Improved FAIR Data Publication Quality in Specialized Environmental Data Portals

Ionut Iosifescu Enescu, Gian-Kasper Plattner, Lucia Espona Pernas, Dominik Haas-Artho, and Rebecca Buchholz
Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

Environmental research data from the Swiss Federal Research Institute WSL, an Institute of the ETH Domain, is published through the environmental data portal EnviDat (https://www.envidat.ch). EnviDat actively implements the FAIR (Findability, Accessibility, Interoperability and Reusability) principles and offers guidance and support to researchers throughout the research data publication process.

WSL strives to increase the fraction of environmental data easily available for reuse in the public domain. At the same time, WSL facilitates the publication of high-quality environmental research datasets by providing an appropriate infrastructure, a formal publication process and by assigning Document Object Identifiers (DOIs) and appropriate citation information.

Within EnviDat, we conceptualize and implement data publishing workflows that include automatic validation, interactive quality checks, and iterative improvement of metadata quality. The data publication workflow encompasses a number of steps, starting from the request for a DOI, to an approval process with a double-checking principle, and the submission of the metadata-record to DataCite for the final data publication. This workflow can be viewed as a decentralized peer-review and quality improvement process for safeguarding the quality of published environmental datasets. The workflow is being further developed and refined together with partner institutions within the ETH Domain.

We have defined and implemented additional features in EnviDat, such as (i) in-depth tracing of data provenance through related datasets; (ii) the ability to augment published research data with additional resources which support open science such as model codes and software; and (iii) a DataCRediT mechanism designed for specifying data authorship (Collection, Validation, Curation, Software, Publication, Supervision).

We foresee that these developments will help to further improve approaches targeted at modern documentation and exchange of scientific information. This is timely given the increasing expectations that institutions and researchers have towards capabilities of research data portals and repositories in the environmental domain.