# FAIR in more detail

Workshop FAIR data and data reuse for ESG researchers – Module 2

October 18, 2022, by Cindy Quik and Luc Steinbuch

WAGENINGEN UNIVERSITY & RESEARCH

https://doi.org/10.4121/21399975

# From concepts to practice

> *Decisions about how to document and share methods should be made when researchers are designing their experiments, not when they are writing their manuscripts*

Teytelman, 2018



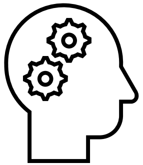
https://doi.org/10.4121/21399975


Teytelman L (2018). No more excuses for non-reproducible methods. Nature, 560, 411.
https://doi.org/10.1038/d41586-018-06008-w

# From concepts to practice: FAIR

- Explanation on how to prepare FAIR datasets

→ Thinking ahead!



- Data management begins before starting your project



https://doi.org/10.4121/21399975

# From concepts to practice: FAIR

*F* Findable

*A* Accessible

*I* Interoperable

*R* Reusable

More background:

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 160018–160019. https://doi.org/10.1038/sdata.2016.18

http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/



WAGENINGEN UNIVERSITY & RESEARCH

https://doi.org/10.4121/21399975

# Findable

- Need an identifier: DOI
- Registered in a searchable resource
- Metadata contain attributes that can be used in searches
- Register paper/report AND dataset in PURE (irrespective of access)

https://www.doi.org/;
https://www.wur.nl/en/library/researchers/research-impact/make-your-wur-output-visible.htm


https://doi.org/10.4121/21399975

# Accessible

- Accessible in time — Long-term (>10 years) storage of data
- Accessible in location — Clear storage location (physical or digitally/online)
- Accessible to whom — Open/restricted/closed access
- Accessible in language?

WAGENINGEN UNIVERSITY & RESEARCH

https://doi.org/10.4121/21399975

# Accessible – to whom

- Open access — To all
- Restricted access — To selected groups
- Closed access — Only accessible in specific cases, GDPR applies

→ Flowchart Data Sharing @ WUR


https://doi.org/10.4121/21399975


https://www.openaccess.nl/; https://www.wur.nl/en/value-creation-cooperation/collaborating-with-wur-1/wdcc/research-data-management-wdcc/finishing/data-sharing-and-guidelines.htm

# Accessible – choosing a data repository

- Consider:
  - Size of your dataset
  - Open access/restricted access
- Options to choose from:
  - 4TU.ResearchData
  - DANS-EASY
  - W: drive

→ Examples will follow in Module 6

https://data.4tu.nl/portal
https://easy.dans.knaw.nl/ui/home



https://doi.org/10.4121/21399975

# Interoperable

- Which file types are still readable many years from now?

→ Preferred file formats

E.g. use:

.txt, .csv, .shp

Do not use:

.xlsx, .docx, etc.

- Think beforehand about...
  - File types you will be using
  - Conversion to which universal file formats is needed



https://doi.org/10.4121/21399975

# Reusable

- Licenses for data
- Licenses for code will be discussed in Module 5
- Understandable documentation (metadata)


https://doi.org/10.4121/21399975

# Reusable – licenses (for data!)

- Creative Commons Licenses

“A standardized way to grant the public permission to use creative work under copyright law”

→ What is the user allowed to do with this work?

https://creativecommons.org/

# Reusable – licenses (for data!)

- CC0
- CC BY
- CC BY-SA
- CC BY-NC
- CC BY-NC-SA
- CC BY-ND
- CC BY-NC-ND
- “all rights reserved”



https://creativecommons.org/; https://commons.wikimedia.org/wiki/File:Copyright.svg

https://doi.org/10.4121/21399975

# Reusable – licenses (for data!)

- CC Creative Commons
- BY Credit must be given to the creator
- SA Adaptations must be shared under the same terms
- NC Only noncommercial uses of the work are permitted
- ND No derivatives or adaptations of the work are permitted

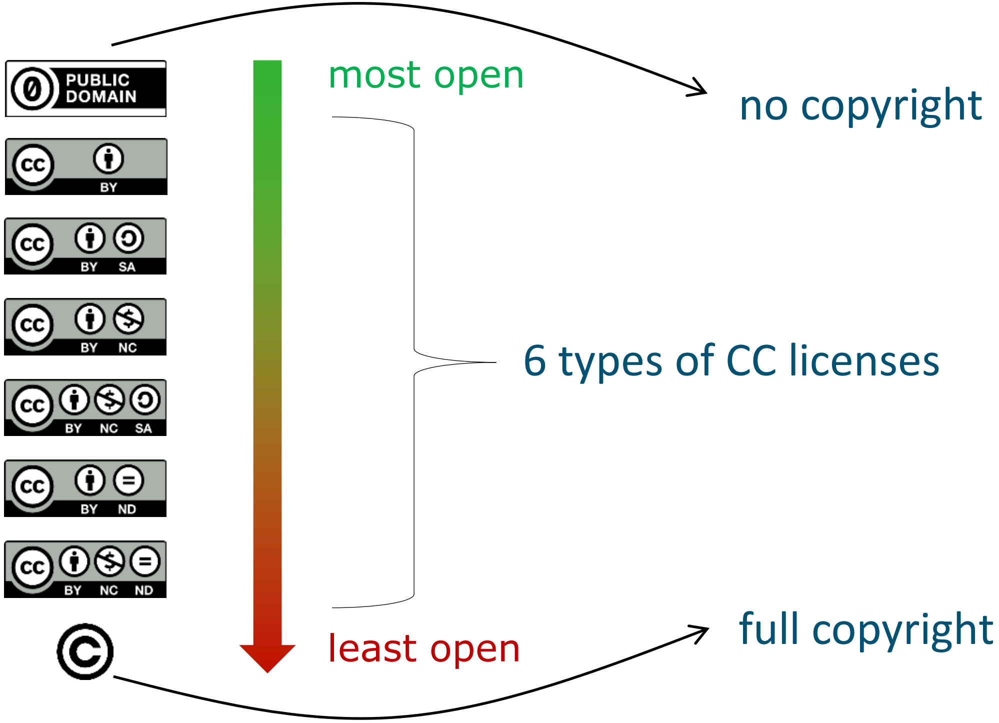


https://creativecommons.org/

https://doi.org/10.4121/21399975

# Reusable – licenses (for data!)

- CC0
- CC BY
- CC BY-SA
- CC BY-NC
- CC BY-NC-SA
- CC BY-ND
- CC BY-NC-ND
- “all rights reserved”




https://creativecommons.org/; https://commons.wikimedia.org/wiki/File:Copyright.svg

# Reusable – licenses (for data!)

- Most often used: CC BY
- Check version (CC BY 4.0)
- If you need help with choosing a license:
  - Creative Commons License Chooser
  - Ask the Wageningen Data Competence Centre for advice
- Keep in mind that:
  - The licenses and CC0 cannot be revoked!
  - You must own or control copyright in the work!

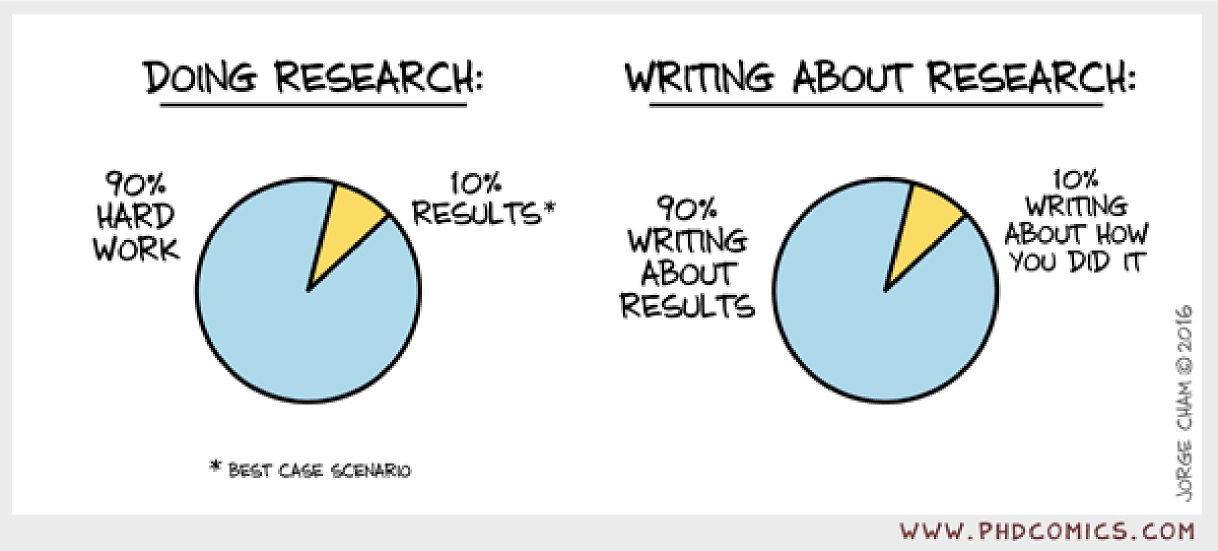
WAGENINGEN
UNIVERSITY & RESEARCH

https://doi.org/10.4121/21399975


https://creativecommons.org/

# Reusable – documentation

- Methodology: what to write down, what to leave out?




http://phdcomics.com/comics.php?f=1887



https://doi.org/10.4121/21399975

# Reusable – documentation

- Understandable documentation
    - In the manuscript/report
    - Metadata

- Stating the obvious, or valuable methodological details....?



https://doi.org/10.4121/21399975

# Reusable – documentation

- Root Mean Square Error, three times different

When the general formula is derived and applied to the control point, a measure of the error—the residual error—is returned. The error is the difference between where the from point ended up as opposed to the actual location that was specified—the to point position. The total error is computed by taking the root mean square (RMS) sum of all the residuals to compute the RMS error. This value describes how consistent the transformation is between the different control points (links). When the error is particularly large, you can remove and add control points to adjust the error.

Although the RMS error is a good assessment of the transformation's accuracy, don't confuse a low RMS error with an accurate registration. For example, the transformation may still contain significant errors due to a poorly entered control point. The more control points of equa

layer is often represented by the total root-mean-square error (RMSE), a metric based in the Pythagorean Theorem and calculated for a coordinate pair by the equation (Slama et al., 1980)

$$\mathrm{RMSE} = \left[(x_s - x_r)^2 + (y_s - y_r)^2\right]^{1/2} \quad (1)$$

where $x_s$ and $y_s$ are geospatial coordinates of the point on the source image; and $x_r$ and $y_r$ are coordinates of the same point on the transformed aerial photo. The RMSE for the whole image is the sum of the RMSE for each coordinate divided by the square root of the


Figure 4. Methodology used in this paper

Overall quantification of the planimetric accuracy can be done by calculating the Mean Positional Error (with accompanying standard deviations). The Mean Positional Error (MPE) is, in this case, defined as $[(\sum v^2)/n-2)]^{1/2}$ for the old map. The factor $v$ is defined as $(vx^2+vy^2)^{1/2}$ with $vx$ and $vy$ the vector distances in x and y – in metres – between the actual points in the old map and the place they would be if the old map was perfect. Alternatively the same calculations could be made in real metres, using the new map and corresponding vector distances.[17] The result should be interpreted as the average distance between a randomly chosen point on the map and its



https://doi.org/10.4121/21399975




References not mentioned for reasons of discretion

# Reusable – documentation

> *It must be written with enough information so that:*
> *(1) the experiment could be repeated by others to evaluate whether the results are reproducible, and*
> *(2) the audience can judge whether the results and conclusions are valid*

Kallet, 2004



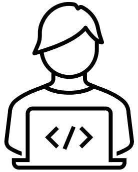


Kallet R.H. (2004). How to write the methods section of a research paper. Respiratory Care, 49(10), 1229-1232.

# Reusable – documentation

- Code

→ More on working with code in Module 5






https://doi.org/10.4121/21399975

# Reusable – documentation

→ ReadMe file

Preparing a ReadMe file before starting research offers a guideline while conducting and finishing your research project!

We will practice with writing a ReadMe file after the next coffee break


https://doi.org/10.4121/21399975

# Thinking ahead

***F*** Findable

***A*** Accessible

***I*** Interoperable

***R*** Reusable

- What will be the size of your dataset?
- Who may access your dataset?
- Choose repository that offers this access type and request DOI
- Which file types will be used in your project?
- Which conversions are needed to obtain universal file formats?
- What is the user allowed to do with your work? Which license?
- Prepare a draft ReadMe file and a draft metadata description

WAGENINGEN
UNIVERSITY & RESEARCH

https://doi.org/10.4121/21399975

# While conducting your research...

- Research journal and ReadMe
- Unprocessed/processed data
- Version management
- Continuously update ReadMe
  and metadata files, ask feedback
- Checklist for finalizing a FAIR dataset





https://doi.org/10.4121/21399975

# The Library can help!

- data.library@wur.nl

WAGENINGEN
UNIVERSITY & RESEARCH



https://doi.org/10.4121/21399975

# Please be back at 11:20

"coffee with milk", an icon from Noun Project, color #0066cc
https://commons.wikimedia.org/wiki/File:Noun_coffee_with_milk_2695655.svg



https://doi.org/10.4121/21399975