FACILE-RS: archival and long term preservation of research software repositories made easy

Jochen Klar [1], **Marie Houillon** [2], **Axel Loewe** [2], **Tomas Stary** [2], and **Ziad Boutanios**[2]

1 Independent Software Developer, Germany 2 Karlsruhe Institute of Technology, Germany

Summary

The FACILE-RS Python package allows to perform tasks around the archival and long term preservation of research software repositories. It consists in a set of Python scripts which facilitate the maintenance of software metadata by automating the generation of metadata in various formats from a unique metadata file that is maintained manually. FACILE-RS also makes it easier to publish and archive software releases according to the Open Science paradigm and the FAIR (Findable, Accessible, Interoperable, Reusable) principles for Research Software, by offering tools to automate the creation of releases and the upload to persistent research data repositories.

In particular, FACILE-RS allows to:

- create a DataCite record based on CodeMeta files present in repositories,
- create a CFF (Citation File Format) file from CodeMeta files,
- create archive packages in the BagIt or BagPack formats,
- create a release on the Dev. lopment platform GitLab using the GitLab API,
- archive software releases using the RADAR service,
- use content from markdown files, bibtex files, or python docstrings to create web pages within the Grav CMS.

While the scripts can be run manually, they are designed to be used within GitLab CI/CD or another workflow automation system, in order to automate the process of maintaining metadata and creating persistent software releases.

Statement of need

Research software development is a fundamental aspect in research (Anzt et al., 2021), and it is now acknowledged that the FAIR principles (Findable, Accessible, Interoperable, Reproducible; (Wilkinson et al., 2016)), historically established for research data, should also be applied to research software (Chue Hong et al., 2021). In particular, reproducible research requires that software and its associated metadata can be found easily by both machines and humans, and that they are retrievable via standardised protocols. In this context, several metadata standards are widely used across the scientific community:

- the Citation File Format (CFF) (Druskat et al., 2021) aims to indicate to users how to cite a software package
- DataCite (DataCite Metadata Working Group, 2021) is a standard Metadata schema for archiving digital assets
- CodeMeta (Jones et al., 2017) is an extension of schema.org created to standardize the exchange of software metadata across repositories and organizations

All of these standards serve specific purposes and several of them are required to cover the whole software lifecycle. However, research software developers should ideally not be burdened with maintaining multiple metadata files in different formats and largely overlapping content. This poses a risk to data consistency and to adoption of good software publication practices.

FACILE-RS aims to overcome this difficulties by making it easy to create and maintain the metadata associated to research software, as well as to publish it according to the FAIR principles.

Functionality

Refer to Table 1

Script	Functionality
create_cff	generates Citation File Format (CFF) metadata file
prepare_release	updates *version* and *dateModified* in metadata
create_release	creates release in GitLab
create_datacite	generates DataCite metadata file
create_bag	creates BagIt package
create_bagpack	adds DataCite XML to BagIt
prepare_radar	reserves DOI on RADAR
create_radar	creates archive and uploads it to RADAR
run_markdown_pipeline	updates Grav CMS website
run_bibtex_pipeline	treats BibTex file for publications on website
run_docstring_pipeline	extracts docstrings from Python scripts

Table 1: Components of openCARP-CI

Figures

Figures can be included like this: Caption for example figure. and referenced from text using section .

Figure sizes can be customized by adding an optional second parameter: Caption for example figure.

Acknowledgements

We acknowledge contributions from Brigitta Sipocz, Syrtis Major, and Semyeong Oh, and support from Kathryn Johnston during the genesis of this project.

References

Anzt, H., Bach, F., Druskat, S., Löffler, F., Loewe, A., Renard, B., Seemann, G., Struck, A., Achhammer, E., Aggarwal, P., Appel, F., Bader, M., Brusch, L., Busse, C., Chourdakis, G., Dabrowski, P., Ebert, P., Flemisch, B., Friedl, S., … Weeber, R. (2021). An environment for sustainable research software in Germany and beyond: Current state, open challenges, and call for action. *F1000Research*, *9*(295). https://doi.org/10.12688/f1000research.23224.2

Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A.-L., Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2021). *FAIR principles for research software (FAIR4RS principles)*. https://doi.org/10.15497/RDA00068

DataCite Metadata Working Group. (2021). *DataCite metadata schema documentation for the publication and citation of research data and other research outputs*. https://doi.org/doi.org/10.14454/3w3z-sa82

Druskat, S., Spaaks, J. H., Chue Hong, N., Haines, R., Baker, J., Bliven, S., Willighagen, E., Pérez-Suárez, D., & Konovalov, O. (2021). *Citation File Format* (Version 1.2.0). https://doi.org/10.5281/zenodo.5171937

Jones, M. B., Boettjiger, C., Mayes, A. C., Smith, A., Slaughter, P., Niemeyer, K., Gil, Y. G., Fenner, M., Nowak, K., Hahnel, M., Coy, L., Allen, A., Crosas, M., Sands, A., Hong, N. C., Cruse, P., Katz, D., & Goble, C. (2017). *CodeMeta: An exchange schema for software metadata. Version 2.0*. https://doi.org/10.5063/schema/codemeta-2.0

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B. da, Bourne, P. E., & others. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18