

Preprocessed inputs for CMIP7 Ozone concentration model runs

Manage topics

git: 3.1 TiB, lfs: 0 B

163 commits 2 branches 0 tags 3.1 TiB

Forgejo-aneksajo: a git-annex/DataLad forge
Apply software development practices to your data-involving projects



My vision of repeatable, reproducible, traceable, discoverable research

Every publication should have an accompanying publishable artifact that fully describes the data, code and computational environment used to produce it. Anyone (with sufficient determination) should be able to go from e.g. a plot in a paper back through the processes used to produce it, be able to follow any intermediate steps, and arrive at the data, or even the data acquisition process.

My vision of repeatable, reproducible, traceable, discoverable research

Three fundamental parts to keep track of:

- Code
- Data
- ... and the ambient environment in which you work

Not just important for publication, but also for yourself or others revisiting a project in the future.

My vision of repeatable, reproducible, traceable, discoverable research

ATRS Issues Pull requests Milestones Explore

ICE-4 / CMIP7-O3-input-preparations Private

Code Issues Pull requests Projects Releases Packages Wiki Activity Actions Settings

Preprocessed inputs for CMIP7 Ozone concentration model runs

Manage topics

163 commits 2 branches 0 tags 3.1 TiB

main Find a file Add file HTTPS SSH https://atris.fz-juelich.de/ICE-4/CMIP7-O3-input-preparations

Search code... Exact Q

Matthias Riße 2100da6bd6 Use checksums for export 5 months ago

.datalad	[DATALAD] new dataset	10 months ago
code	Update scripts to UOEXETER-CMIP-2-2-0	5 months ago
inputs	Add UOEXETER-CMIP-2-2-0 inputs	5 months ago
Makefiles.d/outputs	Update scripts to UOEXETER-CMIP-2-2-0	5 months ago
outputs	[DATALAD RUNCMD] Prepare pi-control stratospheric aerosol optical property LW data	5 months ago
.gitattributes	[DATALAD RUNCMD] Initialize pixi project	10 months ago
.gitignore	Replace height with pressure in optLW/optSW	6 months ago
Makefile	Use checksums for export	5 months ago
pixi.lock	Add script to generate optSW and optLW data files	8 months ago
pixi.toml	Fix incorrect CMIP era in filenames	7 months ago
README.md	Update README	6 months ago

README.md

But you are here to talk about Forgejo-aneksajo?!

If we are already looking at software development practices to organize research projects, let's look at established practices of software development for how to collaborate too.

But you are here to talk about Forgejo-aneksajo?!

The "forge" model as a central "collaboration hub" for projects:

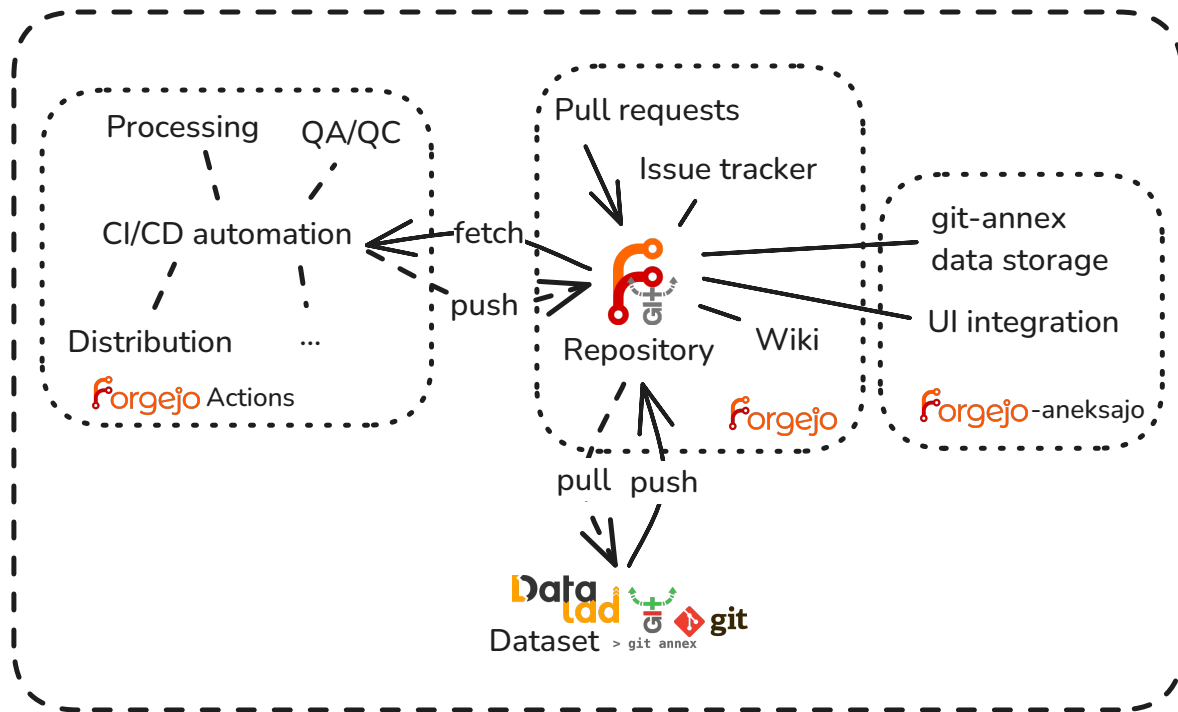
- A place for the canonical repository
- A place to discuss the direction of the project
- Issue trackers for bug reports and feature requests
- A place for change proposals and their review
- An entrypoint into each project

Not the only way to organize, but GitHub has shown that it can work well.

What is Forgejo-aneksajo?

- A soft-fork of Forgejo
 - Forgejo is a self-hostable, GPL'ed git forge
 - Primary goal: integrate git-annex into Forgejo as intuitively as possible
- => Therefore Forgejo-aneksajo is a git-annex/DataLad forge

Feature Overview



How did this come to be?

Ca. April 2022:

- At ICE-4 we started with an instance of G-Node's gogs:
<https://github.com/G-Node/gogs>
- Already dated at that time
- More than just git-annex support
- Use of merging made the history a bit hard to follow
- ... still the backbone of <https://gin.g-node.org/>

Ca. November 2022:

- Searched for an alternative and found a soft-fork of Gitea at NeuroPoly, Université de Montréal: <https://github.com/neuropoly/gitea>
- only git-annex support as the goal
- relatively small and easy to grasp patchset
- shoutout to Mick, kousu on GitHub!
- ... unfortunately, maintenance stopped ~September 2023

How did this come to be?

=> Forked + rebased on Forgejo after Distribits 2024

- Started at v7, now at v13
- Regular rebase with relatively little effort

New developments since then:

- Git-annex p2phttp integration
- Uploads through web interface to annex
- Git-annex-enabled templates
- PRs with annexed content
- Some improvements to the UI integration of annexed content
- A bunch of bug fixes here and there

Git-annex-enabled templates

forgejo / yoda-template Private Template








<> Code Issues Pull requests Projects Releases Packages Wiki Activity Settings

No description
Manage topics

3 commits 2 branches 0 tags 45 KiB

main Find a file Add file Use this template HTTP SSH http://localhost:3000/forgejo/yoda-template.git

Search code... Exact

 Matthias RiBe 17921f820a Template-ify	2 hours ago
 .datalad Template-ify	2 hours ago
 .forgejo Template-ify	2 hours ago
 code Apply YODA dataset setup	3 hours ago
 .gitattributes Apply YODA dataset setup	3 hours ago
 CHANGELOG.md Apply YODA dataset setup	3 hours ago
 README.md Template-ify	2 hours ago

README.md

Project \${REPO_NAME}

Dataset structure

- All inputs (i.e. building blocks from other sources) are located in `inputs/`.
- All custom code is located in `code/`.

Git-annex' p2phttp

- Git-annex operations via http(s)
- Fully integrated with Forgejo's access right management:
- Read/Write/Admin permissions apply to the annex
- Username/password auth, or better yet, access tokens
- Just configure a https remote and use a up-to-date git-annex

Useful if:

- You can't open port 22 / ssh for some reason
- Your clients are forbidden to establish outgoing ssh connections

Useful Forgejo features: git-credential-oauth integration

```
$ git credential-oauth configure
```

=> all future git operations against a Forgejo(-aneksajo) instance via https will use an OAuth2 authorization flow and cache the received access token for 6 hours

- on-demand short-lived access to repositories
- works for plain-git and git-annex p2phttp operations
- Forgejo does not support the OAuth2 device flow for headless systems yet, maybe a topic for the hackathon

Useful Forgejo features: custom file previews

ATRIS Issues Pull requests Milestones Explore

MeteoCloud / ERA5-grib-2024

<> Code Issues Pull requests Projects Releases Packages Wiki Activity Actions Settings

main ERA5-grib-2024 / 01 / 2024010100_ml.grb

Matthias RiBe ae450e1676 [DATALAD RUNCMD] Update dataset last year

1.7 GiB (Stored with Git Annex) Raw Permalink History

```
File format : GRIB2
-1 : Institut Source T Steptype Levels Num Points Num Dtype : Parameter name
 1 : ECMWF unknown v instant 1 1 721200 1 P16 : z
 2 : ECMWF unknown v instant 137 2 721200 1 P16 : t
 3 : ECMWF unknown v instant 137 2 721200 1 P16 : q
 4 : ECMWF unknown v instant 137 2 721200 1 P16 : w
 5 : ECMWF unknown v instant 137 2 721200 1 P16 : vo
 6 : ECMWF unknown v instant 1 1 721200 1 P16 : lnspl
 7 : ECMWF unknown v instant 137 2 721200 1 P16 : d
 8 : ECMWF unknown v instant 137 2 721200 1 P16 : u
 9 : ECMWF unknown v instant 137 2 721200 1 P16 : v
10 : ECMWF unknown v instant 137 2 721200 1 P16 : o3
11 : ECMWF unknown v instant 137 2 721200 1 P16 : clwc
12 : ECMWF unknown v instant 137 2 721200 1 P16 : ciwc
13 : ECMWF unknown v instant 137 2 721200 1 P8 : cc

Grid coordinates :
 1 : lonlat : points=721200 (1200x601)
      lon : 0 to 359.7 by 0.3 degrees_east circular
      lat : 90 to -90 by -0.3 degrees_north

Vertical coordinates :
 1 : hybrid : levels=1
      lev : 1 level
      available : vct
 2 : hybrid : levels=137
      lev : 1 to 137 by 1 level
      available : vct

Time coordinate :
      time : 1 step
      RefTime = 2024-01-01 00:00:00 Units = hours Calendar = proleptic_gregorian
      YYYY-MM-DD hh:mm:ss YYYY-MM-DD hh:mm:ss YYYY-MM-DD hh:mm:ss YYYY-MM-DD hh:mm:ss
      2024-01-01 00:00:00
```

Useful Forgejo features: push-to-create

```
$ datalad create dataset  
$ cd dataset  
$ git remote add origin <url>  
$ datalad push --to origin
```

=> Repository is created on demand iff your user is authorized to create a repository at the specified URL

Forgejo Actions: not just "CI"

Forgejo interface for repository **MeteoCloud / ERA5**. The top navigation bar includes links for ATRIS, Issues, Pull requests, Milestones, and Explore. The repository name **MeteoCloud / ERA5** is displayed, along with icons for Watch (0), Star (0), and Fork (0). The main navigation bar includes links for Code, Issues, Pull requests, Projects, Releases, Packages, Wiki, Activity, **Actions**, and Settings.

The **Actions** tab is active, showing a list of workflows. A sidebar on the left lists "All workflows" and "update-dataset.yml". The main content area displays a table of workflow runs:

	Actor	Status
[DATALAD RUNCMD] Update dataset update-dataset.yml #59 -Scheduled	main	📅 4 days ago 🕒 36m8s
[DATALAD RUNCMD] Update dataset update-dataset.yml #58 -Scheduled	main	📅 last week 🕒 36m10s
[DATALAD RUNCMD] Update dataset update-dataset.yml #57 -Scheduled	main	📅 two weeks ago 🕒 41m18s
[DATALAD RUNCMD] Update dataset update-dataset.yml #56 -Scheduled	main	📅 3 weeks ago 🕒 41m7s
[DATALAD RUNCMD] Update dataset update-dataset.yml #55 -Scheduled	main	📅 last month 🕒 35m49s
[DATALAD RUNCMD] Update dataset update-dataset.yml #54 -Scheduled	main	📅 last month 🕒 37m29s
.forgejo/workflows/update-dataset.yml update-dataset.yml #53 -Commit 5c07897361 pushed by m.risse	main	📅 last month 🕒 38m40s

Forgejo Actions: not just "CI"

ATRS Issues Pull requests Milestones Explore

MeteoCloud / ERA5

Watch 0 Star 0 Fork 0

Code Issues Pull requests Projects Releases Packages Wiki Activity **Actions** Settings

✓ [DATALAD RUNCMD] Update dataset
Commit b36fab3763 pushed by forgejo-actions `main`
Workflow update-dataset.yml

Re-run all jobs

✓ update-dataset 35m30s

update-dataset
Success


- > ✓ Set up job -36s
- > ✓ Set git credentials 0s
- > ✓ Make git use SSH 0s
- > ✓ Make new repositories public by default 0s
- > ✓ git config --global annex.private "true" 0s
- > ✓ actions/checkout@v4 2s
- > ✓ Setup pixi 12s
- > ✓ pixi run datalad get --no-data --recursive . 11m8s
- > ✓ pixi run datalad run -m "Update dataset" --assume-ready both "pixi run --clean-env code/update-dataset" 9m19s
- > ✓ pixi run datalad siblings configure --recursive -s origin --url 'https://atris.fz-juelich.de/MeteoCloud/%NAME.git' 28s
- > ✓ pixi run datalad run -m "Update subdataset URLs" "pixi run --clean-env code/update-subdataset-urls" 2m7s
- > ✓ pixi run datalad push --recursive --to origin 12m37s
- > ✓ Complete job 4s

Forgejo is an active and solid basis


October 14, 2024 - October 14, 2025


Period: 1 year ▾


Overview



2341 active pull requests


1357 active issues

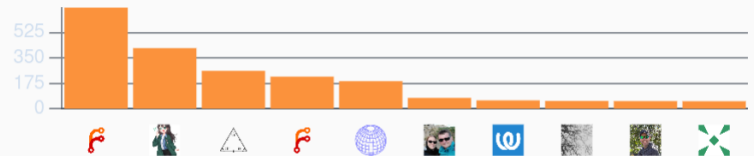
 2271
Merged pull requests


 70
Proposed pull requests

 853
Closed issues

 1129
New issues

Excluding merges, **153 authors** have pushed **1369 commits** to forgejo and **2740 commits** to all branches. On forgejo, **3591 files** have changed and there have been **139876 additions** and **75475 deletions**.



 31 releases published by 3 users

Some stats on the patchset

```
$ git diff --shortstat v12.0.4 v12.0.4-git-annex0  
76 files changed, 4366 insertions(+), 119 deletions(-)
```

```
$ git diff --dirstat v12.0.4 v12.0.4-git-annex0  
3.9% assets/  
4.9% modules/annex/  
3.4% modules/  
8.9% routers/web/repo/  
4.2% services/  
65.0% tests/integration/
```

Lessons from maintaining a soft-fork

- Do not touch the database schema!
- Writing code in an "insertion-only" fashion makes maintenance easier
- A rebase-based workflow keeps your patches together
- Mergraf and rerere remove some tediousness from conflict resolution

Try it out

Demo Instance: <https://demo.forgejo-aneksajo.org/>

- open registration
- enough storage space to play with it
- resets nightly

Run your own, container images available on Codeberg:

<https://codeberg.org/forgejo-aneksajo/-/packages/container/forgejo-aneksajo/forgejo>

Source repository:

<https://codeberg.org/forgejo-aneksajo/forgejo-aneksajo>

Get in contact: m.risse@fz-juelich.de, [@matrssi:matrix.org](https://matrix.org),
[@matrssi@mastodon.social](https://mstdn.social)

Future outlook

Profit from Forgejo improvements:

- Federation! A long way to go, but consistently being worked on.
- Storage quotas - Have gotten some important fixes in v13
- Forgejo Actions runner received a lot of attention lately
- No longer considered "alpha quality", ready for production as of 2025-08-17

Upstream git-annex support into Forgejo proper?

And: Thank you to everyone working on Forgejo!

The End