

EnviDat supports Open Science

The openness of scientific research is increasingly becoming a core requirement for funding agencies. EnviDat can support you in implementing the first steps towards Open Science.

Text: Ionuț Iosifescu Enescu, Gian-Kasper Plattner; Photo: Dominik Haas-Artho/EnviDat

This article is dedicated to the memory of Koni Steffen. Koni initiated EnviDat when he arrived at WSL. He was convinced that through the provision, sharing, and linking of the vast amount of highly relevant environmental data from WSL and beyond, EnviDat could foster new exciting research.

What is Open Science?

Open Science is defined as ‘transparent and accessible knowledge that is shared and developed through collaborative networks’¹. Open Science is also the ‘umbrella term for all efforts aimed at achieving more openness in science and the necessary paradigm shift’². Many research funding agencies, such as the Swiss National Science Foundation, are ‘committed to this goal’ because ‘research funded by the public should be publicly accessible as far as possible and free of charge’². Accordingly, Open Science is being actively promoted by national, European and international research funding agencies, because it represents a systemic change and a paradigm shift ‘from the standard practices of publishing research results in scientific publications towards sharing and using all available knowledge at an earlier stage in the research process’³.

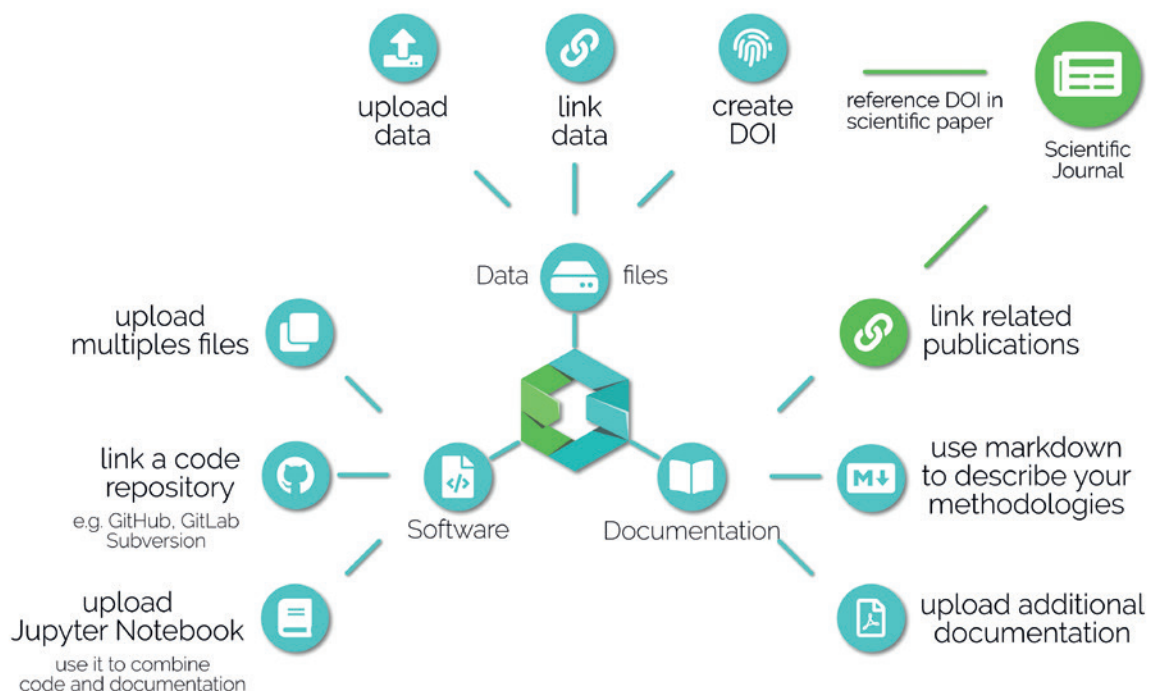
The main challenge and change brought by Open Science is that ‘all available knowledge’ must become accessible. However, it might be quite hard to define what this ‘all available knowledge’ means for an individual research project. Although conforming with Open Science obligations might not be easy, the EnviDat team is here to assist you in every step of the process.

Key activities

Currently EnviDat supports you in your Open Science efforts in the following key activities: (i) making your data accessible, (ii) publishing your entire methodology as open access and (iii) publishing your software as open source. These Open Science activities are completely supported by WSL through EnviDat.

First, EnviDat enables you to publish quality-controlled and well-documented research data with document object identifiers (DOIs) and appropriate citation information anytime during the life-cycle of a research project. You can publish research in EnviDat even at an early stage in the research process, because new files and new versions (e.g. containing corrections to the already published data) can always be added in the future. Assigning a DOI to your publication data, however, comes with some important restrictions, like not being allowed to delete any parts of it – please refer to the EnviDat guidelines for details⁴.

Second, the opening of your methodologies can be largely achieved by publishing your research in open access journals. Additionally, in close collaboration with Lib4RI-DORA, we offer you the opportunity to publish original documentation, articles or even books about your research workflows with formal DOIs



directly in EnviDat, similarly to an open access scientific journal. Please note that all research-related content must be original and that validity, authenticity, quality and respecting publication copyright are your sole responsibility⁵. After publication, your scientific article needs to mention the associated EnviDat dataset and be submitted to Lib4RI-DORA. Your scientific publication should also be included in the 'related publications' field of the EnviDat metadata.

Third, WSL and EnviDat support, where possible and practical, the publication of your models, software or scripts as open source code also with DOIs directly in EnviDat, while advising you about best practices such as code versioning, appropriate code documentation and clean coding. Consequently, in your programming endeavors you can gladly use WSL GitLab (installed and operated by the WSL IT)⁶ or GitHub⁷ as development platforms. We encourage you to write your code in Jupyter notebooks. Jupyter notebooks provide a solution for improved documentation and interactive execution of open code in a wide range of programming languages that are widely used at WSL, such as Python, R, Octave/Matlab, Java or Scala. By publishing your open software as Jupyter notebooks alongside your research data sets in EnviDat, you can combine open source code, detailed text-based descriptions and interactive results output for an even more improved and accessible documentation of your research.

Cooperation with the WSL IT

In close cooperation with the WSL IT, EnviDat offers you a convenient, easy way to run such Jupyter Notebooks on WSL's High-Performance Computing (HPC) Linux Cluster. In the JupyterHub/JupyterLab beta installation, you can choose from a single node instance (with 10 CPU cores) up to a four-node instance (with 40 cores and 32GB RAM) for testing and running Jupyter notebooks on the HPC – please refer to the Hyperion documentation for details⁸. Furthermore, there is also a solution for complex scientific software, scripts and workflows, that cannot be easily transformed into Jupyter notebooks. As long as the original scientific software can run on a HPC/Hyperion, there is also the option of packaging such scientific workflows in Singularity containers⁹. This helps ensure that the software can be run again at a later point in time. For making the interactive execution of your models more convenient for you and your scientific colleagues, we support you to automatically download and extract sample input data directly from EnviDat dataset records inside your code¹⁰.

Conclusion

To summarize, the EnviDat team encourages you to include all your original project knowledge, for which you own the copyright, such as data, links to published papers, additional documentation and the code for your scripts or models (ideally as Jupyter notebooks), in one single place, in EnviDat. WSL and EnviDat provide you with the tools to transform your research into 'transparent and accessible knowledge', while making it easy for your peers to properly cite and reuse your scientific contributions.

The EnviDat team is at your service and ready to support you with more information and individual consulting regarding Open Science. Please do not hesitate to contact us at envi-dat@wsl.ch.

References:

- ¹ Vicente-Saez, R. & Martinez-Fuentes, C. (2018). A systematic literature review for an integrated definition. *Journal of Business Research*, Volume 88, pp. 428-436, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2017.12.043>
- ² Swiss National Science Foundation (2020). Open Science Dossier. <http://www.snf.ch/en/researchinFocus/dossiers/open-science/Pages/default.aspx> (last accessed on August 21, 2020)
- ³ European Commission, Directorate General for Research and Innovation (2016). Open innovation, open science, open to the world. <https://op.europa.eu/en/publication-detail/-/publication/3213b335-1cbc-11e6-ba9a-01aa75ed71a1> (last accessed on August 21, 2020)
- ⁴ EnviDat (2020). EnviDat Guidelines for Publishing Research Data v. 0.1. <https://www.envi-dat.ch/#/guidelines> (last accessed on August 21, 2020)
- ⁵ EnviDat (2020). EnviDat General Policies v. 1.0. <https://www.envi-dat.ch/#/policies> (last accessed on August 21, 2020)
- ⁶ WSL-IT (2020). GitLab. <https://gitlab.wsl.ch/> (last accessed on August 21, 2020)
- ⁷ GitHub (2020). <https://github.com/> (last accessed on August 21, 2020)
- ⁸ WSL IT (2020). Hyperion Jupyter Documentation v. 0.0.1. <https://hyperion.wsl.ch/documentation-master/build/html/Jupyter.html> (available only from inside WSL-Intranet, last accessed on August 21, 2020)
- ⁹ Kurtzer, G.M., Sochat, V., Bauer M.W. (2017). Singularity: Scientific containers for mobility of compute. *PLoS ONE* 12(5): e0177459. <https://doi.org/10.1371/journal.pone.0177459>
- ¹⁰ Iosifescu Enescu, I., et al. (2019). Open science, knowledge sharing and reproducibility as drivers for the adoption of FOSS4G in environmental research. In M. A. Brovelli & A. F. Marin (Eds.), *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: Vol. XLII-4/W14. FOSS4G 2019 – Academic Track* (pp. 107-110). <https://doi.org/10.5194/isprs-archives-XLII-4-W14-107-2019>