Introducing open data management in Dutch space research

Jelle de Plaa, Russ Shipman

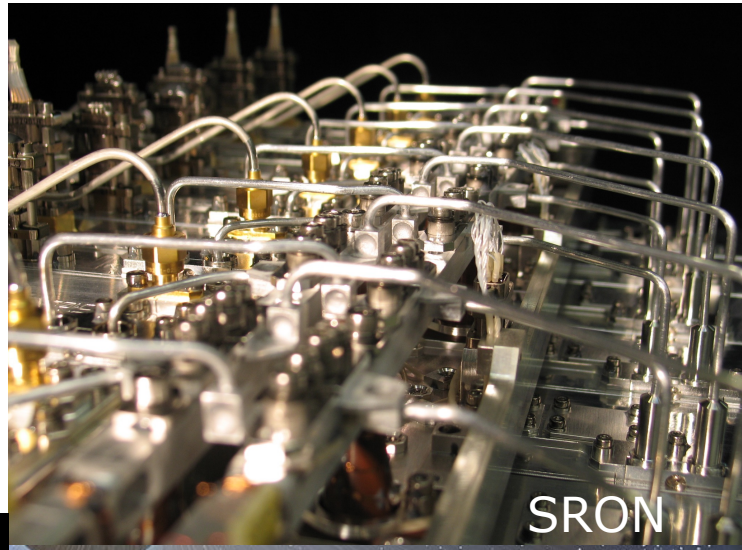
SRON

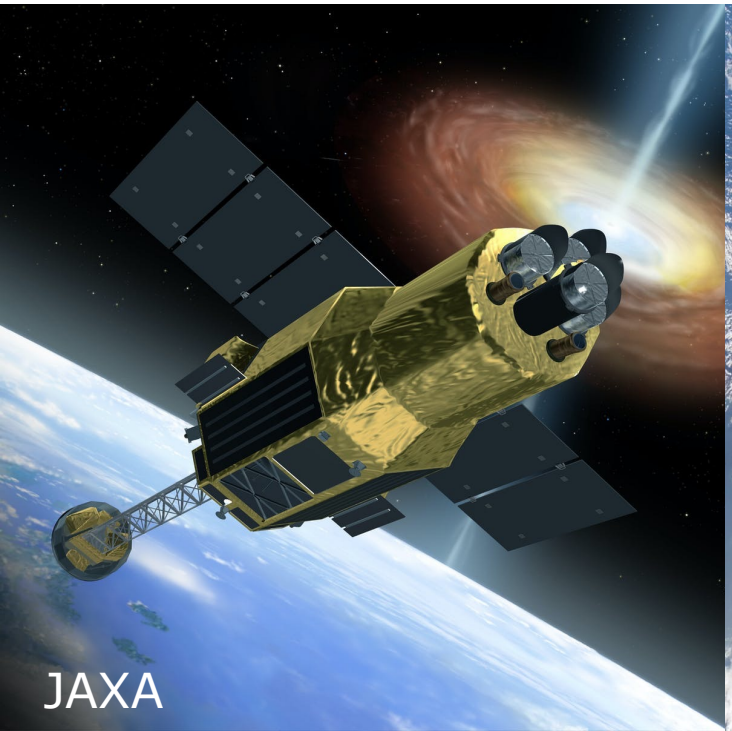
Netherlands Institute for Space Research

Picture: NASA

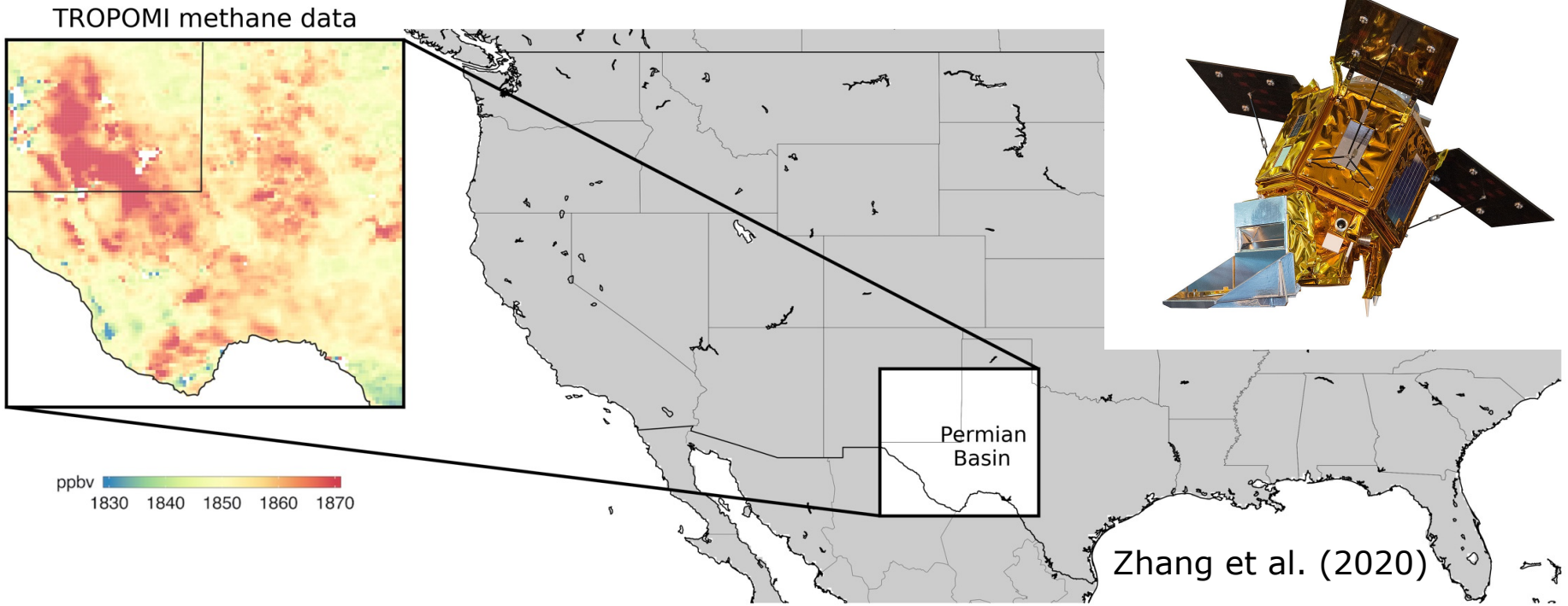
SRON Netherlands Institute for Space Research

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



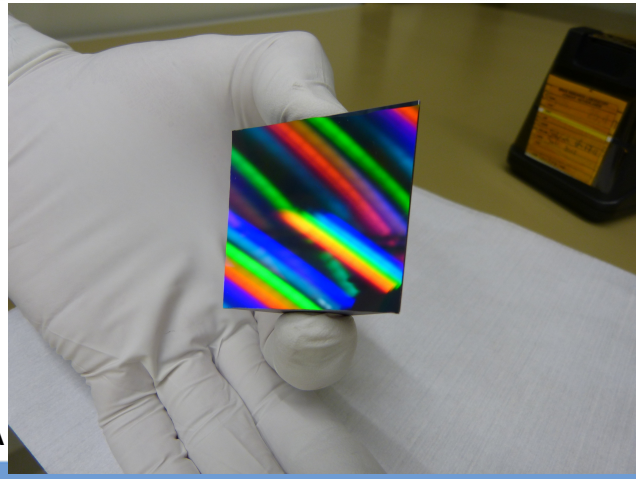


Example: Tropomi Methane measurements



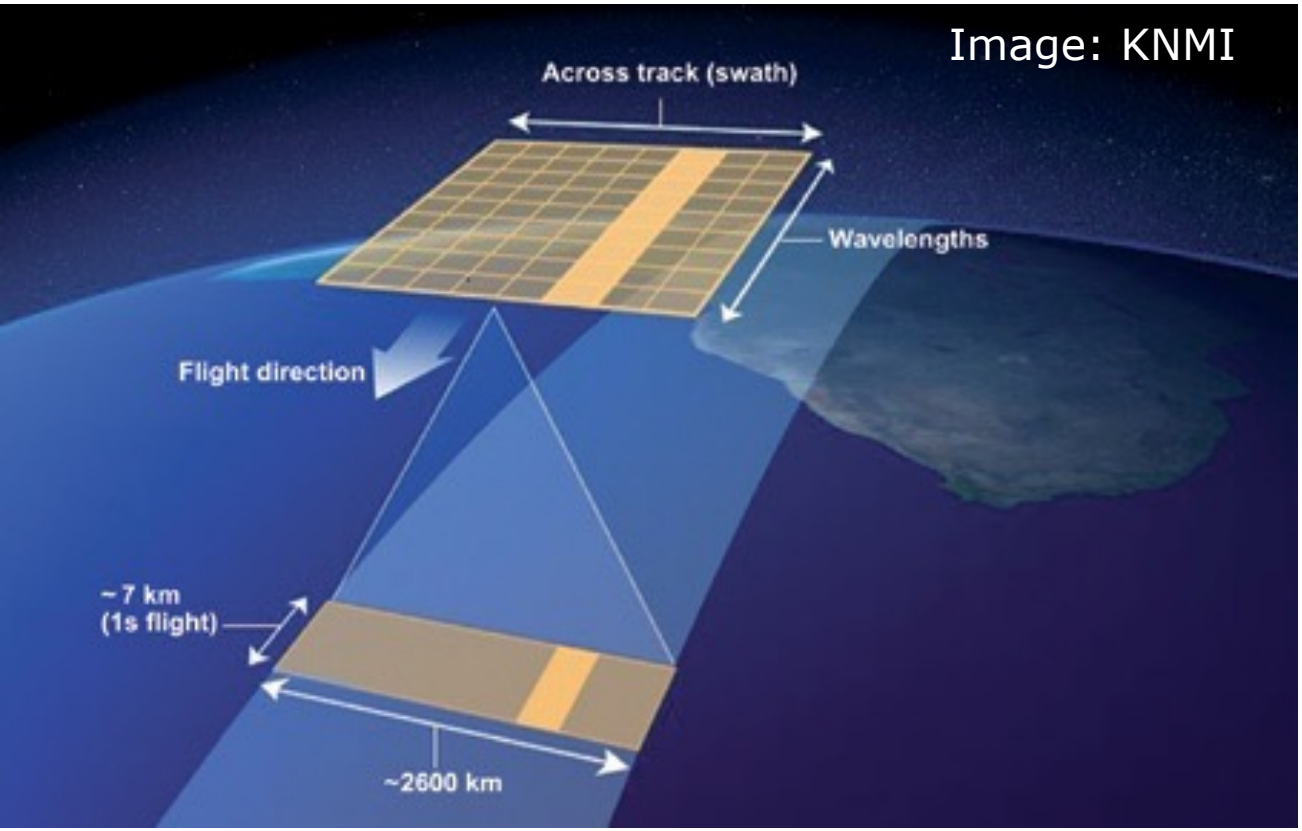
Tropomi measures trace-gas concentrations from space

Using small grating spectrometer

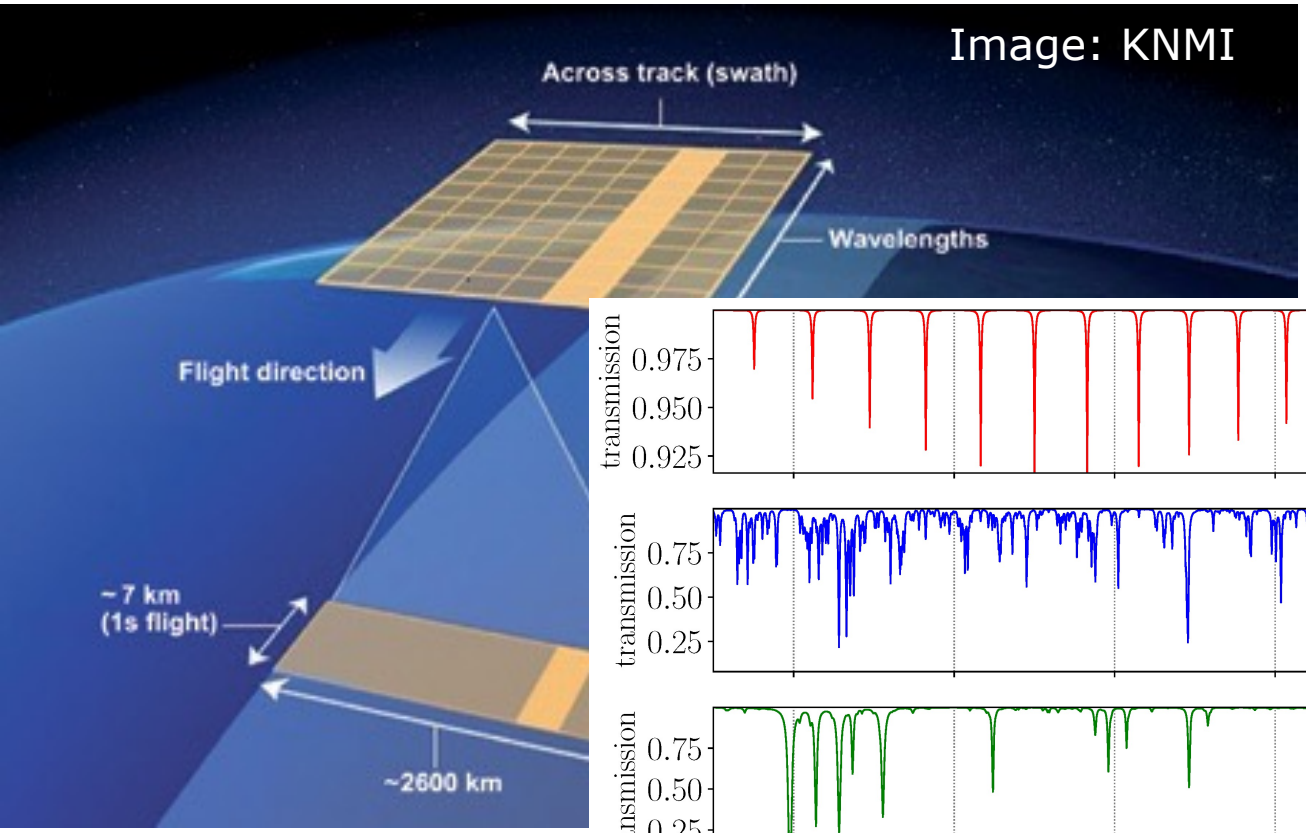


Images: SRON/ESA

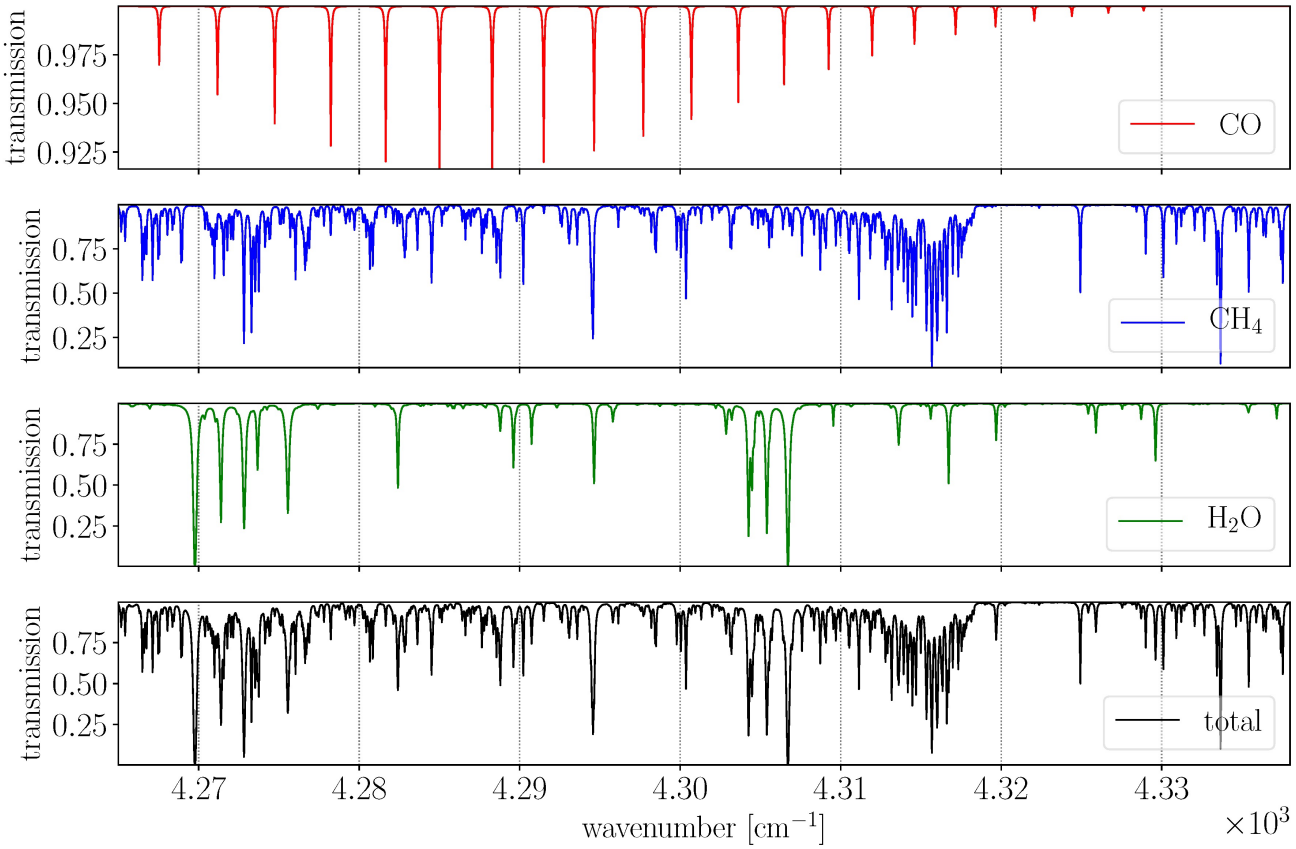
Tropomi spectrum



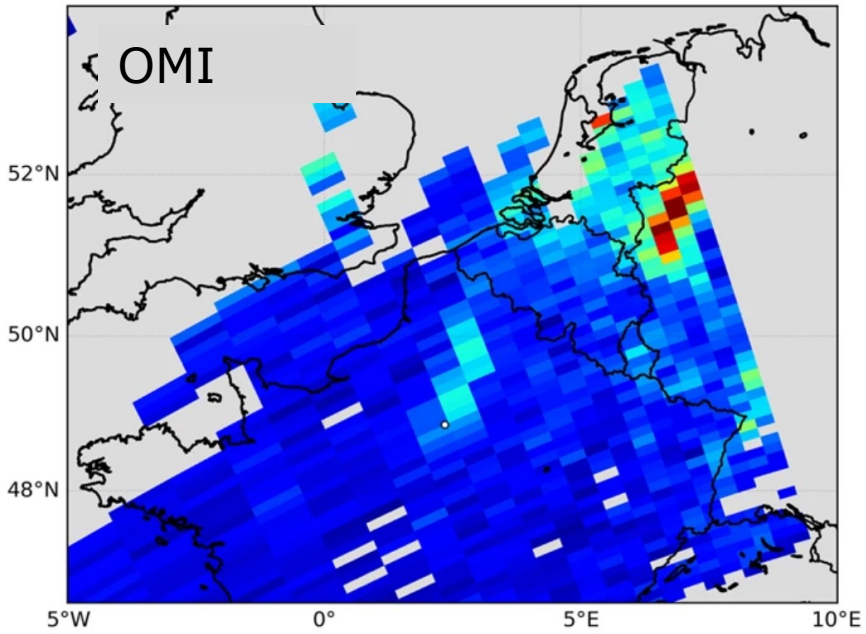
Tropomi spectrum

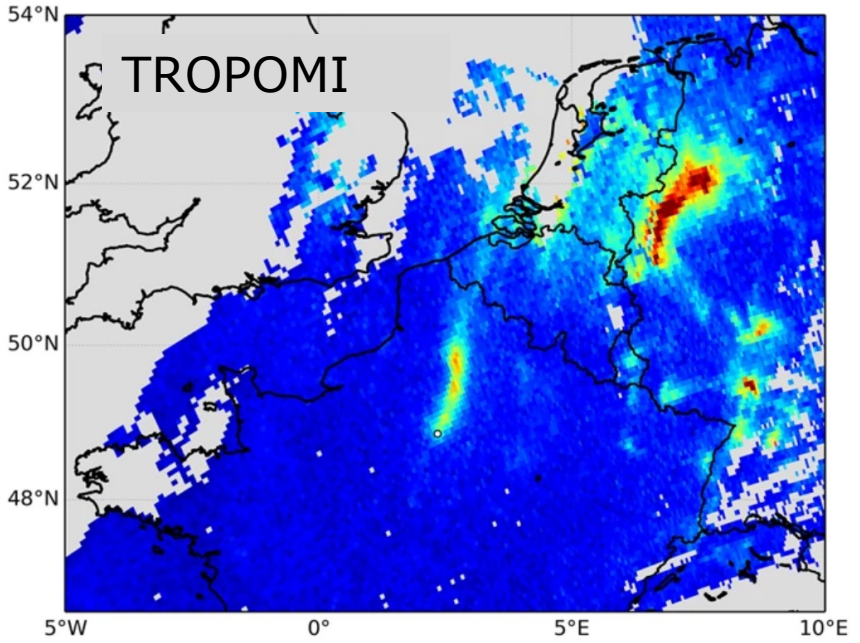


Hochstaffl et al. (2020)



Tropomi NO_2

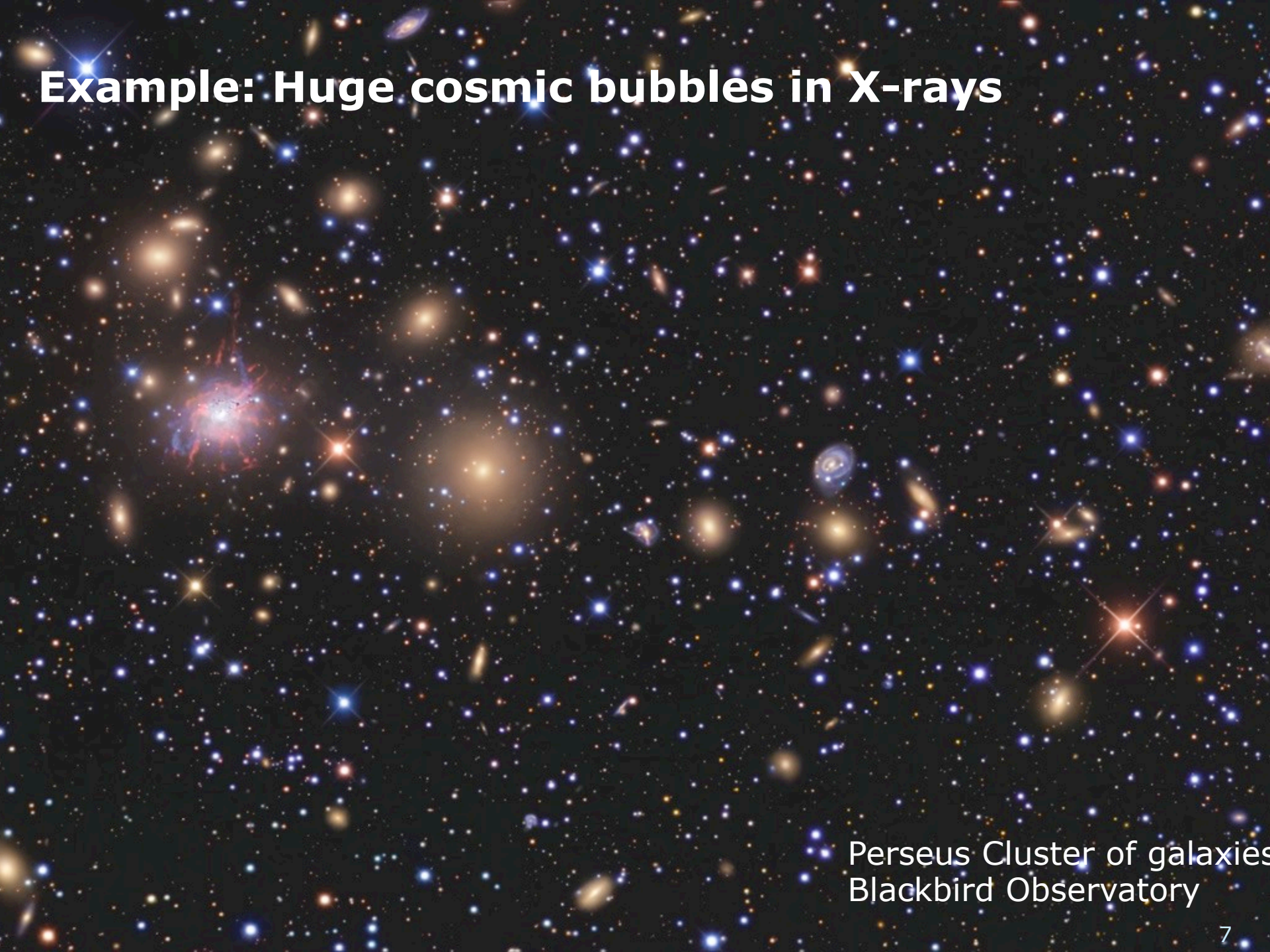




TROPOMI allows us to measure pollution to city level!

Lorente et al. (2019)

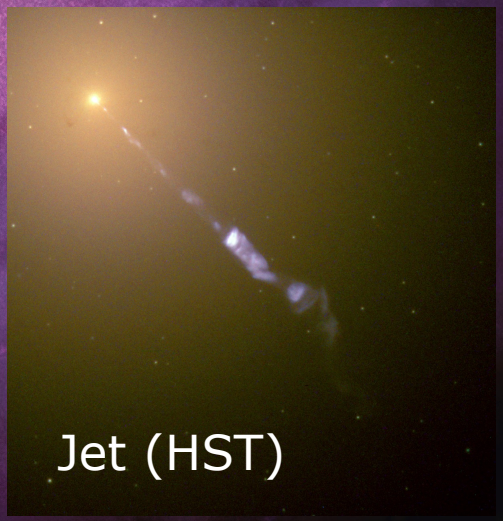
Example: Huge cosmic bubbles in X-rays



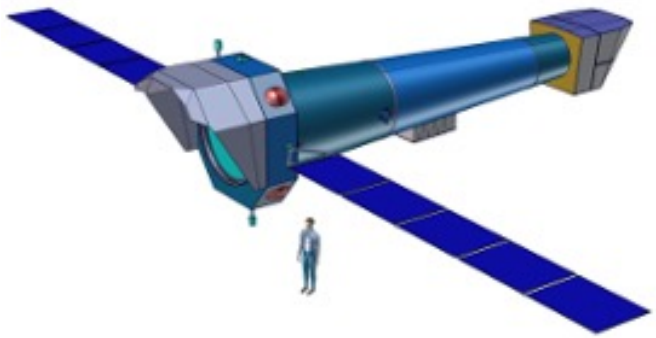
Perseus Cluster of galaxies
Blackbird Observatory

Perseus in X-rays

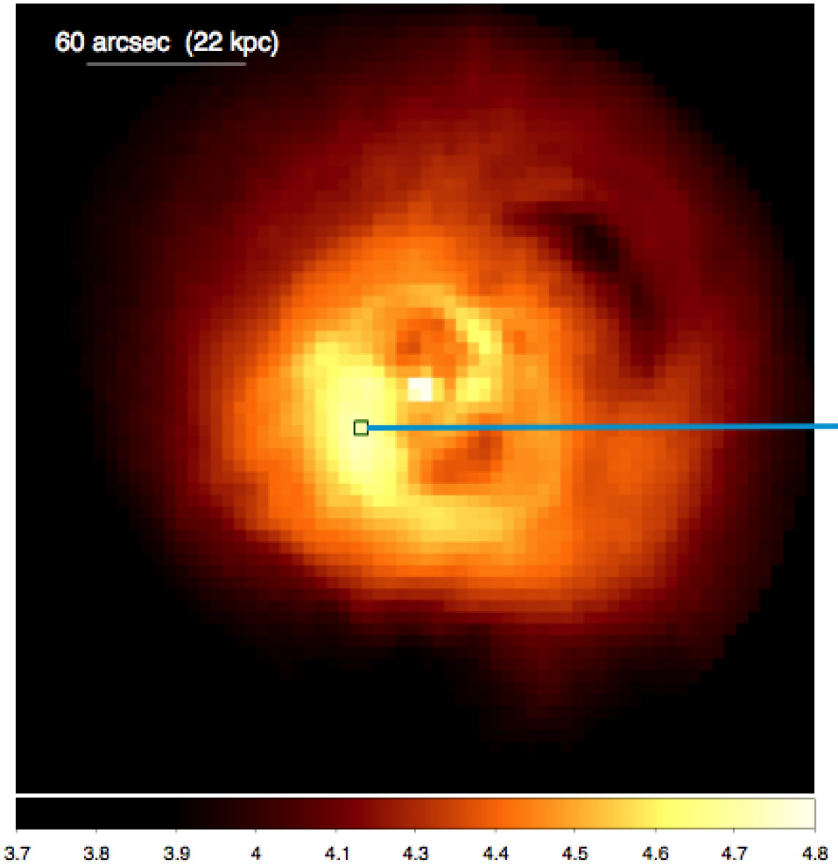
Chandra observatory

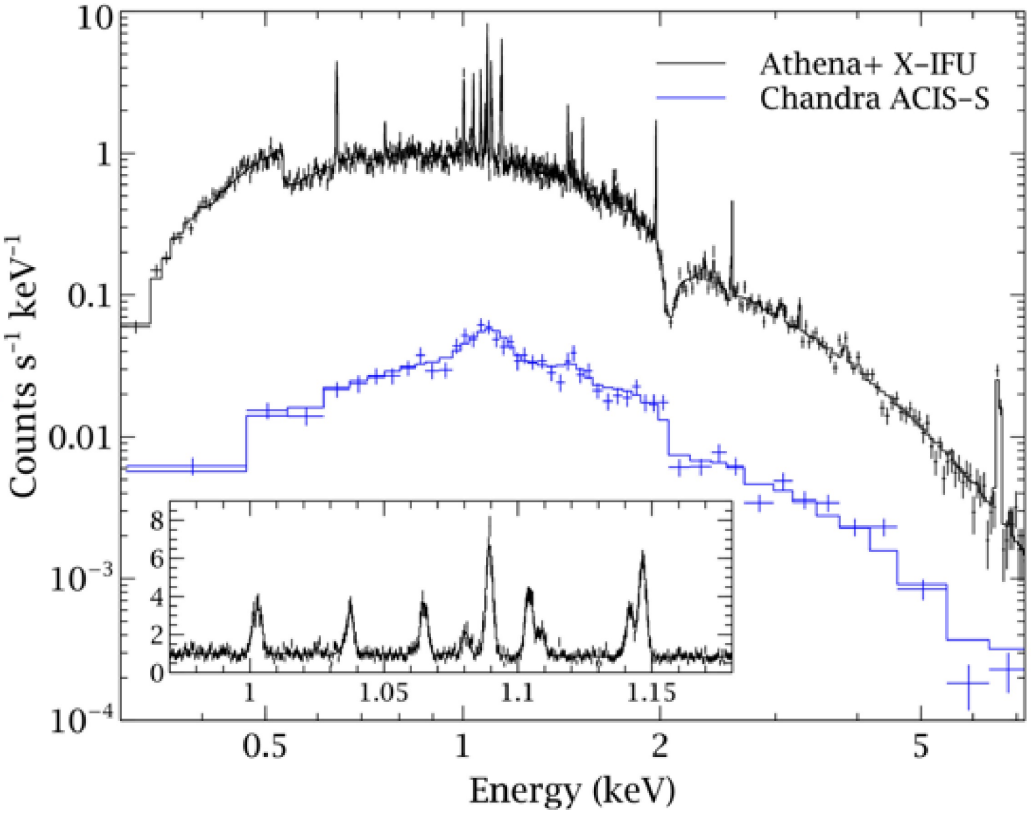


The ATHENA observatory data



New ESA X-ray observatory (launch >2030)

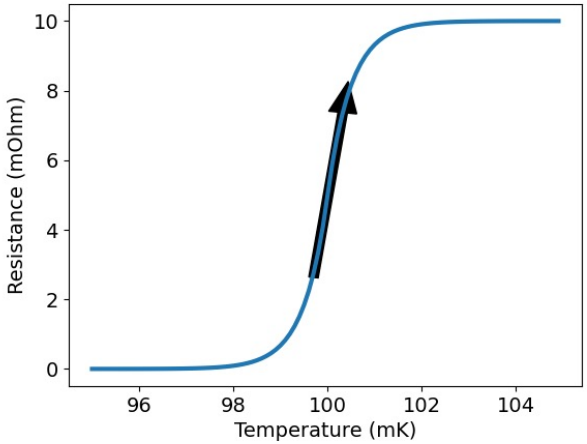




Barret et al. 2013

High-resolution spectrometer development

- Sensor development for ATHENA X-ray spectrometer
- Using superconduction in each pixel to measure energy



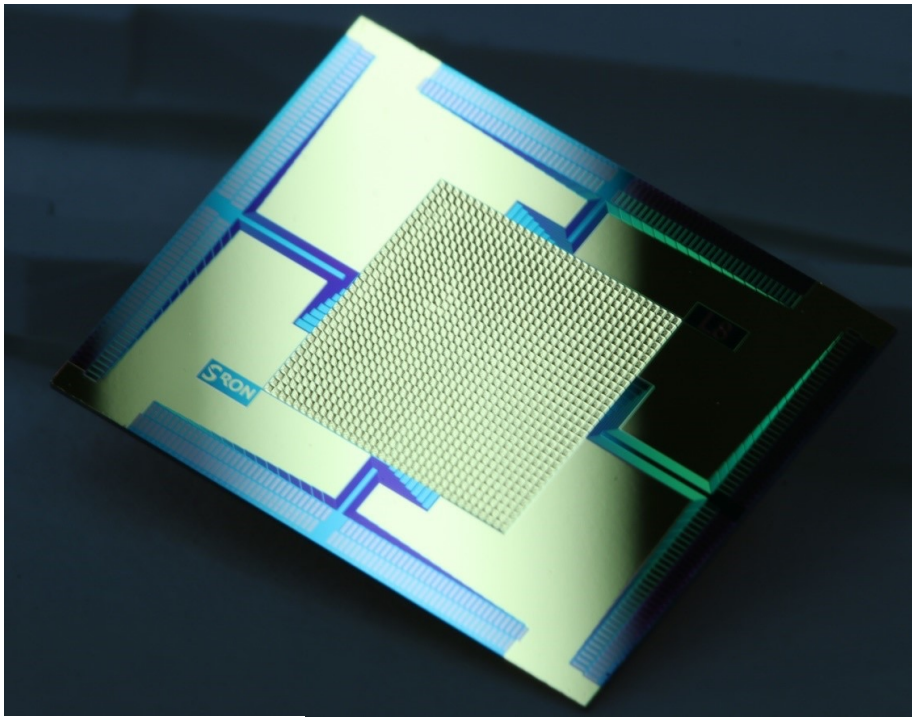
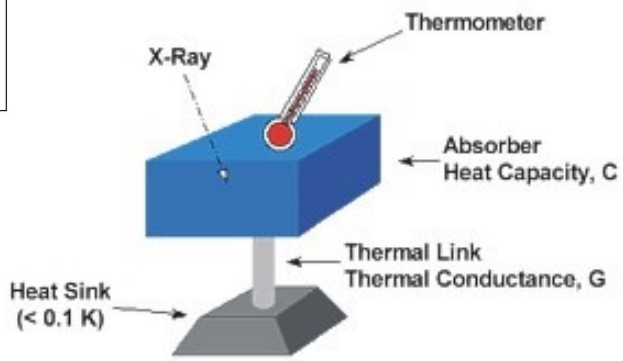


Image: SRON



The Netherlands to Open Science

Netherlands Organisation for Scientific Research (NWO)

- Government science funding agency in NL
- Strategy and policy to move to open science in 2019-2022

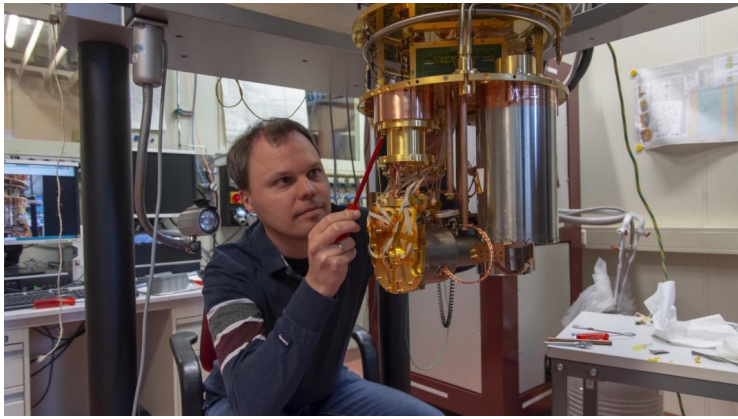
How/why:

- Open Access for journal articles (Plan S)
- Open Data
  - Easy access to publicly funded research data
  - Accelerate science by sharing tools and knowledge
  - Solve reproducibility crisis in science
  - Helps researchers with data management

SRON

Types of data at SRON

- Raw data from satellites
- Data from own experiments
- **Derived data for papers**
  - Images, spectra, fits, figures, etc.



CHEERS: The chemical evolution RGS sample

J. de Plaa[1], J. S. Kaastra[1,2], N. Werner[3,4,5], C. Pinto[7], P. Kosec[6,7], Y-Y. Zhang[8], F. Mernier[1,2], L. Lovisari[8,9], H. Akamatsu[1], G. Schellenberger[8,9], F. Hofmann[10], T. H. Reiprich[8], A. Finoguenov[11,10], J. Ahoranta[11], J.S. Sanders[10], A.C. Fabian[7], O. Pols[12], A. Simionescu[13], Jacco Vink[14], and H. Böhringer[10]

[1] SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands. e-mail: j.de.plaa@sron.nl
[2] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA, Leiden, The Netherlands.
[3] MTA-Eötvös University Lendület Hot Universe Research Group, Pázmány Péter sétány 1/A, Budapest, 1117, Hungary
[4] Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlarská 2, Brno, 611 37, Czech Republic
[5] School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan
[6] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305-4085, USA
[7] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.
[8] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany.
[9] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA.
[10] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany.
[11] Department of Physics, University of Helsinki, FI-00014 Helsinki, Finland.
[12] Dept. of Astrophysics/IMAPP, Radboud University, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands.
[13] Institute of Space and Astronautical Science (ISAS), JAXA, 3-1-1 Yoshinodai, Chuo-ku, Sagamihara, Kanagawa, 252-5210, Japan.
[14] Anton Pannekoek Institute/GRAPPA, University of Amsterdam, PO Box 94249, 1090 GE, Amsterdam, The Netherlands.

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.

How to organize all this?

Challenges:

- Researchers use multiple different programming languages
- Researchers do not want to spend a lot of time
- Researchers do not like to change the way they work (by a lot)

Our solution:

- Data stewards help researchers with open data issues
- Data and software training
- **A simple reproduction package template** (for Zenodo)

A simple reproduction package template

Basic directory structure:

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

Readme.md

Data: derived data products and analyis 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

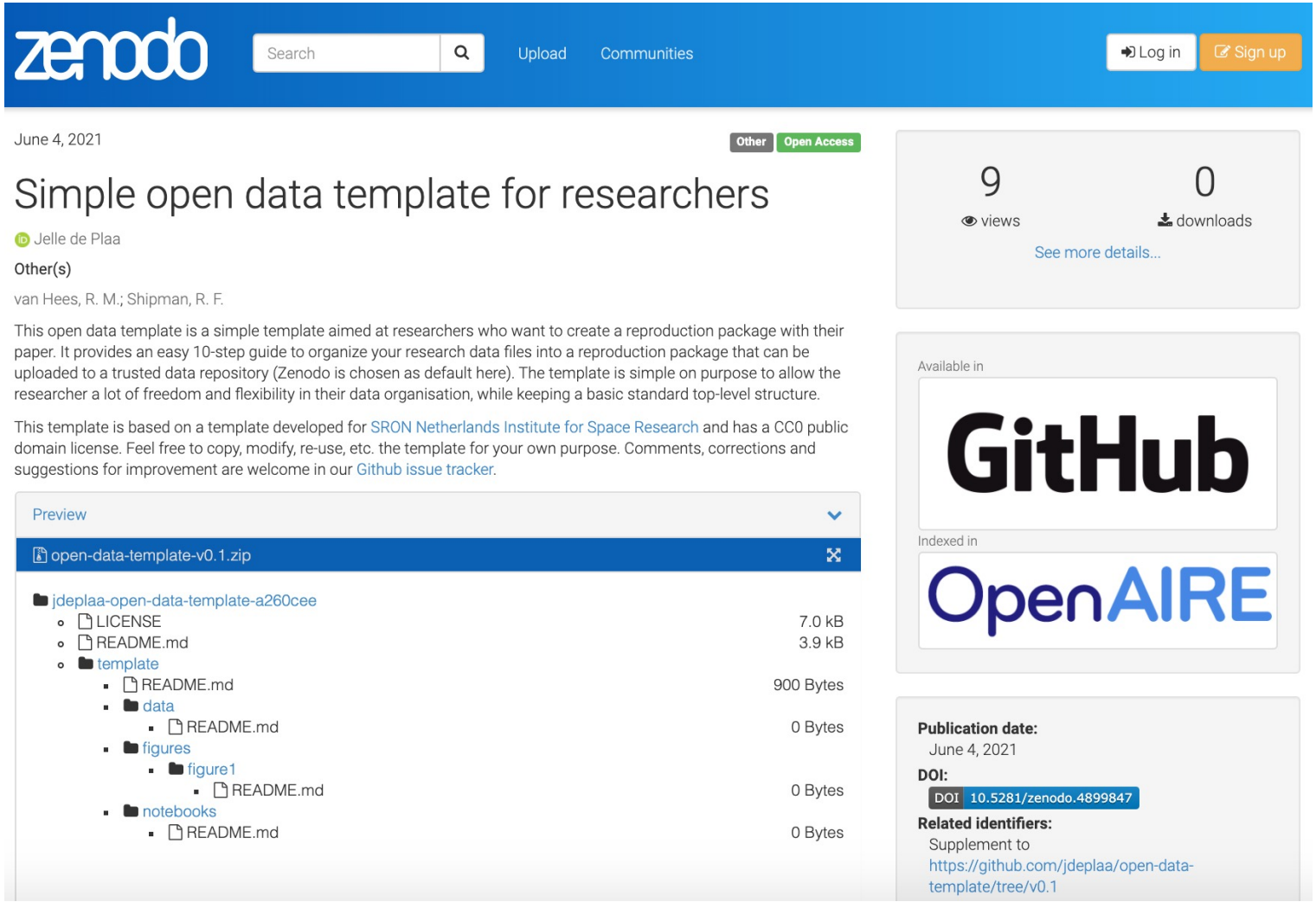
Disadvantages:

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

10 steps toward a reproduction package

1. Create a directory on your machine for your package
2. Copy and document jupyter notebooks (optional)
3. Copy figures and their creation scripts, and create Readme
4. Copy data products and analysis scripts, and create Readme
5. Finish/fill out the top-level Readme.md file
6. Show the package to a colleague or supervisor
7. Compress directories to Zip or Tar files (not top Readme.md)
8. Do a test upload in the data repository (Zenodo sandbox)
9. Do the final upload to the data repository (Zenodo)
10. Cite the DOI link in your paper (data availability statement)

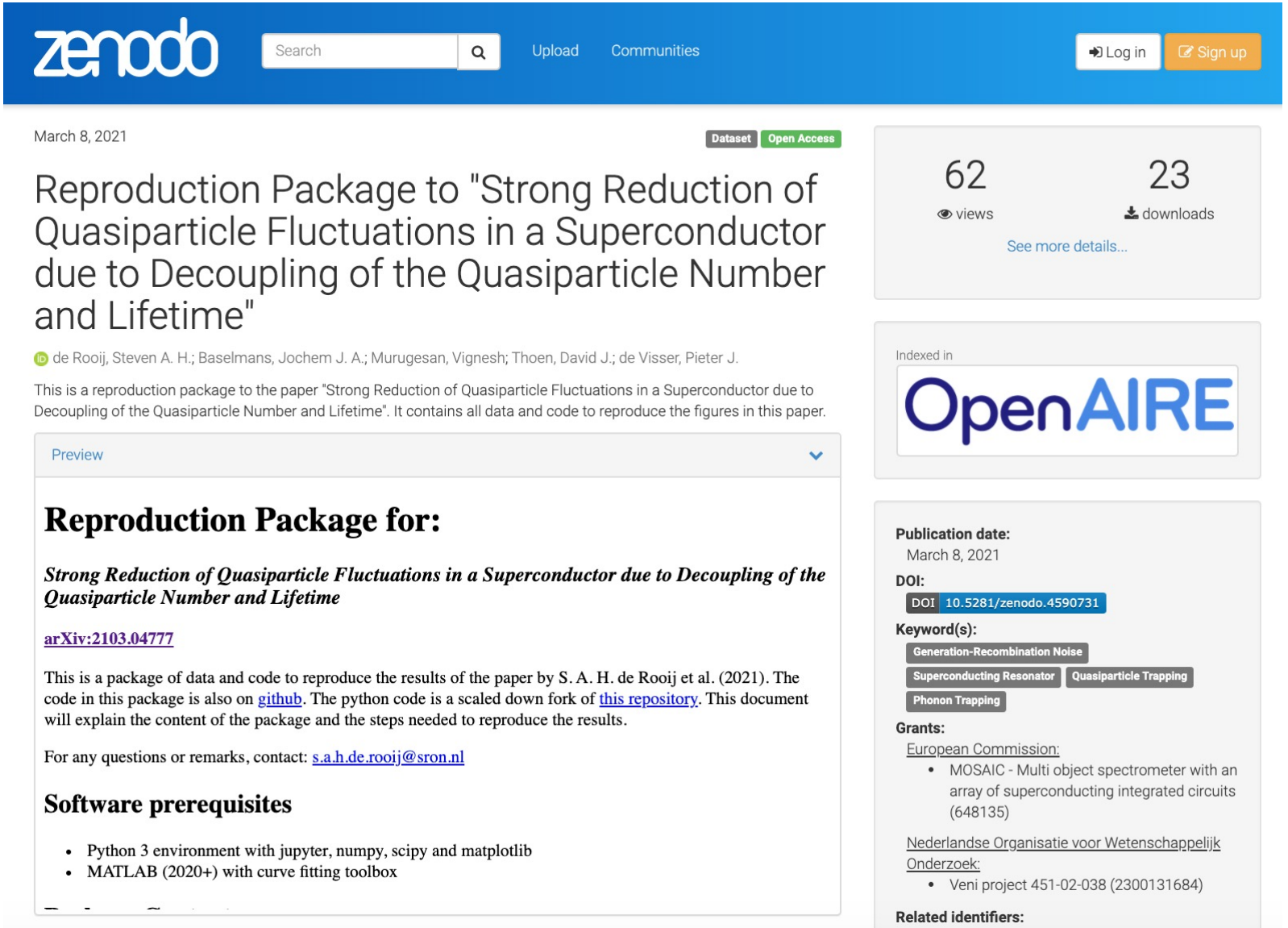
Template on Github/Zenodo



https://github.com/jdeplaa/open-data-template
http://doi.org/10.5281/zenodo.4899847

SRON

Example reproduction package



SRON

http://doi.org/10.5281/zenodo.4590731

Conclusions

- Created very simple template to help researchers create a reproduction package for their paper

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