Fostering Open Science at WSL with the EnviDat Environmental Data Portal

EnviDat is the institutional research data portal of the Swiss Federal Institute for Forest, Snow and Landscape WSL. The portal is designed to provide solutions for efficient, unified and managed access to the WSL's comprehensive reservoir of monitoring and research data, in accordance with the WSL data policy. Through EnviDat, WSL is fostering open science, making curated, quality-controlled, publication-ready research data accessible. Data producers can document author contributions for a particular data set through the EnviDat-DataCRediT taxonomy. The publication of research data sets can be complemented with additional digital resources, such as, e.g., supplementary documentation, processing software or detailed descriptions of code (i.e. as Jupyter Notebooks). The EnviDat Team is working towards generic solutions for enhancing open science, in line with WSL's commitment to accessible research data.

1 # Fostering Open Science at WSL with the EnviDat
2 # Environmental Data Portal
3

4 Ionuț Iosifescu Enescu[1], Marielle Fraefel[1], Gian-Kasper Plattner[1], Lucia Espona-Pernas[1], Dominik
5 Haas-Artho[1], Michael Lehning[2,3], Konrad Steffen[1,3,4]
6
7 [1] Swiss Federal Institute for Forest, Snow and Landscape WSL, Birmensdorf, Switzerland
8 [2] WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
9 [3] École Polytechnique Fédérale de Lausanne EPFL, Lausanne, Switzerland
10 [4] ETH Zurich, Zurich, Switzerland
11
12 Corresponding Author:
13 Ionuț Iosifescu Enescu[1]
14 [1]Eidg. Forschungsanstalt WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
15 Email address: ionut.iosifescu@wsl.ch
16

17 ABSTRACT

18 EnviDat is the institutional research data portal of the Swiss Federal Institute for Forest, Snow
19 and Landscape WSL. The portal is designed to provide solutions for efficient, unified and
20 managed access to the WSL's comprehensive reservoir of monitoring and research data, in
21 accordance with the WSL data policy. Through EnviDat, WSL is fostering open science, making
22 curated, quality-controlled, publication-ready research data accessible. Data producers can
23 document author contributions for a particular data set through the EnviDat-DataCRediT
24 taxonomy. The publication of research data sets can be complemented with additional digital
25 resources, such as, e.g., supplementary documentation, processing software or detailed
26 descriptions of code (i.e. as Jupyter Notebooks). The EnviDat Team is working towards generic
27 solutions for enhancing open science, in line with WSL's commitment to accessible research
28 data.
29
30

31 KEYWORDS

32 *Open Science, EnviDat, Environmental Science, Data Sharing, Research Data Management*
33
34
35
36
37
38
39

INTRODUCTION

The Swiss Federal Institute for Forest, Snow and Landscape WSL, a research institute of the Swiss national network of federal institutes of technology and research institutions (ETH Domain), is developing the institutional environmental data portal EnviDat (Figure 1). EnviDat is designed to provide unified and managed access to WSL's environmental monitoring and research data. WSL has a long tradition in data collection. WSL research datasets cover research themes ranging from forest ecosystems, snow and ice, landscape, biodiversity to natural hazards and include long-term monitoring datasets spanning over a century. Such datasets are particularly valuable for studying the terrestrial environment and for obtaining an integrated view of the Earth System.
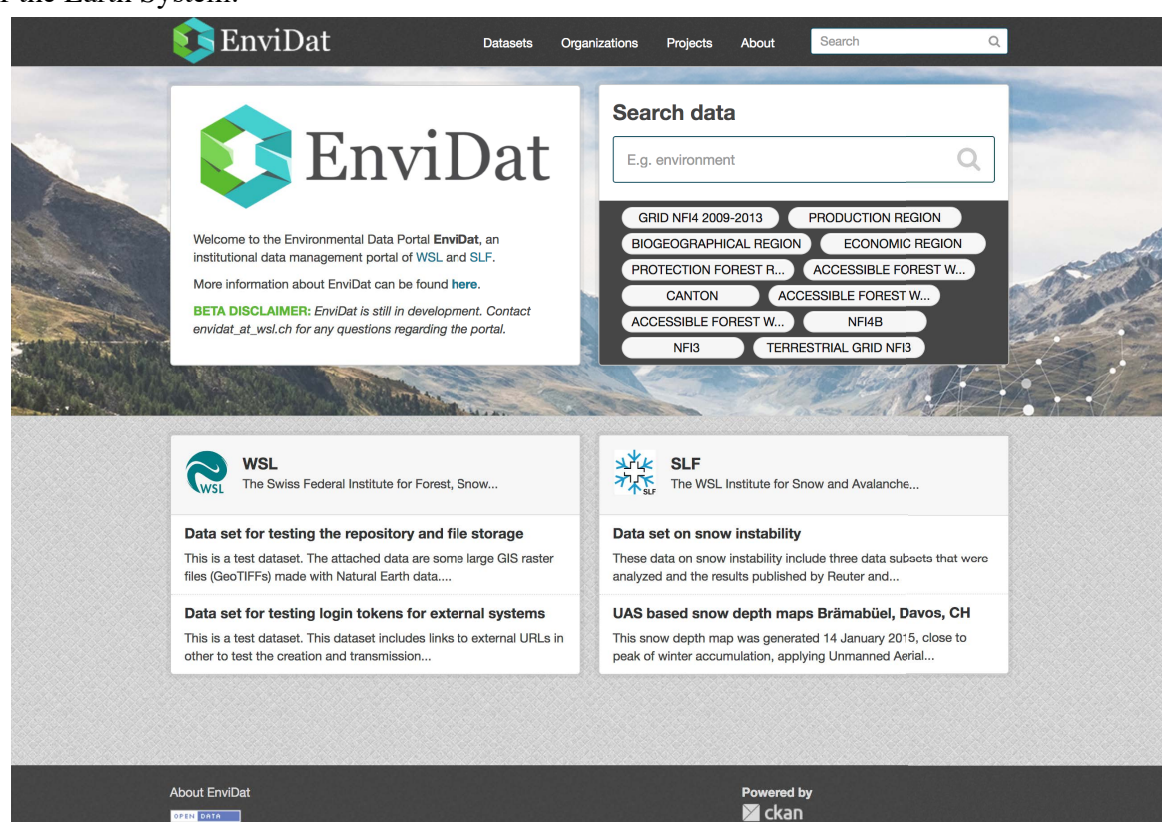

Figure 1: The EnviDat Homepage (September 2018)

The EnviDat environmental data portal has been selected by the WSL board of directors as one of three strategic initiatives of WSL for the planning period 2017-2020, highlighting the importance of research data management at the institutional level. WSL recently also put in place an overarching data policy with the goal to make research data accessible within two years after the completion of a research project (or a program phase for long-term research programs and monitoring projects). Consequently, the EnviDat strategic initiative is revolving around developing solutions for efficient, unified and managed access to the WSL's comprehensive reservoir of monitoring and research data, in accordance with the WSL data policy.

OPEN SCIENCE THROUGH ENVIDAT

61
62   WSL is fostering open science through EnviDat (a) by making research data accessible and (b) by
63 offering WSL data producers to publish the datasets with a DOI.
64 EnviDat services in the area of research data management focus on nine major activity areas.
65 Seven out of nine are currently active, namely (1) development of a concept for EnviDat, and by
66 extension, a concept for the research data management at WSL, (2) data publication with
67 provision of metadata and Document Object Identifiers (DOIs), (3) maintenance of an efficient
68 repository for publication data, (4) integration of existing operational systems into EnviDat, such
69 as the Swiss Long-term Forests Ecosystem Research Program (LWF) and the Swiss National
70 Forest Inventory (NFI), (5) guidance to researchers regarding options for publishing curated
71 data, (6) provision of versioning capabilities for datasets, and (7) contacts and exchange with the
72 international RDM community. Furthermore, it is planned to add (8) solutions for integrated
73 management, validation visualization and publication of streaming sensor data (such as
74 meteorological measurements) and (9) solutions for visualization and management of GIS data.
75
76 WSL's goal to foster open science is supported by several of these services. In EnviDat, WSL
77 scientists can register metadata about the data and provide access to the actual data. For the latter,
78 they can upload the data to the EnviDat data repository or link it from an operational information
79 management system or database. In September 2018, EnviDat achieved the publication of more
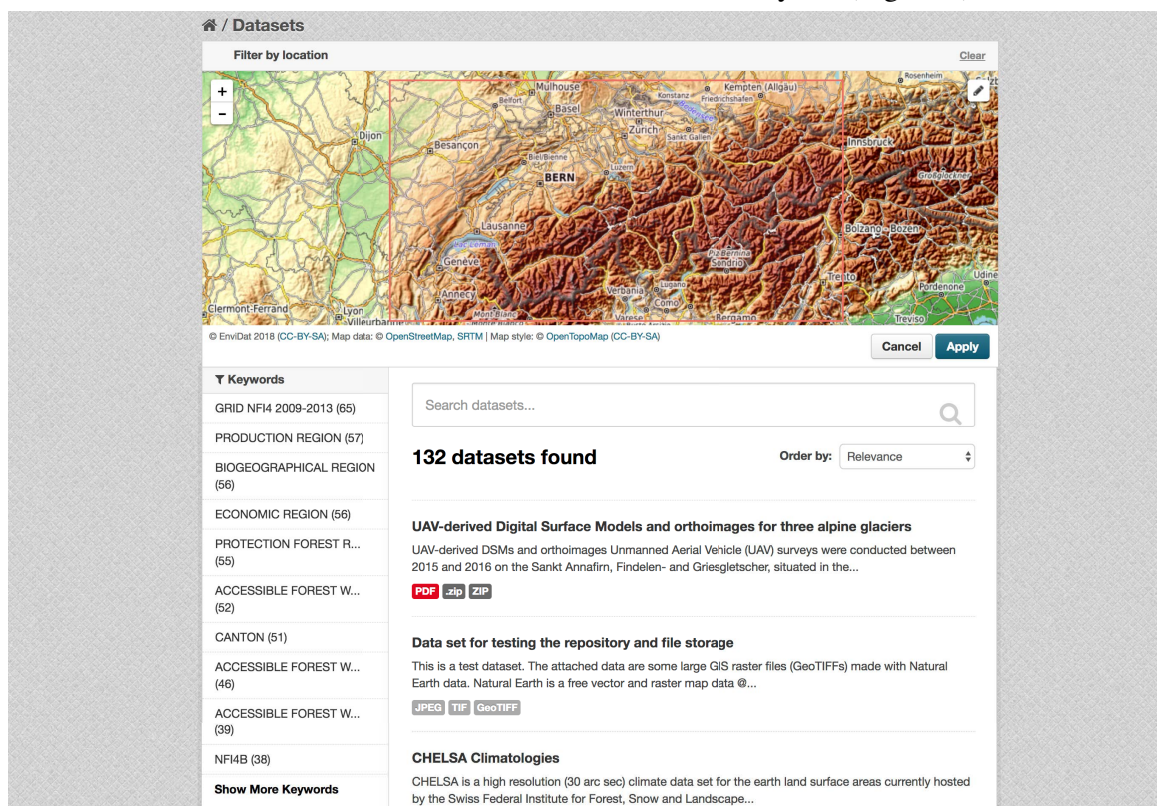80 than one hundred environmental datasets for Switzerland and beyond (Figure 2).

81
82   Figure 2: The EnviDat Datasets Page (September 2018)

83 WSL takes an approach of distributed data management where the responsibility to curate research
84 data remains with the experts and original data providers. EnviDat requires that data providers
85 register curated, quality-controlled, publication-ready data sets, that are ideally conforming to the
86 FAIR (Findability, Accessibility, Interoperability and Reusability) principles. Inspired by the
87 CASRAI Contributor Roles Taxonomy, we have defined the EnviDat-DataCRediT taxonomy[1] for
88 data authorship, explicitly allowing the documentation of individual author contributions involved
89 in the publication of a particular data set in EnviDat.
90
91 As mentioned above, EnviDat is designed to provide unified and managed access to research data.
92 However, publishing data with appropriate metadata is not always enough to effectively foster
93 open science. There is still the problem of sharing and reproducing the computations needed to
94 process and visualize the data, comparable e.g. to the methodologies presented in research papers.
95 EnviDat already allows and recommends to WSL scientists to complement data publication with
96 additional useful resources in digital form, such as supplementary documentations or processing
97 software. In this context, Jupyter Notebooks emerged as a solution for documenting code in a wide
98 range of programming languages[2] often used by environmental researchers, such as Python, R,
99 Octave, Scilab, Matlab, C, Java or Scala.
100
101

SHOWCASE AND RESULTS

103 To showcase the integration and usefulness of Jupyter Notebooks in a data portal, we present a
104 geospatial analysis example for road density calculation. In ecology, the presence of roads is often
105 used as an indicator for the degree of human disturbance in a region. The effect of roads on
106 movement patterns and habitat selection for different species and on different scales has been
107 analysed in numerous studies (Fahrig & Rytwinski, 2009). While roads may act as barriers to the
108 movement of certain animal species (habitat fragmentation) and directly increase mortality by
109 traffic, they can also facilitate movement or provide food for other species. In our example, we
110 examine road densities in the neighbourhood of sample locations using Python. The data used
111 (point sample locations and road geometries) can be retrieved from EnviDat, while the Python
112 script is made available as a Jupyter notebook on EnviDat (Figure 3).
113
114 Jupyter notebooks are an open document format based on JSON that mix the code with descriptive
115 narrative text, as well as rich output. Consequently, similar to the above Python Notebook
116 example, WSL scientists and EnviDat data providers are now encouraged to share detailed
117 descriptions of their code with the community. In the coming years, EnviDat aims for a deeper
118 integration of user-uploaded Jupyter Notebooks. Small scale, proof-of-concept work has been

---

[1] https://www.wsl.ch/datacredit/

[2] https://github.com/jupyter/jupyter/wiki/Jupyter-kernels

119 started for opening Jupyter notebooks hosted in EnviDat in an executable environment based on
120 BinderHub[3]. Yet, the development of a generic solution that would offer WSL researchers the
121 opportunity to access their code from anywhere on the existing EnviDat infrastructure is still
122 underway.

123

124 Figure 3: Example of Jupyter notebooks in EnviDat

125

DISCUSSION AND CONCLUSIONS

127 Beyond the challenges related to the technical implementation, additional important challenges
128 are the proper integration of proprietary software libraries (e.g. ArcGIS) and the limited IT
129 infrastructure resources available for running Jupyter Notebooks as a service. In conclusion,
130 EnviDat is working on a generic forward-looking solution, that is equally applicable to the
131 geospatial domain, for enhancing Open Science support, following WSL's commitment to
132 accessible research data in order to advance science.

133

134

---

[3] https://github.com/jupyterhub/binderhub

ACKNOWLEDGEMENTS

REFERENCES

Fahrig, L. & Rytwinski, T. 2009. Effects of Roads on Animal Abundance: an Empirical Review and Synthesis. Ecology and Society, 14.