Karlsruhe Institute of Technology

Institute of Biomedical Engineering
www.ibt.kit.edu

Automated Software Metadata Conversion and Publication Based on CodeMeta

Marie Houillon, Jochen Klar, Tomas Stary, Axel Loewe

Introduction

Reproducible research requires publication of datasets with appropriate metadata. The automation through continuous integration and deployment (CI/CD) pipelines in modern version control systems can make this process maintainable and sustainable [1]. We present pipelines that are successfully used for the openCARP project [2] and can be readily adapted for other research software projects.

Prerequisites

- Repository in git version control
- Instance of GitLab to host the repo
- GitLab runner configured for docker

Optional:

- RADAR4KIT credentials
- websites in Grav CMS (MD + bibtex)

Metadata:

- `codemeta.json` (maintained)
- Citation File Format (generated)
- DataCite (generated)

Using in Your Project

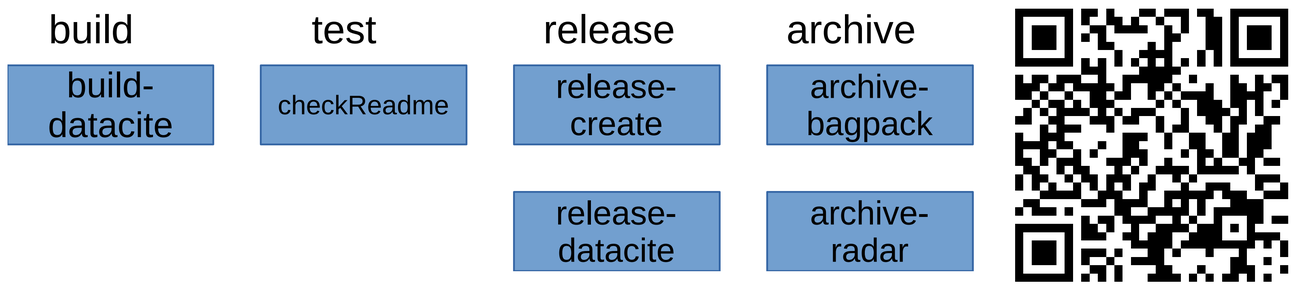
- In your GitLab repository register (or verify existence) of a gitlab-runner able to run docker (Settings → CI/CD → Runners)
- Copy minimal example of CI/CD pipelines from `.gitlab/` and `.gitlab-ci.yml` in:

  `git clone https://git.opencarp.org/openCARP/openCARP-CI.git`
- Generate Access Tokens for `api` and `write_repository`, copy it and paste into repository Settings → CI/CD → Variables as:

  `PUSH_TOKEN=<copied value>, PRIVATE_TOKEN=$PUSH_TOKEN`
- Example of Continuous Integration (CI) job:

```yaml
build-datacite:
  stage: build
  image: python:3.9
  before_script:
  - pip install git+https://git.opencarp.org/openCARP/openCARP-CI.git
  script:
  - create_datacite
  artifacts:
    paths:
    - $DATACITE_PATH
    expire_in: 2 hrs
```

- Workflow: create `pre-vX.Y` tag, the CI job updates metadata and commits them with the `vX.Y` tag
- The jobs can optionally submit to www.radar-service.eu storage and update information on the Grav CMS websites



Components

Table: functions for pipeline jobs

script	makes
create_cff	Citation File Format (CFF) metadata file
prepare_release	update version and dateModified in metadata
create_release	release in GitLab
create_datacite	datacite metadata from `codemeta.json` file
create_bag	BagIt package
create_bagpack	similar BagIt with DataCite XML
prepare_radar	empty archive in Radar service
create_radar	archive and upload it to Radar service
run_markdown_pipeline	update Grav CMS websites
run_bibtex_pipeline	compile bibtex file to put on the website
run_docstring_pipeline	extract docstrings from Python scripts

Conclusions

OpenCARP-CI provides tools for automatic software publication integrated in CI/CD pipelines. After the initial setup, the user maintains a single metadata file `codemeta.json`. The releases and supporting files are archived automatically for every new version.

Acknowledgements

- DFG-507828355
- DFG-391128822
- EuroHPC-955495
- KIT FAIR Research Software

References

[1] Anzt et al., F1000Research (2021), https://doi.org/10.12688/f1000research.23224.2

[2] Plank et al., Computer Methods and Programs in Biomedicine (2020), https://doi.org/10.1016/j.cmpb.2021.106223

[3] Bach et al., Bausteine Forschungsdatenmanagement (2022), https://doi.org/10.17192/bfdm.2022.1.8368

CaMo @KIT

marie.houillon@kit.edu