HERMES | <HMC>

HELMHOLTZ RICH METADATA SOFTWARE PUBLICATIONS

deRSE23 - Conference for Research Software Engineering in Germany, Paderborn, 2023-02-20

HERMES: Easing the path to FAIR software publications

Stephan Druskat

German Aerospace Center (DLR), Institute for Software Technology

with Oliver Bertuch[1], Oliver Knodel[2], Michael Meinel[3], Guido Juckeland[2], Tobias Schlauch[3], Jeffrey Kelling[2] , David Pape[2]

[1]Forschungszentrum Jülich

[2]Helmholtz-Zentrum Dresden-Rossendorf

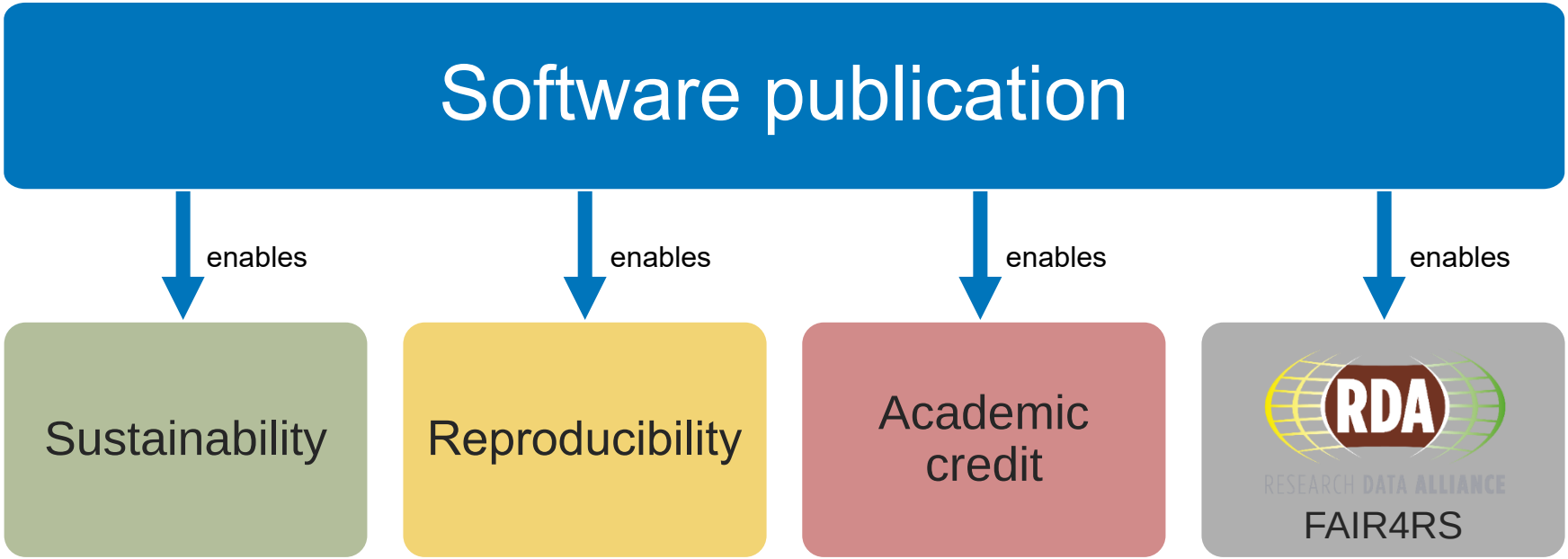
[3]German Aerospace Center (DLR)

DOI 10.5281/zenodo.7654819

License CC-BY-4.0 International

Motivation

Software publication



State of the art

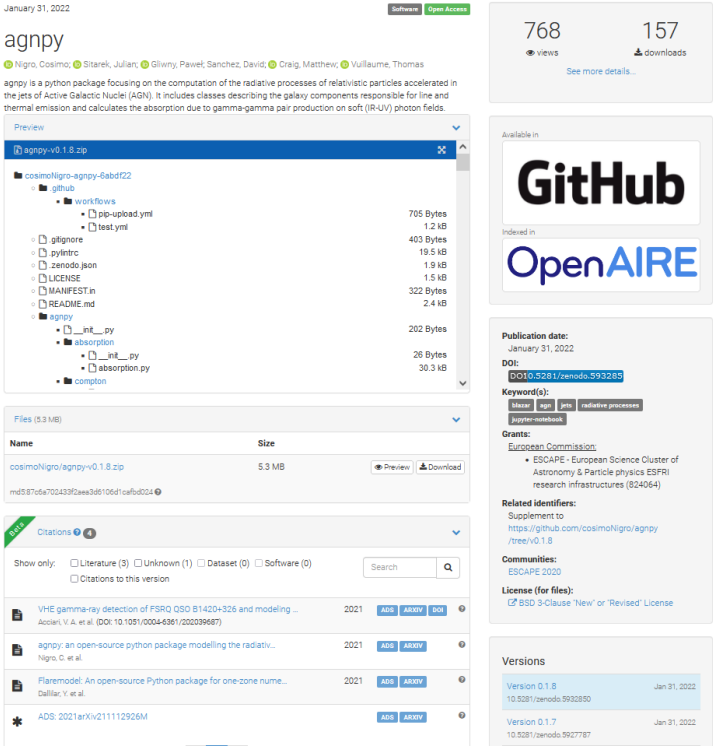
Software publication

```
model = getattr(spectra, spectrum_dict["type"])

if norm.unit in (u.Unit("erg"), u.Unit("erg cm-3")) and norm_type != "integral":
    raise NameError(
        "Normalisation different than 'integral' available only for 'spectrum_norm' in cm-3"
    )

# check the units of the normalisation
# cm-3 is the only one allowing more than one normalisation type
if norm.unit == u.Unit("cm-3"):
    if norm_type == "differential":
        final_model = model(norm, **spectrum_dict["parameters"])
    elif norm_type == "gamma=1":
        final_model = model.from_norm_at_gamma_1(
```

```
{
    "@context": "https://doi.org/10.5063/schema/codemeta-2.0",
    "@type": "SoftwareSourceCode",
    "license": "https://spdx.org/licenses/BSD-3-Clause",
    "codeRepository": "https://github.com/cosimoNigro/agnpy",
    "contIntegration": "https://github.com/cosimoNigro/agnpy/actions",
    "dateCreated": "2019-12-17",
    "datePublished": "2022-01-31",
    "dateModified": "2021-08-02",
    "downloadUrl": "https://github.com/cosimoNigro/agnpy/releases/tag/v0.1.6",
    "issueTracker": "https://github.com/cosimoNigro/agnpy/issues",
    "name": "agnpy",
    "version": "0.1.8",
    "identifier": "10.5281/zenodo.4055175",
    "description": "agnpy is a python package focusing on the computation of the
    "applicationCategory": "astrophysics",
    "funding": "ESCAPE EU H2020 824064",
    "developmentStatus": "active",
    "isPartOf": "https://www.astropy.org/affiliated/#affiliated-packages",
```

Software metadata + (software artifacts)

PID + landing page (metadata)



HERMES project

Scope

- 07/2021-06/2023
- Support RSEs in automatedly publishing software with rich metadata

Software publications with rich metadata

State of the art, automated workflows and HERMES concept

Druskat, Stephan [1,*], Bertuch, Oliver [2], Juckeland, Guido [3], Knodel, Oliver [3], and Schlauch, Tobias [1]

[1]Deutsches Zentrum für Luft- und Raumfahrt e. V.
[2]Forschungszentrum Jülich GmbH
[3]Helmholtz-Zentrum Dresden-Rossendorf e. V.
[*]Corresponding Author (software-metadata.pub)

January 25, 2022
Version 1

arXiv:2201.09015v1 [cs.SE] 22 Jan 2022

Abstract

To satisfy the principles of FAIR software, software sustainability and software citation, research software must be formally published. Publication repositories make this possible and provide published software versions with unique and persistent identifiers. However, software publication is still a tedious, mostly manual process.

To streamline software publication, HERMES, a project funded by the Helmholtz Metadata Collaboration, develops automated workflows to publish research software with rich metadata.

The tooling developed by the project utilizes continuous integration solutions to retrieve, collate, and process existing metadata in source repositories, and publish them on publication repositories, including checks against existing metadata requirements. To accompany the tooling and enable researchers to easily reuse it, the project also provides comprehensive documentation and templates for widely used CI solutions. In this paper, we outline the concept for these workflows, and describe how our solution advances the state of the art in research software publication.

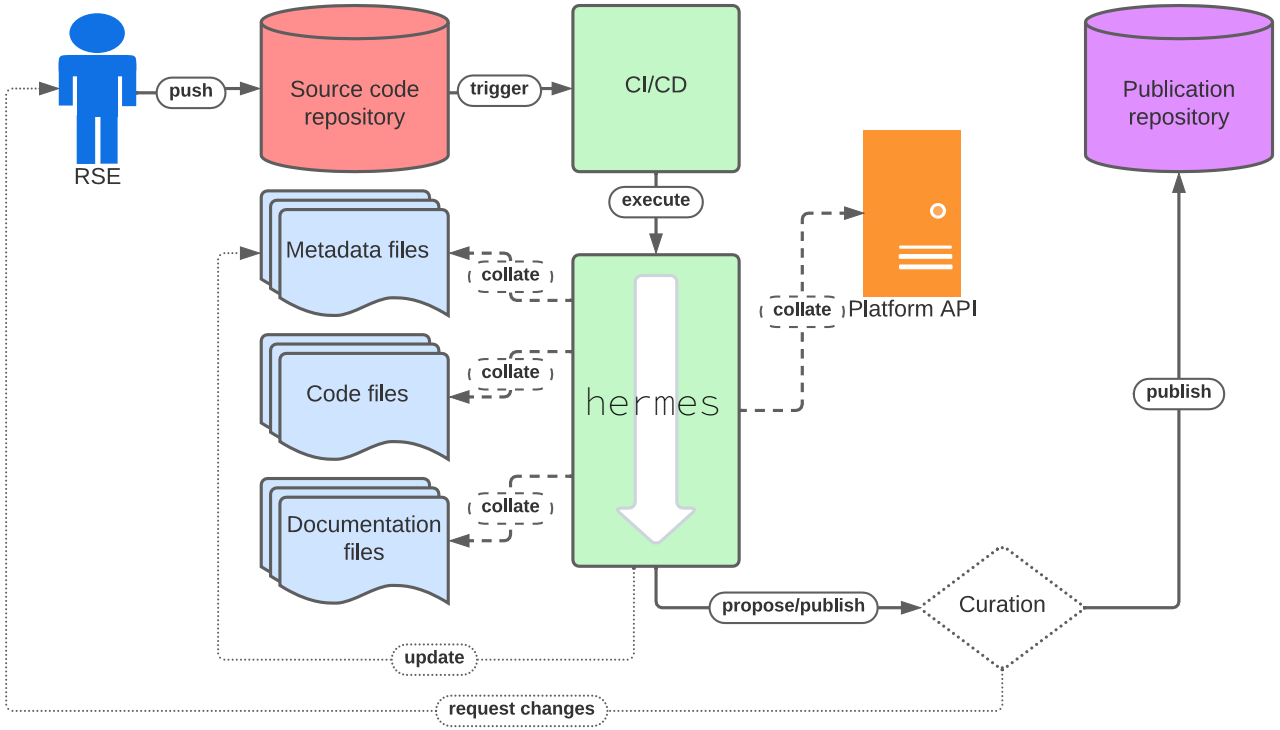
This work is licensed under CC BY 4.0.

The user receives assistance in depositing software in an automated fashion. This may be used to create publications purely with rich metadata (to be at least FAIR [5], even for closed source software) or with attached artifacts like source code, executables, etc. (to be more easily reusable). To achieve this, HERMES provides

- an extensible, configurable and automatable toolchain with capability to be executed for[15]
  - N software publications in
  - M target publication repositories
  - from the same origin
  - as configured by the user,
- initially harvesting and collating statically available metadata from formerly described metadata sources and
- initially targeting
  - InvenioRDM and
  - Dataverse project
- for deposits of metadata and artifacts according to curator-defined requirements
- and output of the respective metadata in a structured format (e.g., CodeMeta files) for further reuse.

[arXiv:2201.09015 | PubPeer] | software-metadata.pub

HERMES
Concept



HERMES
Metadata

- **Metadata**
  - Differences in generation, scope, mode, aspects
  - Generic software metadata vs. software-specific metadata

- **Metadata formats**
  - Metadata files, snippets, third-party systems, API responses
  - Structured vs. unstructured

- **Sources**
  - Collectable structured metadata
  - (Metadata from minable structured data)
  - (Metadata from minable unstructured data)

HERMES
Components

- **Software**
  - Software for software publication workflow automation (workflow runner + modular pipelines)
- **CI templates**
  - GitLab CI, GitHub Actions
- **Improved research software-readiness in publication repositories**
  - Position paper “research software-ready libraries”
  - Respective contributions to Dataverse + InvenioRDM (data models, UI)
- **Training materials**
  - Adaption of open Helmholtz training materials (HIFIS) to include workflow usage
- **Project website**
  - One-stop shop for information and documentation

HERMES
Where are we?

- Proof of concept prototype `hermes`

  - **Harvesting**
    Citation File Format, CodeMeta, Git metadata
  - **Processing**
    Add deduplication, validation.
  - **Curation/Deposition**
    Metadata mapping and deposit for Invenio done. Enable more targets and user feedback.
  - **Post-processing**
    Update `CITATION.cff`/`codemeta.json`.

HERMES

Next steps & future research

- Finalize beta demonstrator (06/2023)
  - More metadata sources
  - Fully support Dataverse + InvenioRDM targets
- Improve usability …
  - … for RSEs: e.g., better user docs/configuration, CI templates
  - … for infrastructure providers: e.g., better developer docs, plugin templates
- Improve quality
  - Improve current development practice, e.g., formalized code reviews
  - Implement standardized provenance tracking for metadata and processing
- Foster future development
  - E.g., clarify incentives (software citation), mentoring

Thank you!

- software-metadata.pub
- team@software-metadata.pub
- github.com/hermes-hmc

HERMES is funded under grant **ZT-I-PF-3-006** from the Initiative and Networking Fund of the Helmholtz Association in the framework of the **Helmholtz Metadata Collaboration** call for projects 2020.

Find us here at #deRSE23!

S. Druskat, M. Meinel, T. Schlauch
Deutsches Zentrum für Luft- und Raumfahrt

J. Kelling, O. Knodel, D. Pape, G. Juckeland
Helmholtz-Zentrum Dresden-Rossendorf

O. Bertuch
Forschungszentrum Jülich