

# Research software engineering (RSE) at UiT in 10-15 minutes

1: The RSE role

[research-software.uit.no](https://research-software.uit.no)

2: Toolset

3: Courses

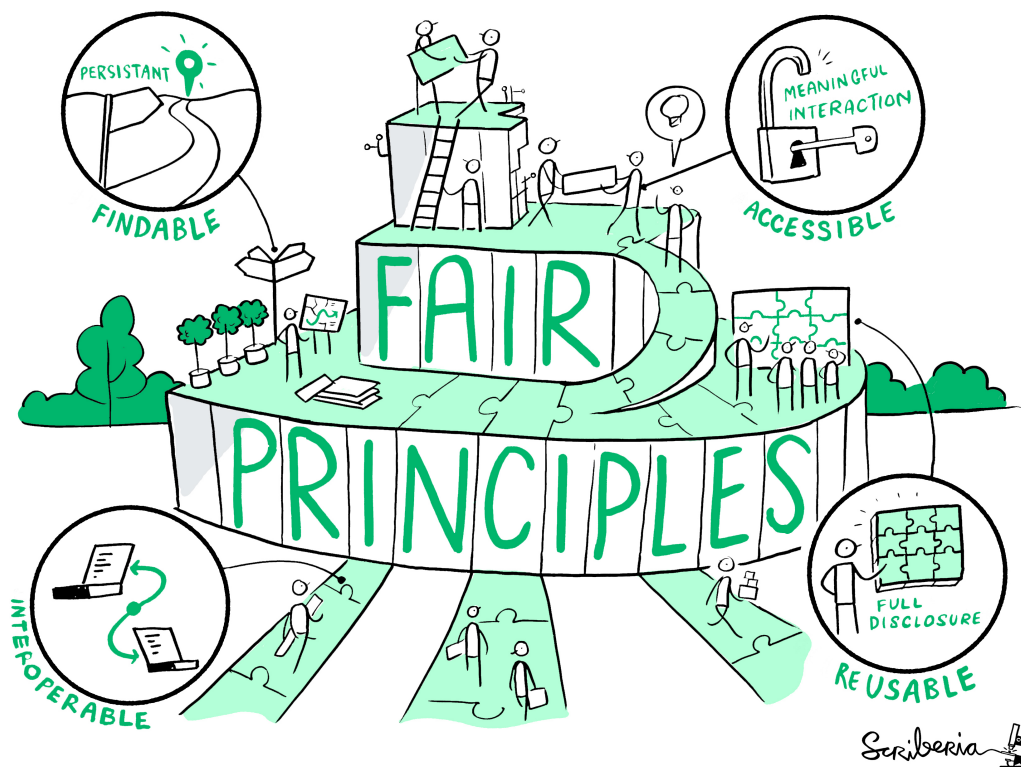
4: Use cases

5: Vision



# FAIR principles and software

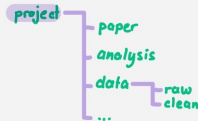
Researchers need to navigate many tools and concepts.



# REPRODUCIBLE RESEARCH

6 helpful steps

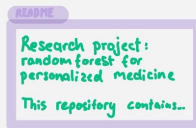
- 1 Get your files + folders in order



- 2 Use good names for files, folders, functions, ...

`6-steps-reproducibility.pdf`  `clean.data <- function(...) { ... }`

- 3 Document with care:  
README, Metadata, code comments, ...



CC-BY 4.0 Heidi Seibold  
@HeidiBaya

- 4 Version control code, text, ...



- 5 Stabilize computing environment and software



- 6 Publish your research outputs:  
Code, data, documents, ...



# Research software engineers

.. are people who combine professional software expertise with an understanding of research [\[https://researchsoftware.org/\]](https://researchsoftware.org/)

- Often people **who grew up in research** and liked computing and programming
- ... or people **who come from software development** drawn towards meaningful and impactful work of academia

## Resources

- [Society of Research Software Engineering](#)
- Recent conference: [RSECon 2023](#)
- <https://nordic-rse.org/>
- [Nordic-RSE unconference 2023](#)

# Our team

- We grew up in research (chemistry and biology)
- Interested in software development and computing
- Employed by IT department
- We are part of [Norwegian Research Infrastructure Services](#)

# What we offer

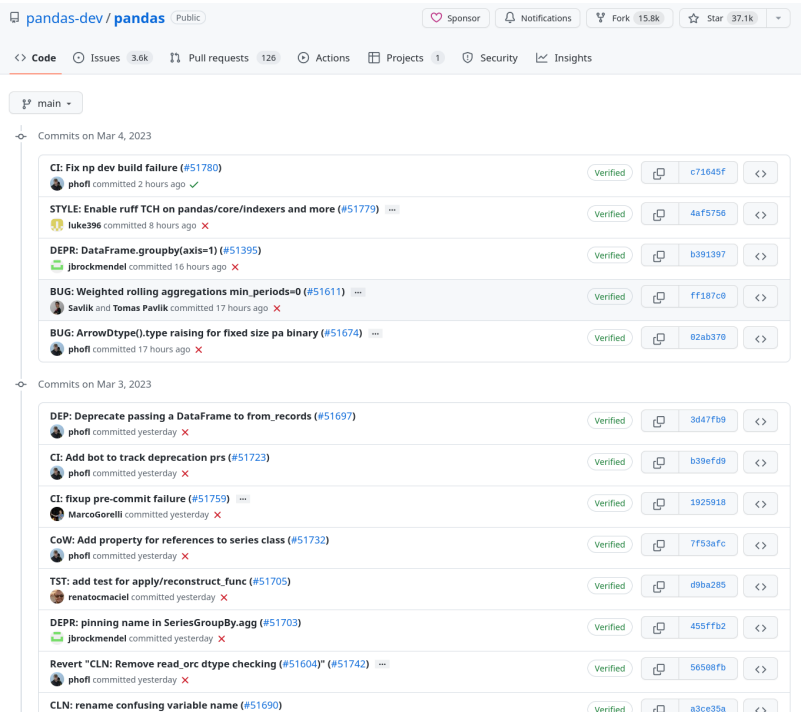
# Help with improving your scripts/code

- Code review (we discuss code in a constructive way)
- Making code more reusable
- Good practices for documentation

|                                                                    |                                                                  |
|--------------------------------------------------------------------|------------------------------------------------------------------|
| 205 - def groups(account):                                         | 222 + def collect_groups(account):                               |
| 206     l = shell_command(f"id -Gn {account}").split()             | 223     l = shell_command(f"id -Gn {account}").split()           |
| 207     # removing these two because we treat them separately      | 224     # removing these two because we treat them separately    |
| outside                                                            | outside                                                          |
| 208     l.remove(account)                                          | 225     l.remove(account)                                        |
| @@ -221,13 +238,12 @@ def dont_colorize(text, color):              |                                                                  |
| 221     return text                                                | 238     return text                                              |
| 222                                                                | 239                                                              |
| 223                                                                | 240                                                              |
| 224 - user = getpass.getuser()                                     | 241 + default_user = getpass.getuser()                           |
| 225                                                                | 242                                                              |
| 226                                                                | 243                                                              |
| 227     @click.command()                                           | 244     @click.command()                                         |
| 228 - @click.option(                                               | 245 + @click.option("-u", "--user", help=f"The username to check |
| 229     "-u", "--user", default=user, help=f"The username to check | (default: {default_user}).")                                     |
| 230     )                                                          | 246 + @click.option("-p", "--project", help=f"The allocation     |
|                                                                    | project.)"                                                       |
| 231     @click.option(                                             | 247     @click.option(                                           |
| 232         "--csv",                                               | 248         "--csv",                                             |
| 233         is_flag=True,                                          | 249         is_flag=True,                                        |
| @@ -238,7 +254,7 @@ def dont_colorize(text, color):                |                                                                  |
| 238     is_flag=True,                                              | 254     is_flag=True,                                            |
| 239     help="Disable colors.",                                    | 255     help="Disable colors.",                                  |
| 240     )                                                          | 256     )                                                        |
| 241 - def main(user, csv, no_colors):                              | 257 + def main(user, project, csv, no_colors):                   |
| 242     cf.update_palette({"blue": "#2e54ff"})                     | 258     cf.update_palette({"blue": "#2e54ff"})                   |
| 243     cf.update_palette({"green": "#08a91e"})                    | 259     cf.update_palette({"green": "#08a91e"})                  |
| 244     cf.update_palette({"orange": "#ff5733"})                   | 260     cf.update_palette({"orange": "#ff5733"})                 |
| @@ -247,12 +263,24 @@ def main(user, csv, no colors):              |                                                                  |

# Help with organising your code

- Git, GitHub, and GitLab
- Moving your work/project/code/data to Git
- Modularizing your code
- Organization of reusable Python/R notebooks



The screenshot displays the GitHub interface for the `pandas-dev/pandas` repository. At the top, navigation links include `Code`, `Issues` (3.6k), `Pull requests` (126), `Actions`, `Projects` (1), `Security`, and `Insights`. Below these, a section titled "Commits on Mar 4, 2023" lists several commits:

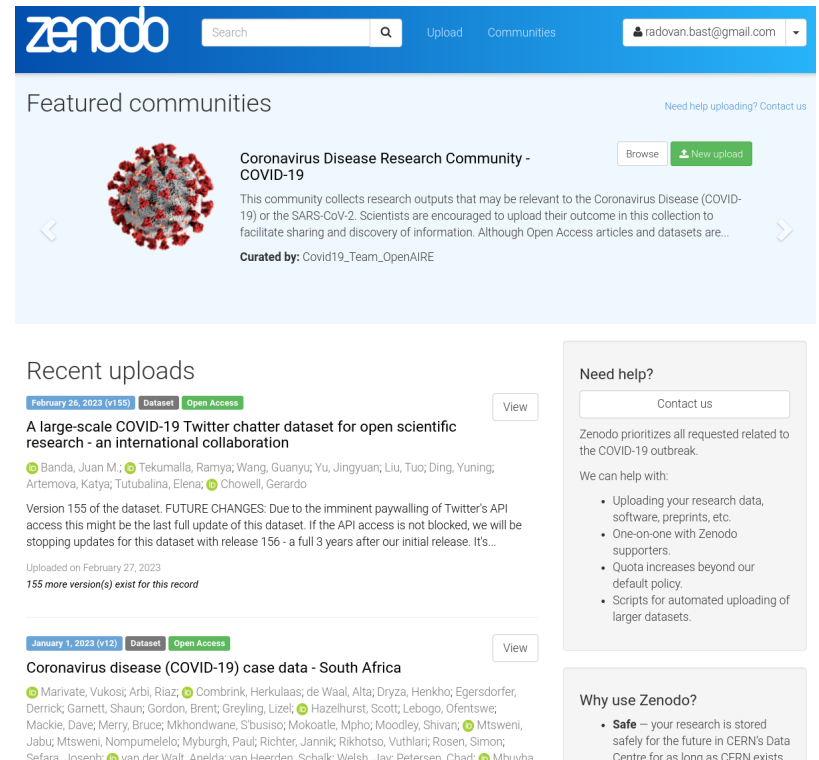
- `CI: Fix np dev build failure (#51780)` by `phofi` committed 2 hours ago. Status: Verified. SHA: `c71645f`.
- `STYLE: Enable ruff TCH on pandas/core/indexers and more (#51779)` by `luke396` committed 8 hours ago. Status: Verified. SHA: `4af5756`.
- `DEPR: DataFrame.groupby(axis=1) (#51395)` by `jbrockmenden` committed 16 hours ago. Status: Verified. SHA: `b391397`.
- `BUG: Weighted rolling aggregations min_periods=0 (#51611)` by `Savlik` and `Tomas Pavlik` committed 17 hours ago. Status: Verified. SHA: `ff187c0`.
- `BUG: ArrowDtype.type raising for fixed size pa binary (#51674)` by `phofi` committed 17 hours ago. Status: Verified. SHA: `82ab378`.

Below this, a section titled "Commits on Mar 3, 2023" lists more commits:

- `DEP: Deprecate passing a DataFrame to from_records (#51697)` by `phofi` committed yesterday. Status: Verified. SHA: `3d47fb9`.
- `CI: Add bot to track deprecation prs (#51723)` by `phofi` committed yesterday. Status: Verified. SHA: `b39efd9`.
- `CI: fixup pre-commit failure (#51759)` by `MarcoGorelli` committed yesterday. Status: Verified. SHA: `1925918`.
- `CoW: Add property for references to series class (#51732)` by `phofi` committed yesterday. Status: Verified. SHA: `7f53afc`.
- `TST: add test for apply/reconstruct_func (#51705)` by `renatocmaciel` committed yesterday. Status: Verified. SHA: `d9ba285`.
- `DEPR: pinning name in SeriesGroupBy.agg (#51703)` by `jbrockmenden` committed yesterday. Status: Verified. SHA: `455ff62`.
- `Revert "CLN: Remove read_orc dtype checking (#51604)" (#51742)` by `phofi` committed yesterday. Status: Verified. SHA: `56588fb`.
- `CLN: rename confusing variable name (#51690)` by `phofi` committed yesterday. Status: Verified. SHA: `a3ce35a`.

# Help with sharing your code

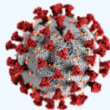
- Help with software licenses and open sourcing
- Publishing code
- Packaging and sharing software
- Containerization (Singularity, Docker)
- PyPI and Conda



The screenshot shows the Zenodo website interface. At the top is a blue header with the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. A user profile for 'radovan.bast@gmail.com' is visible in the top right. Below the header, the 'Featured communities' section highlights the 'Coronavirus Disease Research Community - COVID-19', which includes a description of its purpose and a 'New upload' button. The 'Recent uploads' section features two datasets: 'A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration' (dated February 26, 2023) and 'Coronavirus disease (COVID-19) case data - South Africa' (dated January 1, 2023). Each dataset entry includes a list of authors and a 'View' button. On the right side, there are two informational boxes: 'Need help?' with a 'Contact us' button and a description of Zenodo's priorities, and 'Why use Zenodo?' which lists benefits like safety and long-term preservation.

zenodo Search Upload Communities radovan.bast@gmail.com

Featured communities [Need help uploading? Contact us](#)

 **Coronavirus Disease Research Community - COVID-19** [Browse](#) [New upload](#)

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are...

**Curated by:** Covid19\_Team\_OpenAIRE

Recent uploads

[February 26, 2023 \(v155\)](#) [Dataset](#) [Open Access](#) [View](#)

**A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration**

[Banda, Juan M.](#); [Tekumalla, Ramya](#); [Wang, Guanyu](#); [Yu, Jingyuan](#); [Liu, Tuo](#); [Ding, Yuning](#); [Artemova, Katya](#); [Tutubalina, Elena](#); [Chowell, Gerardo](#)

Version 155 of the dataset. FUTURE CHANGES: Due to the imminent paywalling of Twitter's API access this might be the last full update of this dataset. If the API access is not blocked, we will be stopping updates for this dataset with release 156 - a full 3 years after our initial release. It's...

Uploaded on February 27, 2023  
*155 more version(s) exist for this record*

[January 1, 2023 \(v12\)](#) [Dataset](#) [Open Access](#) [View](#)

**Coronavirus disease (COVID-19) case data - South Africa**

[Marivate, Vukosi](#); [Arbi, Riya](#); [Combrink, Herkulaas](#); [de Waal, Alta](#); [Dryza, Henkko](#); [Egersdorfer, Derrick](#); [Garnett, Shaun](#); [Gordon, Brent](#); [Greyling, Lizet](#); [Hazelhurst, Scott](#); [Lebogo, Ofentswe](#); [Mackie, Dave](#); [Merry, Bruce](#); [Mkhondwane, S'busiso](#); [Mokoatle, Mpho](#); [Moodley, Shivan](#); [Mtsweni, Jabu](#); [Mtsweni, Nompumelelo](#); [Myburgh, Paul](#); [Richter, Jannik](#); [Rikhotso, Vuthlari](#); [Rosen, Simon](#); [Sefara, Joseph](#); [van der Walt, Anelda](#); [van Heerden, Schalk](#); [Welsh, Jay](#); [Petersen, Chad](#); [Mbuvha,](#)

**Need help?** [Contact us](#)

Zenodo prioritizes all requested related to the COVID-19 outbreak.

We can help with:

- Uploading your research data, software, preprints, etc.
- One-on-one with Zenodo supporters.
- Quota increases beyond our default policy.
- Scripts for automated uploading of larger datasets.

**Why use Zenodo?**

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.

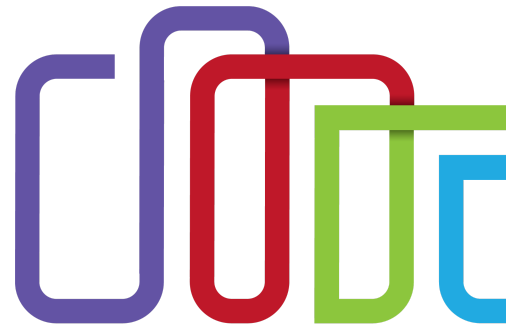
# Getting results sooner

- Improving scaling, CPU, and memory optimization
- Porting to GPU
- Moving from local computer to cloud or HPC
- Helping with running independent steps in parallel



[Midjourney, CC-BY-NC 4.0]

- Version control
- Collaboration using Git
- Testing
- Documentation
- Notebooks
- Modular code development
- Reproducible research
- Software licensing
- How to share and publish code



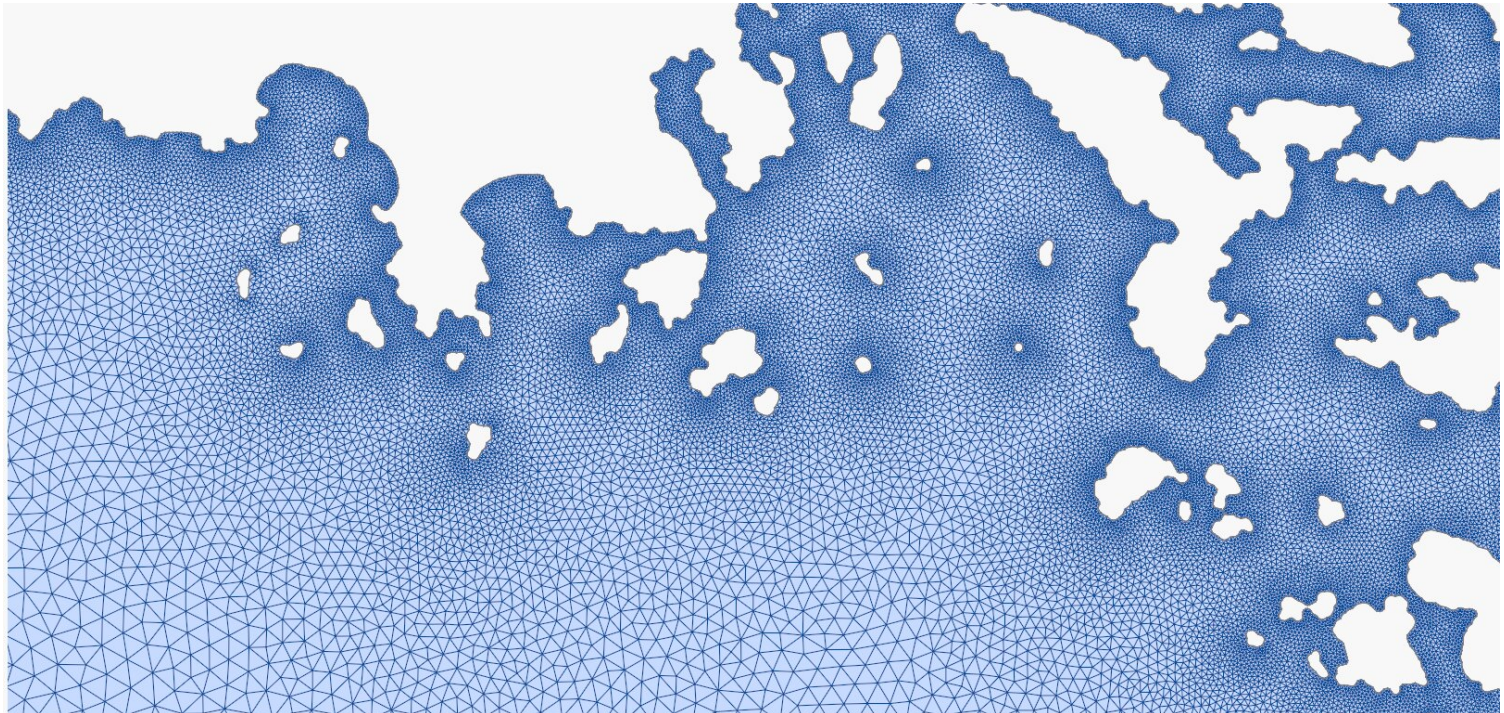
**CODE REFINERY**

**Typical format:** 6 half-days, [twice per year](#), online, free, live-streamed, recorded, archived asynchronous Q&A in collaborative document

**Lessons and recordings:**  
<https://coderefinery.org/lessons/>

# Use case 1: optimization

Oceanography: speed-up of grid mesh generation from days to seconds by a code rewrite from Matlab to Python+Rust using a more optimal algorithm



# Use case 2: web development

- [SMARTool: The Strategic Mastery of Russian Tool](#), Using GitHub Pages, JavaScript, and a CSV-based lightweight "database" to simplify maintenance, reuse, and contributions.
- [Constructicon: Searchable database of multiword grammatical constructions of Russian](#). Moved a project from a dedicated web server and custom database to GitHub Pages, JavaScript, and a YAML-based lightweight "database" to simplify maintenance, reuse, and contributions.

# Use case 3: rewrite instead of buying a 30 GB hard disk

Problematic if data is 30 GB big

```
result = 0.0
with open("data.txt", "r") as f:
    lines = f.readlines()
    for line in lines:
        result += analyze(line)
```

Better

```
result = 0.0
with open("data.txt", "r") as f:
    for line in f:
        result += analyze(line)
```

# Projects

<https://research-software.uit.no/projects/>

# Our vision

# Our vision

## Short term

- Provide consulting: good advice is **not** expensive
- Code review sessions
- Work on "smaller" projects and document use cases
- Focus on UiT
- Attract more projects

## Longer term

- Be part of funding applications
- Research groups have access to best in class RSE services
- Hire more staff who collaborate on projects
- "Proper" application procedure
- Going beyond UiT
- Career path opportunities

Office hours: Wednesday 13:00 - 15:00

<https://research-software.uit.no/contact/>

These slides: <https://doi.org/10.5281/zenodo.8368302>