



2nd European Photon & Neutron EOSC Symposium

26 October 2021

Describe data by scripts for future reuse

Author: Petr Čermák

Affiliation:  <https://mgml.eu/>, Charles University

26. October 2021

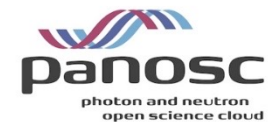


PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.

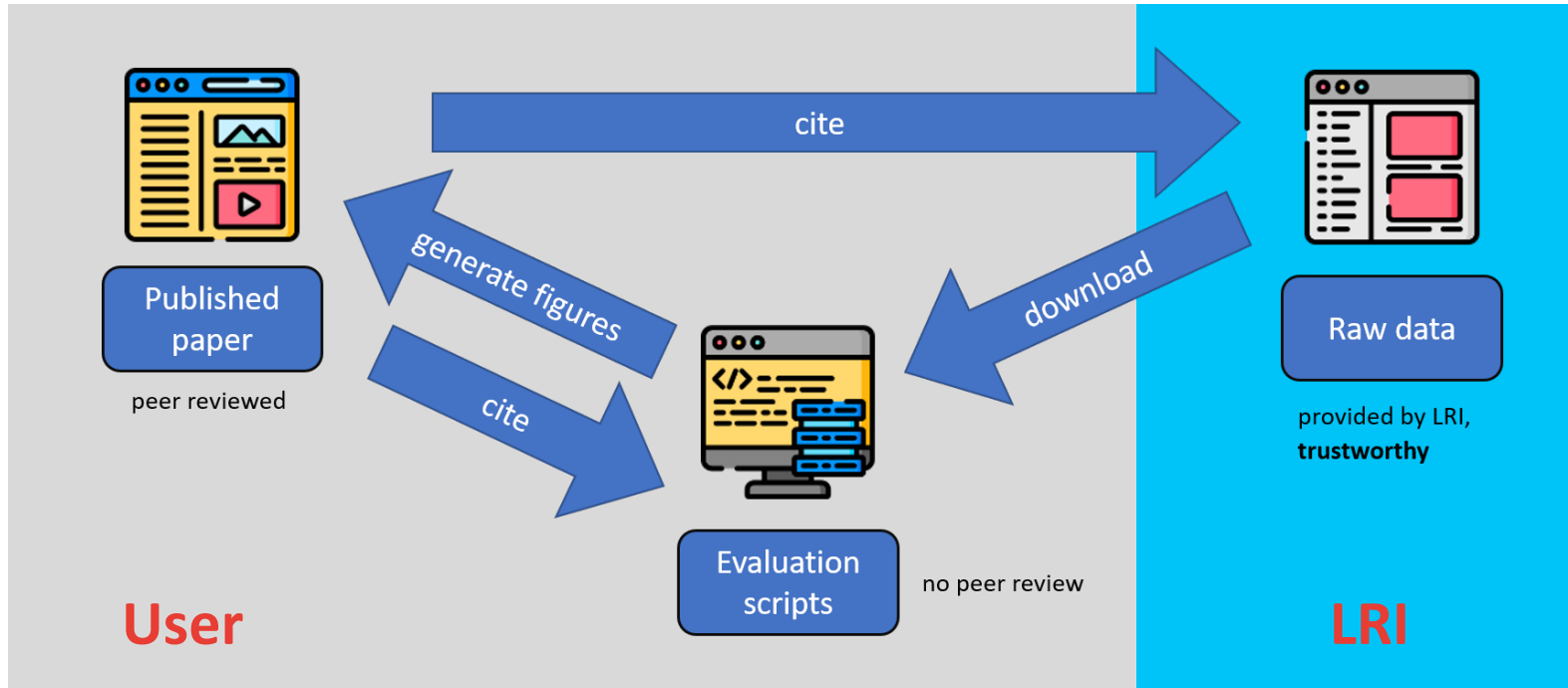
F.A.I.R. and more

Lets do a quick test
of Fairness

- ✓ Findable
- ✓ Accessible
- ? Interoperable
- ? Reusable
- ✓ Trustworthy



Ideal world



- Measured data are published in full form
- Raw data can't be modified → control checksum
- Whole process is automatic

1) Treat data by scripts

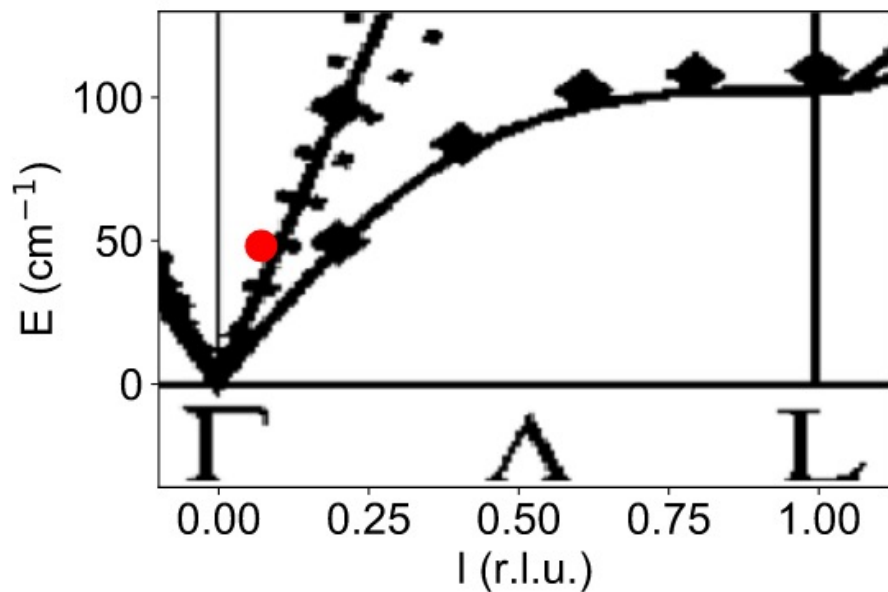
Python, Julia, R ...

Download the data

```
In [15]: # import libraries i will need
# for downloading files from internet

In [2]: # create a model of a simple Gaussian peak with given initial guess
model = Background() + Gauss('p1', pos=0.1, ampl=100, fwhm=0.1) + Gauss('p2', pos='-p1_pos', ampl=100, fwhm=0.1)

In [5]: img = imread("phonon.png")
fig, ax = subplots()
errorbar([result.paramvalues['p1_pos']], [energy], xerr=result.paramerrors['p1_pos'], fmt='ro', ms = 14)
ax.imshow(img, extent=[-0.1, 1.12, -36, 130], aspect='auto')
xlabel("l (r.l.u.)")
ylabel("E (cm$^{-1}$)")
savefig("dispersion.eps", bbox_inches='tight')
show()
```



QL



This proposal is publicly available since 09/19/2019

Title
HSC17 Hercules practical course

Abstract

Metadata

Identifier

DOI doi:10.5291/ILL-DATA.TEST-2385

Authors

- STEFFENS Paul
- DELLEA Greta
- DENG Yue
- DIETL Christopher
- DJURADO David
- GAMBINO Marianna
- HEPTING Matthias (ORCID)
- INKINEN JUHO
- JAFARI Atefeh
- LEFRANCOIS Emilie
- LOPES SELVATI Ana Carolina
- PANAHI Hamed
- PEDERSEN Martin Nors
- PRADIP Ramu
- RANIERI Umbertolua
- ROSSI Matteo
- SCHATTE Sarah
- STANA Markus
- TIMOSHENKO Janis
- VONESHEN David
- ZBIRI Mohamed



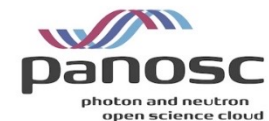
In the year 2525

if man is still alive
If woman can survive,
they may find... **your data**



- PyPI is a repository of software for Python
- `requirements.txt` will tell, which packages to use
- Same exist for Julia (`Project.toml`) or R (`install.R`)

```
≡ requirements.txt  
1 requests==2.24.0  
2 beautifulsoup4==4.9.3  
3 ufit==1.4.2  
4
```



Publish the script

GitHub -> figshare

GitHub is perfect for tracking changes

figshare can clone actual state of repository

figshare will make scripts citable (doi)

There should be:

- xyz.ipynb
(download data, treat data, generate figures)
- requirements.txt
(do be used in 2525)
- Bonus: add paper.tex 🦹

The screenshot shows a GitHub repository page for 'me2d09 / silicon'. The repository is public and has 4 unwatched items, 2 stars, and 1 fork. The main branch is 'main' with 12 branches and 1 tag. The repository contains several files and folders, including .github/workflows, rawdata, .gitignore, LICENSE, README.md, bibliography.bib, dispersion.eps, energy-scan.eps, main.tex, phonon.png, requirements.txt, and silicon.ipynb. The README.md file is open, showing the title 'An Exercise in Open Data: Triple Axis Data on Si single crystal' and a DOI link: 10.6084/m9.figshare.13130270. The repository also includes a 'Cite' button, a 'Download' button (296.79 kB), and options to share, embed, and collect. The repository is categorized under 'Condensed Matter Physics' and has 208 views, 24 downloads, and 0 citations.

The screenshot shows the Figshare page for the repository 'An Exercise in Open Data: Triple Axis Data on Si single crystal'. The page includes a 'Cite' button, a 'Download' button (296.79 kB), and options to share, embed, and collect. The repository is categorized under 'Condensed Matter Physics' and has 208 views, 24 downloads, and 0 citations. The page also includes a 'Usage Metrics' section and a 'Keywords' section with the keyword 'Open Data'.



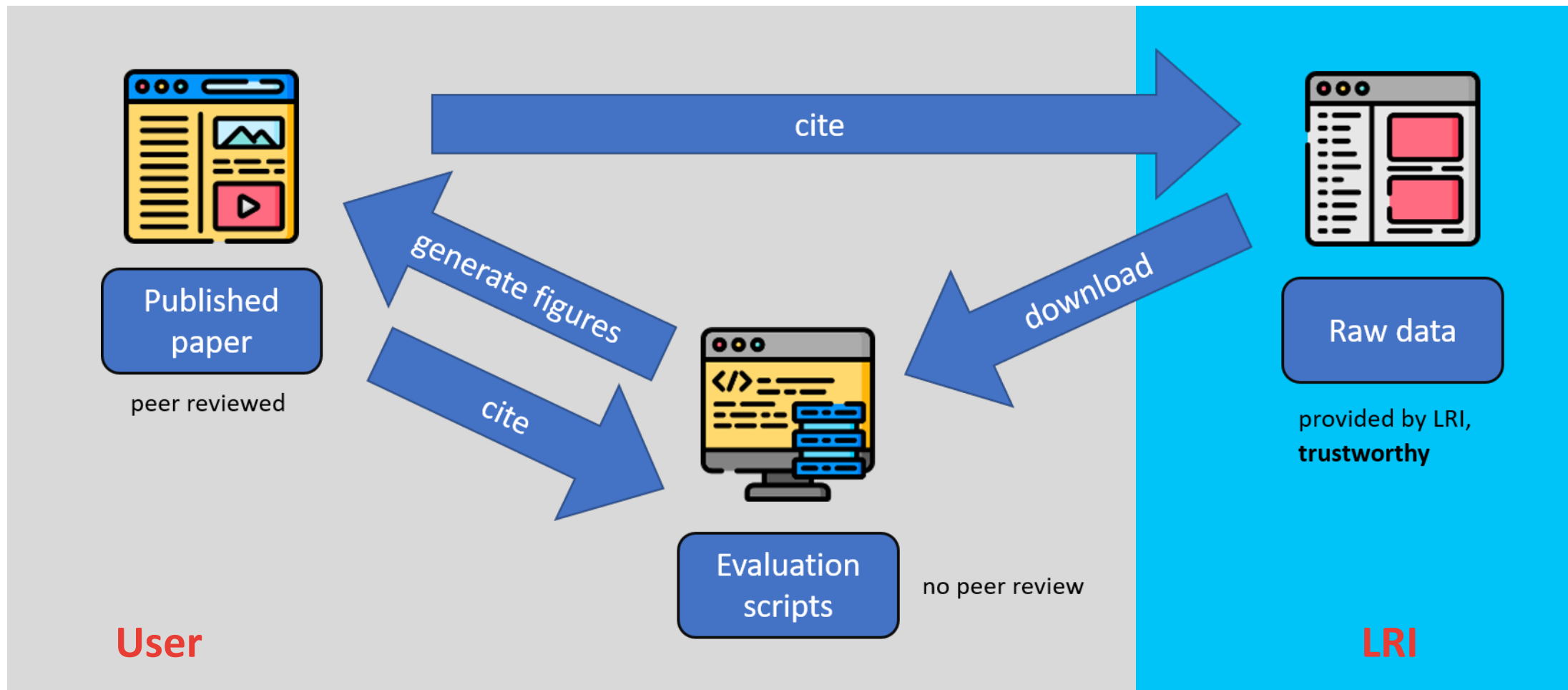


... magic will happen

<https://mybinder.org/v2/gh/me2d09/silicon/main?filepath=silicon.ipynb>



Conclusions





2nd European Photon & Neutron EOSC Symposium

26 October 2021

Thank you



Petr Čermák

 @petrscience

cermak@mag.mff.cuni.cz

Presentation: [doi:10.6084/m9.figshare.16869467](https://doi.org/10.6084/m9.figshare.16869467)

Data: [doi:10.5291/ILL-DATA.TEST-2385](https://doi.org/10.5291/ILL-DATA.TEST-2385)

Scripts: [doi:10.6084/m9.figshare.13130270.v1](https://doi.org/10.6084/m9.figshare.13130270.v1)

Github repo: <https://github.com/me2d09/silicon>

Paper: <https://arxiv.org/abs/2010.12086>

Icons made by Freepik from www.flaticon.com.



PaNOSC and ExPaNDS projects have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements 823852 and 857641, respectively.