Collaborative Document day 1: https://edu.nl/qvm9f

2025-11-05-ds-rss-Collaborative Document Day 1

Welcome to the Collaborative Document of the Research Software Support Training..

This Document is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

This is the Document for today: https://edu.nl/qvm9f

Collaborative Document day 2: https://edu.nl/yyqbr

Teaching materials (Research Software Support Platform): https://edu.nl/wptbh

Workshop website: link (https://esciencecenter-digital-skills.github.io/2025-11-05-ds-rss-masterclass/)

Code of Conduct

Participants are expected to follow these guidelines:

- Use welcoming and inclusive language.
- Be respectful of different viewpoints and experiences.
- Gracefully accept constructive criticism.
- Focus on what is best for the community.
- Show courtesy and respect towards other community members.

If you feel that the code of conduct is breached, please talk to one of the instructors (if the complaint is for one of the participants) or send an email to training@esciencecenter.nl (if the complaint is for one of the instructors).

⚖ License

Getting help

To ask a question, just raise your hand.

If you need help from a helper, place a pink post-it note on your laptop lid. A helper will come to assist you as soon as possible.

Instructors

Fenne Riemslagh, Jaro Camphuijsen, Ole Mussmann, Ewan Cahen

□□□□□□□□□□♂□ Roll Call

Name/ pronouns (optional) / job, role / social media (twitter, github, ...) / background or interests (optional) / city

-

□ Agenda

Time	Topic
9:30	Welcome and icebreaker
10:00	Research Software
11:00	**Break**
11:15	Software in the Research Landscape
12:15	**Lunch**
13:15	Landscape exercise
14:15	**Break**
14:30	Software Management Plans
15:30	**Break**
15:45	Introduce homework exercises
16:15	Wrap-up
16:30	**End**

Location logistics

- Coffee is in the kitchen and toilets are in the hallway, just outside of the classroom.
- As you probably already noticed we have an improvised barrier in the hallway. This is because the glass door of the neighboring room is broken and tiny pieces of glass are coming off. Please don't cross the barrier.
- If you leave the building, be sure to be accompanied by someone from the escience center to let you back in through the groundfloor door
- For access to this floor you might need to ring the doorbell so someone can let you in
- In case of an emergency, you can exit our floor using the main staircase. Or follow green light signs at the ceiling to the emergency staircase.
- **Wifi**: Eduroam should work. Otherwise use the 'matrixbuilding' network, password should be printed out and available somewhere in the room.

Certificate of attendance

If you attend the full workshop you can request a certificate of attendance by emailing to training@esciencecenter.nl.
Please request your certificate within 8 months after the workshop, as we will delete all

personal identifyable information after this period.

Icebreaker

What is something you are proud of that you made or something that has recently brought you joy?

Exercises

What is (Research) Software?

Is it research software or software used in research?

These answers are not absolute but based on general consensus.

- MacOS is software used in research.
- Microsoft Excel is software used in research.
- R is
    - research software because it was made for data analysis
    - software used in research because it is just a programming language
- SciPy is research software (but could be considered software used in research because it is just a library). An important difference is that Excel is made and funded by Microsoft, but ScyPi is made by scientists and everyone can contribute.
- AlphaFold (an AI system) is research software because it is designed to answer research questions (but could be considered software used in research because reproducibility is an issue).
- A specific analysis (Jupyter) notebook is research software.
- ChatGPT is software used in research (but some people do research on creating these LLMs).

Software people use on a regular basis

- SciPy (https://scipy.org/)
- R (https://www.r-project.org/)
- NetworkX (https://networkx.org/)
- Bash (https://www.w3schools.com/bash/)
- Python (https://www.python.org/)
- Open Global Glacier Model (https://oggm.org/)
- INLA (https://www.r-inla.org/)
- Git (https://git-scm.com/)
- LISFLOOD model (https://web.jrc.ec.europa.eu/policy-model-inventory/explore/models/model-lisflood/)
- TypeScript (https://www.typescriptlang.org/)
- pec calculator (https://www.efsa.europa.eu/en/supporting/pub/en-8967)
- Markdown (https://daringfireball.net/projects/markdown/)

The Research Software Landscape

Exercise: Study and present (20 - 30 minutes for 2-5 groups)

Link (https://esciencecenter-digital-skills.github.io/research-software-support/modules/softwarelandscape/exercise1)

Group 1 - The one

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
    - Software heritage
2. What is the core purpose or central thought behind your chosen concept?
    - Collect, preserve and share all software. Software is a central part of our lives and it is fragile. Why? Heritage, important part of human production. Science, reproducibility. A fragile reflection of the human condition.
3. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
    - Globally active. In collaboration with UNESCO, founded by Inria, so tied to the french goverment.
4. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?
    - Started in 2016 but it's archives history. Still an ongoing process.
5. How can the scientific community benefit from this concept?
    - Reproducability, ensure aivlability and tracability. Also if we go to space.
6. Are there other concepts related to your chosen concept?
    - Wayback machine.
7. What else can you tell? Spicy details, fun facts?
    - There is some controversy about not being able to opt out from being scraped. They have an identifier that allows you to reference a single file and even a single line of code.

Group 2 : Dutch Reproducibility Network

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
    - National consortium, aimed at quality and efficiency by strenghtening reproducibility and transparancy of scholary disciplines.
2. What is the core purpose or central thought behind your chosen concept?
    - The NLRN facilitates initiatives that foster reproducible and transparent research in all scholarly disciplines in the Netherlands. Our two main goals are to 1) promote the large-scale implementation of transparent and reproducible workflows and 2) assist the exchange and further development of innovations in research on reproducibility.

3. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
    - This is a national initiative.
4. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?
    - This is a currently active concept, funded by NWO in 2023, with the project duration 2023-2026. This is an addtion to the Dutch of Open Science Communities (OSC-NL), which was established in 2018.
5. How can the scientific community benefit from this concept?
    - The network promotes best practices when it comes to reporducible reseearch. It aims to teach both the way to make research (more) reproducible, but also to show how reproducible research benefits the community and the single researcher.
6. Are there other concepts related to your chosen concept?
    - FAIR principles, which also emphasis on the reproducibility of data and software.
7. What else can you tell? Spicy details, fun facts?
    - Part of their workshop is to reproduce a LEGO scupture from instructions. The website crashed after more people in the course connected to it.

Group 3; to reproduce group 2

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
    - Netherlands Reproducibility Network (NLRN)
2. What is the core purpose or central thought behind your chosen concept?
    - increase the quality and efficiency of research in the Netherlands by coordinating, supporting and strengthening initiatives on reproducibility and transparency in all scholarly disciplines.
3. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
    - Mostly Dutch, also funded by NWO also related to OSNL
4. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?
    - in the timeline we had internet connections, so it was slow
5. How can the scientific community benefit from this concept?
    - materials and metadata
6. Are there other concepts related to your chosen concept?

7. What else can you tell? Spicy details, fun facts?

Exercise: Mapping the FAIR principles

Link (https://esciencecenter-digital-skills.github.io/research-software-support/modules/softwarelandscape/exercise2)

Group 1 : Comet

Comet is a command-line tool and desktop application for tandem mass spectrometry sequence database search6.

- Comet is licensed under the Apache 2.0 open source licence. R1.1
- Comet is registered in the bio.tools catalogue of bioinformatics tools, where it has a globally unique and persistent identifier, and rich metadata that includes the identifier and is searchable and indexable. F1-F4
- The metadata in bio.tools is independent from the Comet repository, and will stay accessible should the software itself become inaccessible. A2
- The publicly accessible project repository on GitHub includes detailed information about the development of Comet.R1.2
- The code includes dependencies to external software packages, such as Thermo Scientific's MSFileReader library. I2 since machines need to be able to run this R2 because humans need to know what to do.
- Comet can be downloaded via the browser following the links provided in the metadata using https. A1.1
- Comet uses standard data types from the proteomics domain for its input and output data that are documented in the metadata as functional annotations. I1 and I2

Group 2 Pure

[FAR] PureGoMe can be downloaded from the project repository, while metadata is accessible independently from the registry.

[R1.2]PuReGoMe's GitHub repository has detailed records of the development history.

[I2, R2]The code includes dependencies to other software, such as various Python libraries.

[F1.2, A1] PuReGoMe has a (versioned) DOI from Zenodo.

[R1.1] PuReGoMe uses the Apache 2.0 open source licence.

[I1] PuReGoMe uses standard file formats (e.g., CSV files) for data exchange.

[I2] PuReGoMe refers to other objects such as websites.

[F2-4, A2, R1] PureGoMe is registered in the Research Software Directory that captures the most relevant metadata, including the identifier, in searchable and indexable form.

from the paper
PuReGoMe is a project aimed at understanding Dutch public sentiment during the COVID-19 outbreak period by analysing real-time Twitter data7. It provides a collection of Python scripts and Jupyter notebooks for this purpose. PuReGoMe has a (versioned) DOI from Zenodo (F1) and is registered in the Research Software Directory that captures the most relevant metadata (F2), including the identifier (F3), in searchable and indexable form (F4). The software can be downloaded from the project repository (A1), while metadata is accessible independently from

the registry (A2). PuReGoMe uses standard file formats (e.g., CSV files) for data exchange (I1) and refers to other objects such as websites (I2). The project uses the Apache 2.0 open source licence, and the GitHub repository has detailed records of the development history (R1). The code includes dependencies to other software, such as various Python libraries (R2).

Group 3 : Shiny - The Group Previously Known as the One

Findable

F1. Software is assigned a globally unique and persistent identifier.

- *Yes.*

— F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.

- *No, but doesn't seem applicable.*

— F1.2. Different versions of the software are assigned distinct identifiers.

- *Different versions have different identifiers.*

F2. Software is described with rich metadata.

- *Yes. README.md (https://github.com/crp2a/gammaShiny/blob/v0.2.0/README.md), license, metadata json file.*

F3. Metadata clearly and explicitly include the identifier of the software they describe.

- *Yes.*

F4. Metadata are FAIR, searchable and indexable.

- *json file, so yes.*

Accessible

A1. Software is retrievable by its identifier using a standardized communications protocol.

- *https, so yes.*

— A1.1. The protocol is open, free, and universally implementable.

- *Yes.*

— A1.2. The protocol allows for an authentication and authorization procedure, where necessary.

- *N/A.*

A2. Metadata are accessible, even when the software is no longer available.

- *Zenodo. So within reason, yes.*

Interoperability

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

- *Difficult for us to ascertain, the nature article says yes in tiny letters.*

I2. Software includes qualified references to other objects.

- *Yes, the namespace file. But doesn't seem to include versions, but since its published on CRAN, probably.*

Reusable

R1. Software is described with a plurality of accurate and relevant attributes.

- *Seems to have usage examples.*

— R1.1. Software is given a clear and accessible license.

- *Yes.*

— R1.2. Software is associated with detailed provenance.

- *Yes. Some role descriptions*

R2. Software includes qualified references to other software.

- *Sort of, we can't find versions of that other software.*

R3. Software meets domain-relevant community standards.

- *We think so. Shiny makes it very easy to use.*

Software Management Plans (SMPs)

Exploring the SMP Questionnaire (15 minutes)

Link (https://esciencecenter-digital-skills.github.io/research-software-support/modules/softwaremanagementplans/smp_questionnaire)

What's in it for me? Convince your colleague on why an SMP is important (10 minutes)

Link (https://esciencecenter-digital-skills.github.io/research-software-support/modules/softwaremanagementplans/smp_importance)

Group 1: The Shiny Ones

What arguments could you collect *in favor* of SMPs?

- Some funders want a SMP! Like NWO.
- Also obligatory for some journals?
- Your Phd's can do it!
- There is training for it.
- Prevent technical debt.
- Explain what is needed to maintain it!
- Make sure your software stays up to date and relevant.

- Prevent waste of funding.
- Learn lessons for the future - what if you need to make a new version?

What arguments *against* its use?

- Already filled in DMP.
- No time.
-

Was the researcher convinced by the arguments?
Yes, by abusing phd students.

Group 2

What arguments could you collect *in favor* of SMPs?

- (Re)thinking my software/project
- Fill in quick, skip some parts
- find things I didn't not anticipate for
- quickly onboarding of PhD's

What arguments *against* its use?

- I allready have documentation, of some sort
- Time consuming
- Who will use this

Was the researcher convinced by the arguments?

Group 3

What arguments could you collect *in favor* of SMPs?

- to show off your important work
- it might be good to make it visible what you did (hinting to funding)
- it is really not that much work

What arguments *against* its use?

- it is not needed
- it will not be re-used
- I have no evidence that the replication package ever gets downloaded
- we already have a DMP, why need a SMP

Was the researcher convinced by the arguments?

Collaborative Notes

What is (Research) Software?

Software now vs software historically

- Old software was physically present on punch cards, modern software is stored electronically on a disk
- Software in the modern age is everywhere

- Modern software uses higher level abstractions, old software makes calls to memory directly

Does Docker guarantee that software will never break? Docker (https://www.docker.com/) does not guarantee that software will always run the same. Causes are that Docker itself is software that changes, Docker user the kernel of the host, which varies per computer. Dockerfiles are recipes, but the outcome might be different every time you build it. It might also depend on certain files being present and values of environment variables. Also have a look at Nix (https://nixos.org/guides/how-nix-works/).

The Research Software Landscape

What is the difference between an RSE and a researcher who writes research software? The distinction is not fully clear, the term RSE is growing. See e.g. "The Netherlands eScience Center Releases Job Profile and Role Description for Research Software Engineers" (https://www.esciencecenter.nl/news/the-netherlands-escience-center-releases-job-profile-and-role-description-for-research-software-engineers/), with those descriptions on Zenodo (https://zenodo.org/records/7994286).

How does FAIR improve the quality of your software? In short, it doesn't (that isn't the goal), but it makes it easier to use for others.

Software Management Plans (SMPs)

Would researchers have done all steps of the SMP by themselves anyway without writing it down? Probably not, making an SMP forces you to think about more things.

Projects have both data and software. Can a SMP and a Data Management Plan (DMP) be combined? Two forms might be too much. Yes, it would be for the best to have everything in one place, combining the two. NWO is working on this. You can also justifying having two forms by pointing out the two forms have two different scopes.

How do different institutions have different regulations with respect to SMPs? Licenses to use, tools to use.

Feedback

Resources

- Playlist Crash Course Computer Science (https://www.youtube.com/watch?v=tpIctyqH29Q&list=PL8dPuuaLjXtNlUrzyH5r6jN9ulIgZBpdo) (infotainment, but very informative)
- Docker (https://docs.docker.com/get-started/docker-overview/)
- Nix (https://nixos.org/guides/how-nix-works/)
- What is a DOI (Digital Object Identifier)? (https://www.doi.org/the-identifier/what-is-a-doi/)
- SciPy (https://scipy.org/)
- AlphaFold (https://deepmind.google/science/alphafold/): open source implementation (https://github.com/Ligo-Biosciences/AlphaFold3)
- Netherlands eScience Center job profiles and role descriptions for Research Software Engineers (https://zenodo.org/records/7805870)
- Professionalizing the role of Research Software Engineers in the Netherlands

(https://zenodo.org/records/15019998)
- Software Management Plan tool (https://SMP.research.software)
- Python template (https://github.com/NLeSC/python-template)

<!-- ## Note-taker's section -->

<!-- Exercice copy&paste -->

<!-- Module: Software in the Scientific Landscape

Mapping the Principles

Software 1: Comet

Software 2: PureGoMe

Software 3: gammaShiny

-->

<!-- Module: Software in the Scientific Landscape

Study and present

"Thing" 1

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
2. What is the core purpose or central thought behind your chosen concept?
3. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
4. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?
5. How can the scientific community benefit from this concept?
6. Are there other concepts related to your chosen concept?
7. What else can you tell? Spicy details, fun facts?

"Thing" 2

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
2. What is the core purpose or central thought behind your chosen concept?
3. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
4. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?
5. How can the scientific community benefit from this concept?
6. Are there other concepts related to your chosen concept?
7. What else can you tell? Spicy details, fun facts?

"Thing" 3

1. What kind of "thing" did you choose? (e.g. an organization, group of people, document, set of guidelines, etc.)
   The NL Reproducibility Network (NLRN)

2. What is the core purpose or central thought behind your chosen concept?
   to increase the quality and efficiency of research in the Netherlands by coordinating, supporting and strengthening initiatives on reproducibility and transparency in all scholarly disciplines.

1. Is your chosen concept globally active, bound to a country or somehow geographically located somewhere?
   Probably mostly Dutch, seeing the name and beeing funded by NWO

2. Describe the timeline of your chosen concept: any relevant history, its inception, any updates, is it still active?

3. How can the scientific community benefit from this concept?

4. Are there other concepts related to your chosen concept?

5. What else can you tell? Spicy details, fun facts?
   -->

<!-- Module: Software Management Plans (SMPs)

What's in it for me? Convince your colleague on why an SMP is important

Group 1

1. Which software did you choose?
2. What arguments could you collect in favor of SMPs?
3. What arguments against its use?
4. Was the researcher convinced by the arguments?

Group 2

1. Which software did you choose?
2. What arguments could you collect in favor of SMPs?
3. What arguments against its use?
4. Was the researcher convinced by the arguments?

Group 3

1. Which software did you choose?
2. What arguments could you collect in favor of SMPs?
3. What arguments against its use?
4. Was the researcher convinced by the arguments?
   -->

<!-- Module: Software Management Plans (SMPs)

SMP vs FAIR

- SMP
    - ...
- FAIR
    - ...
- SMP/FAIR
    - ...
- neither
    - ...
      -->

<!-- Module: Software Management Plans (SMPs)

Exploring the SMP Questionnaire

- Agree
    - ...
- Disagree
    - ...
      -->