

Research software engineering (RSE) på 10-15 minutter

1: Hva er problemet?

research-software.uit.no

2: Hvem er vi?

3: Hva tilbyr vi?

4: Hva har vi gjort?

5: Hva planlegger vi?



Hva er problemet?

Hva er problemet?

- Nesten alle forskere og studenter må kode
- De fleste er ikke utdannet for det
 - Kanskje et 2 ukers Python eller R kurs
 - Ofte ganske abstrakt
- Fokus på forskning og ikke program utvikling
- Flere og flere forventninger på forskerne
 - FAIR prinsipper (Findable, Accessible, Interoperable, Reusable)
 - Deling av kode og resultater (OpenSource / OpenScience)
 - Reproduserbar forskning (Git, tidy data, package management)
- Mange prosjekter har dårlig dokumentasjon -> **dør en langsom død**

Hva slags hjelp er tilgjengelig?

- Online
 - Mixed quality
 - Tar tid å finne
 - Ofte ikke aktuelt
- Kolleger
 - Hyppig i same situasjon
 - Mer erfaring med feltet, men vanlig også ingen utdanning
- Dokumentasjon
 - Ofte høye kvalitet
 - Mye jargon og vanskelig å forstå

Hva mangler?

- Enkelt tilgjengelig og direkte hjelp
- Kontakt som forstår spesielle problemer forskeren har
- Mentoring mht. programmering og utvikling

RSE @ UiT

RSE som løsning

- Kombinerer programvareekspertise med en forståelse av forskning
- Brukerstøtte for forsker og studenter
- Alle slags problemstillinger
- Lav terskel tilbud (epost, åpen skranke)

Hvem er vi?

- Aktuell 4 ansatte med kjemi og bio bakgrunn
 - Erfaring og interesse i koding
 - Alle del av ITA & NRIS
- > Kombinasjon av forståelse av forskning og profesjonelt programutvikling



Hva tilbyr vi?

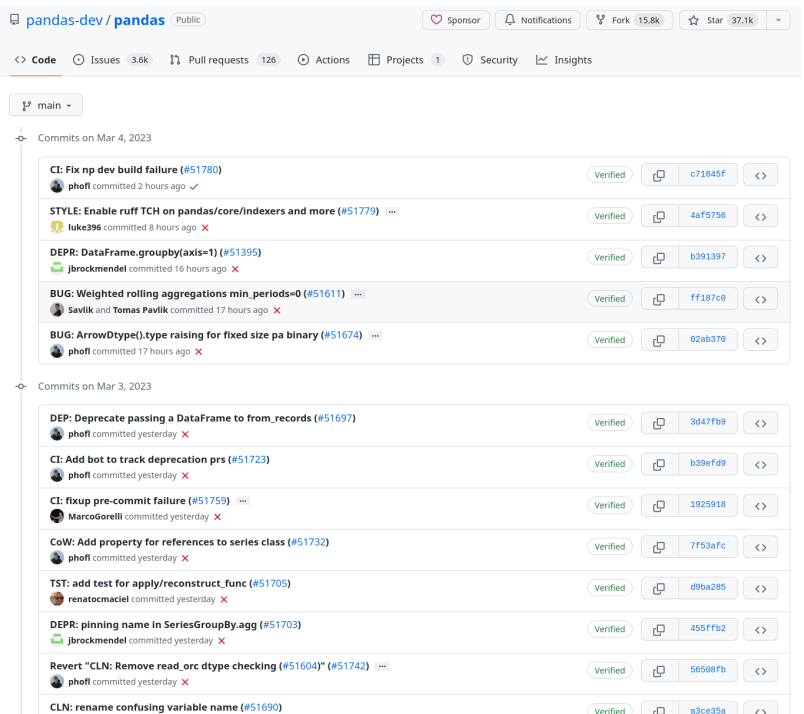
Hjelp med forbedring av skripter

- Code review: vi snakker om koden på en konstruktiv måte
- Lage koden mer gjenbrukbar
- Tips og anbefalinger med hesnikt til dokumentasjon

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 - (default: {user})."	246 + @click.option("-p", "--project", help=f"The allocation
231 -)	project.)"
232	247 @click.option(
233 "--csv",	248 "--csv",
234 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):	

Hjelp med kodeorganisering

- Git, GitHub, and GitLab
- Flytte prosjektet til Git
- Modularisering av kode og skripter
- Organisasjon av gjenbrukbare Python/R notebooker



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 dropdown menu shows `main`. The main content area is titled `Commits on Mar 4, 2023` and lists several commits:

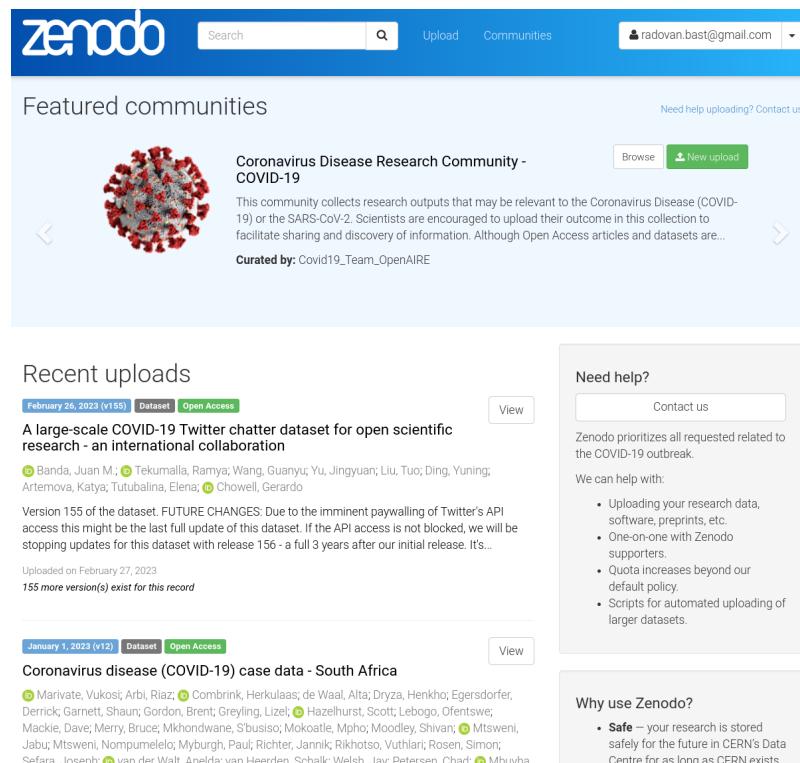
- CI: Fix np dev build failure (#51780)** by `phofi` committed 2 hours ago. Status: Verified. Commit hash: `c71645f`.
- STYLE: Enable ruff TCH on pandas/core/indexers and more (#51779)** by `luke396` committed 8 hours ago. Status: Verified. Commit hash: `4af5756`.
- DEPR: DataFrame.groupby(axis=1) (#51395)**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

Below this, the `Commits on Mar 3, 2023` section is visible, showing additional commits:

- DEP: Deprecate passing a DataFrame to from_records (#51697)** by `phofi` committed yesterday. Status: Verified. Commit hash: `3d47fb9`.
- CI: Add bot to track deprecation prs (#51723)** by `phofi` committed yesterday. Status: Verified. Commit hash: `b39efd9`.
- CI: fixup pre-commit failure (#51759)** by `MarcoGorelli` committed yesterday. Status: Verified. Commit hash: `1925918`.
- CoW: Add property for references to series class (#51732)** by `phofi` committed yesterday. Status: Verified. Commit hash: `7f53afc`.
- TST: add test for apply/reconstruct_func (#51705)** by `renatocmaciel` committed yesterday. Status: Verified. Commit hash: `d9ba285`.
- DEPR: pinning name in SeriesGroupBy.agg (#51703)** by `jbrockmendel` committed yesterday. Status: Verified. Commit hash: `455ffb2`.
- Revert "CLN: Remove read_orc dtype checking (#51604)" (#51742)** by `phofi` committed yesterday. Status: Verified. Commit hash: `56588fb`.
- CLN: rename confusing variable name (#51690)** by `phofi` committed yesterday. Status: Verified. Commit hash: `a3ce35a`.

Hjelp med deling av kode

- Hjelp med programvarelisenser og open source
- Publisering av kode
- Pakking og deling av programvare
- 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 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'. It includes a 3D virus model icon and a description: '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...'. A 'Curated by: Covid19_Team_OpenAIRE' note is present. To the right of the community name are 'Browse' and 'New upload' buttons. Below this, the 'Recent uploads' section lists two datasets. The first is 'A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration', dated February 26, 2023 (v155), with 'Dataset' and 'Open Access' tags. It lists authors: Banda, Juan M.; Tekumalla, Ramya; Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Artemova, Katya; Tutubalina, Elena; Chowell, Gerardo. A note mentions '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...'. The second dataset is 'Coronavirus disease (COVID-19) case data - South Africa', dated January 1, 2023 (v12), also with 'Dataset' and 'Open Access' tags. It lists a large number of authors including Marivate, Vukosi; Arbi, Riaz; Combrink, Herkulaas; de Waal, Alta; Dryza, Henkx; Eggersdorfer, 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. On the right side of the page, there is a 'Need help?' section with a 'Contact us' button and a list of services: 'Uploading your research data, software, preprints, etc.', 'One-on-one with Zenodo supporters', 'Quota increases beyond our default policy', and 'Scripts for automated uploading of larger datasets'. Below that is a 'Why use Zenodo?' section with a bullet point: 'Safe — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.'

Hjelp med akselerere resultater

- Forbedre skalering, CPU, and memory optimization
- Overføring til GPU
- Flytte fra lokal datamaskin til sky eller HPC
- Hjelp med parallellisering av uavhengige trinn



[Midjourney, CC-BY-NC 4.0]

Hva har vi gjort?

Noen eksempler av vårt arbeid

Mer på <https://research-software.uit.no/projects/> og
<https://research-software.uit.no/blog/>

Hjelp med utvikling av en GUI

- Forsker med en eksisterende Python programvare
- **Koden fungerer**, men bare command line interface -> vanskelig å bruke
- Øke brukervennlighet med en grafisk interface
- Sammen med forskeren utviklet vi et PyQt rammeverk

 Screenshot of <https://pythonpyqt.com/pythonpyqtlineedit/>

Hjelp med R programmering

- Biologi master student
- Har problem med R skripter for masteroppgave
- En-til-en code review
- Takle akutt problemer
- Men også diskutere generell skriptstruktur

Hjelp med transkribere og modernisere av gammel kode

- Visual Basic 6 programvare
- Mer enn 15000 linjer
- Stor andel av kodeduplikat
- Skissere en nytt struktur på object-oriented python
- Følge opp med jevnlige code review møter

 Taken from
<https://devhumor.com/med-found-some-of-my-old-code>

Hva planlegger vi?

Vår visjon

Kortsiktig

- Tilby rådgivning: Godt råd er ikke dyrt
- Code reviews
- Arbeide med "mindre" prosjekter og dokumentere brukstilfeller
- Trekke til seg flere prosjekter
- Bli kjent på veileder- og studentnivå

Langsiktig

- Delta i søknader om finansiering
- Tilby førsteklasses RSE-tjenester til forskningsgrupper
- Ansette mer personale som samarbeider på prosjekter
- Etablere en "riktig" søknadsprosedyre
- Gå utover UiT
- Karrieremuligheter

Åpent kontor: Onsdager 13-15

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