoded for storage, e.g., filename extension (pdf, csv, txt, etc.)

- Justify with, e.g.,
  - Staff expertise within the host organisation
  - Preference for open formats
  - Standards accepted by data repositories
  - Widespread usage within the research community
  - Software or equipment that will be used
- Prefer **open** and **standard formats**
  to facilitate sharing and long-term re-use of data
  (several repositories provide lists of such ‘preferred formats’)

Detail the **volumes**, in

- Storage space required (bytes)
- Numbers of objects, files, rows, and columns

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

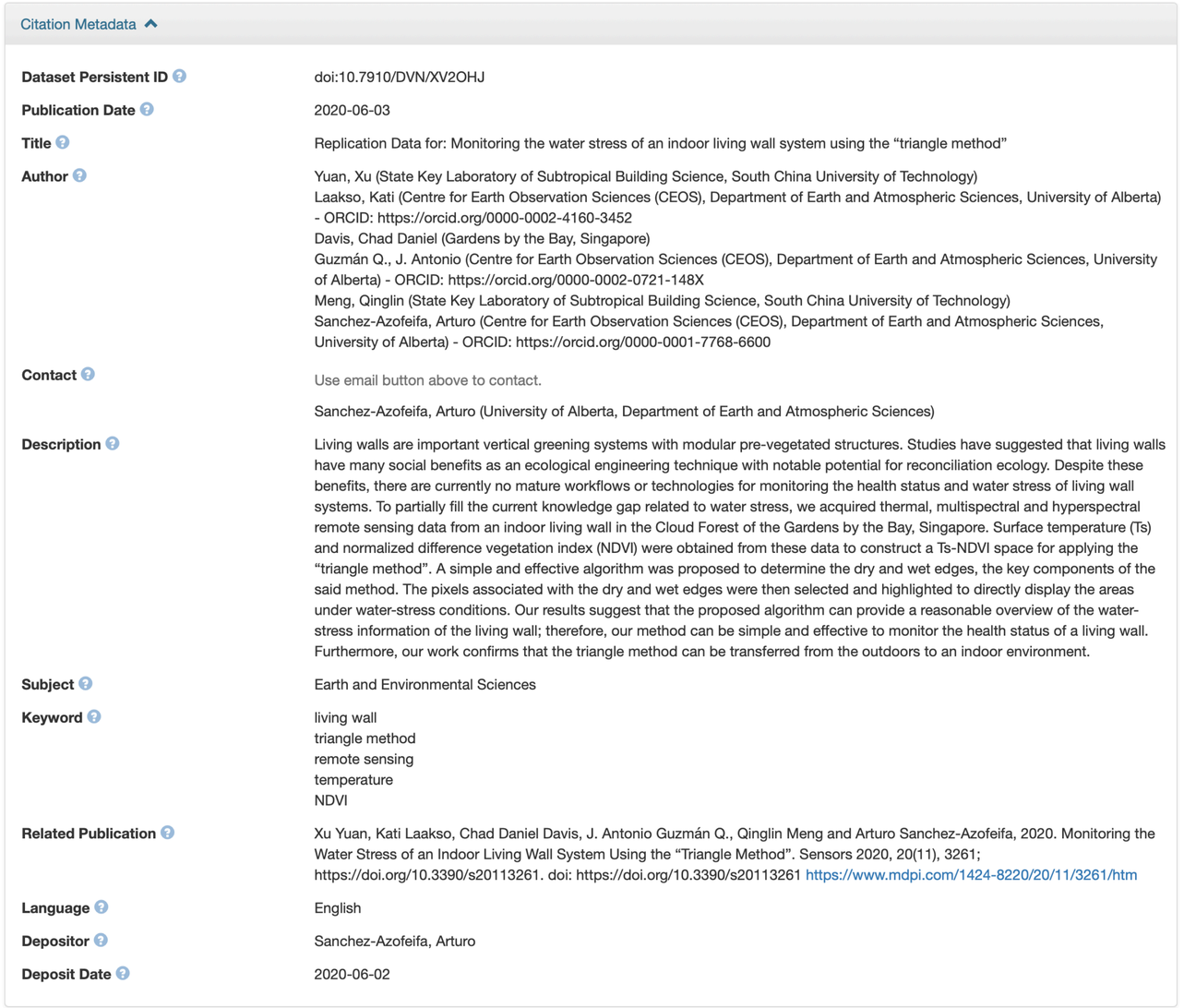
Data management responsibilities and resources

Documentation and data quality

- Describe the **metadata** accompanying your data
  → Help others identify and discover the data
- Detail metadata to be provided:
  who, *why*, *when*, etc. of your research data

→ Different kind of metadata
E.g., citation, geospatial, journal, etc.

Example: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XV2OHJ



Citation Metadata

Dataset Persistent ID	doi:10.7910/DVN/XV2OHJ
Publication Date	2020-06-03
Title	Replication Data for: Monitoring the water stress of an indoor living wall system using the “triangle method”
Author	Yuan, Xu (State Key Laboratory of Subtropical Building Science, South China University of Technology) Laakso, Kati (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0002-4160-3452 Davis, Chad Daniel (Gardens by the Bay, Singapore) Guzmán Q., J. Antonio (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0002-0721-148X Meng, Qinglin (State Key Laboratory of Subtropical Building Science, South China University of Technology) Sanchez-Azofeifa, Arturo (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0001-7768-6600
Contact	Use email button above to contact. Sanchez-Azofeifa, Arturo (University of Alberta, Department of Earth and Atmospheric Sciences)
Description	Living walls are important vertical greening systems with modular pre-vegetated structures. Studies have suggested that living walls have many social benefits as an ecological engineering technique with notable potential for reconciliation ecology. Despite these benefits, there are currently no mature workflows or technologies for monitoring the health status and water stress of living wall systems. To partially fill the current knowledge gap related to water stress, we acquired thermal, multispectral and hyperspectral remote sensing data from an indoor living wall in the Cloud Forest of the Gardens by the Bay, Singapore. Surface temperature (Ts) and normalized difference vegetation index (NDVI) were obtained from these data to construct a Ts-NDVI space for applying the “triangle method”. A simple and effective algorithm was proposed to determine the dry and wet edges, the key components of the said method. The pixels associated with the dry and wet edges were then selected and highlighted to directly display the areas under water-stress conditions. Our results suggest that the proposed algorithm can provide a reasonable overview of the water-stress information of the living wall; therefore, our method can be simple and effective to monitor the health status of a living wall. Furthermore, our work confirms that the triangle method can be transferred from the outdoors to an indoor environment.
Subject	Earth and Environmental Sciences
Keyword	living wall triangle method remote sensing temperature NDVI
Related Publication	Xu Yuan, Kati Laakso, Chad Daniel Davis, J. Antonio Guzmán Q., Qinglin Meng and Arturo Sanchez-Azofeifa, 2020. Monitoring the Water Stress of an Indoor Living Wall System Using the “Triangle Method”. Sensors 2020, 20(11), 3261; https://doi.org/10.3390/s20113261. doi: https://doi.org/10.3390/s20113261 https://www.mdpi.com/1424-8220/20/11/3261/htm
Language	English
Depositor	Sanchez-Azofeifa, Arturo
Deposit Date	2020-06-02

Geospatial Metadata

Geographic Coverage	Singapore, Singapore
Geographic Bounding Box	103° 52’ E 103° 52’ E 1° 17’ N 1° 17’ N

Journal Metadata

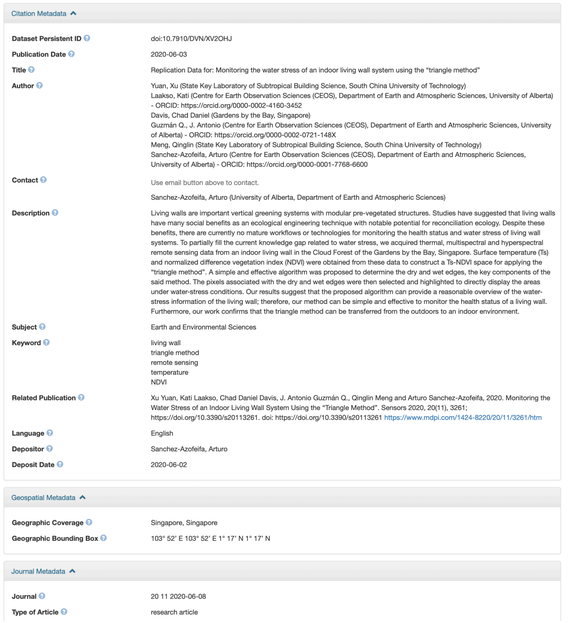
Journal	20 11 2020-06-08
Type of Article	research article

Documentation and data quality

Detail the metadata standard(s) to be used

BibTeX CSL DataCite Dublin Core DCAT JSON JSON-LD GeoJSON MARCXML Mendeley

Dublin Core
DDI
DataCite
DDI HTML Codebook
JSON
OAI_ORE
OpenAIRE
Schema.org JSON-LD



Citation Metadata

Dataset Persistent ID	doi:10.7910/DVN/XV2OHJ
Publication Date	2020-06-03
Title	Replication Data for: Monitoring the water stress of an indoor living wall system using the "triangle method"
Author	Yuan, Xu (State Key Laboratory of Subtropical Building Science, South China University of Technology) Laakso, Kati (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0002-4160-3452 Davis, Chad Daniel (Gardens by the Bay, Singapore) Guzmán Q., J. Antonio (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0002-0721-148X Meng, Qinglin (State Key Laboratory of Subtropical Building Science, South China University of Technology) Sanchez-Azofeifa, Arturo (Centre for Earth Observation Sciences (CEOS), Department of Earth and Atmospheric Sciences, University of Alberta) - ORCID: https://orcid.org/0000-0001-7768-6600
Contact	Use email button above to contact. Sanchez-Azofeifa, Arturo (University of Alberta, Department of Earth and Atmospheric Sciences)
Description	Living walls are important vertical greening systems with modular pre-vegetated structures. Studies have suggested that living walls have many social benefits as an ecological engineering technique with notable potential for reconciliation ecology. Despite these benefits, there are currently no mature workflows or technologies for monitoring the health status and water stress of living wall systems. To partially fill the current knowledge gap related to water stress, we acquired thermal, multispectral and hyperspectral remote sensing data from an indoor living wall in the Cloud Forest of the Gardens by the Bay, Singapore. Surface temperature (Ts) and normalized difference vegetation index (NDVI) were obtained from these data to construct a Ts-NDVI space for applying the "triangle method". A simple and effective algorithm was proposed to determine the dry and wet edges, the key components of the said method. The pixels associated with the dry and wet edges were then selected and highlighted to directly display the areas under water-stress conditions. Our results suggest that the proposed algorithm can provide a reasonable overview of the water-stress information of the living wall; therefore, our method can be simple and effective to monitor the health status of a living wall. Furthermore, our work confirms that the triangle method can be transferred from the outdoors to an indoor environment.
Subject	Earth and Environmental Sciences
Keyword	living wall triangle method remote sensing temperature NDVI
Related Publication	Xu Yuan, Kati Laakso, Chad Daniel Davis, J. Antonio Guzmán Q., Qinglin Meng and Arturo Sanchez-Azofeifa, 2020. Monitoring the Water Stress of an Indoor Living Wall System Using the "Triangle Method". Sensors 2020, 20(11), 3261; https://doi.org/10.3390/s20113261. doi: https://doi.org/10.3390/s20113261 https://www.mdpi.com/1424-8220/20/11/3261/htm
Language	English
Depositor	Sanchez-Azofeifa, Arturo
Deposit Date	2020-06-02

Geospatial Metadata

Geographic Coverage	Singapore, Singapore
Geographic Bounding Box	103° 52' E 103° 52' E 1° 17' N 1° 17' N

Journal Metadata

Journal	20 11 2020-06-08
Type of Article	research article

➔ Use community metadata standards where these are in place

Tools to generate metadata:

- For data (~FR, DataCite format): https://doranum.fr/wp-content/uploads/datacite_metadata_generator_4.0.html
- For research software and code: https://codemeta.github.io/codemeta-generator/

Examples: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XV2OHJ and https://zenodo.org/record/3664215#.YXFvVC-l3ew

Basic information required

Digital Object Identifier
e.g. 10.1234/foo.bar
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.

Reserve DOI

Publication date *
2021-10-22
Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

Title *
Required.

Authors *
Family name, given names | Affiliation | ORCID (e.g.: 0000-0002-1825-0097)
Optional.

+ Add another author

Description *
Source
Required.

Usually metadata standards are transparent when depositing data on a repository
➔ required and suggested (text) fields

Version
Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended.
See semver.org for more information on semantic versioning.

Language
e.g.: 'eng', 'fr' or 'Polish'
Optional. Primary language of the record. Start by typing the language's common name in English, or its ISO 639 code (two or three-letter code).
See ISO 639 language codes list for more information.

Keywords

+ Add another keyword

Additional notes
Optional.

Export Metadata
- Dublin Core
- DDI
- DataCite
- DDI HTML Codebook
- JSON
- OAI_ORE
- OpenAIRE
- Schema.org JSON-LD

Export
BibTeX CSL DataCite Dublin Core DCAT
JSON JSON-LD GeoJSON MARCXML
Mendeley

Documentation and data quality

Use controlled vocabularies (if relevant in your domain)

→ Help others identify and discover the data

Lists or Search tools:

- https://bartoc.org
- https://guides.ucf.edu/metadata/thesauri
- https://fairsharing.org/standards/
- https://www.dcc.ac.uk/guidance/standards/metadata

Documentation and data quality

Detail the **documentation accompanying the data**

- Explain the **organization of the data**:
  conventions, version control, folder structures, etc.
  → Aim for consistent & well-ordered research data
  → Help others (and yourself…) to find, understand, and re-use
- Explain what **documentation is needed to enable re-use**, e.g.,
  - Methodology used to collect the data
  - Analytical and procedural information, e.g., software needed for re-use
  - Definitions of variables
  - Units of measurement
  - Etc.

Describe **how** this information will be captured and **where** it will be recorded, e.g.,

- Database with links to each item
- Readme file
- File headers
- Code books
- Lab notebooks (do not need to be shared)
- Etc.

Template file for ReadMe (FR): https://doranum.fr/wp-content/uploads/gabarit_readme.txt

Guidelines for ReadMe: https://data.4tu.nl/info//fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf

Documentation and data quality

- **Data quality control measures**
  Detail the **control** and **documentation** of the **consistency** and **quality** of data collection
  E.g.,
  - Calibration
  - Repeated samples or measurements
  - Standardized data capture
  - Data entry validation
  - Peer review of data
  - Representation with controlled vocabularies

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

Storage and backup during the research process

Storage and backup of the data and metadata during the research process

- **Where**?
  Minimum 3 back-ups on 2 different supports at 2 distant locations
- **How often**?

➔ Give preference to **secure, robust, managed storage with automatic backup**,
such as provided by IT support services of the home institution, e.g.,
(FR) https://gricad.gricad-pages.univ-grenoble-alpes.fr/cellule-data-stewardship/web/stocker/
https://mycore.core-cloud.net/ (CNRS)
https://cloud.univ-grenoble-alpes.fr/ (UGA)

➔ Laptops, Stand-alone hard drives, or External storage devices (e.g., USB sticks)
not recommended for storage

For (specific) request: uga-cellule-data@univ-grenoble-alpes.fr

Storage and backup during the research process

Data security and protection of sensitive data

- Explain **data recovery strategies** in the event of an incident
- **Who will have access to the data** during the research

➔ Controlled, secure access for research partners

➔ Update access rights

- At the end of the project
- If a partner leaves

especially in collaborative partnerships

- **How access to data is controlled**
- **Detail data protection**: describe the main risks and how these will be managed
  ➔ **Encrypt the data**, e.g., if part of the data is personal (and sensitive data), politically sensitive information, or trade secrets
- **Explain institutional data protection policies** in place

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

PERSONAL DATA

?

PERSONAL DATA

- Data that can **directly identify the person**: last name, first name, address, picture, voice recording, etc.
- Data that can ***in*directly identify the person**: Telephone number, cross-referencing information such as the son of the research director, the latter living in Grenoble, etc.
    - **Pseudo-anonymized data** are personal data that can no longer be directly attributed to the data subject. However, by using additional information, such as a correspondence table, it is possible to re-identify the data subject.

➔ **Both type is personal data**

➔ Regulation on the protection of personal data
E.g., General Data Protection Regulation (GDPR) in Europe (FR: RGPD)

- **Irreversibly anonymized data**, which no longer allow for the re-identification of a person, are not subject to the regulation on the protection of personal data.

SENSITIVE PERSONAL DATA

?

SENSITIVE PERSONAL DATA

- Presumed race or ethnicity
- Political opinions
- Philosophical or religious beliefs
- Trade union membership
- Sexual orientation
- Health-related data
- Biometric data that can identify a person
- Genetic data

(+ Special status for *social security number* and data about *law violation and conviction*)

SENSITIVE PERSONAL DATA

- Processing sensitive personal data is forbidden
- Except, for example,
  - after consent of the person,
  - data made manifestly public by the person concerned,
  - important public interest,
  - safeguarding human life

For research:

- (FR) Ask CNIL (Commission Nationale de l'Informatique et des Libertés)
- Organize the security of the data

RESPONSABILITIES

- **Person responsible for data processing**: the lab director in a CNRS joint research unit (e.g., LIG, LJK, etc.)
- **Data protection officer (DPO)**: e.g., pascale poulet for LIG lab. Ask for you lab or company
- **Project coordinator:** should **ensure the compliance with the regulation**
- **Doctoral student:** implements the research complying with the regulation in conjunction with his/her thesis director
- **Subcontractor** (if any) = people that process personal data on your behalf/ instructions
  The **contract must ensure the security and confidentiality** of the data, and the respective commitments for data processing

RESPONSABILITIES

➔ **Inform the lab direction+DPO** about the processing of personal data:

- Name of person in charge of the data processing
- Name and contact details of
    1. the person responsible for data processing (lab director)
    2. the data protection officer Contact of the subcontract (if any)
- Purpose of the data processing
- Type of persons and data
- Recipients of the data
- Information about the use of the data (explain the processing in detail)
- Information about the long-term storage
- Information about the rights of the persons and the information they have

- **Data minimization**
  Collect as little data as necessary
- **Privacy by default**
  Collect as little personal data as necessary

On the contrary
to the researcher's tendency
to collect as much data as possible
to secure the costly data collection process

Legal and ethical requirements, codes of conduct

If personal data are processed

Detail how you will ensure the compliance with legislation (e.g., GDPR) on personal data and on data security

1. When collecting data

➔ **Gain informed consent** for preservation and/or sharing of personal data
E.g., a model of consent form we use in the Human-Computer Interaction research group:
https://cloud.univ-grenoble-alpes.fr/index.php/s/doGZPyQpnEKfdaD
E.g., another model from PACTE Lab: (in the annex) https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf

➔ **Describe the authorization** needed to access/collect the data

➔ **Describe the delay** needed to access/collect the data

https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

2. When storing the data for your own research

➔ **Consider anonymization of personal data** for preservation and/or sharing
Anonymization is irreversible: truly anonymous data are no longer considered personal data

(HowTo FR) https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles
By hand or with tools like Amnesia or ARX

➔ **Consider pseudonymization of personal data**
Pseudonymization is reversible: data are indirectly identifiable (truncated name, id number, etc.)

➔ **Reinforce security of storage, e.g., consider encryption**
the encryption key must be stored separately from the data, e.g., by a trusted third party

➔ **Reinforce access protection** and

Describe how you will collaborate with authorized users

https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

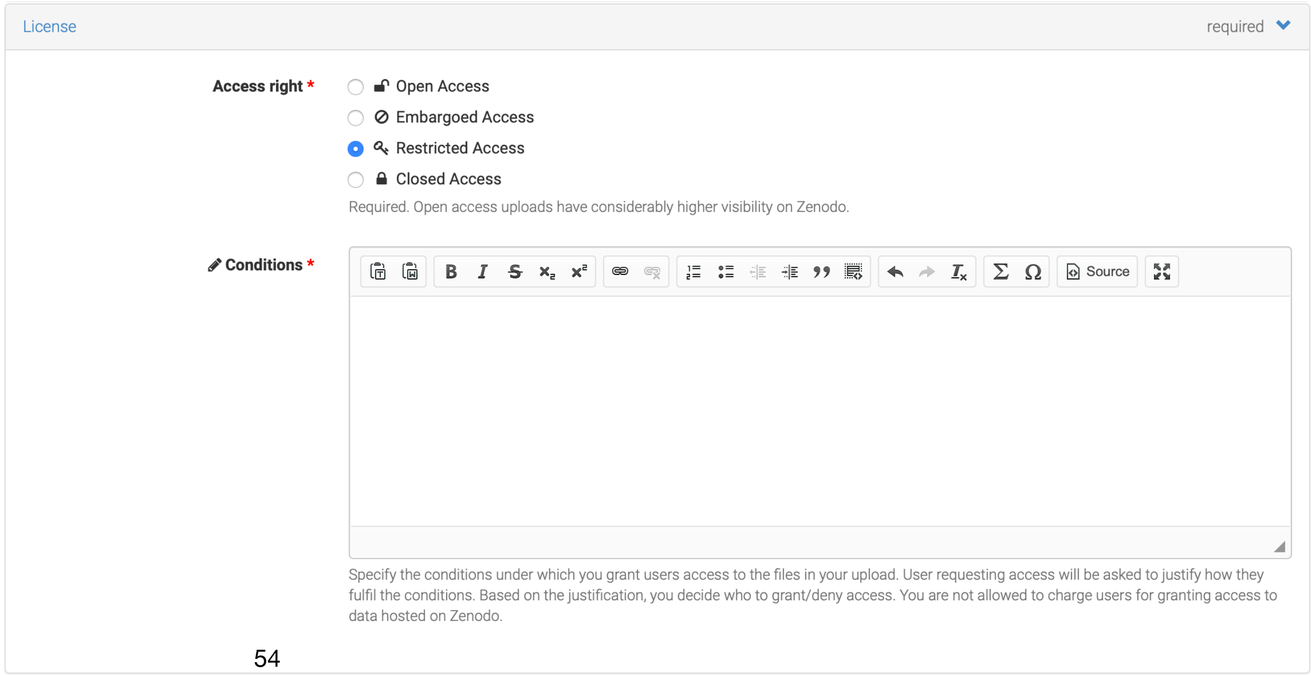
Legal and ethical requirements, codes of conduct

3. When archiving the data for sharing with others

➔ **Limit reuse to predefined usage**, as specified in the informed consent form

➔ **Ensure the recipient can be trusted** before granting access

E.g., http://buff.is.tue.mpg.de/downloads or https://cape.is.tue.mpg.de/dataset.html or on Zenodo



Legal and ethical requirements, codes of conduct

How will you manage other legal issues? What legislation is applicable?

- Name the **owner of the data** = who will have the rights to control access
- Explain access conditions
  ➔ Consider the use of data access and re-use licenses
  - Openly accessible
  - Restricted access ➔ explain access condition
- Ensure that access to data is specified *when multiple partners and data owners* (+ in the consortium agreement too)
- Indicate whether intellectual property rights are affected and explain which and how will they be dealt with
  Ask Innovation and Transfer at UGA, or CNRS, or INRIA, etc.
- Indicate restrictions on the re-use of third-party data

Legal and ethical requirements, codes of conduct

How will you take into account possible **ethical issues**?

Can ethical issues affect:

- Data **collection**
- Data **processing**
- Data **storage**
- Data **transfer**
- Data **access**
- Data **archival**

➔ Demonstrate awareness of these aspects and respective planning

How will you follow **codes of conduct**?

- Follow the national and international codes of conducts
  - (FR) Charte nationale de déontologie des métiers de la recherche
  - (EU) European code of conduct for research integrity
- Follow institutional ethical guidelines if any
- **Ask if you need an ethical review** (for example by an ethics committee) for data collection

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

Data sharing and long-term preservation

- **How and when will you share the data?**
  Are there possible restrictions to data sharing or embargo reasons?

- **How will you select data for preservation**?
  Where will data be preserved long-term? e.g., in a repository or archive

- **What methods or software tools will be needed** to access and use the data?

- How will you **ensure** the application of a **unique and persistent identifier to each data set**?
  e.g., a Digital Object Identifier (DOI) most widely used
  ➔ Often provided by the repository/archive

Where to find a relevant repository?

Tools for searching a repository:

- https://www.re3data.org
- https://fairsharing.org
- OpenDOAR
- Repositories recommended by Nature
- (FR) CAT OPIDoR

Specific domain (e.g., Dryad Digital Repository)
vs. **Multidisciplinary** (e.g., Zenodo or OSF)

Public vs. **Private** (e.g., Figshare)

French/Local/Institutional (e.g., PerSciDo)
vs. **European** (e.g., Zenodo)
vs. **Rest of the world** (e.g., OSF)

Paying (e.g., Dryad Digital Repository:
$120 for first 20 GB
$50 USD for each additional 10 GB)
vs. **Free** (e.g., OSF)

Where to find a relevant repository *for software*?

GitHub, GitLab, etc. are not for long-term preservation
➔ Similar code repositories have been close in the past
e.g., Google Code, Bitbucket, etc.

- Integration of HAL and Software Heritage
- Integration of GitHub and Zenodo

SELECTING A TRUSTWORTHY REPOSITORY

SELECTING A TRUSTWORTHY REPOSITORY

1. Refer to broadly recognized discipline-specific
   or certified repositories
   e.g., https://www.coretrustseal.org/why-certification/certified-repositories/,

2. In cases where no such repository can be identified,
   use the following criteria for the selection of trustworthy repositories

A TRUSTWORTHY REPOSITORY SHOULD

Persistent and Unique Identifiers (PIDs)

- Allow data discovery and identification
- Enable searching, citing, and retrieval of data
- Provide support for data versioning

A TRUSTWORTHY REPOSITORY SHOULD

Metadata

- Enable finding of data
- Enable referencing to related relevant information
  E.g., other data and publications
- Provide information that is publicly available and maintained
  Even for non-published, protected, retracted, or deleted data
- Use metadata standards that are broadly accepted (by the scientific community)
- Ensure that metadata are machine-retrievable

A TRUSTWORTHY REPOSITORY SHOULD

Data access and usage licences

- Enable access to data under well-specified conditions
- Ensure data authenticity and integrity
- Enable retrieval of data
- Provide information about licensing and permissions in ideally machine-readable form
- Ensure confidentiality
  Respect rights of data subjects and creators

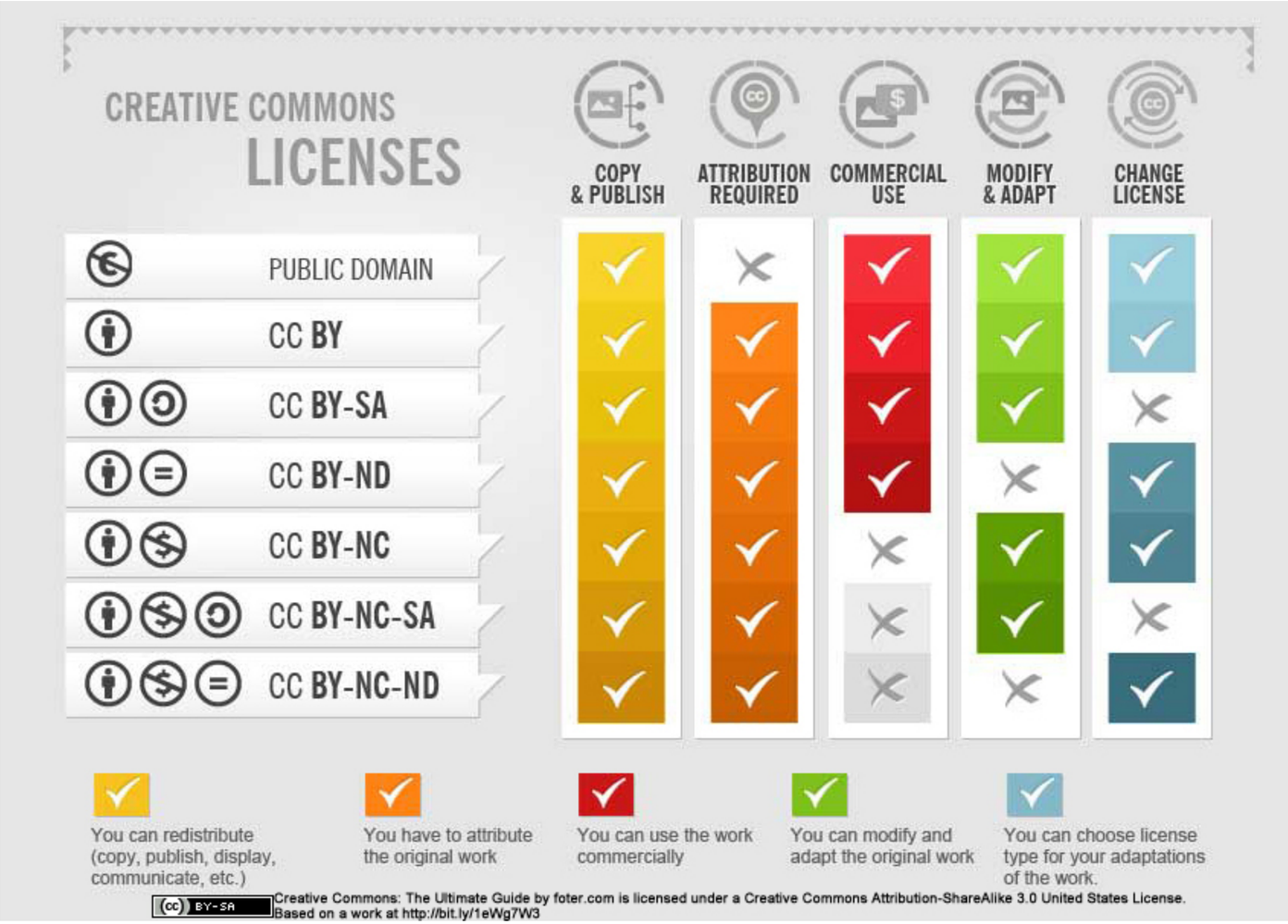
A TRUSTWORTHY REPOSITORY SHOULD

Preservation

- Ensure persistence of metadata and data
- Be transparent about
  - Mission
  - Scope
  - Governance
  - Financial sustainability
  - Preservation policies, including retention period
  - Continuity plan
    In case of unplanned disruption

Decision tool: https://ufal.github.io/public-license-selector/

Open licenses for Data vs. Software



GNU

For data bases: https://opendatacommons.org/licenses/index.html

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

Data management responsibilities and resources

Who will be responsible for data management?

- Not necessarily the project coordinator or PI
  Can be, e.g., an engineer
- Detail name, role, position, and institution

What resources will be dedicated to data management?

- Financial resources
- Time resources

➔ To ensure that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)

Data management responsibilities and resources

Tools to estimate resources:
E.g., for budgeting (eligible for funding)

- OpenAIRE RDM costs
- UK Data Service - Data management costing tool and checklist
- EPFL Library Cost Calculator for Data Management
- Utrecht University Cost of data management
-

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS

Data description and collection or re-use of existing data

Documentation and data quality

Storage and backup during the research process

Legal and ethical requirements, codes of conduct

Data sharing and long-term preservation

Data management responsibilities and resources

+ need to adapt
- to specific domains
- to local legislation
- to funder
- to local institution
  (e.g., university, might provide/enforce safe storage services even if it not necessary for your project)

How to evaluate a DMP?

Europe	(EN) https://www.scienceeurope.org/media/22hpslfl/se-rdm-template-5-guidance-on-the-evaluation-of-data-management-plans.docx (FR) Ancelin-Fabre, Justine. « Grille_evaluation_H2020_fr.docx ». https://drive.google.com/file/d/17kjkq-OEwBre2Z8U7fvwILzlGbaGmlwf/view
France (ANR)	• (FR) Ancelin-Fabre, Justine. « Grille_relecture_PGD_ANR.docx ». https://drive.google.com/file/d/1A7LHW_y1vHmbHxYmmjECpKgmOsicjA01/view • (FR) Doranum. « Grille de relecture de PGD - Modèle ANR ». https://doranum.fr/wp-content/uploads/Grille-relecture-PGD-Modele-ANR-V3.pdf

References

- Science Europe Ressources
  E.g., https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf
- Consortium of European Social Science Data Archives
- (FR) Cécile Arènes, “Rédiger un plan de gestion de données”, https://doi.org/10.5281/zenodo.5559598
- Arnould Pierre-Yves, Jacquemot-Perbal Marie-Christine, “Guide of good practises”, https://doi.org/10.24396/ORDAR-1
- (FR) https://gricad.gricad-pages.univ-grenoble-alpes.fr/cellule-data-stewardship/web/
  ➔ Question about data management: uga-cellule-data@univ-grenoble-alpes.fr