

HOW TO MAKE YOUR DATA FAIR

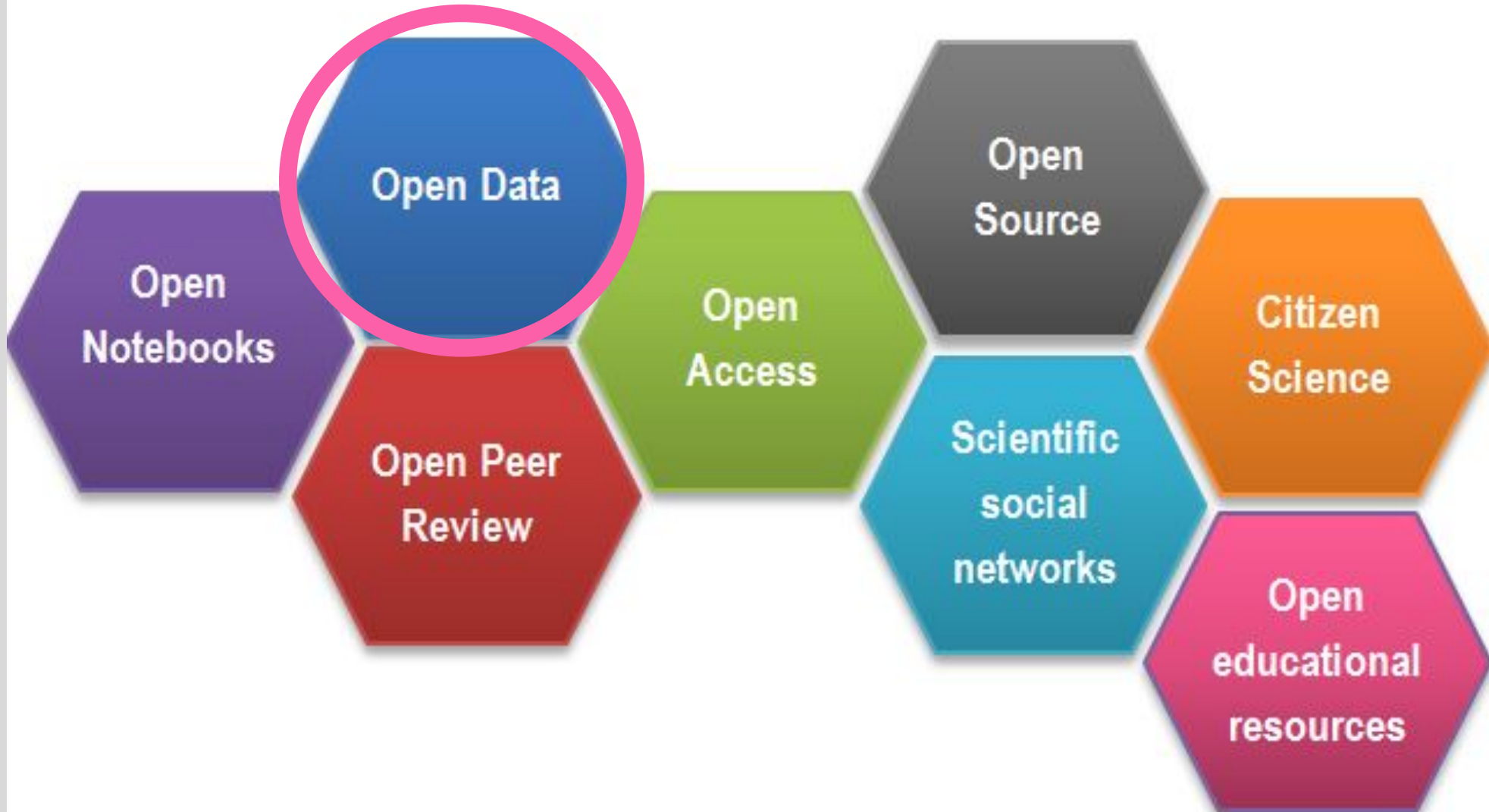
AND WHAT GETS IN THE WAY

AND HOW TO DO IT ANYWAY

Dr. Lena Thöle
Utrecht University Library
l.m.thole@uu.nl

QUESTIONS

WHY OPEN SCIENCE?



WHY OPEN SCIENCE? WHY FAIR DATA?



As open as possible, as closed as necessary

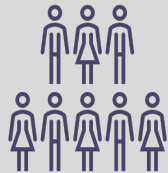
WHY FAIR DATA?



Impact & visibility



Sustainability



Team



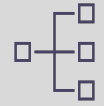
Accessibility



Transparency & accountability



WHY FAIR DATA?



Keep yourself organized

- *File naming, folder structure*
- *Storage & back-ups*



Reproducibility & quality control

- *Version control*
- *Open formats & softwares*



Recognition and visibility

- *Data sharing & preservation*

Collaboration



- *Data sharing & reuse*
- *Metadata & collaboration*



Meet funder mandates & policies

- *Data Management Plans*

Comply to legislation

- *GDPR*



THE FAIR PRINCIPLES

Findable



Accessible



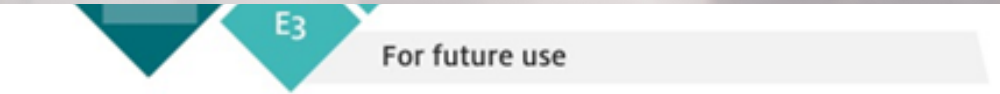
Interoperable



Reusable



THE CARE PRINCIPLES FOR INDIENOUS DATA GOVERNANCE



HOW TO GET TO FAIR DATA: GOOD DATA MANAGEMENT

Folder structure, file naming & version control

Metadata & documentation

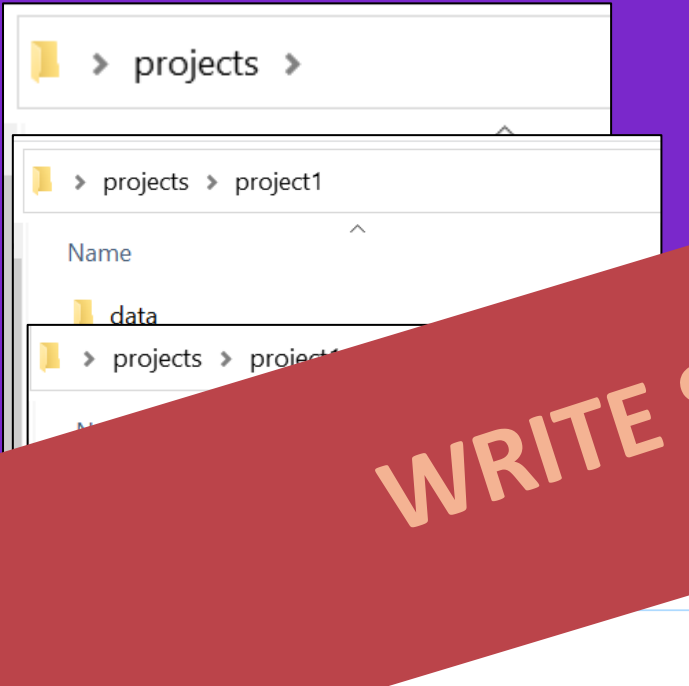
Privacy, Security & Storing

Archiving & Publishing

EXERCISE

THE BASICS: FOLDER STRUCTURE, FILE NAMING & VERSION CONTROL

Folder structure



File naming

- Needs to match
- File

WRITE SEPARATE README FILES

- No special characters, white spaces, periods

Open file format

Version control

- `_v01.docx`
- `_ddmmyy.docx`
- `final_final_final_2.doc`
- Git
- GitHub (GitLab...)



METADATA & DOCUMENTATION

- **Metadata**

- Data on your data: Structured data providing information about one or more aspects of the data
- Descriptive metadata (author, affiliation, language,..)
- Data level metadata (description of variables)

- **Standardized and controlled vocabulary**

- **Documentation**


All contextual information pertaining to your data, data files

- **Methodology**
(i.e methods section of an article)
- Data management plan
- **Codebooks**, controlled vocabularies (data level metadata)
- **READ_ME.txt files**
- Lab book
- Etc...

METADATA & DOCUMENTATION

CODEBOOKS & README FILES

	A	B	C	D	E	F
1	Variable	Descriptions	Levels	Unit	Notes	
2	T	Temperature	from -20 to +50	C	from thermometer	
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						



Utrecht University **FAIR DATA CHEATSHEET**
uu.nl/rdm-cheatsheets

README.md

Title of your dataset

In this section, provide an overview of your dataset and describe the project and research questions in which the dataset was collected. Explain the data collection methods. Highlight the purpose, scope, and potential uses of your dataset.

Prerequisites

Include any necessary prerequisites for using your dataset, such as specific software or hardware requirements. Provide clear instructions for how to access or download your data files.

Contents

Describe the organization of your data package, including the contents of each folder and the files it contains. Provide and explain the naming convention used for the files in your dataset. Also describe the file format(s) used in your dataset and the software required to open them.

Codebook

Provide a codebook that explains the variables in your dataset. This is particularly important for tabular data, but it is useful for any other type of file as well. Include at least the variable name, a brief description, and the data type.

License

Include a license that allows others to use and share your work. Consider using an open data license, such as a Creative Commons License.

For example: This work is licensed under a CC BY 4.0.

Citation (optional)

Provide clear instructions on how to cite your dataset or related publications in a research paper or publication.

Contact

Include contact information for questions, feedback and suggestions about your dataset.

Interested in more on writing a README?
Visit uu.nl/rdm for guides, workshops, and walk-in hours. Or contact our experts at info.rdm@uu.nl.

Check out the extended version:
<https://bit.ly/fair-cheatsheet-readme>

14

PRIVACY, SECURITY & STORING

PRIVACY – the GDPR



SECURITY

Security of data files

- Encryption
- Pseudonymization
- Minimization
- Abstraction
- Anonymization
- Access control
- Secure transport
- Complete deletion
- 3-2-1 back ups

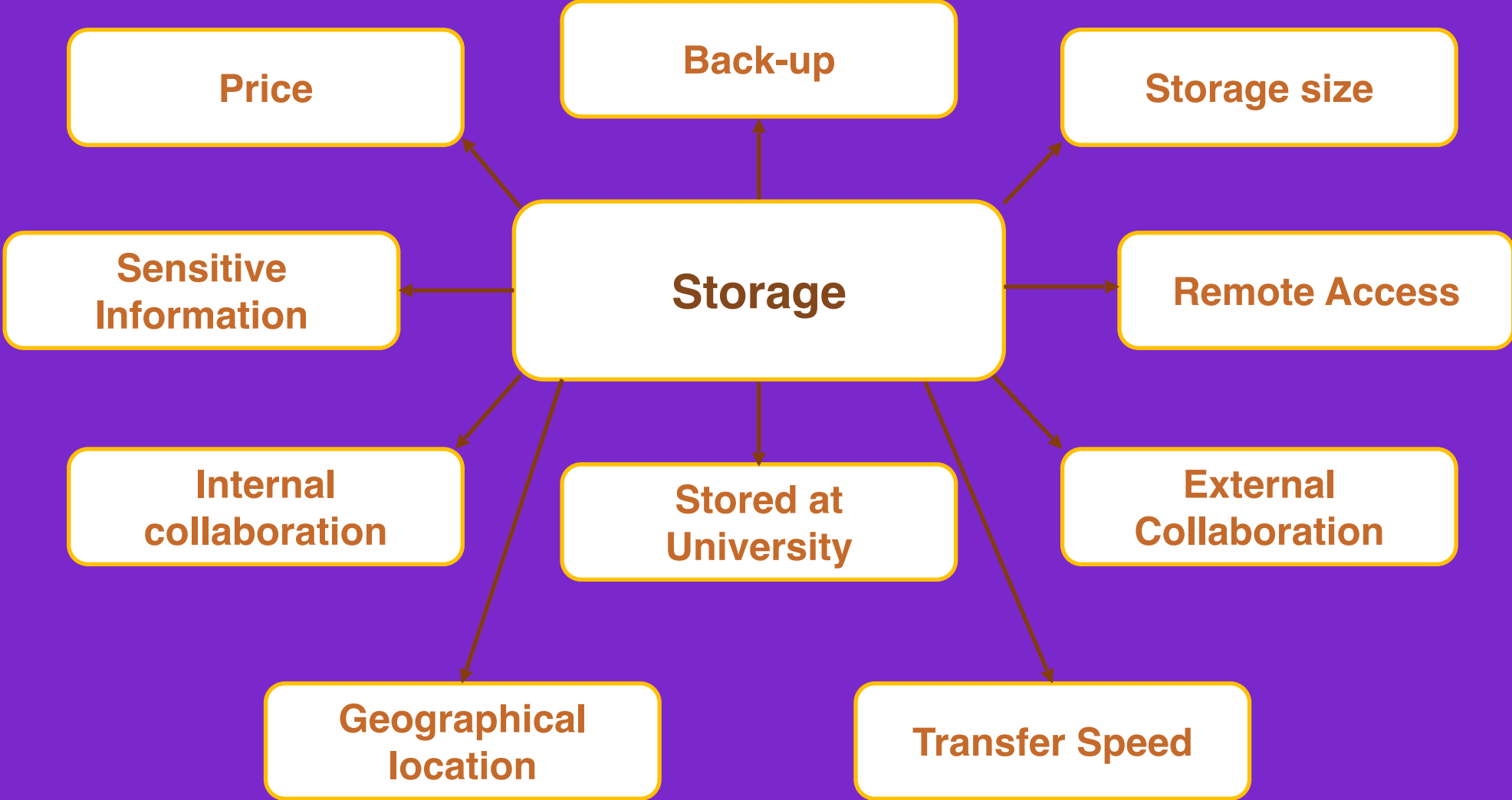
Security of computer system

- Firewall
- Antivirus software
- Installing updates
- Using secured WIFI
- Password-protected
- Device encryption

Physical data security

- Key & Lock
- Don't leave unattended
- Safe transport

STORING – DURING RESEARCH



ARCHIVING & PUBLISHING DATA

PUBLISHING DATA

WHERE? Data repositories!

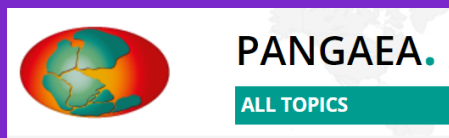
HOW? Licenses & Data agreements

PUBLISHING DATA

WHERE? Data repositories!



domain specific vs institutional vs general repo



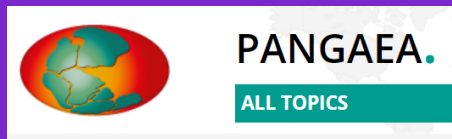
HOW? Licenses & Data agreements

PUBLISHING DATA

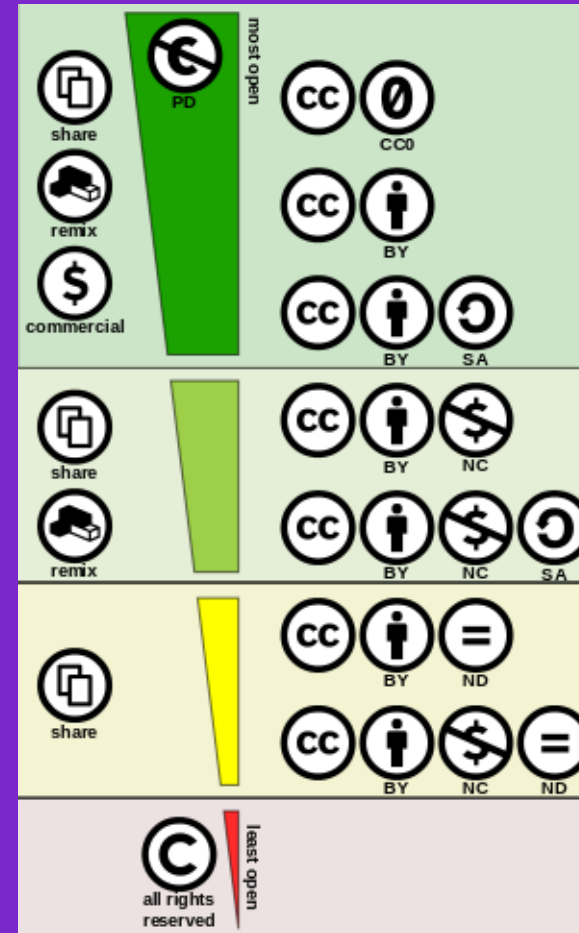
WHERE? Data repositories!



domain specific vs institutional vs general repo



HOW? Licenses & Data agreements



DTA
NDA

WHAT GETS IN THE WAY?

WHAT GETS IN THE WAY?

- This seems trivial
- Who could use my data?
- Extra work
- Misuse/misinterpretation
- Ethical concerns
- Data sensitivity

- Idea taken away!
- Fear of criticism!

WHAT GETS IN THE WAY?

- Idea taken away!
 - Collaborative effort
 - Beneficial for society/policy – think CORONA, WEATHER
 - Publish only metadata
 - Adjust your license
 - Data governance – who does the data belong to?
 - Advancement of knowledge

WHAT GETS IN THE WAY?

- Fear of criticism:
 - Imposter syndrome
 - Fear of mistakes
 - Idea of perfectionism
- Don't be perfect
- Be brave
- Be vulnerable
- Opportunity to grow & learn

Thank you!

Go and be FAIR!

