

How to write a README.*

Workshop FAIR data and data reuse for ESG researchers – Module 3

October 18, 2022, by Cindy Quik and Luc Steinbuch



metadata (“information about your data”) makes your data:

Findable (title, keywords etc.)

Accessible

Interoperable (in development)

Reusable (context & meaning; license)

metadata & README

4TU metadata:

Metadata record only [Link file](#)

needed to publish ●

Title

Authors
 
Search co-authors by name, full email or ORCID. Hit enter after each.

Categories

Item type

Keyword(s)

Description

Tips
Use this form to edit all information related to your data. Please be as descriptive as possible. The file upload is independent from the rest of the form, so you don't need to save an upload. This message will be replaced with helpful tips and suggestions as you begin interacting with the form.

Preview item (private)
Edit timeline
Last edited on 02 Oct 2022 - 15:36

Funding

+ [Add another grant](#)

Resource title

Resource DOI

References

Licence (what's this?)

Publisher

Language

Time coverage

metadata & README

Developments:

- Metadata standards (research field specific)
- Controlled vocabularies (with thesauri), taxonomies, ontologies:
=> standard definition of key concepts + their relationships.
Allows to logically link datasets (“symantic web”)

metadata & README

Example metadata standards

Recommended Metadata Element Sets

General Purpose:

- [DublinCore](#) (DC) Metadata Element Set is a generic set of 15 properties for describing a wide range of resources.
- [Metadata Object Description Schema](#) (MODS) is a descriptive standard used to describe a variety of types of resources; it is maint

Sciences:

- [Darwin Core](#) (DWC) is used in the biological sciences to describe collections of biological objects or data and includes a glossary of about biological diversity.
- [Access to Biological Collection Data](#) (ABCD) is used in the biological sciences to describe specimens and scientific observations.
- [Astronomy Visualization Metadata Standard](#) (AVM) is used to describe data-derived astronomical images.
- [Ecological Metadata Language](#) (EML) is used to formalize and standardize concepts necessary when describing ecological data.
- [FDGC Federal Geographic Data Committee Standard](#) is a standard for documenting digital geospatial data; it is especially relevant for Geospatial Information Systems.

Social and Behavioral Sciences:

- [Data Documentation Initiative](#) (DDI) is a standard for describing observational and survey data in the social, behavioral, economic

From: <https://pitt.libguides.com/metadatadiscovery/metadata-standards>

Metadata standard for geographic data: ISO19115 etc -> next module!

metadata & README

Example controlled vocabulary

Hague, The (inhabited place)

Coordinates:

Lat: 52 05 00 N *degrees minutes* Lat: 52
Long: 004 18 00 E *degrees minutes* Long: 4

Note: Originally located in woodland called H-1248, Wilhelm II built a castle, which his son economically in the late 14th century when it became the seat of the stadholders and the 19th century, it became an important conference site. The International Court of Justice of the United Nations is located here.

Names:

's-Gravenhage ([preferred](#), [C](#), [V](#), [Dutch](#), [U](#))

Gravenhage, 's- ([C](#), [V](#))

's Gravenhage ([C](#), [V](#))

's Gravenhage ([C](#), [V](#))

Den Haag ([C](#), [O](#), [Dutch](#), [U](#))

~~Haag ([C](#), [O](#), [Dutch](#), [U](#))~~

The Hague ([C](#), [O](#), [English](#), [U](#))

Hague, The ([C](#), [O](#), [English-P](#), [U](#), [N](#))

La Haye ([C](#), [O](#), [Dutch](#), [U](#))

Haye, La ([C](#), [O](#))

Aia ([C](#), [O](#))

~~The Haag ([NA](#), [V](#), [Dutch](#), [U](#))~~

אהא ([C](#), [U](#), [Hebrew](#))

[Research Home](#) ▶ [Tools](#) ▶ [Thesaurus of Geographic Names](#) ▶ [Full Record Display](#)



Getty Thesaurus of Geographic Names® Online
Full Record Display

Hierarchical Position:

-  [World](#) (facet)
-  [Europe](#) (continent) (P)
-  [Netherlands](#) (nation) (P)
-  [South Holland](#) (province) (P)
-  [Hague, The](#) (inhabited place) (P)

Place Types:

- inhabited place ([preferred](#), [C](#)) originally site of the hunting lodge, grew up around castle built in 1248
- city ([C](#))
- ~~national capital ([C](#))~~
- provincial capital ([C](#))

Related: how to name your dataset? I

“Findable”

How would you name your dataset or code?

Related: how to name your dataset? II

Option: Proper description

Examples from research.wur.nl -> Dataset

Documented socio-economic utility of Asia-Pacific mangroves and mangrove-associate species

Ingredient database for the formulation of sustainable, local fish feeds

The percentage of total agricultural area under maize, rice, wheat, vegetables, pulses and fruit

AgroDataCube: A Big Open Data collection for Agri-Food Applications

Survey data on Dutch farmers' perceived resilience, risk management, risk preferences, and risk perceptions

Cone index and surface soil moisture measurements in selected agricultural fields within the

Gridded rainfall maps retrieved from commercial microwave link (CML) data from Sri Lanka

Tomato plant spider mite damage images

Related: how to name your dataset? III

Other option: Title of paper/report with preceding
“Data ...”

Examples from research.wur.nl -> Dataset

Enhancing agroecosystem productivity with woody perennials in semi-arid West Africa: a meta-analysis

Dataset from: On the importance of root traits in seedlings of tropical tree species

Data from observations of the role of water hyacinths in macroplastic transport and accumulation

Data underlying the proceeding: Are low-cost, hand-held NIR sensors suitable to detect

Replication data for: Effectiveness of a behaviorally informed financial education program for

Data belonging to Oram et al. Plant community flood resilience in intensively managed

Related: how to name your dataset? IV

Examples from research.wur.nl -> Research output -> software

With code: similar diversity

Maternal food restriction during pregnancy affects offspring development and swimming performance in a placental live-bearing fish

Diurnal and nocturnal mosquitoes escape looming threats using distinct flight strategies

Algorithm for constrained multi-objective land use allocation optimization under uncertainty

Code for: Harvesting forage fish can prevent fishing-induced population collapses of large piscivorous fish

Data from: In silico study of the role of cell growth factors in photosynthesis using a virtual leaf tissue generator coupled to a microscale photosynthesis gas exchange mod

Python scripts to generate Figure 9.1: Rarefaction curves of twenty biggest genera (i.e., having the most number of available genomes) from the global analysis of 1.2 million BGCs

Software and scripts for "Genetic Polyploid Phasing from Low-Depth Progeny Samples" `microsud/Db-MM-10: Manuscript-version`

R code for reproducing results

`bfast`

Related: how to name your dataset? V

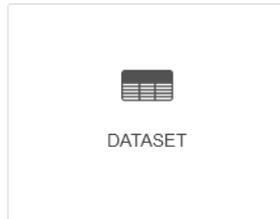
Advice from WUR library:

Title of the dataset:

<https://edepot.wur.nl/534147>

Provide a title that reflects the content of the dataset. If the title of the dataset is the same as the title of your publication, you can add 'Data from:' or 'Data underlying the publication:' to make the dataset easily distinguishable from the publication.

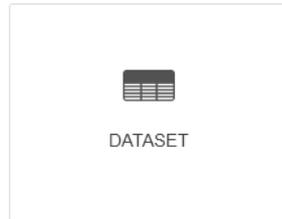
But:



Data underlying the publication "Rapid change of friction ..."

Dataset posted on 29.09.2022

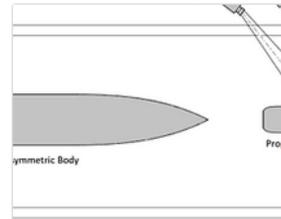
Laurence Willemet ▾



Data underlying the article: Early fatigue damage accumulati...

Dataset posted on 29.09.2022

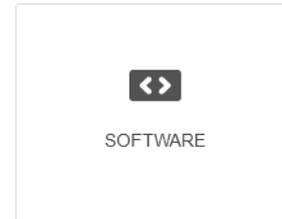
Xi Li ▾



Figures created during the work on "Design of ..."

Dataset posted on 29.09.2022

Sagar Adatrao ▾



MATLAB scripts created during the work on "Design of ..."

Software posted on 29.09.2022

Sagar Adatrao ▾

<https://data.4tu.nl/>, screenshot Oct 3rd, 2022

Related: how to name your dataset? VI

Proposal:

Functional traits shape tree species distribution in the Himalayas (dataset)

On the importance of root traits in seedlings of tropical tree species (dataset)

Diurnal and nocturnal mosquitoes escape looming threats using distinct flight strategies (code with dataset, both in Matlab)

Genetic Polyploid Phasing from Low-Depth Progeny Samples (code in Python)

README in case of code

In Git* repositories, you are asked (almost forced) to create a README.md when starting a software project.

“.md” stands for Markdown, a human readable text with basic formatting options

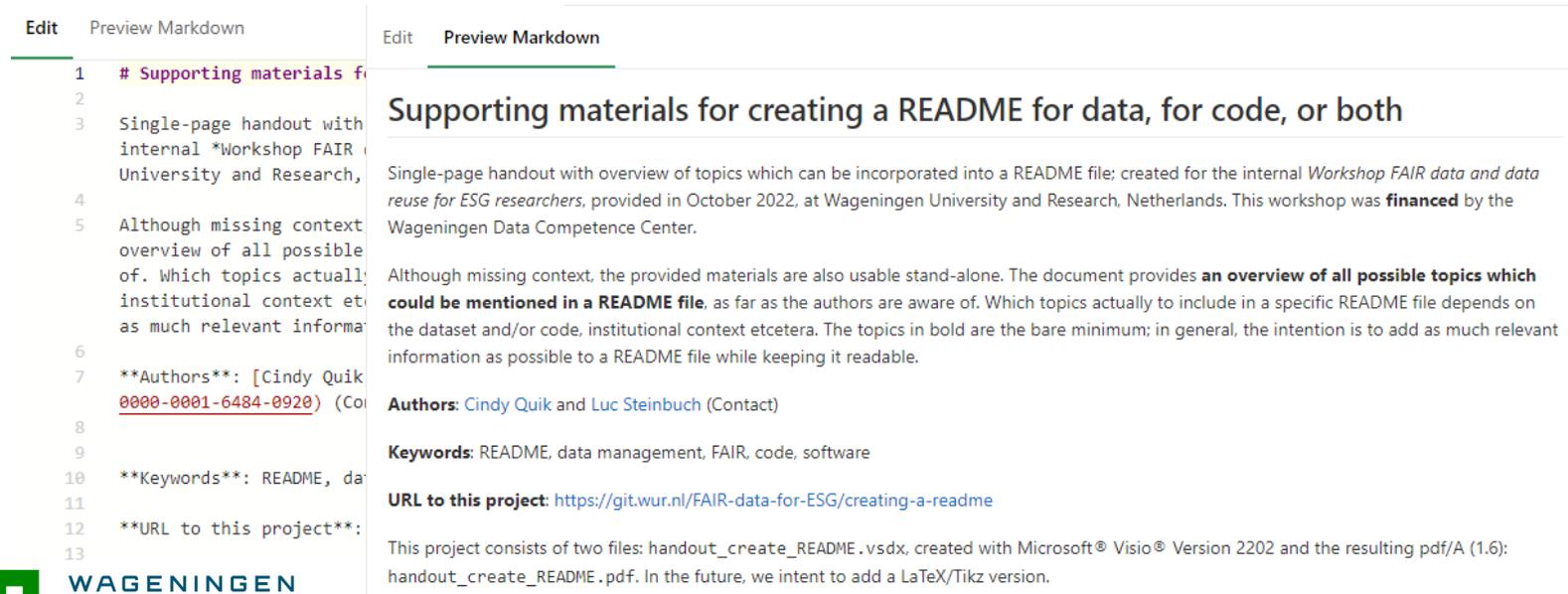
Edit Preview Markdown

```
1 # Supporting materials for creating a README for data, for code, or both
2
3 Single-page handout with overview of topics which can be incorporated into a README file; created for the
  internal *Workshop FAIR data and data reuse for ESG researchers*, provided in October 2022, at Wageningen
  University and Research, Netherlands. This workshop was financed by the Wageningen Data Competence Center
4
5 Although missing context, the provided materials are also usable stand-alone. The document provides an
  overview of all possible topics which could be mentioned in a README file, as far as the authors are aware
  of. Which topics actually to include in a specific README file depends on the dataset and/or code,
  institutional context etcetera. The topics in bold are the bare minimum; in general, the intention is to add
  as much relevant information as possible to a README file while keeping it readable.
6
7 Authors: [Cindy Quik](https://orcid.org/0000-0002-7112-0195) and [Luc Steinbuch](https://orcid.org/0000-0001-6484-0920) (Contact)
```

README in case of code

In Git* repositories, you are asked (almost forced) to create a README.md when starting a software project.

“.md” stands for Markdown, a human readable text with basic formatting options



The image shows a side-by-side comparison of a README.md file. On the left is the raw Markdown source code, and on the right is the rendered HTML preview.

Source Code (Left):

```
1 # Supporting materials for
2
3 Single-page handout with
  internal *Workshop FAIR
  University and Research,
4
5 Although missing context
  overview of all possible
  of. Which topics actually
  institutional context etc
  as much relevant informa
6
7 **Authors**: [Cindy Quik
  0000-0001-6484-0920] (Co
8
9
10 **Keywords**: README, da
11
12 **URL to this project**:
```

Rendered Preview (Right):

Supporting materials for creating a README for data, for code, or both

Single-page handout with overview of topics which can be incorporated into a README file; created for the internal *Workshop FAIR data and data reuse for ESG researchers*, provided in October 2022, at Wageningen University and Research, Netherlands. This workshop was **financed** by the Wageningen Data Competence Center.

Although missing context, the provided materials are also usable stand-alone. The document provides **an overview of all possible topics which could be mentioned in a README file**, as far as the authors are aware of. Which topics actually to include in a specific README file depends on the dataset and/or code, institutional context etcetera. The topics in bold are the bare minimum; in general, the intention is to add as much relevant information as possible to a README file while keeping it readable.

Authors: [Cindy Quik](#) and [Luc Steinbuch](#) (Contact)

Keywords: README, data management, FAIR, code, software

URL to this project: <https://git.wur.nl/FAIR-data-for-ESG/creating-a-readme>

This project consists of two files: `handout_create_README.vsdX`, created with Microsoft® Visio® Version 2202 and the resulting pdf/A (1.6): `handout_create_README.pdf`. In the future, we intent to add a LaTeX/Tikz version.

Possible topics

See handout

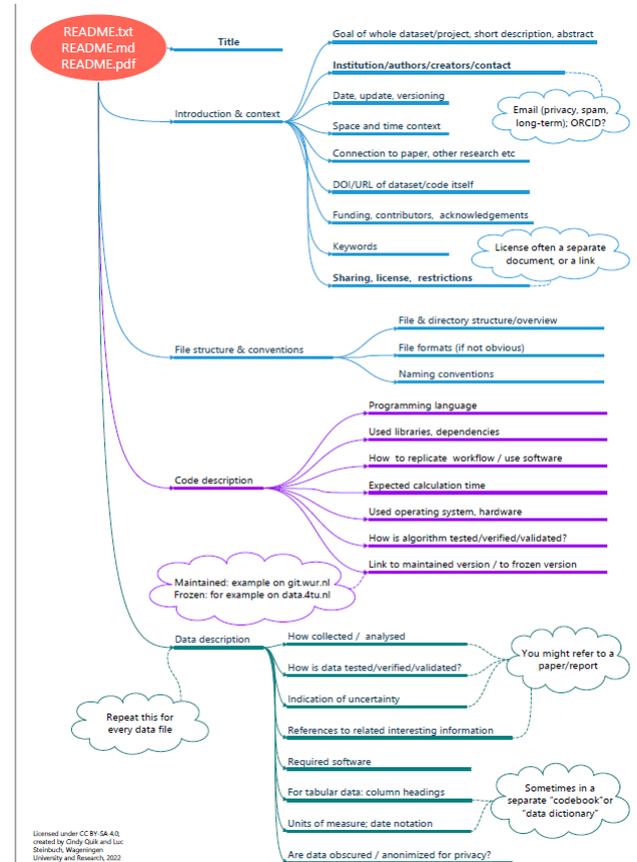
(Go to <https://git.wur.nl/FAIR-data-for-ESG/creating-a-readme>)



edu.nl/8n4f6

.. and then:

Name	Las
README.md	Up
handout_create_README_22_10_...	Up
handout_create_README_22_10_...	Up



README example

How FAIR is this README?

This folder contains the processed data underlying the publication in review in Royal Society Interface entitled "Rapid change of friction causes the illusion of touching a receding surface".

In the human factor experiment, we used a randomized two-alternative forced choice (2-AFC) protocol where participants had to discriminate whether one of the two stimuli moved similarly to a mechanical detent found in computer keyboards. The reference stimulus had a medium friction level, maintained throughout the entire press. This medium friction level was obtained for a vibration amplitude of +/- 1.5 μm . The comparison stimulus was a step change of vibration amplitude that started at +/- 1.5 μm .

The signal was constructed by choosing between falling friction (where the amplitude was going from small to high) and rising friction followed by choosing one of the three levels of friction changes (low, medium, high), that correspond to an amplitude change of: 1, 2 and 3 μm . The change of friction was triggered by measuring in real-time the force amplitude, and if the force reached the threshold of 0.7 N, the controller would change the ultrasonic amplitude.

Each matlab file "S_.mat" contains the processed data of one participant from S1 to S12.

- the data "proba" is a 1x6 vector gathering the probabilities of correctly identifying the comparison stimuli as the detent. The probabilities are computed for the respective stimuli -3, -2, -1, 1, 2, 3 μm .

for the first 4 participants:

"S_.mat" contains also ten 60x2 cell arrays of processed data extracted from their finger skin images:

- brightnessAll = sum of pixels of the contact image (red illumination)
- contactAreaAll = real contact area computed with the thresholding method
- dispTotalAll = mean global displacement of all the points
- forceNAll = normal force interpolated on image time
- forceTAll = tangential force interpolated on image time
- stickRatio = stick ratio along time

Each line corresponds to one trial.

The raw data include images of the participants' fingerprints, which are sensitive, so they will be available only upon request.

Your README

- For one of your projects
- A template which fits your future work (from scratch, or based on existing material)

Questions, remarks?

(Lunch till 12:45)