ReadMe to

WORKSHOP

FAIR DATA AND DATA REUSE

FOR ENVIRONMENTAL SCIENCE GROUP RESEARCHERS

A project funded by the

WUR Data Science & Artificial Intelligence Fellowship Program 2022

TABLE OF CONTENTS

GENERAL INFORMATION

DATASET DOI (SOURCEFILE FOR THIS README)

https://doi.org/10.4121/21399975

BRIEF DESCRIPTION OF THIS DATASET

We designed and organized a one-day workshop, where in the context of FAIR the following themes were discussed and practiced: scientific transparency and reproducibility; how to write a README; data and code licenses; spatial data; programming code; examples of published datasets; data reuse; and discipline and motivation. The intended audience were researchers at the Environmental Science Group of Wageningen University and Research. All workshop materials were designed with further development and reuse in mind and are shared through this dataset.

AUTHORS

Cindy Quik*
Soil Geography and Landscape Group, Wageningen University & Research, Wageningen, the Netherlands
https://orcid.org/0000-0002-7112-0195

Luc Steinbuch*
Soil Geography and Landscape Group, Wageningen University & Research, Wageningen, the Netherlands
https://orcid.org/0000-0001-6484-0920

* These authors contributed equally to the materials included in this dataset, and can be contacted for further questions.

INTENDED USAGE AND DISCLAIMER

Note that the methods of this workshop included interaction and sharing experiences. Therefore, the workshop materials might be less suitable for self-learning. These workshop materials are intended as base for a similar workshop, provided by researchers to fellow-researchers where experiences and expectations of both presenters and participants are shared. We encourage adjusting the learning materials to the actual context. Of course, also single parts of the workshop materials might be reusable in other situations.

Even though all included materials were developed with the utmost care, they might contain errors. In case advice for particular research projects is needed, we urge you to contact specialists from your university library.

TERMS OF USE AND CITING

- Quik C, Steinbuch L. 2022. Materials from: Workshop FAIR data and data reuse for Environmental Science Group researchers. 4TU.Centre for Research Data. DOI: 10.4121/21399975.

FUNDING

This project was funded by the WUR Data Science & Artificial Intelligence Fellowship Program 2022. The lunch during the first edition of the workshop on October 18th, 2022 was funded by the graduate school WIMEK. Both funders are gratefully acknowledged.

ACKNOWLEDGEMENTS

We would like to thank Maarten Storm for his contribution before and during the workshop, and Jakob Wallinga, Theo Jetten, Johan Feenstra, Irene Verhagen, and Shauna Ni-Fhlaithearta for their support in the development of this workshop. We thank Sylvie van der Poll for the pleasant communication with WDCC.

OVERVIEW WORKSHOP PROGRAM

TIME SCHEDULE

9:00 Door open
9:30 Start morning program: Modules 1 – 3 *
12:00 Lunch break
13:15 Start afternoon program: Modules 4 – 7 *
16:30 Wrap-up
17:00 Closure

* 10-min breaks between each Module

WORKSHOP CONTENTS

Module	Topic
Module 1:	A) Introduction to Workshop 'FAIR data and data reuse for ESG researchers' B) Transparency and reproducibility in environmental sciences
Module 2:	FAIR in more detail
Module 3:	How to write a README
Module 4:	Working with spatial data, working with code (domain-specific attention points)
Module 5:	Examples of how to publish datasets/code
Module 6:	A) The other way around: data reuse B) Discipline and motivation
Module 7:	A) Questions, discussion of cases from own work, where to find further information B) Workshop evaluation

OVERVIEW OF WORKSHOP MODULES

MODULE 1

Topic(s):

A) Introduction to Workshop 'FAIR data and data reuse for ESG researchers'
B) Transparency and reproducibility in environmental sciences

Objectives:

- Inform about the significance of data management (reproducibility/transparency of research)
- Introduce the concept of FAIR
- Inventory norms, values, knowledge and experience of participants

Methods and materials:

→ PowerPoint presentation
→ Group discussions based on propositions

MODULE 2

Topic(s):

FAIR in more detail

Objectives:

- Provide more information about the concept of FAIR
- Provide practical suggestions to help participants move from concepts to practice (covering licenses, metadata, file formats).

Methods and materials:

→ PowerPoint presentation
→ Hand-out 'Data sharing WUR'
→ Hand-out 'Overview of CC licenses'
→ Hand-out 'Considerations for working towards a FAIR dataset
→ Hand-out 'Checklist for finalizing a FAIR dataset'

MODULE 3

Topic(s):

How to write a README

Objectives:

- Provide practical suggestions to help participants move from concepts to practice (covering metadata, dataset title, documentation, and the difference between README's for data and for code).

Methods and materials:

→ PowerPoint Presentation
→ Hand-out 'Possible topics README'
→ Additional information for README
→ README general template
→ Small hands-on assignment
→ Guided discussion

MODULE 4

Topic(s):
Working with spatial data, working with code (domain-specific attention points)

Objectives:

→ Provide information about and suggestions to handle domain-specific attention points, addressing storage and documentation of large spatial datasets, privacy issues regarding spatial data, and FAIR practices regarding programming code.

Methods and materials:

→ PowerPoint presentation, partly by ESG Lead Data Steward (Maarten Storm)

MODULE 5

Topic(s):
Examples of how to publish datasets/code

Objectives:

- Showcasing examples that participants probably have affinity with, to make FAIR more accessible as a concept and to lower the threshold for starting publishing FAIR datasets;
- Providing examples of both datasets and code, highlighting points of attention for each;
- Providing examples that illustrate different types of access within different repositories.

Methods and materials:

→ PowerPoint presentation

MODULE 6

Topic(s):

A) The other way around: data reuse

B) Discipline and motivation

Objectives:

For submodule A:

- Showcase issues and points of attention for data reuse;
- Highlight approaches for responsible data reuse.

For submodule B:

- How to motivate yourself

Methods and materials:

For submodule A:

→ PowerPoint presentation

→ Hand-out 'Tips & tricks for data reuse'

For submodule B:

→ PowerPoint presentation

→ Hand-out 'Motivation'

→ Small assignment

→ Guided discussion

MODULE 7

Topic(s):

A) Questions, discussion of cases from own work, where to find further information

B) Workshop evaluation

Objectives:

For submodule A:

- Answer remaining questions of participants where possible;
- Offer assistance with challenges encountered during participants' own work;
- Offer points of reference for further information.

For submodule B:

- Evaluate the workshop and the workshop materials.

Methods and materials:

For submodule A:

For submodule A:
For submodule A:
For submodule A:
For submodule A:

- → For submodule A:Guided discussion
- → Hand-out ‘Further Info’

For submodule B:

- → Time reserved to fill in a hardcopy course evaluation
- → Guided discussion

WORKSHOP MATERIALS

For long-term storage and readability we provide a .pdf of all materials; for easy reuse we also provide the source documents in the native format of Microsoft® Word for Microsoft 365 MSO (Version 2202 Build 16.0.14931.20648) 32 bits; Microsoft® PowerPoint® for Microsoft 365 MSO (Version 2202 Build 16.0.14931.20648) 32 bits; and Microsoft® Visio® MSO (Version 2202 Build 16.0.14931.20648) 32 bits. There is one .txt file, for which the above obviously does not hold.

The filesystem is organised as follows:
Module <1..7>/<filename>, with below the overview of all workshop materials (included in the .zip file to which this ReadMe belongs):

Module	File	Editable version
1	Presentation_Introduction_and_Module1.pptx	Yes
	Presentation_Introduction_and_Module1.pdf	No
2	Presentation_Module2.pptx	Yes
	Presentation_Module2.pdf	No
	Data sharing WUR.pdf	No
	Overview of CC licenses.pdf	No
	Considerations for working towards a FAIR dataset.docx	Yes
	Considerations for working towards a FAIR dataset.pdf	No
	Checklist for finalizing a FAIR dataset. docx	Yes
	Checklist for finalizing a FAIR dataset.pdf	No
3	Presentation_Module3_HowToWriteAReadme.pptx	Yes
	Presentation_Module3_HowToWriteAReadme.pdf	No
	Transcript_Module3.docx	Yes
	Transcript_Module3.pdf	No
	Possible_topics_README.vsdx	Yes
	Possible_topics_README.pdf	No
	Additional_information_for_README.docx	Yes
	Additional_information_for_README.pdf	No
	README_template.txt	Yes
4	Presentation_Module4_DomainSpecificAttentionPoints.pptx	Yes
	Presentation_Module4_DomainSpecificAttentionPoints.pdf	No
	Transcript_Module4.docx	Yes
	Transcript_Module4.pdf	No
5	Presentation_Module5.pptx	Yes
	Presentation_Module5.pdf	No

6	Presentation_Module6_Reuse.pptx	Yes
	Presentation_Module6_Reuse.pdf	No
	Tips tricks for data reuse.docx	Yes
	Tips tricks for data reuse.pdf	No
	Presentation_Module6_Motivation.pptx	Yes
	Presentation_Module6_Motivation.pdf	No
	Motivation.pptx	Yes
	Motivation.pdf	No
	Transcript_Module6_motivation.docx	Yes
	Transcript_Module6_motivation.pdf	No
7	Further Info.docx	Yes
	Further Info.pdf	No

RECOMMENDATIONS

IMPROVEMENTS

Based on questions that were raised during the workshop and suggestions included in the evaluation forms, several modifications could help to improve potential future editions of the workshop. We have divided ideas for improvement in three categories below, distinguishing content additions, activities and interaction, and miscellaneous. Please note that in the text below "ESG" means: Environmental Science Group of Wageningen University and Research.

Content additions:

- In Module 2, information on student supervision in relation to copyright and FAIR could be added. This is probably relevant for the majority of participants, as many supervise thesis or internship students.
- In Module 3, an ESG/WUR ReadMe template could be useful in addition to the ReadMe handout. Also, some examples of well-prepared ReadMe files could be added to the course materials (the latter was already done after the workshop and shared with the participants). Also explanation of "ORCID" could be added.
- In Module 4 or 5, more examples on FAIR in code writing could be included.
- In Module 5, where examples of FAIR datasets and code are presented, some information and examples of data management plans could be added as well (in a way of 'Begin with the end in mind').

Activities and interaction:

- In the evaluation form one participant replied the question 'Which part of the workshop did you like best?' with "The interaction during the whole course." In addition, one participant ended the evaluation form with "Great structure, with the breaks and all: I've never felt so fresh after a day of workshop." We were pleased with these comments. However, we noticed that after lunch energy levels dropped temporarily, and Module 4 was a slightly longer module. We think the afternoon structure of the workshop could be improved by adding a refresher after lunch (perhaps a group assignment that involves physical movement such as the group discussion of propositions in Module 1).
- One participant suggested to "[Add] a practical assignment in the afternoon." This would be indeed helpful to break up the theory of the afternoon program, and could perhaps be combined with the point above.
- Another suggestion that was included on one of the evaluation forms was "Let participants prepare own cases with questions." We had already tried to include this, but this yielded no response. We requested each participant a week before the workshop to share a case and questions with us, and reserved time for discussion of cases from participants' own work in Module 7. Unfortunately, we did not receive any cases beforehand, and also during the workshop no cases were raised. One participant suggested on the evaluation form that "We can make case together and give some comments during workshop." Perhaps this idea has more potential, as it does not require participants to prepare a case beforehand. As an afternoon exercise participants could discuss a case in small groups, after which conclusions can be shared collectively.

Miscellaneous:

- One participant indicated that he/she/they found a whole-day workshop fine, but heard from colleagues that they considered it long and did not sign up for this reason. We can imagine that for

some it might be difficult to find the time for a whole-day workshop. However, we are convinced that reducing the workshop length to attract a bigger audience is detrimental to the workshop contents.

POTENTIAL FOR CONTINUATION AND ADAPTED VERSIONS

Of the 13 participants who filled out the evaluation form (3 participants had to leave early), all 13 would recommend the workshop to their colleagues.

A few quotes from the answers to the question 'Would you recommend the workshop to colleagues?':

- "ESG should provide more workshops like this one to employees and researchers."
- "Yes definitely. I cannot say enough how useful this was, especially for new PhDs who often don't have these skills. This workshop should 100% continue!"
- "100%."

The examples that we included in the workshop were all part of ESG research or had affinity with ESG's research domain. In addition, Module 4 covered some domain-specific attention points. Participant feedback indicates that the workshop was considered very relevant for ESG but also has the potential to be adapted rather easily to editions for WUR's other departments:

- "This workshop was very useful for all type of data use in ESG research. But I think it can be easily adapted to other groups."
- According to one participant from outside ESG: "As a plant scientist, everything was applicable except for the 'geographic data' bit."

Based on the received feedback and overall very positive response to the workshop, we can recommend future editions of this workshop for ESG and adapted versions for other research departments.

A few participants came in their role as data steward. One of the participating data stewards stated on the evaluation form that the workshop would be useful for all data stewards. A version of the workshop directed especially at data stewards is possible, for which we would advise some slight changes. In this case we would recommend to provide more self-explaining materials, which can be dispersed by data stewards to colleagues in their team or chair group and used specifically when data stewards assist their colleagues. In addition, we would include an interactive discussion about the type of questions and challenges data stewards encounter.