

Workshop FAIR data and data reuse for ESG researchers

October 18, 2022, by Cindy Quik and Luc Steinbuch



Who are we?

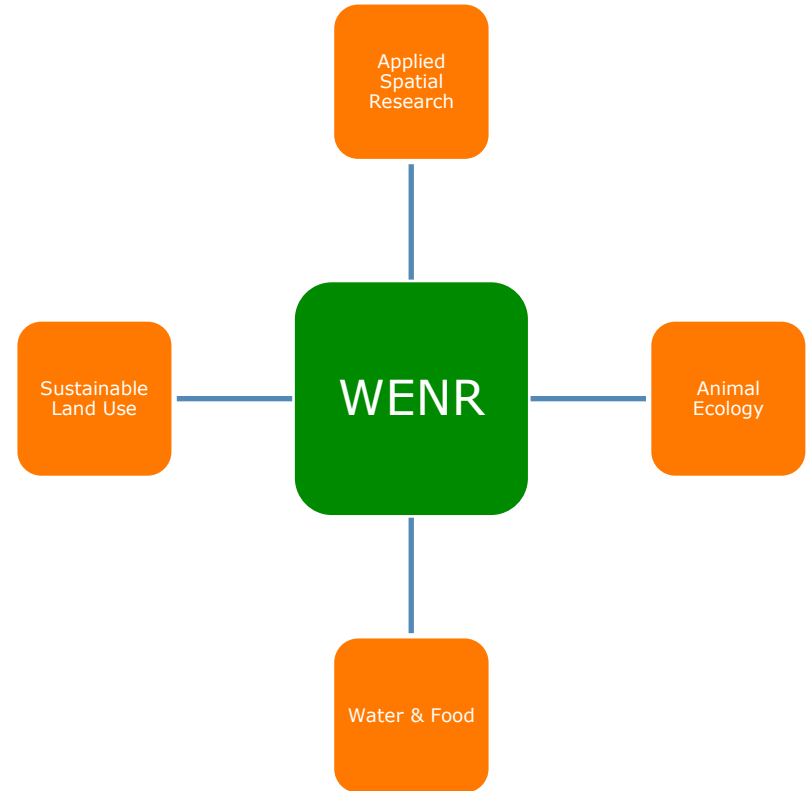
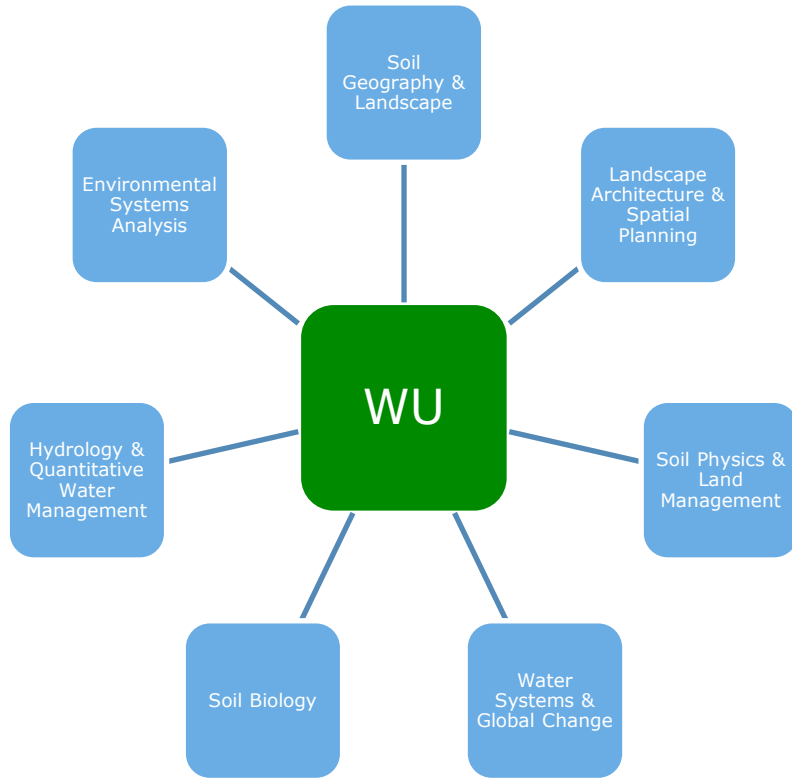
- Luc Steinbuch

- Data steward and researcher at Soil Geography & Landscape (ESG)

- Cindy Quik

- PhD candidate at Soil Geography & Landscape (ESG)

Welcome!



Why this workshop?

- UNESCO's Recommendation for Open Science:

To make scientific knowledge freely available to everyone, accessible and reusable and to promote scientific cooperation and the exchange of information for the benefit of science and society

- The Dutch National Programme Open Science endorses the UNESCO goals
- WUR embraced Open Science by establishing the Open Science and Education (OSE) Programme in July 2019

WUR's Open Science and Education Programme



FAIR data

“Our research data are as open and reusable as possible”

WUR's aim is to make all data FAIR by 2025⁽¹⁾

Challenges

- Research@WUR:
Datasets (FAIR/non-FAIR) comprise just over 1% of all research output at ESG
- Many questions on how-to, often domain-specific challenges

Development of this workshop

- WUR's Data Science & Artificial Intelligence Fellowship Program 2022
- Additional funding from WIMEK for lunch
- Support from consortium:
 - Jakob Wallinga (Chairholder SGL)
 - Maarten Storm (Lead Data Steward ESG)
 - Irene Verhagen (Data Management Support, WUR Library)
 - Theo Jetten (PE&RC Secretary + Information Specialist WUR Library)
 - Johan Feenstra (WIMEK Secretary)

Workshop aims

- Help creating awareness
- Increase knowledge of/experience with 'best practices' regarding FAIR data publishing and data reuse
- Focus on ESG's research domain

→ Contribute to increased implementation among ESG researchers

Program

9:30	Start morning program: Modules 1 – 3 *
12:00	Lunch break
12:45	Start afternoon program: Modules 4 – 7 *
16:30	Wrap-up

* 10-min breaks between each Module

Program

Module 1:	Introduction to Workshop 'FAIR data and data reuse for ESG researchers' 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:	The other way around: data reuse Discipline and motivation
Module 7:	Questions + discuss cases from own work

Transparency and reproducibility in environmental sciences

Workshop FAIR data and data reuse for ESG researchers – Module 1

October 18, 2022, by Cindy Quik and Luc Steinbuch



Concepts of the scientific method

- Transparency
- Reproducibility

Transparency

operating in such a way that it is easy for others to see what actions are performed

Wikipedia, accessed 19-09-2022

the quality of being done in an open way without secrets

Cambridge Dictionary, accessed 19-09-2022

Reproducibility

able to be shown, done, or made again

Cambridge Dictionary, accessed 2022

*agreement between the results of measurements
[...] carried out with the same methodology*

Wikipedia, accessed 28-01-2019

Why

- Trustworthiness of science (prevention of fraud)
- To increase the value of your research efforts
 - Poorly documented science → gets lost on dusty (digital) bookshelves
 - Well documented science → reused
- Reuse:
 - Meta-analyses
 - To support future generations of scientists

From concepts to practice

- Other people (either everybody or selected groups) need:
 - Access to full methodological details
 - Access to data
 - Access to code that generates/processes data
- This requires solid data management from start to end of a project

From concepts to practice: FAIR

- Enabling FAIR data:

Findable, Accessible, Interoperable, Reusable

- F* Dataset with own DOI
- A* Long-term storage of data (open access or restricted access)
- I* Preferred file formats
- R* Copyright & data licenses

Let's discuss

If you tend
to **agree**
stand left



If you tend
to **disagree**
stand right

- 1) Publishing open access is only for fellow scientists, not for society.
- 2) Publishing data is not necessary, people can always ask for it later.
- 3) Publishing data should be obligatory.
- 4) The option of adding a supplement to a paper makes data repositories superfluous.

If you tend
to **agree**
stand left



If you tend
to **disagree**
stand right

- 1) Scripts/code should always be published (either internally or externally).
- 2) Scripts/code should be reviewed for errors and readability.
- 3) The use of open source software is preferable.
- 4) Any published dataset/code should count as true output (for a PhD, tenure track, H-index, other types of evaluation).

If you tend
to **agree**
stand left



If you tend
to **disagree**
stand right

- 1) Publishing datasets/code is only feasible if the funder requires it, otherwise it cannot be given priority.
- 2) The main challenge for publishing datasets is time allocation/financial constraints.
- 3) The main challenge for publishing datasets is limited knowledge and/or lack of suitable tools/facilities.

If you tend
to **agree**
stand left



If you tend
to **disagree**
stand right

Let's discuss

If applicable to
you stand left



If not applicable to
to you stand right

- 1) I keep a research journal/project journal.
- 2) I use a file naming system for management of previous versions.
- 3) I have published Open Access and am familiar with CC licenses.
- 4) I have published a dataset with ReadMe before (either published internally or externally).
- 5) I have published data in a data repository.
- 6) I reuse existing data or code.





Please be back
at 10:30