Linking Environmental Data into European Scale RIs

Jan Bumberger[1], Thomas Schnicke[1], Steffen Zacharias[1], Michael Mirtl[1]
Supported by the eLTER Information Management Team
[1]Helmholtz Centre for Environmental Research - UFZ Leipzig, Germany

Leipzig, July 22, 2020

Abstract

In order to enable targeted measures to mitigate and adapt to global change analysis and assessment of the effects of climate change and environmental pressures on ecosystem processes and biodiversity are need and access to long-term monitoring as well as experimental data are needed (Mirtl et al. 2018, Mollenhauer et al. 2018). While eLTER strives for harmonization of measurements at its sites, existing environmental and biological monitoring time series are not always easy to combine in large-scale assessments because of historic differences in measurement protocols and varying access mechanisms and policies (Haase et al. 2018; Hoffmann et al. 2014). The pilot aims to enable access to data from long-term monitoring by
(a) identifying and align key data sources and variables across a range of different stakeholders and establish common workflows for the documentation, harmonization and integration of bio-geochemistry data supporting near real time data provision;
(b) establishing and adopt data documentation, data curation and data access procedures aligned with key stakeholders on European as well as global scale ensuring data ownership and data licensing and leverages new observed as well as legacy data.
Within the pilot, we will prototype the workflows to link environmental observation data from selected national data providers (e.g. TERENO, MOSES) using common services as provided by the EOSC and showcase issues on applying a FAIR data management. The pilot will involve a broad range of communty actors including the scientific community as well as data providers in order to ensure a user tailored implementation of data quality control and analysis.
Providing FAIR environmental data from selected eLTER sites will an important asset with the national research data infrastructure enabling the linkage of site based observation data in the analysis by leveraging access to data from long-term sites. This contributes to the uptake and implementation of the European Open Science Cloud (EOSC) by research infrastructures as well as the scientific community.

I. Introduction

Investigating interactions of carbon and nitrogen biogeochemical cycles across the European ecosystem types and biogeographic regions are core activities of eLTER. Biogeochemical cycles link all spheres of ecosystems (atmo-, bio-, pedo-, hydrosphere). They are of key importance for a quantitative understanding of ecosystem services and functions such as the net balance of greenhouse gases, removal of pollutants, biodiversity, provisioning of nutrients and food and water security. Climate change (e.g. changes in rainfall and temperature patterns), agricultural management (e.g. fertilization, tillage, irrigation), timber harvest and the concentrations of airborne acidifying and eutrophying substances are important drivers of biogeochemical

processes. The eLTER information cluster sites together with ICOS and other networks provides a unique opportunity of a whole system analysis of the impact of external drivers (such as drought) on biogeochemical process. Applying the FAIR (findable, accessible, interoperable, re-useable) principles on the management and provision of environmental data is crucial to ensure data sustainability. Data provided through the different data channels will serve a wide range of users and will support the integrated analysis fostered by the eLTER RI extend by common services. eLTER RI works towards the adoption and implementation of common protocols and standards for the provision and consumption of environmental data in near real time. Infrastructure elements like the central data node (CDN) supporting OGC SOS 2.0 as well as the linking data catalogues within the eLTER DIP (Data Integration Portal). Standard services (e.g. OGC CSW, OGC SOS) are used to exchange information between the different components of the information system. This also allows easy machine-readable access to metadata and data streams produced by the observation networks and thus the extension of the data network. This site and national efforts are important to support and enable European scale analysis of environmental data needs. Data mobilization needs to be addressed and fostered providing benefits to the data provider. Key aspects in the linking and use existing data streams is supporting automated data quality assurance (e.g. by using AI) or ingesting to data into analytical workflows This needs the use of analytical frameworks (e.g DataLabs) but also computing resources as provided by the EOSC. Access to these technologies and knowledge about the usage is not widespread in the LTER community and should be fostered by the pilot prototyping solutions based on existing data streams (e.g. as provided by TERENO). This could include the application of new techniques, as e.g. AI, for the integration of data from multiple in-situ sensors with EO based radar data. This also requires new methods of quality assurance from sensor networks and sensor services. By this, results of analysis will strengthen the data providers and create incentives for data mobilization.

II. Pilot description

The pilot will support the implementation of the European eLTER RI and foster the integration of data within the national LTER network in Germany. This will also foster the adoption of common services and the development of common interfaces between national data hubs and Ris using services provided by the EOSC. The use case will focus on leveraging and streamlining the data flows and interfaces between national local data hubs and the European scale RI (e.g. eLTER RI) enabling service based access to data. It will focus on prototyping workflows and services for data quality assurance and data analysis building on existing data services and resources. Inter alia, data streams being based on OGC Sensor Observation Services or the emerging SensorThing should be taken into account being ready for machine-to-machine access to data.
Generic services provided by the EOSC (e.g. EUDAT as storage service or EGI as computing services) provided by national and European scale e-infrastructure will be evaluated to form the service backbone for data curation and analysis. In addition, virtual research environments as e.g. DataLabs using JupyterNotebooks will be evaluated. Proposed solution is linking data sources and analytical workflows by the use of common services provided e.g. by EOSC.

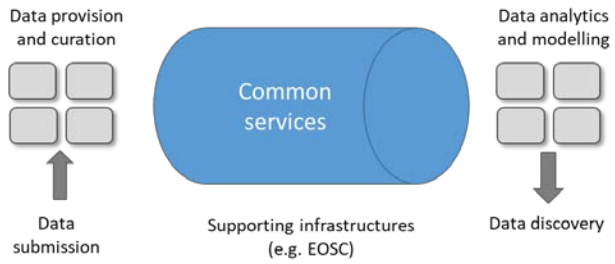

Fig. 1 Conceptual model for the pilot

The pilot will focus on prototyping common workflows (services) for data quality assurance and integration being of general use for the community using virtual research environments (as e.g. DataLabs) and services provided by existing e-infrastructures in the EOSC context. Also new techniques (e.g. AI) will be evaluated. In the long, this will allow to establish national scale data products based on curated and quality checked data feeding into models and analytical workflows. The evaluation of modelling results requires the existence of large data quality controlled datasets and computing resources as provided by EOSC. By this, the pilot will foster innovation and knowledge sharing within the different communities tackled. In detail this adresses

- the testing of new methods of data cureation and data quality assessment (e.g. based on AI) for sensor data from a range of eLTER sites
- the establishment of workflows for quality controlled large scale datasets as input to modelling excerises
- the provisionof a middle layer for the access and use EOSC services for the wider community by embedding community services and workflows

III. Relevance for the NFDI4Earth

The linking of NFDI4Earth with research infrastructures on a European scale is essential for the connectivity, in particular to open up the data streams for larger user groups. Data mobilization for a number of stakeholders and users is to be promoted within the project. The NFDI context provides the use of eLTER and EOSC services for data users. The resulting provision of workflows for quality-assured datasets is essential for the entire community of Earth System Sciences. The application of the FAIR principles (discoverable, accessible, interoperable, reusable) is essential for the management and provision of environmental data in order to ensure the sustainability of the data. Data provided through the various data channels will serve a wide range of users and support the integrated analysis that is enhanced by the expansion of eLTER RI through shared services. The connection with European RIs ensured that the reusability with FAIR principles is ensured in the long term. The use of the data streams can be connected to other platforms via the central data node (CDN) and the eLTER DIP (Data Integration Portal) via various standard services in order to enable subsequent use of the data streams.

IV. Deliverables

List the names and specifications of the two deliverables:

- D.1 Piloting Quality Assurance for Sensor Data based on common services (prototype of Jupyter Notebook, tested EOSC services,)

- D.2 Roadmap for integration of EOSC services for environmental data curation and analysis

V. Work Plan & Requested funding

	Q1	Q2	Q3	Q4
Identification of key data sources and providers				
Identification of common workflows and transformation routines with focus on data quality assurance				
Evaluate generic services provided by EOSC				
Define QA workflows from sensor networks using generic services provided by EOSC				
Prototype development of integrated data products from a range of data sources (e.g. TERENO)				
Define Roadmap for the implementation				

Resources requested: 1 FTE/year

VI. References

Mirtl, M., Borer, E. T., Djukic, I., Forsius, M., Haubold, H., Hugo, W., Jourdan, J., Lindenmayer, D., Mcdowell, W. H., Muraoka, H., Orenstein, D. E., Pauw, J. C., Peterseil, J., Shibata, H., Wohner, C., Yu, X., & Haase, P. (2018). Genesis, goals and achievements of Long-Term Ecological Research at the global scale: A critical review of ILTER and future directions. . Science of the Total Environment, Volume 626, 1439-1462. https://doi.org/10.1016/j.scitotenv.2017.12.001

Haase, P., Tonkin, J. D., Stoll, S., Burkhard, B., Frenzel, M., Geijzendorffer, I. R., Häuser, C., Klotz, S., Kühn, I., McDowell, W. H., Mirtl, M., Müller, F., Musche, M., Penner, J., Zacharias, S., & Schmeller, D. S. (2018). The next generation of site-based long-term ecological monitoring: Linking essential biodiversity variables and ecosystem integrity. Science of the Total Environment, Volumes 613–614, 1376–1384. https://doi.org/10.1016/j.scitotenv.2017.08.111

Hoffmann, A., Penner, J., Vohland, K., Cramer,W., Doubleday, R., Henle, K., Kõljalg, U., Kühn, I., Kunin,W.,Negro, J.J., Penev, L., Rodríguez,C., Saarenmaa,H., Schmeller, D.S., Stoev, P., Sutherland, W., Tuama, É.Ó., Wetzel, F., Häuser, C.L. (2014). The need for an integrated biodiversity policy support process – building the European contribution to a global biodiversity observation network (EU BON). J. Nat. Conserv. 6:49–65. https://doi.org/10.3897/natureconservation.6.6498

Mollenhauer, H., Kasner, M., Haase, P., Peterseil, J., Wohner, C., Frenzel, M., Mirtl, M., Schima, R., Bumberger, J., Zacharias, S. (2018). Long-term environmental monitoring infrastructures in Europe: observations, measurements, scales, and socio-ecological representativeness. Science of The Total Environment, Volume 624, 968-978, https://doi.org/10.1016/j.scitotenv.2017.12.095