

# How to make research accessible and reproducible in your institute

Jelle de Plaa, SRON data stewards

**SRON**

Netherlands Institute for Space Research

Picture: NASA

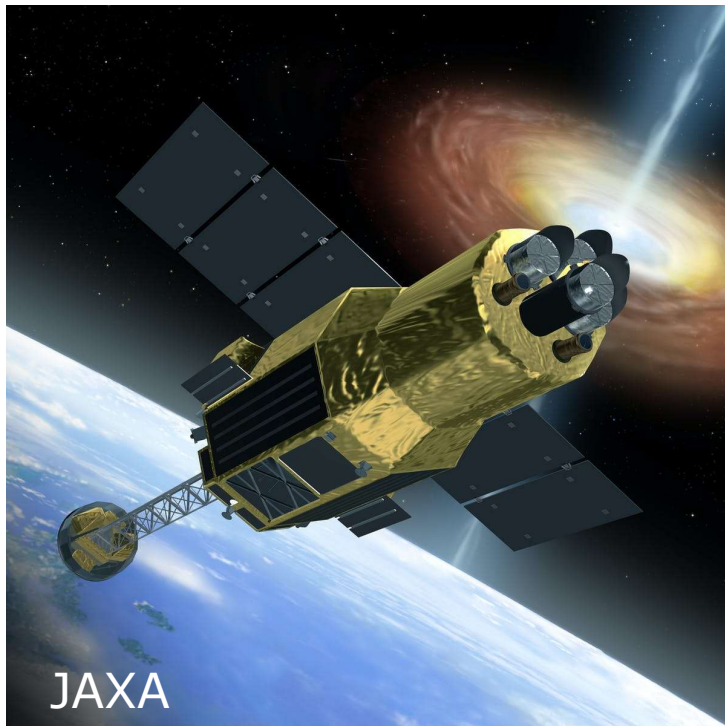
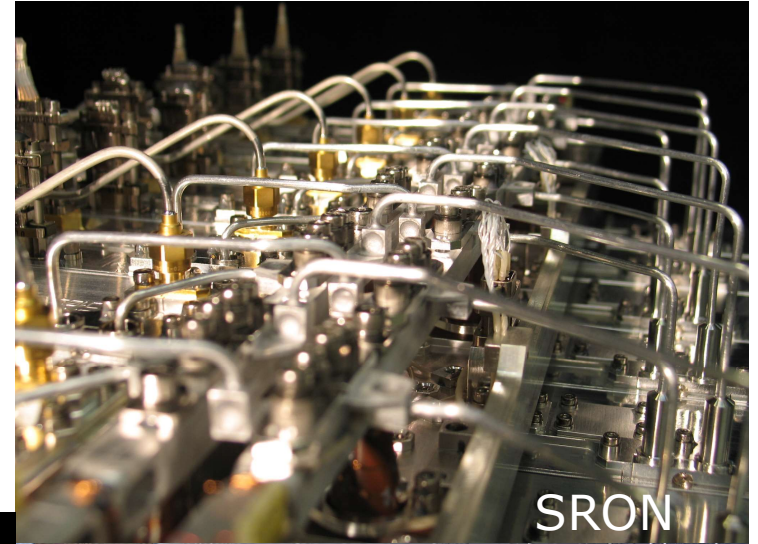
## In this talk

- The institute SRON
- Our open data policy
- How we help our scientists
- Lessons learned and recommendations



# SRON Netherlands Institute for Space Research

- Develop pioneering technology and advanced space instruments
- Use them to pursue fundamental astrophysical research, Earth science and exoplanetary research



# National ambitions regarding open science

- SRON is an NWO institute.  
NWO is the science foundation of the Netherlands
- NWO 2019-2022 strategy: implement open science together with Dutch universities, supported by government.
- SRON formulated its own open data policy in 2019.
- SRON policy in effect since 2020.



# The SRON data management plan

SRON has written its own data management plan.

In short, there are three types of data with responsibilities:

**Type A:** Satellite mission data and software. Space agency responsibility.

**Type B:** Data from other instruments/experiments. Project responsible.

**Type C:** Data connected to publications (derived data from type A and own software/scripts). Scientist responsible.

Type C data is most relevant for Astronomy

-> create reproduction package for each paper

## How we prepared our scientists

- Create a group of data stewards (helpdesk)
- Organize talks about open data policy for each department
- Create a reproduction package template and instructions
- Organize research data management workshops



## Main message

*Open data may seem like a burden.*

*However, by organizing your work you do not only help others, but also your (future) self.*

## How to create reproduction packages?

*What would a student need to reproduce my result?*

For instance:

- Links to the raw data (ESA/NASA)
- Scripts that derive products from raw data
- The derived products (images, spectra, etc.)
- Analysis/optimization scripts that derive results
- Result tables
- Scripts to re-create the figures
- For each of the above a README file explaining how to use it.



# A simple reproduction package template

Basic directory structure:

- data
- figures
  - figure 1
  - ...
- notebooks (optional)

Readme.md

**Data:** derived data products and analysis scripts

**Figures:** data and script(s) for creating figures

**Notebooks:** Jupyter notebooks (optional)

**Readme.md:**

In all directories, one adds a Readme.md file to explain what the folder contains and how to use it. (in Markdown format)

# A simple reproduction package template II

Top level Readme.md:

## Reproduction package for ...

---

### Software prerequisites

---

To run the scripts in this package, the following software was used:

- [XMM-Newton SAS v18](#)
- [Astropy v4.0](#)
- etc.

### Data prerequisites

---

### Package contents

---

Jupyter notebooks

Data products and results

Figures

---

# A simple reproduction package template

## pro/cons

### Advantages:

- High flexibility/freedom
- Any type of data/software fits in the template
- Built-in guidance what to put in repro package
- Readme.md file shows nicely on Zenodo
- Load directly from git

### Disadvantages:

- Not machine readable
- No built-in enforcement of FAIR requirements
- No built-in inclusion of software dependencies (but docker image could be included/not ideal)

## Public template

If you are interested:

- Feel free to fork and adapt the template to your needs.
- Suggestions for improvement are welcome:
  - Just create an issue on Github

The template has a CC0 license: free to do anything you want!

Link:

<https://github.com/jdeplaa/open-data-template>

# Template on Github/Zenodo

The screenshot shows the Zenodo interface for a repository titled "Simple open data template for researchers". The repository is dated June 4, 2021, and has 9 views and 0 downloads. It is available in GitHub and indexed in OpenAIRE. The repository is a zip file named "open-data-template-v0.1.zip" and contains a directory "jdeplaa-open-data-template-a260cee" with subdirectories for LICENSE, README.md, and a "template" folder. The "template" folder contains subfolders for "data", "figures", and "notebooks", each with its own README.md file.

zenodo Search Upload Communities Log in Sign up

June 4, 2021 Other Open Access

## Simple open data template for researchers

Jelle de Plaa

Other(s)

van Hees, R. M.; Shipman, R. F.

This open data template is a simple template aimed at researchers who want to create a reproduction package with their paper. It provides an easy 10-step guide to organize your research data files into a reproduction package that can be uploaded to a trusted data repository (Zenodo is chosen as default here). The template is simple on purpose to allow the researcher a lot of freedom and flexibility in their data organisation, while keeping a basic standard top-level structure.

This template is based on a template developed for [SRON Netherlands Institute for Space Research](#) and has a CC0 public domain license. Feel free to copy, modify, re-use, etc. the template for your own purpose. Comments, corrections and suggestions for improvement are welcome in our [Github issue tracker](#).

Preview

open-data-template-v0.1.zip

- jdeplaa-open-data-template-a260cee
  - LICENSE 7.0 kB
  - README.md 3.9 kB
  - template
    - README.md 900 Bytes
    - data
      - README.md 0 Bytes
    - figures
      - figure1
        - README.md 0 Bytes
    - notebooks
      - README.md 0 Bytes

Available in

GitHub

Indexed in

OpenAIRE

Publication date: June 4, 2021

DOI: DOI 10.5281/zenodo.4899847

Related identifiers: Supplement to <https://github.com/jdeplaa/open-data-template/tree/v0.1>

# Example reproduction package

The screenshot shows a Zenodo dataset page. At the top is a blue navigation bar with the Zenodo logo, a search bar, and links for 'Upload' and 'Communities'. On the right side of the bar are 'Log in' and 'Sign up' buttons. Below the navigation bar, the dataset is dated 'March 8, 2021' and has 'Dataset' and 'Open Access' labels. The main title is 'Reproduction Package to "Strong Reduction of Quasiparticle Fluctuations in a Superconductor due to Decoupling of the Quasiparticle Number and Lifetime"'. It shows 62 views and 23 downloads, with a link to 'See more details...'. The authors listed are de Rooij, Steven A. H.; Baselmans, Jochem J. A.; Murugesan, Vignesh; Thoen, David J.; de Visser, Pieter J. A short description follows, stating it's a reproduction package for a paper. Below this is a 'Preview' section with a dropdown arrow. The preview content includes the title 'Reproduction Package for: Strong Reduction of Quasiparticle Fluctuations in a Superconductor due to Decoupling of the Quasiparticle Number and Lifetime', the arXiv ID 'arXiv:2103.04777', a paragraph describing the package's purpose, contact information 's.a.h.de.rooij@sron.nl', and a 'Software prerequisites' section listing Python 3 and MATLAB. To the right of the preview are two boxes: 'Indexed in OpenAIRE' and a 'Publication date' box showing 'March 8, 2021', DOI '10.5281/zenodo.4590731', and keywords 'Generation-Recombination Noise', 'Superconducting Resonator', 'Quasiparticle Trapping', and 'Phonon Trapping'. Below these are 'Grants' from the European Commission and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, and 'Related identifiers'.

zenodo Search Upload Communities Log in Sign up

March 8, 2021 Dataset Open Access

## Reproduction Package to "Strong Reduction of Quasiparticle Fluctuations in a Superconductor due to Decoupling of the Quasiparticle Number and Lifetime"

62 views 23 downloads See more details...

de Rooij, Steven A. H.; Baselmans, Jochem J. A.; Murugesan, Vignesh; Thoen, David J.; de Visser, Pieter J.

This is a reproduction package to the paper "Strong Reduction of Quasiparticle Fluctuations in a Superconductor due to Decoupling of the Quasiparticle Number and Lifetime". It contains all data and code to reproduce the figures in this paper.

Preview

### Reproduction Package for:

*Strong Reduction of Quasiparticle Fluctuations in a Superconductor due to Decoupling of the Quasiparticle Number and Lifetime*

[arXiv:2103.04777](#)

This is a package of data and code to reproduce the results of the paper by S. A. H. de Rooij et al. (2021). The code in this package is also on [github](#). The python code is a scaled down fork of [this repository](#). This document will explain the content of the package and the steps needed to reproduce the results.

For any questions or remarks, contact: [s.a.h.de.rooij@sron.nl](mailto:s.a.h.de.rooij@sron.nl)

### Software prerequisites

- Python 3 environment with jupyter, numpy, scipy and matplotlib
- MATLAB (2020+) with curve fitting toolbox

Indexed in OpenAIRE

Publication date: March 8, 2021  
DOI: [10.5281/zenodo.4590731](https://doi.org/10.5281/zenodo.4590731)

Keyword(s):  
Generation-Recombination Noise  
Superconducting Resonator Quasiparticle Trapping  
Phonon Trapping

Grants:  
European Commission:

- MOSAIC - Multi object spectrometer with an array of superconducting integrated circuits (648135)

Nederlandse Organisatie voor Wetenschappelijk Onderzoek:

- Veni project 451-02-038 (2300131684)

Related identifiers:

## Idea for improvement

- Start using cookiecutter templates at start of the project
- Cookiecutter is an easy way to set up a new project directory
- Many templates available for different purposes
  - TIP: Cookie cutter data science template  
<https://drivendata.github.io/cookiecutter-data-science/>



<https://cookiecutter.readthedocs.io/en/1.7.2/>

## Introduce good practices



OpenClipart/Public domain

**SRON**

Teach scientist to:

- Create a clear project directory structure
- Use version control
- Document your data and code



## Open data workshops

We started workshops based on Software/Data Carpentries

Topics:

- Git version control + Github
- Introduction to jupyter notebooks/numeric python
- Software development workflow (advanced Github)

And more workshops to come.

These workshops will also be held at other NWO institutes.

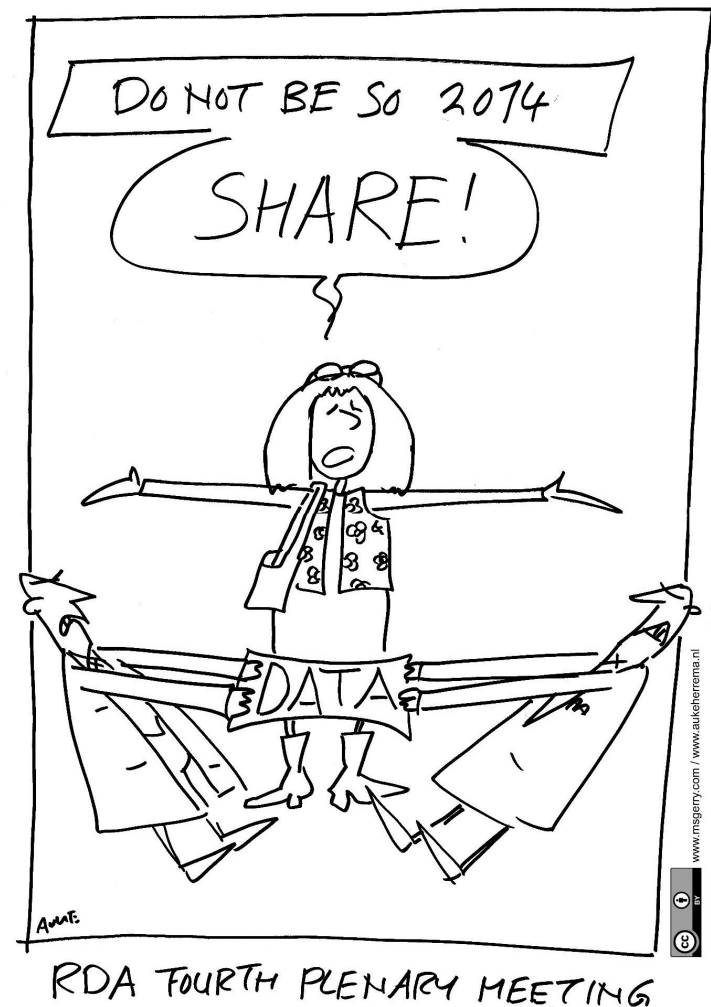
## Lessons learned

- Scientists generally perceive open data as more work
  - Because they aim for perfection
- Workshops help to teach scientists reproducible science and work more efficiently
- Having data stewards/local expertise in your institute helps a lot



## Lessons learned II

- Some scientists are afraid of competition
- Community adoption of open science needed
- Give credits to published data and software
- You need to keep open data on the agenda



## Recommendations to the Astronomy community

- Discuss general requirements for reproduction packages
  - How to deal with diversity of data analysis platforms?
  - What makes a reproduction package acceptable/good?
  - Make concrete what level of detail/documentation is acceptable.
- Cite software and data publications
  - Give credit to people who did a lot of work to share their software and/or data.
- For publishers, A&A, ApJ, MNRAS:
  1. Add data availability statement to paper template.
  2. Add reproduction package to referee process.

## Useful links

SRON reproduction package template:

<https://github.com/jdeplaa/open-data-template>

SRON Open data instructions page:

<https://personal.sron.nl/~jellep/openscience/>

SRON Open data page:

<https://www.sron.nl/data>

Feel free to contact me on Slack, Twitter (@JdePlaa) or Github.

## End credits

Thanks to the SRON data steward team:

- Russ Shipman
- Richard van Hees
- Daniela Huppenkothen
- Stephen Yates

Cartoons by: <https://aukeherrema.nl/> (CC-BY)