Plan Overview

A Data Management Plan created using DMPonline

Title: Automatising the publication of CEUR-WS.org workshop proceedings

Creator: Ilaria Tiddi

Principal Investigator: Ilaria Tiddi, Michael Cochez

Data Manager: Michael Cochez

Project Administrator: Michael Cochez

Affiliation: Vrije Universiteit Amsterdam

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO (September 2020)

ORCID iD: 0000-0001-7116-9338

Project abstract:
This project aims at automatising the open-access publication workflow of workshop proceedings of CEUR-WS.org. The current workflow is manual, and requires significant efforts at all levels, from the paper authors, to workshop organisers who prepare the proceedings, to the editorial board who assesses the submissions. Instead, we propose to develop an interface to create volumes with less effort. This new interface will allow the collection of information about the open-access publications, like authorship, dataset information, used software etc, in a structured, machine-readable form. This will facilitate the permanent publication of the proceedings, and ultimately reduce effort for quality assessment.
Note: this is *not* a usual research project. No research is performed. No data, nor research outputs are generated. The output consists of the implementation of the software which will be deployed by CEUR-WS.

ID: 135081

Start date: 01-01-2024

End date: 31-08-2024

Last modified: 14-12-2023

Grant number / URL: OSF23.1.010

as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Automatising the publication of CEUR-WS.org workshop proceedings

General Information

Name applicant and project number

Ilaria Tiddi - OSF23.1.010

Name of data management support staff consulted during the preparation of this plan and date of consultation.

Brett Olivier, Data Steward Faculty of Science, 13/12/2023
rdm.beta@vu.nl

1. What data will be collected or produced, and what existing data will be re-used?

1.1 Will you re-use existing data for this research?

If yes: explain which existing data you will re-use and under which terms of use.

- Yes

We will reuse the existing metadata provided by CEUR-WS.org published under CC0 licence. CEUR-WS.org's data is free for re-use, so we do not foresee any constraints.

1.2 If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.

In this project we will produce software in the form of python source code and documentation in the form of markdown files.

1.3. How much data storage will your project require in total?

- 0 – 10 GB

2. What metadata and documentation will accompany the data?

2.1 Indicate what documentation will accompany the data.

The software will be documented in two ways.

First, there will be documentation on the code itself, in the form of a specification of inputs and outputs.

Then, we will create user documentation for three audiences: a guide for volume editors, one for authors and one for the associate editors. This documentation will be written as markdown files, which will be compiled to html for publication on the CEUR-WS web page.

2.2 Indicate which metadata will be provided to help others identify and discover the data.

The software repository will include a CITATION.cff file with up to date information about the software.

3. How will data and metadata be stored and backed up during the research?

3.1 Describe where the data and metadata will be stored and backed up during the project.

- Institution networked research storage

The code will be stored on github and on locally kept on the teams' machines during prototype development.

We will configure the github action https://github.com/marketplace/actions/zenodo-upload to automatically upload the releases to Zenodo.

The documentation is also part of the same github repository and will hence also be backed up.

3.2 How will data security and protection of sensitive data be taken care of during the research?

- Not applicable (no sensitive data)

4. How will you handle issues regarding the processing of personal information and intellectual property rights and ownership?

4.1 Will you process and/or store personal data during your project?

If yes, how will compliance with legislation and (institutional) regulation on personal data

be ensured?

- No

4.2 How will ownership of the data and intellectual property rights to the data be managed?

The copyright for the software will be held by the authors / the institution. However, it will be released under an open source license.

5. How and when will data be shared and preserved for the long term?

5.1 How will data be selected for long-term preservation?

- All data resulting from the project will be preserved for at least 10 years

As mentioned, the source code and documentation will be archived on Zenodo.

5.2 Are there any (legal, IP, privacy related, security related) reasons to restrict access to the data once made publicly available, to limit which data will be made publicly available, or to not make part of the data publicly available?

If yes, please explain.

- No

5.3 What data will be made available for re-use?

- All data resulting from the project will be made available

5.4 When will the data be available for re-use, and for how long will the data be available?

- Data available upon completion of the project

5.5 In which repository will the data be archived and made available for re-use, and under which license?

Releases will be automatically archived on Zenodo. These will be under the Apache Licence, version 2.0

5.6 Describe your strategy for publishing the analysis software that will be generated in this project.

There is software generated during this project, but it is not analysis software.

That software will be published on github as part of the CEUR-WS organization, under the Apache License, Version 2.0

In our repository, we work with trunk based development (https://trunkbaseddevelopment.com/). We do continuous integration. This includes three types of checks. First, we use a linter called flake8. Then we use mypy for type checking of the code. Finally, we ensure that all test for the code pass before code is merged into the main code branch. We will also include a information on our security policy, and enable security advisories for our repositories.

After this software is releases, the long term maintenance will be taken over by the CEUR-WS team. This includes enabling periodical scanning of the code and its dependencies to find new security vulnerabilities that have been discovered.

6. Data management costs

6.1 What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Support in the implementation and resources for prototyping will be provided in-kind by the Computer Science department (Dr. Tiddi/Dr. Cochez, co-supervising the scientific programmer).