NEW PROJECT, NEW TOOLS?!

How granular should our tools and data-processing workflows be to allow for reusability?

openmod workshop Grenoble, March 2024

DLR

What makes small tools reusable?

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages
DLR

- Documentation with tutorials and ready to use example datasets
- Simple installation, few dependencies, up to date code
- Based on common basic software (e.g. you are used pandas)
- Active community and support
- functionality is separated from GUI or config files, clear API
- Tested code
- Readable source code (use speaking names)
- No implicit assumptions inside the code, make it transparent
- Use of repository templates

What makes large workflows reusable?

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages
DLR

- All in one suite
  - easy installation, few dependencies
  - few manual steps, e.g. open excel, then do 1, then open QGIS, then run y
- Repository structure follows cookiecutter templates from the domain
  - Similar structure for all your projects?
- Publish small parts of your workflows, that may be of use for others as own library
  - e.g. PyPSA -> PyPSA, atlite, linopy
  - oemof -> demandlib, feedinlib, oemof-solph, oemof-tabular, …

How to write reusable code

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages
DLR

- Regular interaction with colleagues and peers
- Refactor your monoliths regularly
- Publish your code repository together with the paper
  - Responsiblity does not end with paper acceptance
- Adhere to design patterns
  - io, controller, model, views, …
- Follow basic coding principles
  - Don‘t repeat yourself, integration operation separation principle, single responsibility principle, …
- Test driven development (wish for a tutorial for openmod 2025)
- Read blogs/essays
- We need to be better software engineers really

Some Resources

Gefördert durch:
Bundesministerium für Wirtschaft und Klimaschutz
aufgrund eines Beschlusses des Deutschen Bundestages
DLR

- Blogs/Books/Essays
  - Tons of js libraries: https://evolutionjobs.com/exchange/why-are-there-so-many-javascript-frameworks/
  - Mini js libraries. They do one job and they do it well: https://medium.com/@PepsRyuu/tips-on-writing-good-javascript-libraries-e3c3068ec705
  - Microservices granularity: https://medium.com/@lviazrnio/how-to-choose-wisely-when-defining-microservices-granularity-8e223072636c
  - Start a monolith, break it down to small pieces: https://martinfowler.com/bliki/MonolithFirst.html
  - The Cathedral and the Bazaar (more of a systemic essay on open source, but with some touches to granularization, which was a success factor in this case): http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/index.html
  - https://martinfowler.com
  - https://medium.com (Freedium for firefox users ;))
  - “Clean Code” by Robert Martin
  - “The art of UNIX programming”

- Papers/Research
  - Rescience: https://rescience.github.io/
  - Good scientific software: https://arxiv.org/pdf/1210.0530.pdf
  - Reproducible computational research: https://doi.org/10.1371/journal.pcbi.1003285
  - Workflow management: Snakemake – A framework for reproducible data analysis | Hacker News
  - Workflows and data used for a specific paper: https://github.com/fneum/spatial-sector/
  - Explanations and reasoning for desings of workflow management: https://www.softxjournal.com/article/S2352-7110(21)00188-6/fulltext

- Tutorials/Best practices
  - Scientific workflow management in julia (principles apply for other languages): https://juliadynamics.github.io/DrWatson.jl/
  - Introduction of basic coding principles: https://aeturrell.github.io/coding-for-economists/
  - https://lfenergy.org/training-certification/ for open source model / quality training
  - https://coderefinery.org/ for software modularity training
  - https://software-carpentry.org/
  - https://www.workingsoftware.dev/the-ultimate-list-of-domain-driven-design-books-in-2024
  - https://www.hello-startup.net/
  - https://packaging.python.org/en/latest/

Imprint

Gefördert durch:

Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

DLR

Date: 2024-03-28

Author: the openmod community