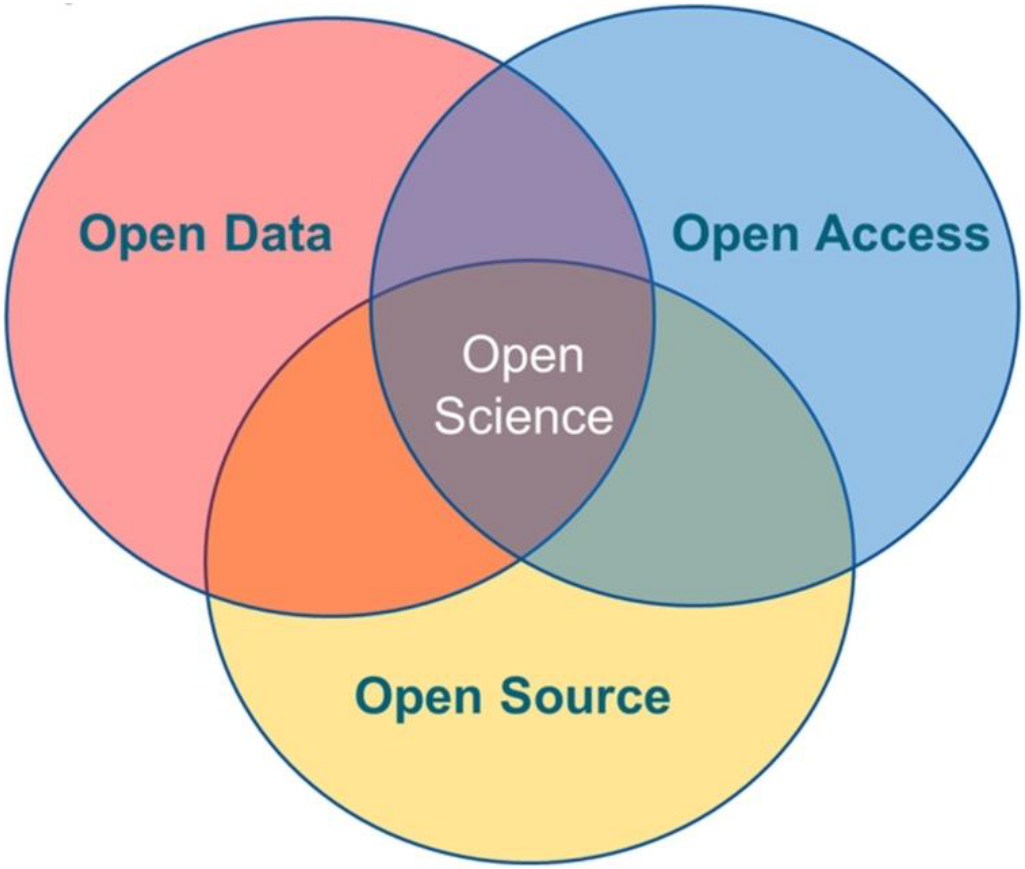
Open Science

What is it all about?



Russell Shipman/Jelle de Plaa

SRON

Netherlands Institute for Space Research

Netherlands Organisation for Scientific Research (NWO)

Table of Contents

Open Science

“Open science is the practising of science in a sustainable manner which gives others the opportunity to

- *work with,*
- *contribute to*
- *and make use of the scientific process.*

This allows users 'from outside of the science world' to influence the research world with questions and ideas and help gather research data.”

https://www.openscience.nl/en/open-science OCW (2017, 9 February). National Plan Open Science.
doi: 10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65

Closer to home…

Open science is a practice to make your scientific endeavors more transparent, accessible, and reproduceable.

SRON

Why Open Science?

- A survey of over 1500 researches by Nature (2016) about reproducibility of science results showed that
  - More than 70% were unable to reproduce science results of others researchers
  - Half of the scientists could not reproduce their own results

- Some (data related) reasons given:
  - Data unavailable
  - Methods / code unavailable

https://www.nature.com/news/1500-scientists-lift-the-lid-on-reproducibility-1.19970

NWO and EU

- Open Science is part of the NWO strategy 2019-2022

- NWO calls for Open Science following the EU and worldwide initiatives
  - Data management plans in proposals get more weight
  - Part of institute's evaluation
  - "Scientific" code-of-conduct
  - Open access (publications and data) required by funding agencies

- Part of Dutch National Research Agenda
  http://www.wetenschapsagenda.nl/

SRON

Open Data

- Improve infrastructure supporting the reuse of scholarly data: the FAIR principles
  - Findable, Accessible, Interoperable, Reusable
  - Recognizes multiple stakeholders: researchers, research fields, institutions, and society as a whole.

- Data management:
  - **Type A**: Satellite/Observatory mission data and software. Space agency or Institution responsibility.
  - **Type B**: Data from other instruments/experiments. Project responsible.
  - **Type C:** Data connected to publications (derived data from type A/B and own software/scripts). Scientist responsible.

Publishing data

When publishing an article also provide relevant data and software.

Data can be published in a ’trusted repository’ and findable with a DOI (Digital Object Identifier) .

For now, Zenodo **(https://zenodo.org/)** , which is run by CERN, is chosen as the default repository for SRON. We have started a ‘community’ there to group the uploads.

The data is called a ‘reproduction package’ that contains (**if reasonably feasible**):

- Relevant Raw data or link(s) to the raw data used (link to Type A data)
- Scripts/programs used to process data and derived products
- Description how to use the scripts and derive the products
- Software/data to reproduce the figures in the publication.
- Help is available from **Data Stewards**: Jelle de Plaa and Russell Shipman for SRON-ASTRO Utrecht/Groningen .

A trusted repository…

zenodo Search Upload Communities r.f.shipman@sron.nl

December 17, 2020 Poster Open Access

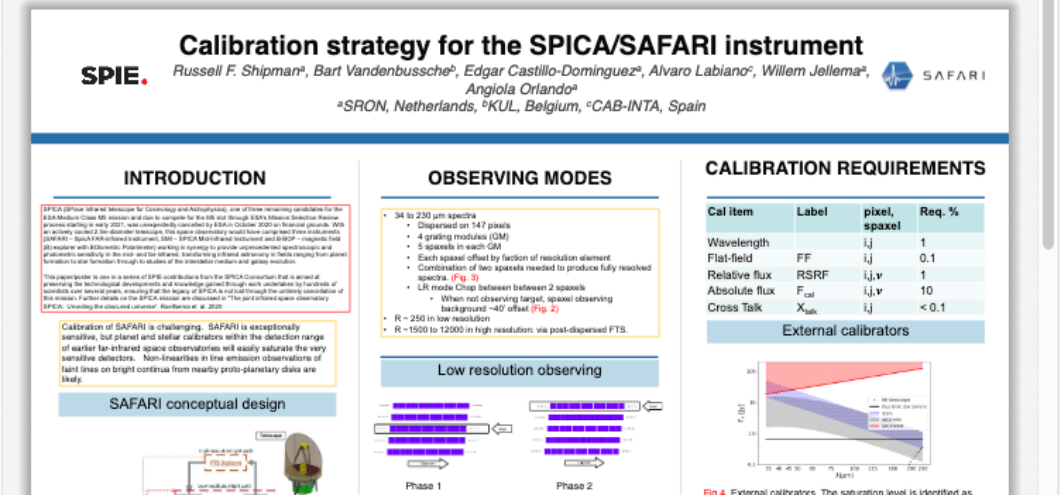
Calibration strategy for the SPICA/SAFARI instrument

Russell Shipman; Bart Vandenbussche; Edgar Castillo-Domínguez; Alvaro Labiano; Willem Jellema; Angiola Orlando

Poster presentation and proceedings for SPIE: Astronomical Telescopes + Instrumentation December 14-18 2000

Preview

Page: 1 of 1 Automatic Zoom



Edit

Communities

SRON Netherlands Institute for Space Research Remove

5 views 7 downloads

See more details...

Indexed in OpenAIRE

Publication date:
December 17, 2020

DOI:
DOI 10.1117/12.2562071

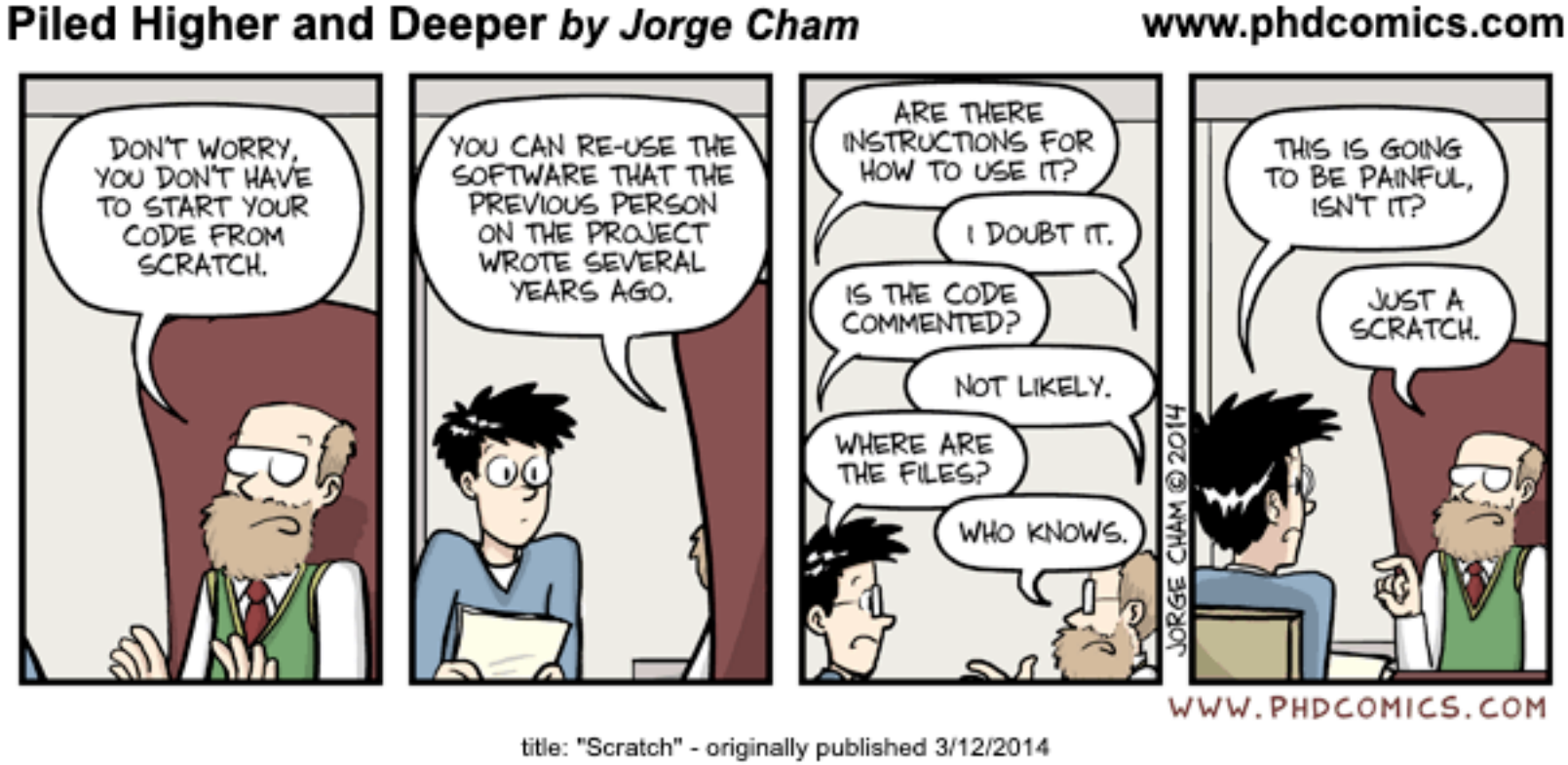
SRON

Open Source

Moving to Open Source software

- Open science means the following for your software/scripts:
  - Your scripts/software will be published at some point
  - Other people need to be able to read/understand your software
  - There needs to be some basic documentation

SRON



title: "Scratch" - originally published 3/12/2014

- Working and programming in a clear and organized way will also make it easier for colleagues to continue previous work by other colleagues.
- Answering questions about your previous work (sometimes years ago), will be easier if the research is well documented.

Adopt good software practices

Writing software tips

- Write software for humans, not computers (so use comments)
- Organize software such that tools can be re-used.
- Use version control systems, like svn or git.
- Source code documentation can help (like sphinx, doxygen)
- Review each others software

More tips in: "Best practices for scientific computing"
http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745

There are specific courses available for PhD students (and others), like software carpentry to learn and practice these best practices.

SRON

License your software and documentation

Attach a license to your software and documentation. This makes clear to potential users how ownership is arranged, how to cite the work and what adaptations are allowed. Repositories, like Zenodo, accommodate this.

For software:

- In the spirit of Open Science, open source licenses are,
  - MIT (https://en.wikipedia.org/wiki/MIT_License) or
  - GPL (https://en.wikipedia.org/wiki/GNU_General_Public_License) type licenses.
- Unfortunately, there are no easy to use non-open source licenses that I know of.

For everything else:

- Creative commons licenses are very easy to understand and use. Choosing a license is made easy: https://creativecommons.org/choose/

NWO does not (yet) provide guidance on which licenses to use.

SRON

Open Access

From January 1 2020, all NWO financed publications should be freely available online.

Green Open Access:

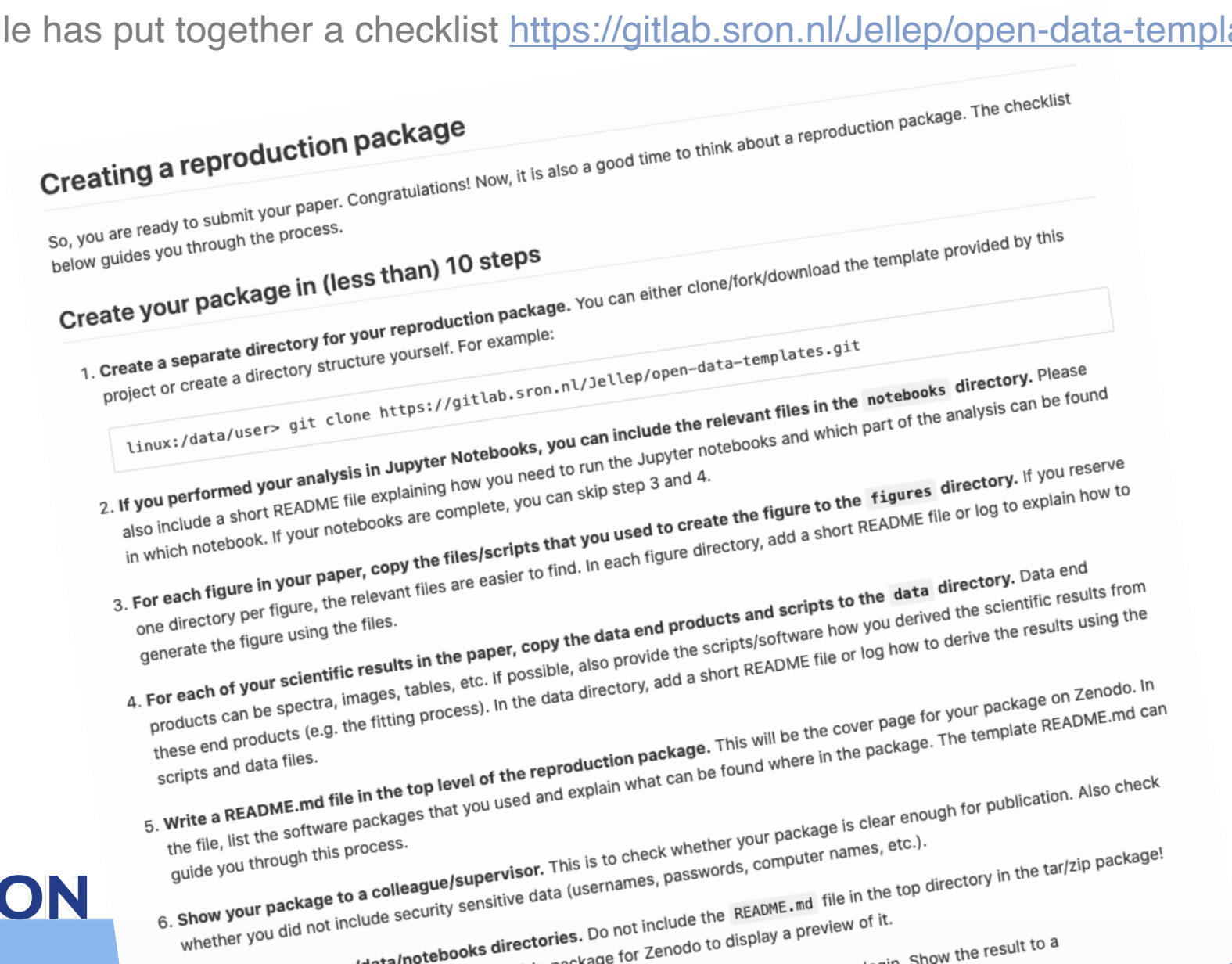
- Author publishes in restricted journal, but also archives a copy in a public archive (arXiv). **Copyright remains with author!**

Gold Open Access:

- Author publishes in open access or hybrid journal (for a fee). The journal will host a freely available version of the paper online.

Reproduction Package

- Jelle has put together a checklist https://gitlab.sron.nl/Jellep/open-data-templates.



Creating a reproduction package

So, you are ready to submit your paper. Congratulations! Now, it is also a good time to think about a reproduction package. The checklist below guides you through the process.

Create your package in (less than) 10 steps

1. **Create a separate directory for your reproduction package.** You can either clone/fork/download the template provided by this project or create a directory structure yourself. For example:

```
linux:/data/user> git clone https://gitlab.sron.nl/Jellep/open-data-templates.git
```

2. **If you performed your analysis in Jupyter Notebooks, you can include the relevant files in the `notebooks` directory.** Please also include a short README file explaining how you need to run the Jupyter notebooks and which part of the analysis can be found in which notebook. If your notebooks are complete, you can skip step 3 and 4.

3. **For each figure in your paper, copy the files/scripts that you used to create the figure to the `figures` directory.** If you reserve one directory per figure, the relevant files are easier to find. In each figure directory, add a short README file or log to explain how to generate the figure using the files.

4. **For each of your scientific results in the paper, copy the data end products and scripts to the `data` directory.** Data end products can be spectra, images, tables, etc. If possible, also provide the scripts/software how you derived the scientific results from these end products (e.g. the fitting process). In the data directory, add a short README file or log how to derive the results using the scripts and data files.

5. **Write a README.md file in the top level of the reproduction package.** This will be the cover page for your package on Zenodo. In the file, list the software packages that you used and explain what can be found where in the package. The template README.md can guide you through this process.

6. **Show your package to a colleague/supervisor.** This is to check whether your package is clear enough for publication. Also check whether you did not include security sensitive data (usernames, passwords, computer names, etc.).

/data/notebooks directories. Do not include the `README.md` file in the top directory in the tar/zip package! package for Zenodo to display a preview of it.

login. Show the result to a

Open Science in practice

Getting started:

No recipe or tool works for all situations!
Best practices in Open Science are still to be established per field.

You are the people who will put this into practice, so:

You can help establish the standard!

Think about how to do things in the coming period and discuss them with colleagues.

Some final advice

Open Science is NOT asking for the impossible.

- Think about what information and data should be there to be reasonably able to reproduce the results, but do NOT spend huge amounts of time on it. Be practical.
- Start working on new projects with this in mind and organize it such that data can be made public later, no need to do this for old or nearly finished projects if that is too much work.
- Do NOT re-write your entire legacy IDL code into an open language. Publishing IDL code is fine.

Everything you do more for open science than now will be regarded as a positive thing!

SRON

Questions