European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS

Workflow for retrieval and harmonisation of legacy data

Deliverable D4.1

31 March 2022

Author(s) and affiliations:
Johannes Peterseil[1], Ika Djukic[1], Alessandro Oggioni[2], Martina Zilioli[2], Paolo Tagliolato[2], Will Bolton[3], Christoph Wohner[1], Carmela Marangi[2], Christian Poppe[4], Antonello Provenzale[2], Alice Baronetti[2], Angelica Parisi[2], Marie-Noëlle Pons[5,6]

[1] Umweltbundesamt GmbH, Austria
[2] CNR, Italy
[3] UK Centre of Ecology and Hydrology, UK
[4] Research Centre Jülich, Germany
[5] Université de Lorraine, CNRS, France
[6] eLTER, France

Contributing authors and affiliation
(in alphabetical order)

Dominique Joly (CNRS, France), Hanna Koivula (CSC, Finland), Jan Dick (UK-CEH, UK), John Watkins (UK-CEH, UK), Lauren Gillespie (BOKU, Austria), Lubos Halada (ILE-SAS, Slovakia), Mariasilvia Giamberini (CNR-IGG, Italy), Mathilde Emery (CNRS, France), Ricarda Sinn (EAA, internship, Austria), Sarah Venier (EAA, Austria), Susannah Rennie (UK-CEH, UK), Virginie Girard (Université Grenoble, eLTER France, France), Katharina Beer (EAA, internship, Austria), Leo Kunze (EAA, internship, Austria)

Contribution to the work and the report

All people listed as authors contributed to the textual description of the report. Chapter authors are given, where appropriate. Authors and contributing authors contributed to the implementation of the task in collecting the information on the data sources, participation at working meetings and in defining the scope of the report and the demonstrators.

Internal reviewers and affiliation
(in alphabetical order)

Eugenio Diaz-Pines (BOKU)
Ulf Mallast (UFZ)
Vladan Minić (BioSense)

Prepared under contract from the European Commission
Grant agreement No. 871128
EU Horizon 2020 Research and Innovation action

Project acronym:	**eLTER PLUS**
Project full title:	European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS
Start of the project:	Feb 2020
Duration:	60 months
Website:	https://www.lter-europe.net/projects/elter-plus
Deliverable title:	Workflow for retrieval and harmonisation of legacy data
Deliverable n°:	D4.1
Nature of the deliverable:	[Report, Demonstrator]
Dissemination level:	[Public]
Citation:	Peterseil, J., Djukic, I., Oggioni, A., Zilioli, M., Tagliolato, P., Bolton, W., Wohner, C., Marangi, C., Poppe, C., Provenzale, A., Baronetti, A., Parisi, A., Pons, M.-N. (2022). Workflows for retrieval and harmonisation of legacy data. Deliverable D4.1 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 62pp + Annexes (1-5).

Deliverable status:

Version	Status	Date	Author(s)
0.1	Draft	03 March 2021	Johannes Peterseil (EAA)
0.9	Draft for review	29 March 2022	Johannes Peterseil and co-authors
1.0	Final	13 April 2022	Johannes Peterseil and co-authors

Table of Content

Glossary

API	Application Programming Interface
B2DROP	EUDAT B2DROP service, based on NextCloud
B2SHARE	EUDAT B2SHARE publication service, based on Invenio issuing DOIs
BDL	Research challenge Biodiversity loss
BGC	Research challenge Bio-geochemical processes
CDN	eLTER Central Data Node
CLC	Corine Land Cover
csv	Comma-separated values
CWF	Research challenge Climate-Water-Food-Nexus
DAR	eLTER Digital Asset Registry
dataLab	eLTER Data Laboratory
DEIMS.iD	DEIMS-SDR (persistent site) identifier
DEIMS-SDR	Dynamic Ecological Information System - Site and Dataset Registry, being the central site catalogue for eLTER
DIP	eLTER Data Integration Platform
DOI	Digital Object Identifier
eLTER	Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research
eLTER	Long Term Ecosystem, critical zone and socio-ecological Research
EnvThes	Environmental thesaurus
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
IC	eLTER Information Clusters
ICT	Information and Communication Technology
IS	eLTER Information System
netCDF	netCDF file
PID	(Unique) Persistent Identifier
RC	Research challenges
RI	Research Infrastructure
SC	Science cases
SES	Research challenge Socio-ecological systems
SO	eLTER Standard Observations
SOS	OGC Sensor Observation Service
SWE	OGC Sensor Web Enablement
TRL	Technology Readiness Levels
UML	Unified Modelling Language

Summary

Progress in understanding, managing, and securing current and future ecosystem functions and services is challenged by fragmented and dispersed ecosystem research. As the topic is often approached using narrow disciplinary perspectives, a holistic understanding of complex eco- and socio-ecological systems is hampered and prevented. The emerging European Long-Term Ecosystem, critical zone and socio-ecological systems Research Infrastructure (eLTER RI) aims to overcome this challenge by addressing this issue in the ecosystem and biodiversity domain and thereby closing this gap in the European RI landscape. With its concept of the '*Information Clusters*' eLTER aims to provide a framework to lower the barrier to information access and exchange. The main idea behind the concept is to simplify the harvesting and user uptake of data from multiple information sources, facilitating the integration with eLTER data by making use of existing services, like Copernicus or statistical information. The selection of sources and content of relevant data layers is the result of an internal discussion where the Research Challenges (RC) play the main role by identifying the current requirements for environmental research and the ensuing demand for external data. The overarching framework of the eLTER Standard Observations (SOs) informs this process. In order to achieve the implementation of 'Information Clusters', three different data sources have been identified to complement eLTER observations and analysis: (a) *in-situ* legacy and third party data, (b) data from official statistics, and (c) remote sensing data and products.

The activities described in the report focus on the collection and exemplary retrieval of relevant *in-situ* legacy data, which we identified as complementary data sources and could play an important role within the planned eLTER data analysis workflows. This is relevant to (a) get additional data for data analysis or visualisation, (b) retrieve data from eLTER sites provided by national level catalogues, and (c) retrieve data from eLTER sites provided to other relevant RIs or monitoring networks. The aim of task 4.1 was to develop and test workflows for access and basic level harmonisation of relevant *in-situ* data sources on global, continental and national scale. We focused on data requirements defined both by the RC addressed in the eLTER PLUS project as well as the needs for supporting the implementation of data flows defined by the eLTER SOs. We identified 176 legacy and third party data sources which could be assigned to a respective eLTER SO and which sufficiently cover each component of the Ecological Integrity concept.

Based on a generic workflow described in the report we tested through demonstrators exemplary data extraction workflows being of relevance in the project context. This demonstrators focused on: (a) retrieve occurrence biodiversity data based on API access, (b) retrieve harmonised site gas flux observation data based on downloads, (c) retrieve data from E-OBS historic data (Copernicus Climate Change Service, 2020) to calculate climate diagrams for sites, (d) retrieve data from gridded and modelled data (e.g. E-OBS) based on the site extent, and (e) retrieve earth observation data products based on site extent.

It could be shown that the selected workflows are, at least on a prototype level, operational and are useful for the eLTER PLUS users. We applied a co-design process including the respective RC leads and Science Case (SC) contributors in the design and implementation phase on a regular basis. However, eLTER needs to decide if eLTER Information Clusters focus on on-demand services for extracting information sources or pre-calculated datasets. The results of the work done in task 4.1 provide input to the design and architecture of the extended eLTER Information System led by WP11 and the further definition of workflows towards the eLTER Standard Data Products led by WP10.

The report summarises the work done with respect to define and prototype workflows for the retrieval and harmonisation of legacy data. It specifically focuses on priority variables defined by the eLTER SO and aims to support Research Challenge related Science Cases at both, site and network scale. The first section describes the context of the work done, also in relation to the 'Information Clusters'

concept, which aims to enhance findability and accessibility of relevant data sources in the eLTER context. The second section lists identified relevant data sources relevant in this context and provides demonstrators for data retrieval and harmonisation in the third part. We finally discuss and provide recommendations for the eLTER Information Clusters that focus on thematic prioritisation, structural and legal interoperability as well as outline next steps for the implementation. The annexes provide detailed information shown in the report only in aggregated format.

1 Introduction

Johannes Peterseil

Access to and availability of quality controlled and trusted *in-situ* data is key for scientific analysis and to provide a sound basis for evidence-based decision-making. In this respect, the main aim of the emerging eLTER RI and the supporting LTER Europe network is to provide long-term, quality controlled and harmonised data to analyse and understand ecosystem processes and their vulnerability to global changes (e.g. climate or land use change). In order to achieve this goal, observations on the different components of the ecosystem need to be integrated in a whole-system approach (Mirtl et al., 2018). This results in a wide range of data being collected, which in future are centrally provided via the eLTER Information System. Nevertheless, this information can be complemented and combined with data stemming from other sources, such as other monitoring networks or research infrastructures, which provide added value to eLTER data streams and analysis.

In the eLTER PLUS project, WP 4 aims to: (1) develop the workflows for statistical data harvesting and user uptake from different information sources and integrate them with eLTER data, leading to eLTER Information Clusters; (2) make best use of Copernicus and other Remote Sensing (RS) data based services to provide information on broader scale social and environmental conditions, including targeted downstream service improvements, and (3) pilot workflows of data harvesting and incorporation into the eLTER Data Integration Portal (DIP).

The background of WP4 and its links to other parts of the eLTER PLUS project is defined in the WP4 description of action (eLTER PLUS, 2019) as "*striving for whole system research, the usage of data from all available sources is needed to broaden the thematic, spatial and temporal scope and enable gap filling of information gathered directly at eLTER* in-situ *facilities. WP4 aims to define strategies and workflows and to elicit prior knowledge for integration of data from multiple sources, including eLTER legacy data and external information sources (e.g. for gridded data, other* in-situ *data, Remote Sensing products, and official statistics). WP4 will focus on priority variables and information in the catalogue of eLTER Standard Observations (WP3) and support the WP8 and WP9 Case Studies. WP4 will trigger the service development carried out by WP11 both in terms of establishing the structures for incorporating information into eLTER DIP (WP11.2 and considering them in discovery and visualisation (WP11.3).*"

In this context, task 4.1 focuses (a) on the identification of *in-situ* data sources provided on European scale which are relevant to both, implement the eLTER PLUS project aims and also in the frame of the emerging eLTER RI; and (b) on the demonstration for selected data sources access and harmonisation. The task aimed to evaluate and develop procedures to provide third party as well as legacy data[1] of relevance for the eLTER SO (Zacharias et al., 2021) in a FAIR and digital format. In the context of the task, we define 'legacy *in-situ* data' as non-harmonised historic data collected at eLTER sites being held in local repositories (e.g. historic observations) or provided via a third-party service (e.g. monitoring networks (like ICP FOREST) or data initiatives (like FLUXNET). 'Third party *in-situ* data' are defined as any data source outside the eLTER network. The activity should screen relevant *in-situ* data sources from eLTER sites as well as third parties.

[1] Legacy data - old information that an organization has, especially information stored in an old-fashioned way (https://www.ldoceonline.com/dictionary/legacy-data)

1.1 Project context

Progress in understanding, managing, and securing current and future ecosystem functions and services is challenged by fragmented and dispersed ecosystem research. As the topic is often approached using narrow disciplinary perspectives, a holistic understanding of complex eco- and socio-ecological systems is hampered and sometimes prevented. The European Strategy Forum on Research Infrastructures (ESFRI) evaluated the emerging European Long-Term Ecosystem, critical zone and socio-ecological systems Research Infrastructure (eLTER RI) [2] as having high potential for addressing this issue in the ecosystem and biodiversity domain, thereby closing this gap in the European RI landscape. The primary objective of the eLTER PLUS project is to open and expand the research capacities and impact of eLTER by engaging current and new users and facilitating cross- and transdisciplinary research, exemplified in eLTER Site and Platform design and the RI's Standard Observation framework. Its Whole-Systems approach will derive meaningful scientific and policy relevant information via co-designed, transdisciplinary research in collaboration with diverse stakeholders at local, regional and EU-scales. Concerted actions also focus on collaboration with peer RIs to maximise synergies, increase efficiencies and catalyse holistic understanding of ecosystem function, and on development of virtual laboratories where *in-situ* site data are linked with other data sources, e.g. Copernicus. It will further advance community building and provisioning of services in close cooperation with the eLTER PPP project running in parallel. In this respect, eLTER PLUS will test proposed RI services and assess and strengthen its operations in real time.

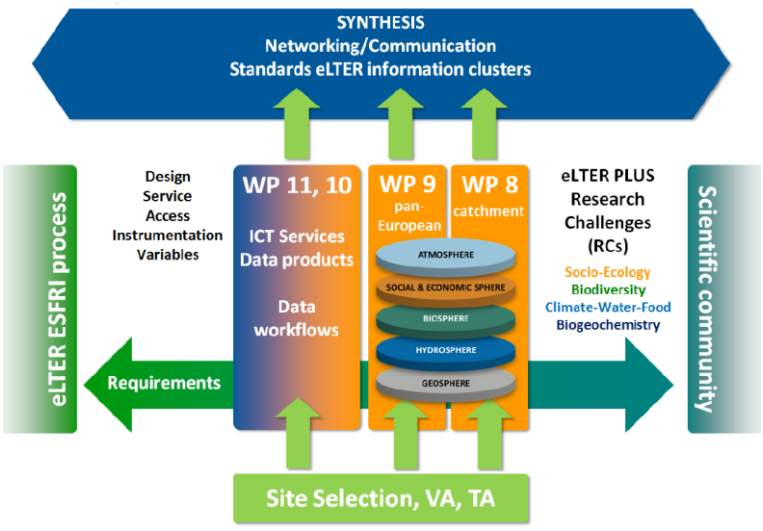

Figure 1: The core concept of eLTER PLUS (Source: eLTER PLUS Proposal)

This is implemented by close interactions between the different activities defined in the project design (Figure 1). The right side of the figure highlights the input by scientific and other user communities driving exemplary thematic Research Challenges (RCs) and specific Science Cases (SCs). The project outcomes, related to communities' requirements and developments feeds into the eLTER ESFRI process, as represented on the left side. Synthesis and networking activities (e.g. WP2, 3, 5, and 6) link project outcomes, related communities' requirements and the development of eLTER Information Clusters (WP4), as shown in the top part, within the eLTER PLUS project. These activities occur in parallel to a strategically driven site selection process based on the existing eLTER site pool for the use

[2] See https://elter-ri.eu/

of the site infrastructure for Transnational Access (TA) and Remote Access (RA) research and data-use opportunities, as well as providing access to data and services from these sites via Virtual Access (VA).

For the selected science cases, a Whole System Approach is realised by cross-theme collaboration within WP8 (site/platform scale) and WP9 (catchment scale). The work packages WP10 and WP11 aim to translate user requirements and research products across themes into tool and data product specifications and perform service/tools piloting and implementation.

1.2 Information cluster concept

Carmela Marangi

eLTER Sites and Platforms conduct a wide variety of ecological and socio-ecological studies, facilitated by the increasing availability of resources in terms of sensors, data, tools and processing power and by the development of new modelling approaches. Among those resources, data coming from external sources becomes more and more relevant to complement, integrate or surrogate *in-situ* observations. The integration of third party's data enables the scientific community to broaden the thematic, spatial and temporal scope of ecological and socio-ecological research conducted both, at site and across-site scales. This is especially true for cross-site research activities, which aim to perform analyses and compare results within or between bioclimatic regions. This allows improving reproducibility and robustness of the findings and investigating contrasting patterns of ecological characteristics and dynamics. Therefore, the demand for data from multiple and heterogeneous information sources, as well as the strategies for their integration at different scales, evolve in quantity and quality at an increasing pace. Within the eLTER community, this is amplified by applying a holistic ecosystem view pursued through the whole system approach (Mirtl et al., 2018).

Looking from a wider perspective, an additional challenge is posed by the recent Destination Earth initiative[3], which aims at building digital models or 'twins' of the Earth systems to enable better understanding and control of the factors that influence nature conservation and consequently human health and well-being (Bauer et al., 2021b). As for the whole system research, the new paradigm represented by digital twins can be applied only if we can guarantee a continuous flow of standardised and harmonised information stemming from a variety of research domains, networks and infrastructures (Bauer et al., 2021a, Nativi et al., 2021). That notwithstanding, a closer look at the kaleidoscopic landscape of data provisioning for environmental research, suggests that the exploitation of such information is not a trivial task. Each data source provides its products with different access rules, formats, licensing conditions, coverage as well as resolution in time and space. As a result, the collection of multiple sources' data and its integration is often perceived as a barrier to further scientific development, e.g. in performing research synthesis for nature conservation management (Wyborn et al., 2018).

With the concept of the 'Information Clusters' (see Figure 2) eLTER aims to address these challenges and provide a framework for the emerging RI. The main idea behind the concept is to simplify the harvesting and user uptake of data from multiple information sources, thereby facilitating the integration with eLTER data by making use of existing data and services.

[3] see https://digital-strategy.ec.europa.eu/en/policies/destination-earth

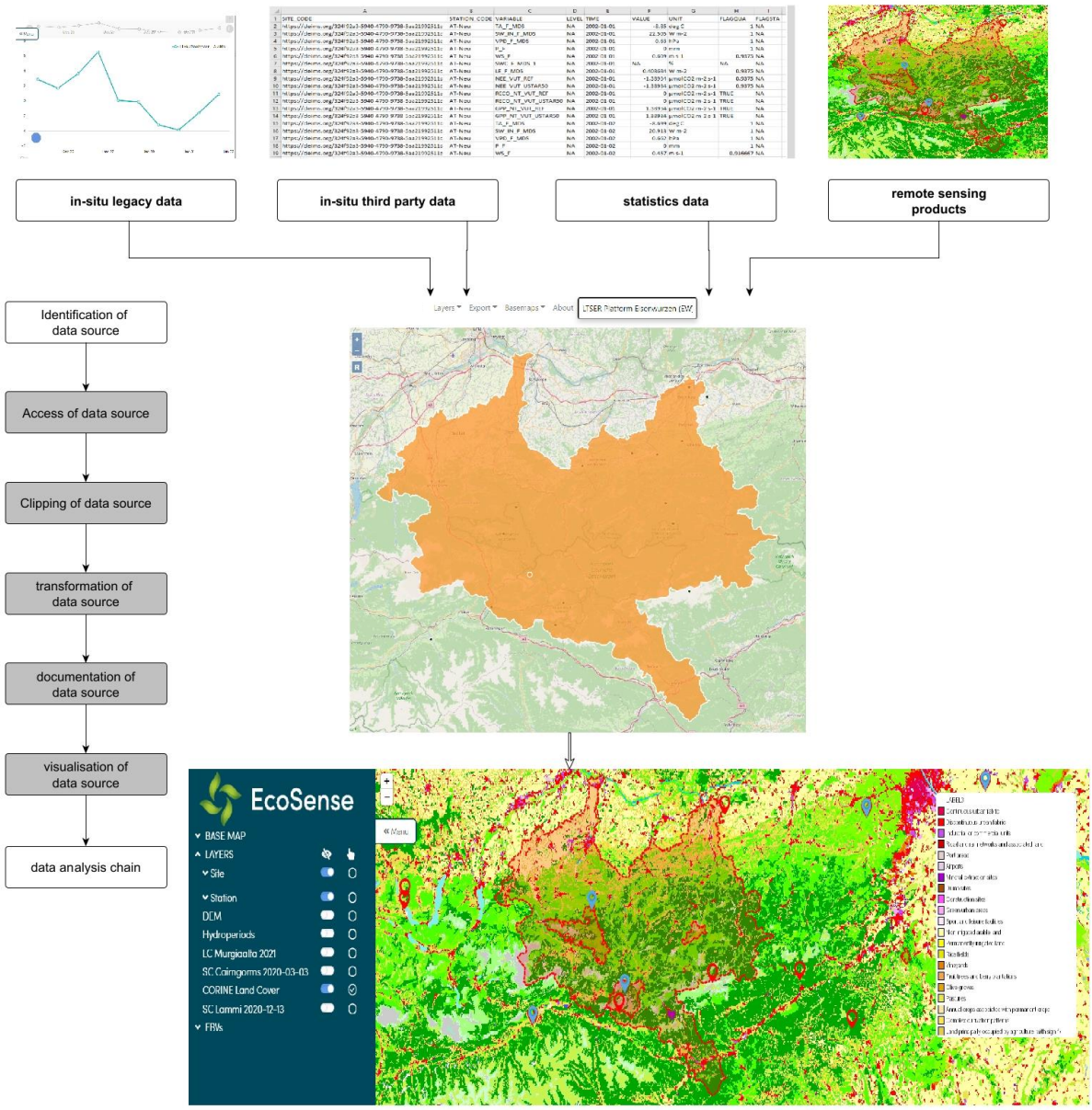

Figure 2: Schematic concept of 'Information Clusters' in the eLTER context including complementing data sources for visualisation and analysis with data from eLTER Sites

The selection of relevant data sources constituting the Information Clusters is the result of an internal discussion where the RCs play a main role by identifying the current requirements for environmental research and the ensuing demand for external data. The overarching framework of eLTER SOs (Zacharias et al. 2021) informs this process.

As outlined in chapter 3, describing a generic workflow for data source identification and retrieval, the identification of data types and sources is just the first step in the construction of Information Clusters. Following the data harvesting, the data layers are cropped to the boundary of the eLTER Sites or Platforms, to make the information available for successive use by research activities carried out at site and across-site scales. Cropping as well as pre-processing tools are developed and prototyped in the project context as outlined by Halada et al. (2022) or in this report, in order to enable Information Clusters for the site as well as to overcome the issues related to different native temporal and spatial resolutions of multiple data layers.

The design and pilot implementation of the workflows is intended to assist the development of new services, by establishing the structures for incorporating information into the eLTER Data Integration

Platform (DIP), being the central visualisation and access point for eLTER information. We expect that such services will contribute to (a) broaden the scope of site-based research, (b) facilitate the access to relevant information also overcoming technical barriers among different knowledge domains, (c) provide relevant processing tools and (d) improve the sustainability of the eLTER service provision to various stakeholder categories.

2 eLTER Data Requirements

Johannes Peterseil

An important first step of the work was to define the thematic scope for the data source collection and access. The general requirements for assessing the availability and accessibility of *in-situ* legacy and third-party data in the context of the Information Clusters were driven by the project requirements addressed by the RC and SC (Skiba et al., 2020, Dirnböck et al., 2020) as well as by the collection of ICT and data requirements (Peterseil et al., 2020). In addition, the work on defining the SO for the emerging eLTER RI (Zacharias et al. 2021) provided an important input. This work is also related to a number of milestones achieved within the eLTER PLUS project (see Table 1).

Table 1: Schedule of relevant project milestones.

WP	Milestone	Milestone title	Lead	Due date
4	MS8	Key variables and sites for eLTER Information Clusters pilots	UK-CEH	31.01.2021
10	MS23	Harmonisation workshop across the RCs	EAA	31.01.2021
10	MS24	Key data workflows	EAA	31.07.2021

eLTER data requirements have been taken into account when selecting *in-situ* data sources of interest. In this respect, we used the project internal requirements defined by the RCs as well as the SOs, focusing on the long-term implementation plan of the eLTER RI.

2.1 eLTER PLUS data requirements

An important part of the eLTER PLUS project is to improve the key infrastructure services of the LTER networks over the coming years. This is implemented alongside four scientific RCs with the key objective to assess the added value of using existing LTER data complemented by third party data focusing on cross-site and cross-scale analyses. These RCs are (a) biodiversity trends and losses (BDL), (b) bio-geochemical processes (BGC), (c) climate water food nexus (CWF), and (d) socio-ecological processes (SES) carried out in the project frame. Based on the SCs, data requirements were defined and summarised (Skiba et al., 2020, Dirnböck et al., 2020, Peterseil et al., 2020) identifying 98 observation variables of interest (see Table 2).

Table 2: Number of observation variables grouped by data type identified by eLTER PLUS RCs and SCs.

Variable groups	count	Variable groups	count
TimeSeries data		**TimeSeries data**	
Air quality	3	Soil-atmosphere gas exchange	4
Atmospheric deposition	10	Solid soil chemistry	7
Biodiversity*	1	**Descriptive site and plot data**	
Biomass	8	Habitat diversity	2
Climate	10	Land cover	1
Groundwater	3	Land use	3
Land use	8	Soil physics	9
Runoff, streams and standing water	11	Topography	3
Socio-economy	4	Vegetation	1
Soil climate	2	Land cover	1
Soil water	5	**Remote sensing**	3

** variable group covering all taxonomic groups*

Full details on the variables selected can be found in Peterseil et al. (2020). In addition, this information was also used in the selection and definition of the eLTER SO (Mollenhauer et al., 2018, Zacharias et al., 2021). We used the list of prioritised variables to perform a first screening and selection of relevant data sources.

2.2 eLTER Standard Observations

Based on eLTER's whole system approach[4], the eLTER Standard Observations (SO) will include a minimum set of variables and the associated method protocols that allow adequately characterising the state and future trends of Earth systems (Zacharias et al., 2020). They aim to provide a reference for the national eLTER ESFRI processes towards the future implementation of the RI on the European as well as on the national scale and are also defining the core data streams for the eLTER RI. Based on the Ecosystem integrity (EI) concept - the system's capacity to maintain structure and ecosystem functions using processes and elements characteristic for its ecoregion (Dorren et al., 2004; Haase et al., 2018) - a total of 173 SOs were identified in a first iteration and will be discussed in expert rounds during subsequent iterations. Final SOs are considered relevant and unique for eLTER. The SOs are organised along the six EI components, namely: (1) Abiotic characteristics, (2) Biotic heterogeneity, (3) Energy budget, (4) Matter budget, (5) Water balance, and (6) Socio-Ecology.

Table 3: Number of proposed eLTER Standard Observations per EI component (SO, Zacharias et al., 2021).

Component	high priority	optional	total
Abiotic characteristics	17	1	18
Biotic heterogeneity	8	16	24
Energy budget	5	12	17
Water balance	12	2	14
Matter budget	7	47	54
Socio-ecology	24	22	46

Table 3 provides an overview on the number of variables per EI component. A full list with details is provided by Zacharias et al. (2021) in the report on "*Discussion per on eLTER Standard Observations (eLTER SOs)*".

[4] Following the WAILS concept

3 Generic workflow for building Information Clusters

Johannes Peterseil

As outlined in the introduction, the retrieval of *in-situ* legacy data and complementary data sources is an intrinsic part of the eLTER data workflows, which will be setup within the eLTER RI. This is relevant to (a) get additional data for data analysis or visualisation, (b) to retrieve data from eLTER sites provided through national level catalogues, or (c) to retrieve data from eLTER sites provided to other sister RIs (e.g. ICOS) or monitoring networks (e.g. ICP). Basic requirements for eLTER specific workflow components have been collected within the eLTER PLUS project (Peterseil et al., 2020) and are further detailed in the current report addressing specifically the identification and access of relevant data sources from legacy and third-party data sources.

In Halada et al. (2022) a basic workflow has been outlined focusing on data flows from national statistics and prototyped for 6 LTSER platforms covering different bio-geographic regions. This workflow detailed (a) specification of data needs followed by (b) identification of data source, (c) preparation of metadata, and (d) data retrieval. Based on this workflow we further developed the concept for legacy *in-situ* data and modelling results. In this context, the generic term ***data source*** is used which applies to both specific datasets or data series, as well as for comprehensive data portals and catalogues. We use the term 'input data' for any legacy and third-party data provided by a selected data source and the term 'output data' for the resulting dataset after the extraction and harmonisation process.

Figure 3 provides an overview on the extended workflow for these data types. In the figure, we used the following colour coding: orange for workflow components, blue for eLTER Information System components, and green for specific workflow result sets as output data.

We identified the three components of the generic workflow, which map and are in-line with the steps defined by Halada et al. (2022).

First, in the **Identification of data source** the thematic scope for data source screening is defined, e.g. by the eLTER SOs as well as the specific data requirements resulting from research questions to be addressed (see chapter 2). Nevertheless, the thematic scope might change and be extended in future based on additional needs of the eLTER RI as well as the LTER network. The thematic scope is the filter for selecting relevant input data sources. This results in a list of relevant data sources being subject to the further assessment steps.

The **Data Source Assessment (assess)** includes two sub-components. The first sub-component '***Assessment Coverage***' focuses on the fitness-for-purpose addressing the thematic, spatial and temporal coverage of the data source of interest. Currently, we do this step manually by screening and assessing relevant data sources using information provided by the metadata, websites and personal communication. If the 'fitness-for-purpose' fails, the data source is not considered as relevant. The second sub-component '***Assessment Structure***' focuses on selected elements of the FAIR principles (Wilkinson et al., 2016) dealing with the findability, accessibility, interoperability, and re-usability. In the current context, we focused on elements like: (a) data format and standard supported, (b) data access mechanisms, (c) metadata formats and standards supported, (d) metadata access mechanisms, and (e) data licence. In addition, in future procedures also data citation (e.g. availability of persistent identification for static and dynamic datasets) as well as semantic references should be taken into account. If a data source is not accessible by either direct machine-to-machine interaction (e.g. API), personal contact or download, the data source is not considered as relevant.

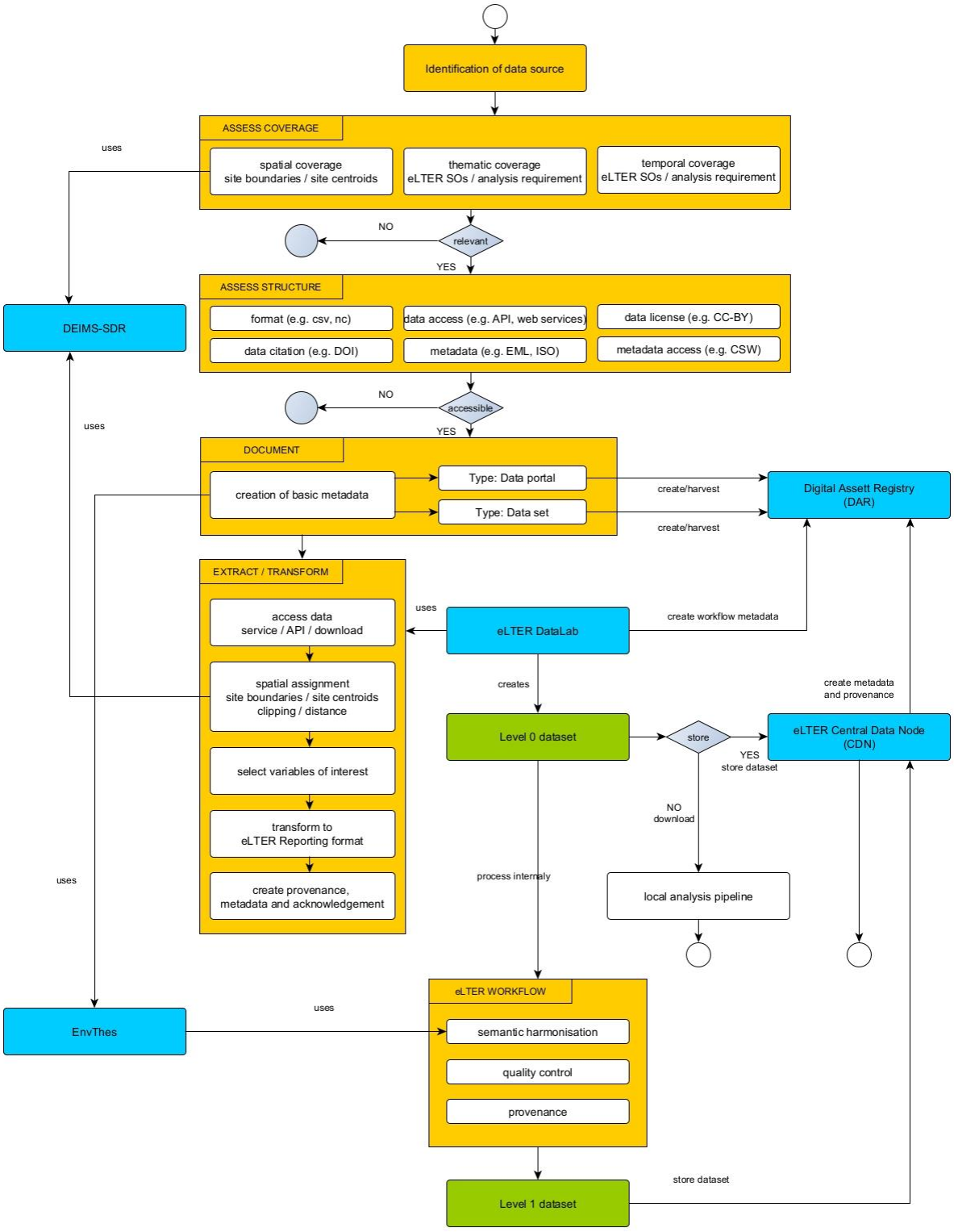

Figure 3: Generic workflow retrieving in-situ legacy data. Boxes in orange reflect workflows, boxes in green resulting datasets, and boxes in blue the respective components of the eLTER Information System.

Once a data source is selected as being relevant, basic descriptive metadata on the data source are generated or retrieved if available in a machine-readable format. This refers to **Data Source Documentation (document)** in the figure. These links to workflows of the eLTER Central Data Node (CDN) generating sign-posts for datasets or metadata for relevant data portals. Halada et al. (2022)

defined basic metadata elements for the documentation which have partly been followed and extended by the current work, which has been done in parallel. Basic metadata has been collected and are provided in Annex 9.3. Table 4 provides an overview on the meta-information collected for the data sources.

Table 4: Possible metadata elements used to document data sources.

Metadata element	Description	Format
Name/Title	Name or title of the data source	text
Abstract	general description of the data source	text
URL link	URL link to the data source main page	URL
Language	specifying the language of the data source (e.g. en)	code
Spatial scope	spatial extent and coverage of the data source (e.g. global, regional, national)	code
Country	if national, reference to the specific county or selection of countries	text
Thematic scope	main ecosystem type or ecosystem component covered by the data source	code
Temporal scope	start and end time of the observation data provided	date
Temporal resolution	temporal resolution of the observation data provided (e.g. annually)	text
Data licence	description of the data usage and licence, if available also the link to the data licence	text
Data format	listing of the supported data formats	text
Online distribution link	if available, URL link to the data source access or download page	URL
Metadata standard	metadata standard supported	text
Metadata catalogue	if available, URL link to a machine-readable metadata catalogue endpoint	URL
Thematic mapping SO	mapping of the thematic content of the data source to the eLTER SO	code
Thematic mapping int	mapping of the thematic content of the data source to the data requirements defined by eLTER PLUS WP8/9	code

Currently we are in the process to extend EnvThes (Schentz et al., 2011, 2013) to include the eLTER SO as terms, which will be used to enable the semantic mapping in the Digital Asset Registry (DAR).

The **Data Source Extraction (extract/transform)** is following the preceding step and can be split into different workflow steps of general applicability. The overall workflow results in a level 0 dataset, which is subject to further workflows on the interface to the eLTER Information System. This dataset can either be ingested and documented within the eLTER Central Data Node or is locally used in an analysis pipeline. A preliminary definition of the data levels addressed in eLTER is given by Peterseil et al. (2020) and further detailed by Koivula et al. (2022). For the implementation of these sub-workflows e.g. the eLTER DataLabs could be and are used in the current eLTER process. The demonstrators described in the following chapter are focusing on this part of the workflow. The Data Source Extraction workflow could be implemented by different solutions. Examples are e.g. the ReLTER package (Oggioni, Silver, Ranghetti & Tagliolato, 2021), the prototyped FLUXNET/ICOS workflow or the PhenoAPP prototype developed in task 4.4.

The extraction workflow starts with *accessing the data* which can either be via a machine readable interface (e.g. API, OGC SOS), direct download of the dataset from data portal (e.g. ICOS Carbon Portal) or indirect download after application for data retrieval (e.g. ICP Forest). Depending on the data source

authentication is needed to access the datasets or not. The current demonstrators aimed to cover these cases.

Next in the extraction is the *spatial assignment* of the datasets to a relevant eLTER site for which the data should be used. This can either be done by clipping (see Halada et al., 2022) for areal covering datasets (e.g. land cover) or by mapping of observation stations (e.g. FLUXNET) to eLTER Sites based on the distance for point data sources. DEIMS-SDR (Wohner et al., 2019, 2020) provides relevant location information (centroid as well as the boundaries of the eLTER Sites) via a REST-API as well as OGC WFS layers. Based on this spatial analysis, which needs to be verified manually, a mapping of input datasets to eLTER Sites can be done enabling an easy integration of datasets in an analysis pipeline. The original reference to the observation station from the input dataset needs to be kept. For spatial mapping sometimes additional data sources are needed (e.g. station catalogue or list) if the station location (e.g. latitude and longitude) is not provided directly in the input dataset.

Finally data are *transformed and harmonised* to enable better interoperability in the eLTER analysis workflows. For eLTER a basic data reporting format has been defined (Peterseil & Geiger, 2020) which should enable an easy integration of data applying a defined common data structure (example see Table 5). Still further discussion is needed in the frame of the eLTER RI implementation to finally decide on specific data formats. This might lead to adaptations of the current proposal for data reporting. The transformation and harmonisation step encompasses a first level harmonisation focusing on the data structure. Common field names should be used to identify e.g. the column containing the name of the variables.

Table 5: Example of basic data table following the eLTER Data Reporting Format (version 1.2)

SITE_CODE	STATION_CODE	SUBST	LEVEL	TIME	VALUE	UNIT	FLAGQUA	FLAGSTA
deims.id	IP1	TEMP	200	2016-03-15	5.5	°C		X
deims.id	IP1	PREC	100	2016-03-03	10.2	MM		S
deims.id	IP1	TEMP	200	2016-02-15	2.5	°C		X
deims.id	IP1	NH4N	100	2016-03	5.5	mg N/l		W
deims.id	IP1	SO4S	100	2016-03	10.2	mg S/l		W
deims.id	IP1	CA	100	2016-03	2.5	Mg/l	L	W
	…	…	…	…	…	…	…	…

In addition basic *metadata and provenance information* should be provided for the extracted dataset including (a) reference to the input dataset, (b) data licence and use conditions of the input dataset, (c) link to the extraction workflow, (d) input parameters set in the extraction workflow (e.g. selected variables, selected sites), and (e) version of the extraction workflow.

The resulting **Output dataset** is defined as *Level 0* for the eLTER workflow as only a structural harmonisation to the eLTER Data Reporting specification is performed. By this, Level 0 applies even if data quality control has been applied by the data provider. The usage level, e.g. externally open or internally restricted, if defined by the input dataset data licence and needs to be passed to the following steps. These issues regarding the governance are not addressed in the current report. The output dataset could be used in local analysis or pipelined into eLTER internal data production pipelines creating higher level datasets. The transition to Level 1 where central quality control as well as content harmonisation as part of the eLTER data pipelines is not further described in the current report and is subject of the WP10 activities of the eLTER PLUS project.

4 Identification and assessment of relevant *in-situ* data sources

Ika Djukic, Marie-Noëlle Pons & Johannes Peterseil

4.1 Identify relevant Data Sources

The first step in forming the Information Clusters is the identification of relevant *in-situ* data sources. This was done in close alignment with (a) data needs supporting the eLTER SOs, (b) data needs specified by WP8 (eLTER Whole System Approach at site and catchment scale), WP9 (optimization of the eLTER Network design at the Pan-European scale), and WP10 (prioritisation of core data and their data workflow), but also taking into account (c) data needs coming from new arising research questions that might be of relevance for the eLTER RI but has not been assigned yet. We provide an overview description of the requirements in chapter 2.

Hence, provided information on *in-situ* data sources should be considered as a "living" document, undergoing continuous adjustment in the course of the project and development of eLTER RI. In the long-term, this information should be shared and published via the eLTER Digital Asset Registry (DAR) providing a central catalogue for digital assets in the RI context.

In total, when writing the report we identified 176 *in-situ* data sources that are of potential relevance for eLTER RI covering various scales ranging from global, regional, to national. It needs to be stressed that the research work related to the socio-ecology *in-situ* data was dealt with succinctly, as information on available socio-ecology data has been collected within Task 4.2 (Halada et al. 2022).

4.2 Assess Data Sources

The next step was to assess and document the identified relevant data sources to enable a consistent description for each data source and its thematic scope. In the current work, we focused on the following aspects of the data sources:

- spatial extent of collected data (i.e. national, regional, global)
- temporal and spatial resolution of collected data
- metadata standards
- data accessibility and formats
- data policy and licenses
- language in which data are available

In addition, we collected basic information on the data sources, like data source title, description, web site or contact information. This information formed the basic metadata. We performed desk research screening the relevant data sources as well as contacted directly the data providers via email for more detailed information when information was missing on the websites.

Following the first assessment based on the prioritised observation variables, in a second turn the current version of the SOs was used to assign the thematic relevance of the identified data sources of interest. This was done based on expert knowledge dividing the list of data sources according to the expertise and knowledge of the task team. Table 6 shows the total number of SOs identified for the selected data sources. So for 'Abiotic characteristics' 13 out of 18 variables listed in the eLTER SOs could be found. For 'Matter budget' we identified 40 out of 54 variables listed in the eLTER SOs (compare with Table 3). For 'Socio-ecology' we found only 5 of 24 highly prioritised SO. This is due to the fact that we did not focus on socio-ecological variables, as this has been addressed by Halada et al. (2022). The code 'other' indicates if variables from the input data sources are of interest and could

be assigned to an EI Component and Compartment, but are not specifically linked to a specific SO. Thus, this data are complementary for the SOs.

Table 6: Number of identified eLTER Standard Observations (SO, Zacharias et al., 2021) in the data sources

SO Component	SO number*	SO Identified in data sources			
		high priority	optional	other	total
Abiotic characteristics	18	13	0	9	22
Biotic heterogeneity	24	4	12	9	25
Energy budget	17	3	8	2	13
Water balance	14	7	1	5	13
Matter budget	54	7	33	7	47
Socio-ecology	46	5	2	4	11
Other		0	0	7	7

* number of SOs according to Table 3

With regard to the ecological integrity components, most reviewed data sources represent variables that were assigned to the abiotic characteristics (52 counts), followed by biotic heterogeneity (50 counts), matter budget and water balance (each 29 counts), socio-ecology (23 counts), and energy budget (9 counts; Figure 4). It is important to mention, that one data source can provide variables for different ecological integrity components. We provide a detailed overview on the contribution of the data source to the ecological integrity components in Annex 2.

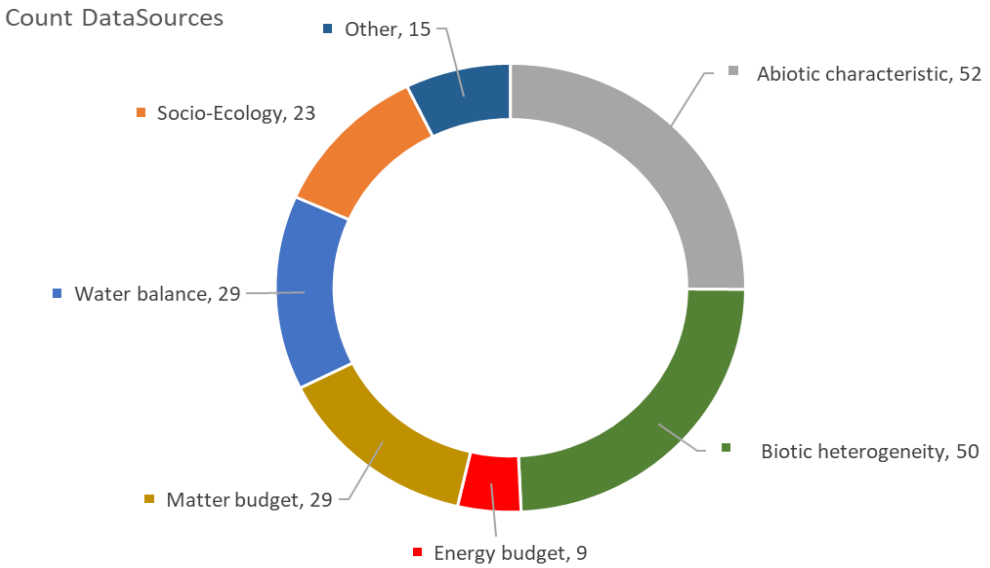

Figure 4: Count of data sources contributing to SO (level of Ecological Integrity Component) n=176 data sources considered (multiple assignments possible)

Hence, the available information spectrum on relevant data can serve to address a range of questions defined within science cases in WP8 (site/platform scale) and WP9 (catchment scale). An overview on recorded data sources at different scales is provided in the following subchapters and details on the thematic mapping of the data sources is given in Annex 4.

4.3 Data source description

We divided the identified relevant data sources addressing the eLTER data needs into four groups, which are described in the following chapters by their reference. This are: (a) data sources providing data on global scale, (b) data sources providing data on continental and regional scale with European focus, (c) data sources at national scale, and (d) catalogues provided by national eLTER networks. Whereas the first three reflect third-party data, the latter is in-alia addressing *in-situ* legacy data from eLTER.

We collected information on the data sources on the selected criteria for the assessment and description. Whereas, common description a thematic scope could mostly easily be accessed, information on metadata standards and catalogue endpoints was not easy to retrieve. In depth analysis will be needed, when developing process pipelines in future.

4.3.1 Global scale data sources

Based on our review, we provide a list of global data sources relevant for eLTER RI (Table 7). Detailed description of these data sources is provided at the specific websites as well as in Annex 3 of this report. In brief, we recorded data platforms both with more comprehensive data sets (e.g. PANGAEA or observatory networks such as FLUXNET) as well as data specific for the given data sources (e.g. SoilTemp). Temporal and spatial resolution varies depending on the dataset. The recorded information to different compartments is available in following order: terrestrial<climatic<aquatic<experiments<socio-ecology<diverse.

Table 7: List of relevant global scale data sources

Data source	Link	Compartments
FLUXNET	https://FLUXNET.org/	Terrestrial
SRDB v4.0 Global Database of Soil Respiration	https://daac.ornl.gov/SOILS/guides/SRDB_V4.html	Terrestrial
PhenoCam	https://phenocam.sr.unh.edu/webcam/	Terrestrial
INTERACT	https://eu-interact.org/	Terrestrial
GLORIA	https://www.gloria.ac.at/scope/aims	Terrestrial
ForestREplot	http://www.forestreplot.ugent.be/about.html	Terrestrial
Global Ants Database	http://globalants.org/	Terrestrial
BDD link to Global Mammal Assessment	https://globalmammal.org/activities/data-sets/	Terrestrial
BDD link to Global Amphibian Assessment	https://amphibiaweb.org/data/access.html	Terrestrial
MoveBank	https://www.movebank.org/cms/movebank-content/about-movebank	Terrestrial
FungalRoot	https://www.gbif.org/dataset/744edc21-8dd2-474e-8a0b-b8c3d56a3c2d	Terrestrial
SoilBON	https://geobon.org/bons/thematic-bon/soil-bon/	Terrestrial
TeaComposition	https://www.teacomposition.org/	Terrestrial

Data source	Link	Compartments
Global Carbon Project	https://www.globalcarbonproject.org/index.htm	Terrestrial
Global soil partnership	http://www.fao.org/global-soil-partnership/en/	Terrestrial
INI-International Nitrogen Initiative	https://initrogen.org/	Terrestrial
Global Partnership on Nutrient Management	https://www.unep.org/science-data	Terrestrial
SoilTemp	https://soiltemp.weebly.com/	Terrestrial
International Soil Moisture Network	https://ismn.geo.tuwien.ac.at/en/	Terrestrial
GBIF	https://www.gbif.org/dataset/search	Terrestrial
DRILOBase	http://geo.drilobase.org/	Terrestrial
Copernicus Global Land Services LAI300	https://land.copernicus.eu/global/products/lai	Terrestrial
ESA CCI BIOMASS v3.0	https://climate.esa.int/en/projects/biomass/	Terrestrial
ESA CCI LAND COVER v2.0.7	https://climate.esa.int/en/projects/land-cover/	Terrestrial
Land Use Harmonisation 2	https://luh.umd.edu/	Terrestrial
ESA CCI SNOW	https://climate.esa.int/en/projects/snow/	Terrestrial
ESA CCI SM v6.1	https://www.esa-soilmoisture-cci.org/index.php?q=node	Terrestrial
GLASS	http://www.glass.umd.edu/Overview.html	Terrestrial
SMOS Soil Moisture and Ocean Salinity	https://smos-diss.eo.esa.int/oads/access/	Terrestrial
MODIS	http://www.ntsg.umt.edu/project/modis/mod17.php	Terrestrial
CRYOBS-CLIM	http://data.cryobsclim.fr	Terrestrial
GLACIOCLIM	https://glacioclim.osug.fr/-Acces-aux-donnees-79-	Terrestrial
GLIMS	http://glims.colorado.edu/glacierdata/	Terrestrial
EBAS	http://ebas.nilu.no/	Terrestrial
Critical Zone Observatories	*https://czo-archive.criticalzone.org/national/*	Terrestrial
OBIS API	https://api.obis.org/	Aquatic
GLEON -Global Lake Ecological Observatory Network	https://gleon.org/data	Aquatic
ESA CCI LAKES v1.1	https://climate.esa.int/en/projects/lakes/	Aquatic
TeaComposition H2O	https://www.bluecarbonlab.org/teacomposition-h2o/	Aquatic
COPERNICUS	https://land.copernicus.eu/global/themes/water	Aquatic
ESA-CCI+	https://catalogue.ceda.ac.uk/uuid/3c324bb4ee394d0d876fe2e1db217378	Aquatic
Droughtnet	https://drought-net.colostate.edu/	Experiments
TreeDivNet	http://www.treedivnet.ugent.be/	Experiments
NutNet	https://nutnet.org/	Experiments
CHELSA v2.1	https://chelsa-climate.org/downloads/	Meteorology
WorldClim v2.1	https://worldclim.org/	Meteorology
CHPclim	https://www.chc.ucsb.edu/data/chpclim	Meteorology

Data source	Link	Compartments
GPCC	https://www.dwd.de/EN/ourservices/gpcc/gpcc.html	Meteorology
TRMM	https://climatedataguide.ucar.edu/climate-data/trmm-tropical-rainfall-measuring-mission	Meteorology
ERA-Interim	https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim	Meteorology
ERA5	https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5	Meteorology
HyMeX	https://mistrals.sedoo.fr/HyMeX/	Meteorology
ESGF	https://esgf-data.dkrz.de/projects/esgf-dkrz/	Meteorology
iNaturalist	https://www.inaturalist.org/observations	Socio-Ecology
Teatime4App	https://teatime4schools.at/teatime4app/	Socio-Ecology
Google Earth Engine	https://earthengine.google.com/	Diverse
GBIF API	https://www.gbif.org/developer/summary	Diverse
BioTime	https://biotime.st-andrews.ac.uk/	Diverse
SRTM	http://srtm.csi.cgiar.org/srtmdata/	Diverse

4.3.2 Continental scale data sources (European / regional)

Data available at the European scale originates often from national sources, which data are reported within different programs and hence show a high quality assurance as well as standardised data formats. Data sources for all six ecosystem integrity components are obtainable. A list of pan-European data sources and their links is provided in Table 8. More detailed description of these data sources can be found at the specific websites and as well as in Annex 3 of this report. The recorded information to different compartments is available in following order: terrestrial<climatic<aquatic<experiments<socio-ecology<diverse.

Table 8: List of relevant continental scale data sources

Data source	Link	Compartments
ICP FORESTS	http://icp-forests.net/	Terrestrial
ICP VEGETATION	https://icpvegetation.ceh.ac.uk/	Terrestrial
ICP Integrated Monitoring	https://www.syke.fi/en-US/Research__Development/Nature/Monitoring/Integrated_Monitoring	Terrestrial
ICP Modelling and Mapping	https://www.unece-wge.org/?page_id=23	Terrestrial
AMAP-Arctic monitoring and Assessment Programme working Group	https://www.amap.no/	Terrestrial
EMEP - European Monitoring and Evaluation Programme	https://www.ceip.at/webdab-emission-database http://ebas.nilu.no/ https://www.emep.int/mscw/mscw_moddata.html	Terrestrial
ACTRIS	http://ebas.nilu.no/	Terrestrial
ICOS - Integrated Carbon Observation System	https://www.icos-cp.eu/	Terrestrial
LUCAS 2009 TOPSOIL data	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	Terrestrial

Data source	Link	Compartments
(ESDAC)		
LUCAS 2015 TOPSOIL data (ESDAC)	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	Terrestrial
ESDAC European Soil Data Centre	https://esdac.jrc.ec.europa.eu/resource-type/datasets	Terrestrial
ESDAC	https://esdac.jrc.ec.europa.eu/resource-type/datasets	Terrestrial
Soil Geographical Database of Eurasia	https://esdac.jrc.ec.europa.eu/esbn/EUSIS.html	Terrestrial
PASERA	https://www.isric.org/projects/pan-european-soil-erosion-risk-assessment-pesera	Terrestrial
Art.11 of habitat directive (FFH)	https://ec.europa.eu/environment/nature/knowledge/rep_habitats/index_en.htm	Terrestrial
EUNIS	https://eunis.eea.europa.eu/	Terrestrial
EGDI European Geological Data Infrastructure	https://www.europe-geology.eu/about-egdi/	Terrestrial
The Farmland Bird Index	https://pecbms.info/trends-and-indicators/indicators/	Terrestrial
EuroBirdPortal	https://www.eurobirdportal.org/ebp/en/#home/HIRRUS/r52weeks/CUCCAN/r52weeks/	Terrestrial
EBBC Atlas	https://www.ebcc.info/?ID=34	Terrestrial
Copernicus Global Land Services SSM	https://land.copernicus.eu/global/products/ssm	Terrestrial
Copernicus Global Land Services SWI	https://land.copernicus.eu/global/products/swi	Terrestrial
European Fluxes Database Cluster	http://www.europe-fluxdata.eu/	Terrestrial
INSPIRE	https://inspire-geoportal.ec.europa.eu/	Terrestrial
COPERNICUS HRL Layers	https://land.copernicus.eu/pan-european/high-resolution-layers	Terrestrial
CLC+ work form MC	https://land.copernicus.eu/eagle	Terrestrial
SapFLUXNET Database	https://zenodo.org/record/3971689#.YGxkOugzbcs	Terrestrial
EFAS snow equivalent	https://www.efas.eu/en/data-access	Terrestrial
NitroEurope IP	http://www.nitroeurope.ceh.ac.uk/	Terrestrial
CORINE LANDCOVER (CLMS)	https://land.copernicus.eu/pan-european/corine-land-cover	Terrestrial
Ecosystem types of Europe	https://www.eea.europa.eu/data-and-maps/data/ecosystem-types-of-europe-1	Terrestrial
The Farmland Bird Index	https://pecbms.info/trends-and-indicators/indicators/	Terrestrial
CDDA (National designated areas)	https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-15	Terrestrial
ICP WATERS	http://www.icp-waters.no/	Aquatic
Danubius-RI	https://www.danubius-ri.eu/index.html	Aquatic
MarineStrategyFrameworkDirective	https://ec.europa.eu/environment/marine/eu-coast-and-marine-policy/marine-strategy-framework-directive/index_en.htm	Aquatic/marine
JERICO RI	https://www.jerico-ri.eu/	Aquatic/marine
Water Framework Directive	https://ec.europa.eu/environment/water/water-	Aquatic

Data source	Link	Compartments
	framework/info/intro_en.htm	
WATERBASE -Water quality ICM	https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality-icm-1	Aquatic
WATERBASE -Biology	https://www.eea.europa.eu/data-and-maps/data/waterbase-biology	Aquatic
CCM (Catchment Characterisation and Modelling)	https://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23	Aquatic
UWWTD	https://uwwtd.eu/	Aquatic
BioFresh	http://www.freshwaterplatform.eu/	Aquatic
COPERNICUS	https://land.copernicus.eu/global/themes/water	Aquatic
E-OBS	https://www.ecad.eu/download/ensembles/download.php	Climate
EuroCordex	https://cordex.org/?option=com_content&view=featured&Itemid=476 and https://cordex.org/domains/cordex-region-euro-cordex/	Climate
ECAD dataset (e-obs gridded dataset)	https://www.ecad.eu/dailydata/index.php	Climate
ECAD daily data	https://www.ecad.eu/dailydata/index.php	Climate
AQUACOSM	https://www.aquacosm.eu/	Experiments
EJP SOIL	https://ejpsoil.eu/	Experiments
CensusHub	https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false	Socio-Ecology
CDDA	https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-15	Socio-Ecology
European Social Survey	https://www.europeansocialsurvey.org/	Socio-Ecology

Table 9: List of generic data repositories (domain specific and domain agnostic)

Data source	Link	Compartments
FigShare	https://figshare.com/	diverse
Pangaea	https://www.pangaea.de/	diverse
Zenodo	https://www.eui.eu/Research/Library/ResearchDataServices	diverse
EUDAT B2SHARE	https://b2share.eudat.eu	diverse
Open Government Data	https://www.europeandataportal.eu/en	diverse

4.3.3 National scale data sources

In-situ data collected at the national level are another potentially relevant source of information for the eLTER RI since they cover a wide range of observations that are often collected with higher frequency and over a longer time span as compared to European datasets. Moreover, many collected datasets at national level comply with eLTER SO. Others are complementary and could be implemented during a later adaptation stage of the eLTER SO list. The main challenges for the identification of available national data sets are their site-based distribution within national catalogues and their documentation in national languages. Based on the experience and knowledge within the task team we focused on France as one national example to show the variety of data sources. More work needs to be done for the other European countries.

In France data related to socio-ecosystems taken in their wider range have been collected for many years now. A list of about 40 data sources is given in Annex 5 but it is certainly not exhaustive. Some of the datasets are indeed related to the eLTER SOVs: NAIADES provides information on the quality, in terms of hydrobiology, hydromorphology, physicochemistry and temperature, of aquatic bodies and several millions of data are available, since 1960 for some watersheds. Census data are available on the INSEE website (https://www.insee.fr) for the 35,000 French municipalities since 1876. Many of these datasets were initiated before the concept of SOs but they provide information which could be very complementary to the eLTER SOs. Some French datasets are directly related to European ones but with a finer spatial resolution, which is very important for local stakeholders. For example, the new Hydroportail (https://www.hydro.eaufrance.fr/) collects in real time water height and flowrates on 3000 stations. The main disadvantages are that these databases are relatively dispersed, without homogeneity in the way to search them, which generally requests the use of the French language. Some knowledge of the general geography of France is useful. Effort is being made actually to centralise the access and to modernise the way to question the database: the open platform for French public data (https://www.data.gouv.fr/fr/) is the entry point for some 42,000 datasets, mostly corresponding to social issues. However, the use of APIs is still not generalised. Open access, with or without registration, is more and more the rule, at least for research applications.

Discovery of *in-situ* data and the overview of national networks will be facilitated when well-compiled metadata information is stored in a common structure and shared through the eLTER DAR. An exemplary list of national scale data sources list and their links is provided in Annex 5.

4.3.4 *In-situ* legacy data from national eLTER networks

Most of the national eLTER networks do not provide a central catalogue on national level yet so that management and sharing of publicly available data/metadata is done on site level using e.g. DEIMS-SDR. Beside the complex workflow on getting together relevant data for integrated ecosystem studies, one of the main reasons for lacking the central catalogue is the shortage of resources for developing and maintaining such a catalogue at the national level, as the national LTER work is mostly unfinanced. Recently however, a lot of effort has been put into centralising national data sources that we would like to illustrate with a few examples shown in Table 10. Nevertheless, many provide information and data on part of the national LTER network but not for all sites.

Within the eLTER projects efforts have been started to integrate national scale metadata into the eLTER DIP for common discovery and access. With the current developments of the eLTER DAR this needs to be extended. Recently work has started to map SITES (SE) metadata into the eLTER Information System based on metadata harvesting.

Table 10: Examples for national scale eLTER databases

Data source	***Link and description***	***Modalities***	***Country***	***Language***
TERENO	*TERENO (Terrestrial Observation Network) is embarking on new paths with an interdisciplinary and long-term research programme involving six Helmholtz Association Centers. TERENO spans an Earth observation network across Germany that extends from the North German lowlands to the Bavarian Alps. This unique large-scale project aims to catalogue the longterm ecological, social and economic impact of global change at regional level. Scientists and researchers want to use their findings to show how humankind can best respond to these changes. It is an important element of the LTER Germany network.*	*Experiments*	*Germany*	*en*

Data source	***Link and description***	***Modalities***	***Country***	***Language***
	https://www.tereno.net/ *https://ddp.tereno.net/ddp/*			
GEM Greenland ecosystem monitoring	*https://g-e-m.dk/*	*Terrestrial*	*Greenland*	*en*
LTER Italy Data Nodes	*The LTER Italy Data Node provides data on marine observations covering part of the LTER Italy network mainly focusing on marine observations. Within the Ritmare project Sensor Web enabled data provision was developed and applied. Acri et al. (2020) published a data paper and harmonised marine long-term observation data.* *http://vesk.ve.ismar.cnr.it/* *http://getit.lteritalia.it/* *https://zenodo.org/communities/lter-italy/search?type=dataset*	*All*	*Italy*	*it*
SITES	*Swedish Infrastructure for Ecosystem Science (SITES, https://www.fieldsites.se/en-GB) facilitates national infrastructure for long-term field measurements (from 10 research stations) across terrestrial and aquatic ecosystems. Quality-controlled data are freely available through SITES Data Portal (https://data.fieldsites.se/portal/) and not published data are available on request.* *https://www.fieldsites.se/sv-SE*	*Terrestrial*	*Sweden*	*en*
ECN	*The Environmental Change Network (https://ecn.ac.uk/, UK-ECN) provides a catalogue (UK-ECN catalogue, http://data.ecn.ac.uk/) integrates 137 UK environmental monitoring sites where various physical, chemical, and biological data are collected. Raw data is available on the CEH Environmental Information Platform (https://catalogue.ceh.ac.uk/documents/4971bce4-8a81-4e23-9637-7fcff37c5f21) under an Open Government Licence and/or can be explored through diverse interfaces as Summary data (e.g. http://data.ecn.ac.uk/view.asp).* *http://data.ecn.ac.uk/* *https://catalogue.ceh.ac.uk/eidc/documents*	*Terrestrial / Aquatic*	*United Kingdom*	*en*
THEIA/OZCAR	*https://www.theia-land.fr/en/homepage-en/* *Jointly with the French Continental surface data pole Theia (https://www.theia-land.fr/en/homepage-en/), and in coordination with the French e-infrastructure Data Terra (https://www.data-terra.org/en/), OZCAR-RI (https://www.ozcar-ri.org/) is implementing a unique portal allowing transparent access to metadata and data from all OZCAR-RI's observatories according to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles. So far, 11 out of 21 observatories have implemented the pivot data model that allows their datasets to be visible in the data portal. It is accessible with recent versions of Firefox and Google Chrome (https://in-situ.theia-land.fr/). A web cataloguing service has also been implemented, making datasets on the Theia/OZCAR portal visible on the eLTER-DIP portal (https://dip.lter-europe.net/#/) of eLTER RI. For interoperability purposes, the vocabulary on variables names and categories has been published on the Web (https://in-situ.theia-land.fr/skosmos/theia_ozcar_thesaurus/en/) and links have been provided to ontologies relevant to the domain. Recent evolutions include the adoption of the I-ADOPT standard for variable names (https://i-adopt.github.io/index-en.html). The next steps include supporting all observatories in implementing the pivot data model, developing data exchange services (SensorThings) and*	*Terrestrial*	*France*	*en*

Data source	*Link and description*	*Modalities*	*Country*	*Language*
	providing users with the possibility to download data from the Theia/OZCAR portal in harmonised formats (work plan for 2022). The pivot data model will also be extended to all types of data, as only time series data have been processed so far.			
LTER France Data Node	*http://meta.data-za.org/geonetwork/srv/fre/catalog.search#/home The Zones Ateliers (ZA) focus on a functional unit (a river and its watershed, landscapes - agricultural or urban - and biodiversity, from Antarctica to sub-Saharan Africa, or the coastline, or the environments characterised by chronic radiation of natural or enhanced natural origin) and develop a specific scientific approach based on observations and experiments on workshops sites, to conduct multidisciplinary research in the long term. A ZA is therefore, most often, a network of workshop sites. The RZA has a portal and a geocatalogue (Geonetwork 3.4.2), which is currently being updated. This catalogue harvests the ZAs catalogues, if they exist.*	*All*	*France*	*fr/en*
WaLTER	*The WaLTER database (http://www.walterwaddenmonitor.org/tools/dataportaal/) on Wadden Sea ecosystems in the Netherlands, Germany and Denmark, has established automated extraction of data from several distributed databases containing information on: RWS geoservices (seagrass mapping, habitat mapping, morphology, water quantity and – quality parameters) are available as services and the bold ones are already in the data viewing facility for Datahuis Wadden; Data about birds (link to distributed data services is scheduled for 2022), national society for research on birds; Data about ecological parameters other than benthic, fish, mammals or birds (scheduled for 2022 from distributed data services); Data about benthic, fish and mammals (scheduled for 2022 from distributed data services) (Wageningen Marine Research); and Surveys of commercial fish species are automatically derived from ICES (Demersal Fish Survey, DATRAS data) and made available for reuse*	*marine/transitional*	*Netherlands*	*en*

DEIMS-SDR catalogue

A central catalogue services provided by eLTER is DEIMS-SDR (Dynamic Ecological Information Management System - Site and dataset registry; deims.org; Wohner et al., 2019) which aims to provide information about sites, locations, sensors, datasets, as well as network affiliations. Such information is used to manage the network, identify monitoring gaps and data handling habits at network level (Zilioli et al., 2019), coordinate communication with site principal investigators and field scientists, promote the sites and their scientific work and use this information in data publication workflows (Wohner et al., 2020). DEIMS-SDR is a site registry that provides information about each site's location, ecosystems, facilities, measured parameters, and research themes. It issues a persistent identifier for sites, called DEIMS.ID, which allows querying machine-readable site information through a variety of APIs (REST, CSW, WFS) (Wohner et al. in prep). In addition to sites, sensors can also be described and linked to existing site records, which are also exposed through the REST-API. DEIMS-SDR enables to document datasets from single sites and link to online repositories. Metadata are published using an Open Access licence (CC-BY 4.0 International).

As of March 2022, 615 sites and platforms are documented and registered for the eLTER network[5] and in total 1.009 datasets are described. Metadata on the datasets are managed and curated by single

[5] see https://deims.org/networks/4742ffca-65ac-4aae-815f-83738500a1fc

users mainly being site managers. For the data, an online distribution link to the respective data location is provided or a contact for data access. This online distribution links either point to public repositories (like EUDAT B2SHARE), national repositories, data portals, or services (e.g. SOS). Despite efforts to harmonise data structures applying the eLTER Data Reporting specification, still a number of different data structures exist. DEIMS-SDR enables the link from observations to the respective eLTER site via the deims.id, which is provided as persistent and unique identification. Table 11 shows an overview on the distribution of metadata on datasets across the different national LTER networks building up eLTER.

Table 11: Number of datasets documented on DEIMS-SDR (by 2022-03-23)

eLTER Networks	n datasets	eLTER Network	n datasets
LTER Austria	172	LTER Poland	11
LTER Belgium	7	LTER Portugal	44
LTER Bulgaria	17	LTER Romania	8
LTER Denmark	2	LTER Serbia	4
LTER Finland	18	LTER Slovakia	6
LTER France	2	LTER Slovenia	3
LTER Greece	3	LTER Spain	55
LTER Hungary	19	LTER Sweden	21
LTER Israel	18	LTER Switzerland	18
LTER Italia (Italy)	223	LTER-D (Germany)	95
LTER Latvia	14	Netherlands LTER	2
LTER Lithuania	2	UK ECN	243
LTER Norway	2		

eLTER data published via B2SHARE

DEIMS-SDR is a metadata catalogue managing inter-alia information on sites and the related datasets. It does not provide a repository functionality. This is implemented using the EOSC service B2SHARE provided by EUDAT. Currently, until the full implementation of the eLTER Central Data Node capabilities, the EUDAT data repository service B2SHARE[6] is used which is part of the EOSC service catalogue. B2SHARE is used to store and publish eLTER data in a FAIR manner, if no site or national repository is available. DOIs are issued for the published datasets together with basic metadata. More extensive description on the datasets is provided in DEIMS-SDR including an online distribution link to the respective record on B2SHARE.

In the eLTER context two B2SHARE repositories are used with no overlap in data content. This is due to historic reasons and depends on the preference of the data provider. B2SHARE is used as a service if no trusted site level repository can be provided. Firstly, B2SHARE hosted by CSC (https://b2share.eudat.eu/) and B2SHARE hosted by the Research Centre Jülich (https://b2share.fz-juelich.de/). Both instances allow to assign uploads to the eLTER community with a defined minimum set of metadata elements linked to DEIMS-SDR.

[6] see https://sp.eudat.eu/catalog/resources/709c053d-3b3f-4c54-8ef4-6efeea387816

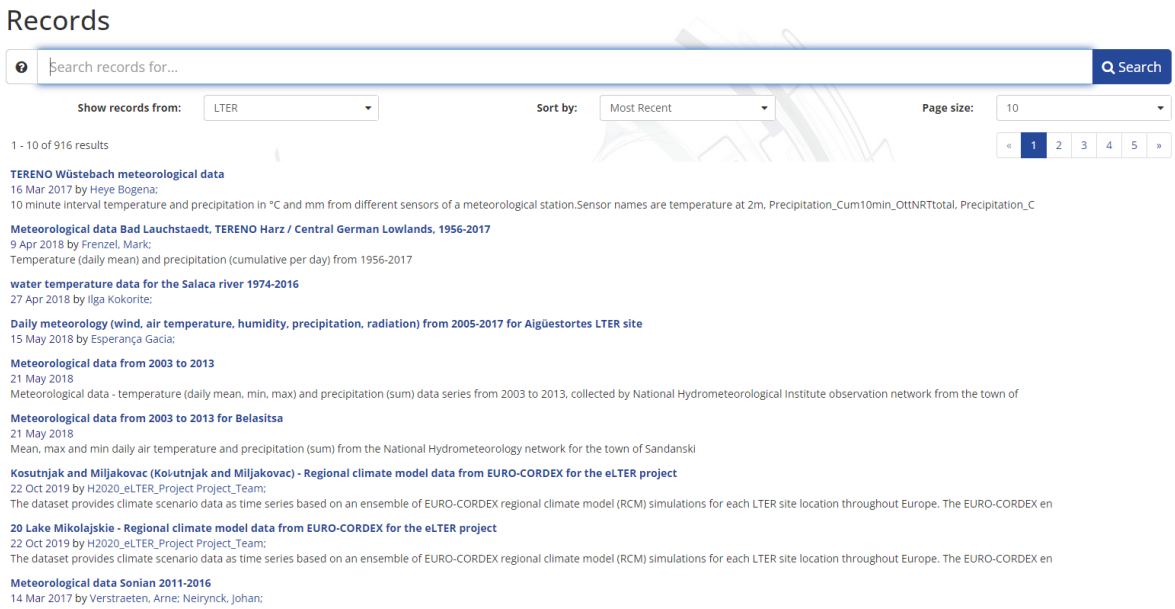

Figure 5: Search page for EUDAT B2SHARE (CSC)

Currently, in total 916 data sets from the eLTER community was shared using the EUDAT repository B2SHARE[7] provided by the central EUDAT services (CSC, see Figure 5) and 229 data on theB2SHARE[8] repository provided by the Research Centre Jülich (FZJ). Still efforts need to be taken to complete metadata linking in DEIMS-SDR and the eLTER DAR to complete the overview on available data and to easy grant access for scientific users.

[7] see https://b2share.eudat.eu/records?community=d952913c-451e-4b5c-817e-d578dc8a4469

[8] see https://b2share.fz-juelich.de/records/?community=d952913c-451e-4b5c-817e-d578dc8a4469

5 Retrieve relevant data sources

Based on the generic workflow described in chapter 3 we tested and prototyped examples for the "data source extraction workflows". As described in the introduction we defined 'legacy *in-situ* data' as non-harmonised historic data collected at eLTER Sites being held in local repositories (e.g. historic observations) or provided via a third-party service (e.g. monitoring networks (like ICP FOREST) or data initiatives (like FLUXNET)). Acri et al. (2020) show examples for the harmonisation of *in-situ* legacy data for marine observations for the North Adriatic Sea (Italy) which includes laborious work and contact to data providers. This process cannot easily be automatized.

In the project context WP7 deals with the collection of site *in-situ* observation and the related SCs in WP8 and WP9 with data processing and analysis. Therefore, we focused on providing demonstrators on additional data sources to complement the data landscape for eLTER. We selected demonstrators based on three criteria. Firstly, its relevance for RC tackled in the eLTER PLUS project. Secondly, prototyping different access modes to the data sources. Finally, prototyping different spatial extents and resolution of data. This resulted in the following demonstrators being described in the following section:

- retrieving occurrence biodiversity data based on API access
- retrieving harmonised site gas flux observation data based on download
- retrieving data from E-OBS historic data to calculate climate diagrams for sites
- retrieving data from gridded and modelled data based on the site extent
- retrieving earth observation data product based on site extent

5.1 Demonstrator Biodiversity Data

Alessandro Oggioni, Martina Zilioli & Paolo Tagliolato

Species occurrences (i.e., presence-only/site occupancy of species) represent useful measures to estimate the population status for selected species (e.g., protected, vulnerable, threatened) across space and time. *In-situ* observations, even if raw, are increasingly exploited for ecological studies and to develop species distribution models and available through data publishers and service providers (e.g., iNaturalist, Global Biodiversity Information Facility (GBIF), Ocean Biodiversity Information System (OBIS)). These data infrastructures, although characterised by different disciplinary scopes (e.g., ocean and marine systems) and contributing communities (e.g., volunteers, museum conservators), represent valuable sources of open legacy data[9]. With respect to the concept of Information Clusters introduced in the report, external data sources can also feed the eLTER biodiversity data pipelines and its integration can be beneficial. However, suitable APIs and IT tools are needed for data users to assist the automatic extraction and harmonisation of legacy data.

In this regard, the ReLTER package (Oggioni, Silver, Ranghetti & Tagliolato, 2021) is a collection of multi-purpose R functions developed to improve access and increase quality of eLTER data. Among these functions, *get_site_speciesOccurrences()* has the principal aim to allow its users to harvest legacy data (i.e., species occurrence records) for a specified eLTER site from iNaturalist, GBIF, and OBIS. By coupling this with other ReLTER package functions (e.g., *map_occ_gbif2elter()*, *save_occ_eLTER_reporting_Archive()*) specifically tailored to harmonise the output, users can then

[9] In this context, we use the term 'legacy data' to indicate the historical data acquired from third party service

automatically structure the harvested records (which are normally organised in tabular data format) in the eLTER Data Reporting template (Peterseil & Geiger, 2020).

ReLTER package code is released through a GitHub repository with GPL-3.0 Licence. Package and functions, including the *get_site_speciesOccurrences()* and related functions, are followed the rOpenSci (rOpenSci et al., 2021) practices and the documentation can be reached at this webpage: https://oggioniale.github.io/ReLTER/. At the time of delivering this document, the ReLTER package is under open peer review[10] of rOpenSci community. Fully functioning Docker images with preinstalled ReLTER package and the RStudio environment are also available at https://hub.docker.com/r/ptagliolato/rocker_relter.

A basic use case for the extraction of biodiversity data with ReLTER can be described using UML common style-textual content representation with the intent to illustrate the actions the ReLTER user is required for retrieving occurrence records from iNaturalist/GBIF/OBIS by exploiting the ReLTER functions.

The key elements of this representation are the Main Success Scenario, which is the sequence of steps, which allow the user to achieve the goal defined by the use case, and the extensions, i.e., variations on the Main Success Scenario. Extensions illustrate alternative interactions between user and system and are referred to the steps numbered in the Main Success Scenario.

Goal of the use case: Species occurrences harvesting from external data service for an eLTER site

Main Success Scenario

1. User downloads R Studio (follow R Studio instructions)
2. User instals ReLTER package "dev__withImprovements" branch[11]
3. User prepares the ReLTER *get_site_speciesOccurrences()* function call defining the input and other parameters for tuning the output:
    a. user sets parameter '**deimsid**' to select the area of interest from which data are to be harvested (i.e., paste DEIMS.ID **of the site** of interest; e.g., https://deims.org/758087d7-231f-4f07-bd7e-6922e0c283fd)
    b. user sets parameter '**list_DS**' to select the external data service to be queried (i.e., 'gbif' or 'obis' or 'inaturalist')
    c. user sets parameter '**show_map**' to get data in map data format in addition to tabular data format (i.e., show_map = TRUE)
    d. user sets parameter '**limit**' to select the records number (i.e., species presence record) to be extracted
4. User runs *get_site_speciesOccurrences()* function obtaining the outputs according to the given parameters
5. User visualises fetched data in the selected data formats (i.e. tabular data/map data)
6. User separates the output of step 4 per originating source and passes them to the mapping functions (*map_occ_gbif2elter()*, *map_occ_inat2elter()*, *map_occ_obis2elter()*)
7. User exports harvested records in eLTER data reporting template zip archives by function *save_occ_eLTER_reporting_Archive()*

[10] https://github.com/ropensci/software-review/issues/485

[11] At the moment of writing, the *get_site_specieOccurrences()* and related functions are not included in the main version of the package. Before following this point of scenario verify the improvements of the code, if you don't find the function *get_site_specieOccurrences()* in the "main" branch of the ReLTER package, install the package from the branch "*dev__withImprovements" with* this R statement: *devtools::install_github('https://github.com/oggioniale/ReLTER',ref = 'dev__withImprovements')*.

Extensions

1a, 2a: User logs into DataLabs and access to “Species Occurrences” Notebooks (Projects > Notebooks or https://datalab.datalabs.ceh.ac.uk/resource/elter/relterspocc)

1b, 2b: User, with a computer with docker, runs the Docker image *Rocker_ReLTER* and operates in the RStudio server through the web browser

4a: If user doesn’t own the DEIMS.ID, user looks for the ‘deimsid’ string by consulting DEIMS-SDR website and by extracting the DEIMS.ID from the site landing page

4b: User can choose more than one external data service (i.e., c("gbif", "inat", "obis")) to get data and combine them

4c: User can choose to leave ’show_map’ parameter without value to show only tabular data

The workflow of the aforementioned process is illustrated as a Unified Modelling Language (UML) activity diagram (Fowler 2004) in Figure 6. User actions trigger the workflow, which aim at (a) selecting the eLTER site to probe and (b) customising the parameters necessary to retrieve the occurrence records according to the user preferences (box 1). Three paths fork from the box since data sources (iNaturalist, GBIF, OBIS) can be queried in parallel (all or in double combination) or one at a time according to the parameter setting. Harvested occurrences are then visualised in a single map and sorted in three tabular data formats specific to the exporting schema of the original data publisher. In the second branching of the workflow activities, the user runs the functions to transform the exporting schemas in the eLTER Data Reporting template. The script generates archives (zip) containing CSV files following the eLTER Data Reporting specification. The user can exploit this data for further analysis or processing providing the reference to the origin as provenance.

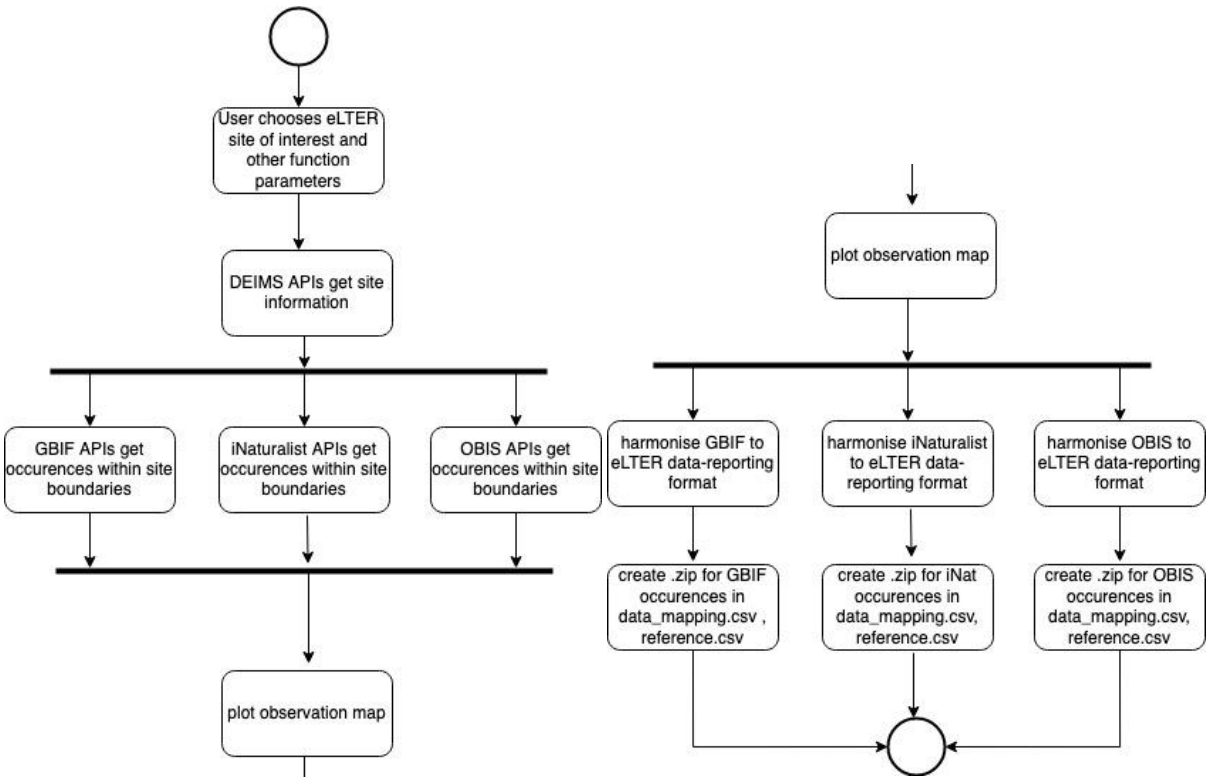

Figure 6: Activity diagram for ReLTER Biodiversity workflow showing the steps from external services to eLTER reporting format. Circles indicate the start and end point of the process. Figure split and continues with ‘plot observation map’ on the right side.

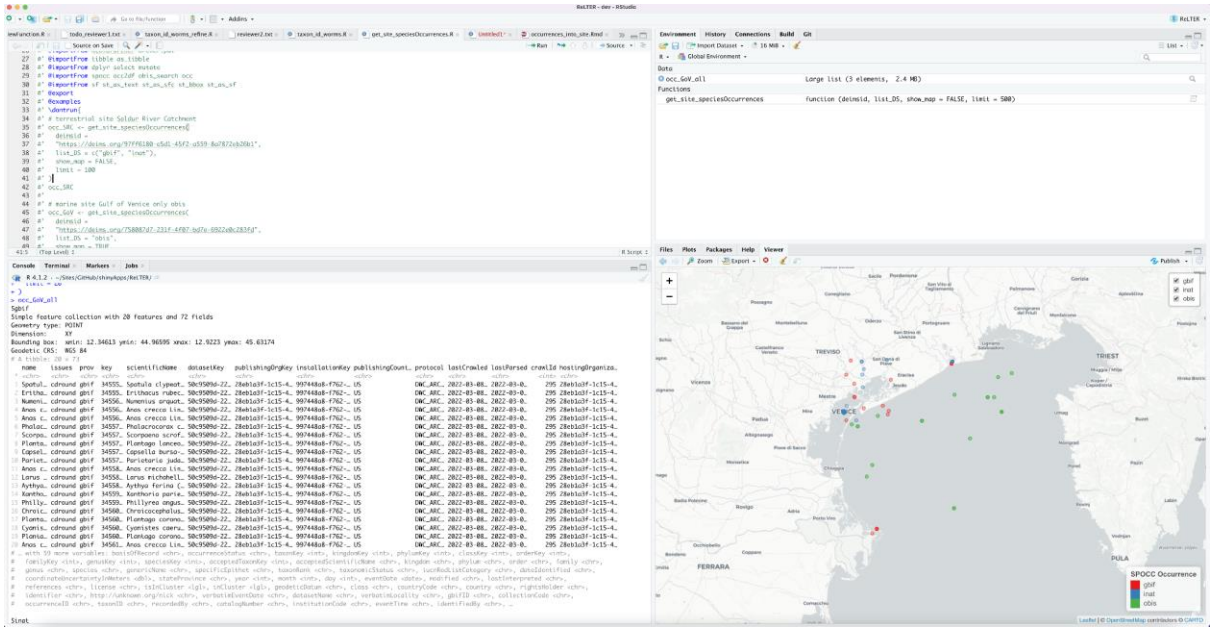

Figure 7: RStudio interface Outputs of the get_site_speciesOccurrences() *ReLTER function are provided in Console and Plots panels respectively. The function is used to retrieve GBIF data for eLTER marine site "Gulf of Venice" (Italy).*

Figure 7 illustrates the RStudio interface after launching the *get_site_speciesOccurrences()* function (steps 1-5). The function is shown in the Script editor panel while Console and Plot panels show the fetched data respectively in tabular and map data formats. Tabular data formats are exposed according to the schema (i.e., columns order, attribute names/headers and data types) of their data publisher.

The structure of species occurrence records reflects the exporting schema of the original data publisher. Fields mapping among the three data source schemas (i.e., GBIF, iNaturalist, OBIS) and the eLTER Data Reporting template (Peterseil & Geiger, 2020) was carried out so as to design three functions (i.e., *map_occ_gbif2elter()*, *map_occ_inat2elter()*, *map_occ_obis2elter()*) to structure fetched data in simplified output data formats based on eLTER data_mapping.csv template[12]. The output data are harmonised according to the eLTER Data Reporting specification and is more readable since only essential attributes are maintained.

Three requirements have been considered for fields mapping:

- assuring whole mapping of the "mandatory" fields of the eLTER data reporting template (i.e., data_mapping.csv) to the fields of each of the three data source schemas
- assuring preservation of relevant (e.g., useful for data re-use) attributes of data sources schemas *(i.e., species name, time, lat, long, provider, licensing)* by introducing additional columns to eLTER data-reporting template; This was performed by following guidelines contained in technical specifications (e.g., adding extended fields, creating new on-purpose fields)
- maximise coverage among fields of the eLTER data-reporting template and of data sources schemas

[12] Among the eLTER data-reporting templates, data_mapping.csv was chosen as its structure is conceived to store occurrences information of an entity, coherent with species occurrence records published by iNaturalist/GBIF/OBIS

The tables below show examples of the output data formats generated through steps 6 and 7 of the use case. Table 12 presents the output data format for GBIF occurrence records. GBIF data have been structured in the mandatory fields of data_mapping.csv template. New columns are added to the template's schema to maintain useful GBIF data. New fields (e.g., record_id) are documented in a designated reference file (Table 13). Also, GBIF data for mandatory fields which need to be further detailed (e.g., taxa, named "Extended fields") are collected in a specific reference file (Table 14).

Table 12: Example of data_mapping.csv output for GBIF data. The output format is featured by "mandatory" eLTER reporting columns and added columns (e.g., record_id, dataset_key)

	A	B	C	D	E	F	G	H	I	J	K
1	SITE_CODE	ABS_POSITION	TIME	VARIABLE	TAXA	VALUE	ORG_NAME	RECORD_ID	DATASET_KEY	INSTITUTION_CODE	LICENSE
2	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	5231240	TRUE	gbif	3525905398	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
3	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	7745240	TRUE	gbif	3535038067	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
4	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	2492521	TRUE	gbif	3543775395	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
5	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	2482473	TRUE	gbif	3544657301	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
6	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	2498326	TRUE	gbif	3629833769	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
7	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	11034996	TRUE	gbif	3638749926	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
8	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	6871058	TRUE	gbif	3639928644	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
9	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	5229168	TRUE	gbif	3642208512	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode
10	https://deims.org/97ff6180-e5d1-45f2-a559-8a7872eb26b1	c(10.539314 46.72105)	2021-05-22TNA	occurrence	5231918	TRUE	gbif	3646646622	4fa7b334-ce0d-4e88-aaae-2e0c138d049e	CLO	http://creativecommons.org/licenses/by/4.0/legalcode

Table 13: Example of reference_VARIABLES.csv for GBIF records. The output Reference is generated in order to document added columns reported in data_mapping.csv .

	A	B	C	D
1	FIELD_NAME	VARIABLE_CODE	VARIABLE_NAME	VARIABLE_DEFINITION
2	VARIABLE	RECORD_ID	gbif record id	
3	VARIABLE	DATASET_KEY	gbif datasetKey	
4	VARIABLE	INSTITUTION_CODE	gbif institution code	
5	VARIABLE	LICENSE	licence applied to each gbif record	

Table 14: Example of reference_TAXA.csv for GBIF records

	A	B	C	D
1	FIELD_NAME	CODE	NAME	CODE_URL
2	TAXA	5231240	Oenanthe oenanthe (Linnaeus 1758)	https://www.gbif.org/species/5231240
3	TAXA	7745240	Lanius collurio Linnaeus 1758	https://www.gbif.org/species/7745240
4	TAXA	2492521	Saxicola rubetra (Linnaeus 1758)	https://www.gbif.org/species/2492521
5	TAXA	2482473	Coloeus monedula (Linnaeus 1758)	https://www.gbif.org/species/2482473
6	TAXA	2498326	Bucephala clangula (Linnaeus 1758)	https://www.gbif.org/species/2498326
7	TAXA	11034996	Sylvia curruca (Linnaeus 1758)	https://www.gbif.org/species/11034996
8	TAXA	6871058	Acanthis flammea cabaret (P.L.Statius Mvºller 1776)	https://www.gbif.org/species/6871058
9	TAXA	5229168	Milvus milvus (Linnaeus 1758)	https://www.gbif.org/species/5229168
10	TAXA	5231918	Cuculus canorus Linnaeus 1758	https://www.gbif.org/species/5231918
11	TAXA	6092830	Spinus spinus (Linnaeus 1758)	https://www.gbif.org/species/6092830

Figure 8 illustrates the RStudio interface after launching the *map_occ_obis2elter* and *save_occ_eLTER_reporting_Archive()* functions (steps 6-7). The functions are shown in the Script editor panel while Console and Plot panels show the fetched data respectively in tabular and map data formats. Tabular data formats are exposed according to the eLTER Data Reporting template requirements.

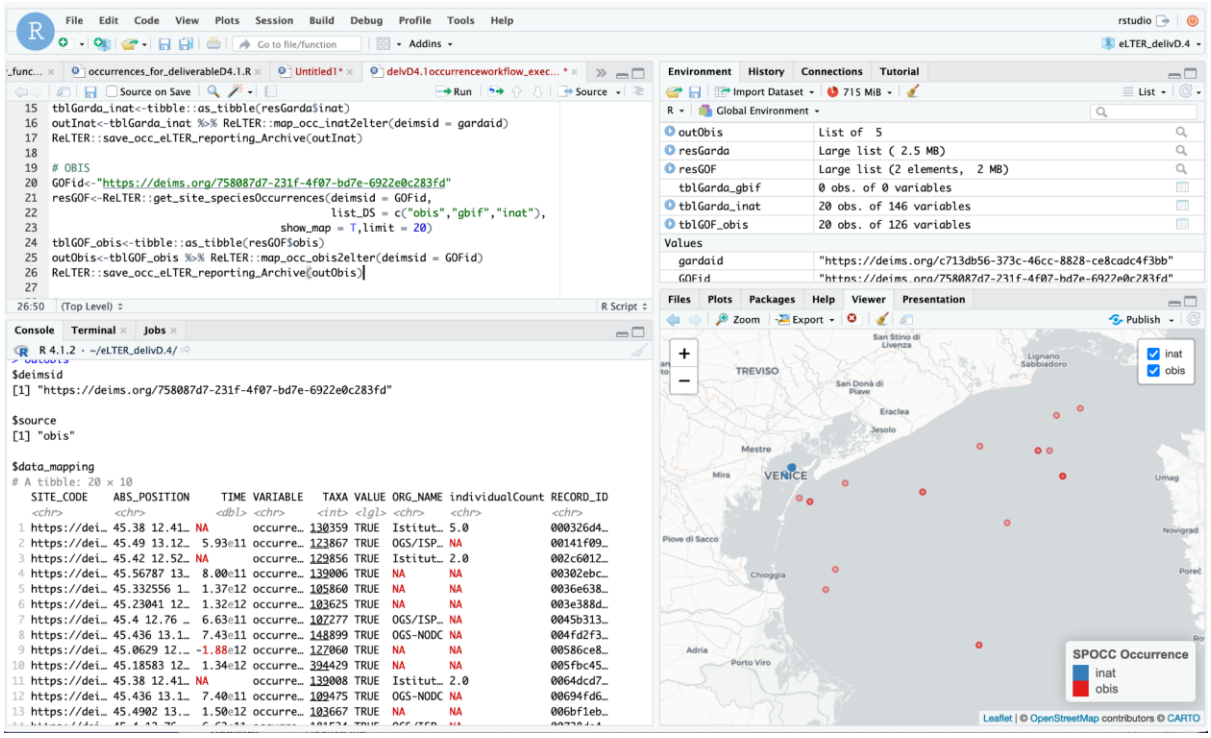

Figure 8: Transform the get_site_speciesOccurrences() *output in eLTER data reporting template and save it as data_mapping.csv*

The *get_site_speciesOccurrences()* function follows the Cookie-cutter approach and it is well suited to be integrated in a Shiny web application. Taking into account both project and users requirements, the task (WP11 - task 11.1) devoted to the data workflow development can consider it as a useful implementation. The Biodiversity Loss workflow, described above, is also recommended for Socio-Ecological Systems (SES). Citizen Science contributions to the knowledge of biodiversity in the eLTER site can be known through the same ReLTER function and the output can be shared within Research Infrastructure using the eLTER data-reporting template by *map_occ_gbif2elter()* and *save_occ_eLTER_reporting_Archive()* ReLTER functions. We provide a guide of the workflow and the demonstration of described functionalities[13,14].

References to the source code:

- https://datalab.datalabs.ceh.ac.uk/resource/elter/relterspocc
- https://github.com/oggioniale/ReLTER - at the time of delivering this document, the ReLTER package is under open peer review[15] of rOpenSci community
- https://hub.docker.com/r/ptagliolato/rocker_relter - fully functioning Docker images with preinstalled ReLTER package and the RStudio environment

[13] See https://rawcdn.githack.com/oggioniale/ReLTER/6713256bc20aed79cbb03a0787dd5c4f4aef4e42/docs/articles/occurrences_into_site.html

[14] See source of the referred page at the (permanent) link: https://github.com/oggioniale/ReLTER/blob/6713256bc20aed79cbb03a0787dd5c4f4aef4e42/docs/articles/occurrences_into_site.html

[15] https://github.com/ropensci/software-review/issues/485

Outlook:

The presented workflow for the biodiversity data demonstrator follows the cookie-cutter approach (Halada et al., 2022) by presenting the eLTER community external data in the context of the eLTER network research sites. We plan to embed the workflow within a standalone web application realized with R-Shiny. This way it will be possible to ease the user not confident with R programming. At the same time, it will be possible to make it available as a new online service for the eLTER network, hosting the Shiny App in the eLTER DataLabs platform. The ReLTER package is an open source collaborative project, and hence we expect that other data sources and functions will be added in the future, and possibly concur in ameliorating the workflow by extending it to other resources found within the web. Some other functions (e.g. *get_site_ODS()*, *get_site_MODIS()*), already implemented in ReLTER package, can be exploited to construct analogous workflows and services.

The workflow could meet the goals of activities framed in the Socio-Ecological Systems (SES) RC. Citizen Science contributions to the knowledge of biodiversity in the eLTER site can be known through the same workflow and the output can be shared within the Research Infrastructure using the eLTER data-reporting template. Finally, we intend to test the ReLTER workflow within the eLTER network and then investigate its usefulness for different RC scientific groups. Feedbacks and use cases collection from the eLTER users could guide further ReLTER workflow improvement.

5.2 Demonstrator Gas Flux Data

Johannes Peterseil, Will Bolton, Christian Poppe & Christoph Wohner

Ecosystem research on hydrological and bio-geochemical processes of ecosystems strongly relies on measurements of *in-situ* gas flux data e.g. to quantify the effect of water deficits on ecosystem functions and health. The need to validate complex ecosystem models increases the importance of easy access, assessment and download of harmonised datasets of multiple observation networks (e.g. FLUXNET and ICOS) that can currently just be accessed individually. Aiming for this, we developed a prototype demonstrator for gas flux data from such networks and stations corresponding to eLTER. The prototype demonstrator takes metadata and selected variables from the gas-flux networks for a chosen eLTER Site, visualises them and offers a direct data download in the eLTER data format (Peterseil & Geiger, 2020). We tested some capabilities, which can be upgraded for the implementation of the Information Clusters.

A basic use case was defined as the basis for the prototype. Here, a series of steps can be defined from a user point of view. The main success story defines the sequence of steps which allow the user to achieve the goal defined by the use case. This information was used to define a workflow diagram (see Figure 9).

Goal of the use case: Retrieve *in-situ* gas-flux data from external data sources / services for a given eLTER site.

Main Success Scenario

1. User logs in to the DataLabs shiny application to execute the workflow
2. User selects the workflow for 'FLUXNET/ICOS'
3. User sets input parameters for the workflow:
    a. user sets parameter '**Data**' to select the external data source to be queried (i.e., 'FLUXNET' for FLUXNET 2015 dataset or 'ICOS' for Drought 2018 dataset)
    b. user sets parameter **'Variable'** to select the list of variables to be mapped and transformed into the output data file (e.g. TA_F_MDS), which automatically sets the respective quality control column

   c. user sets parameter **'eLTER Site'** to select the area of interest from which data are to be retrieved (i.e., select DEIMS.ID of the site of interest; e.g., https://deims.org/758087d7-231f-4f07-bd7e-6922e0c283fd from the list)
4. User executes the download button to download the data file (data.csv) and related provenance information (metadata.txt) as tabular data following the eLTER Data Reporting format
5. User downloads the visualisation of the fetched data as jpg

Building on the experience of the 'cookie-cutter' approach (Lubos et al. 2022) the eLTER DataLabs[16] developed by WP11 were used for data processing. For the pilot phase of the demonstrator, the FLUXNET 2015 dataset[17] (Pastorello et al., 2020) as well as the Drought 2018 dataset[18] provided by ICOS (Drought 2018 Team, and ICOS Ecosystem Thematic Centre., 2020) were used to showcase data retrieval and harmonisation. Both of them follow the FLUXNET data format and can be dealt with in a combined workflow. Data access for both of the datasets have been through the respective portal and downloading the required datasets for further processing manually. For both, B2DROP was used as intermediate storage and DataLabs data store to provide access for further processing. Recently, a python library[19] has been provided by ICOS which will be evaluated in future.

A basic workflow was defined (see Figure 9), retrieving the information on the site of interest from DEIMS-SDR and merging the information with the sites provided by the input datasets.

[16] see https://datalab.datalabs.ceh.ac.uk/projects/elter/info

[17] see https://fluxnet.org/data/fluxnet2015-dataset/

[18] see https://www.icos-cp.eu/data-products/YVR0-4898

[19] See https://github.com/ICOS-Carbon-Portal/pylib

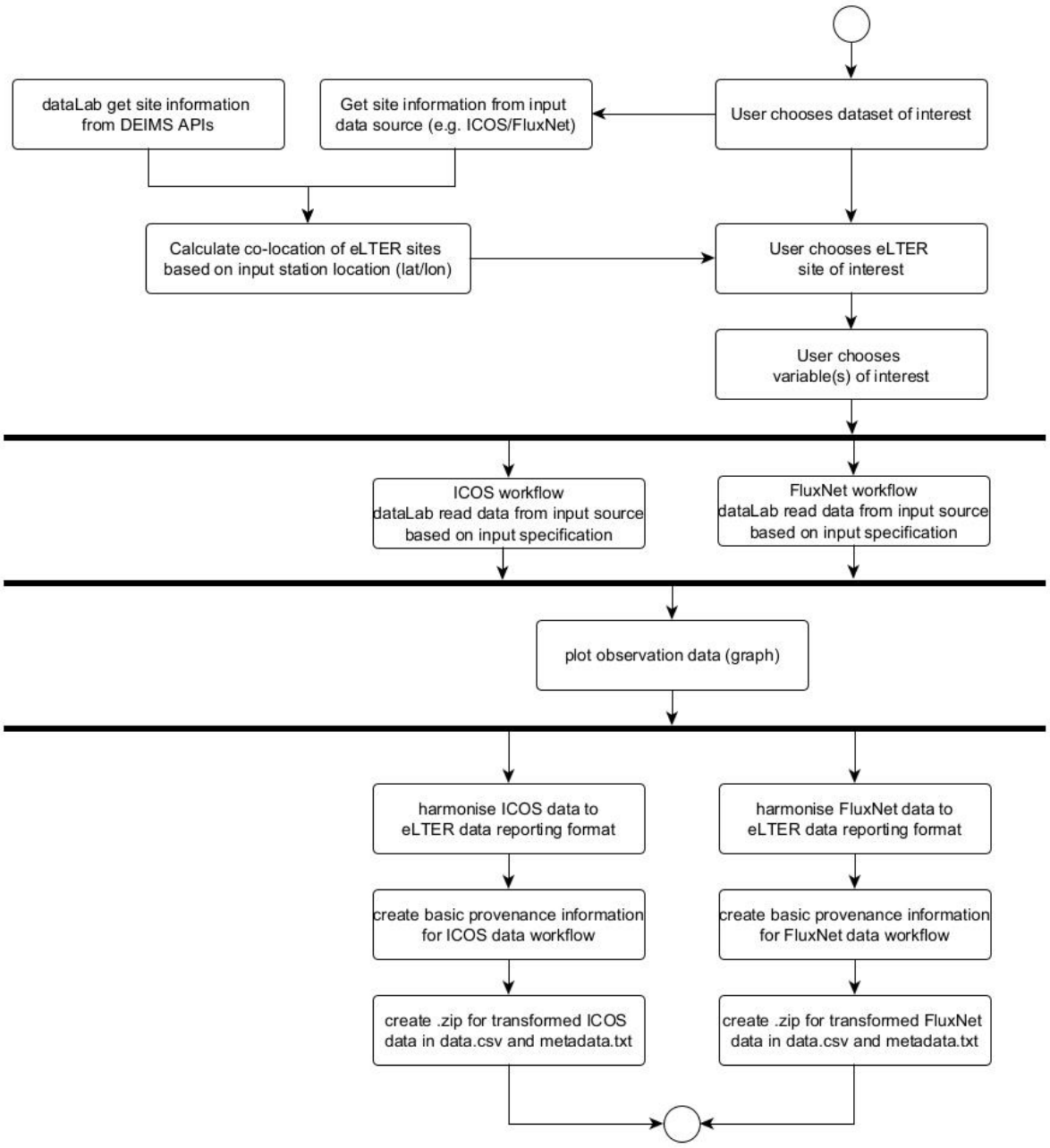

Figure 9: Demonstrator workflow to retrieve and harmonise FLUXNET / ICOS data using the eLTER DataLabs

Assigning appropriate eLTER sites

Co-location of observations at long-term monitoring sites is an important asset. In DEIMS-SDR this is addressed by assigning this type of information within the network affiliation together with the respective identification in the member network or research infrastructure. This information is managed by the site managers and sometimes is not complete or missing, if e.g. new networks are considered. It is important to note, that the naming of the sites and stations in different monitoring networks or research infrastructures could differ. With the DEIMS.ID we tried to address this issue and provide a solution, but not in every catalogue the DEIMS-ID is used (Wohner et al., 2020). Therefore, a mapping based on the location of the sites and stations was implemented to address the issue of co-location.

For this purpose, the DEIMS python package (PyPI 2022), was extended by a new function that allows spatial searches called *getSitesWithinRadius()*. By providing point coordinates and a search radius, the function will return all sites located within the defined search radius sorted by the distance to the input coordinates. This function was used to calculate the three sites on DEIMS-SDR located closest to each FLUXNET/ICOS (see Figure 10, Table 15). We took the results of the function to create a mapping table from FLUXNET/ICOS stations to eLTER sites documented in DEIMS-SDR applying a distance filter for the selection process (e.g. < 1,5 km). The mapping table is used as a first step of the extraction workflow.

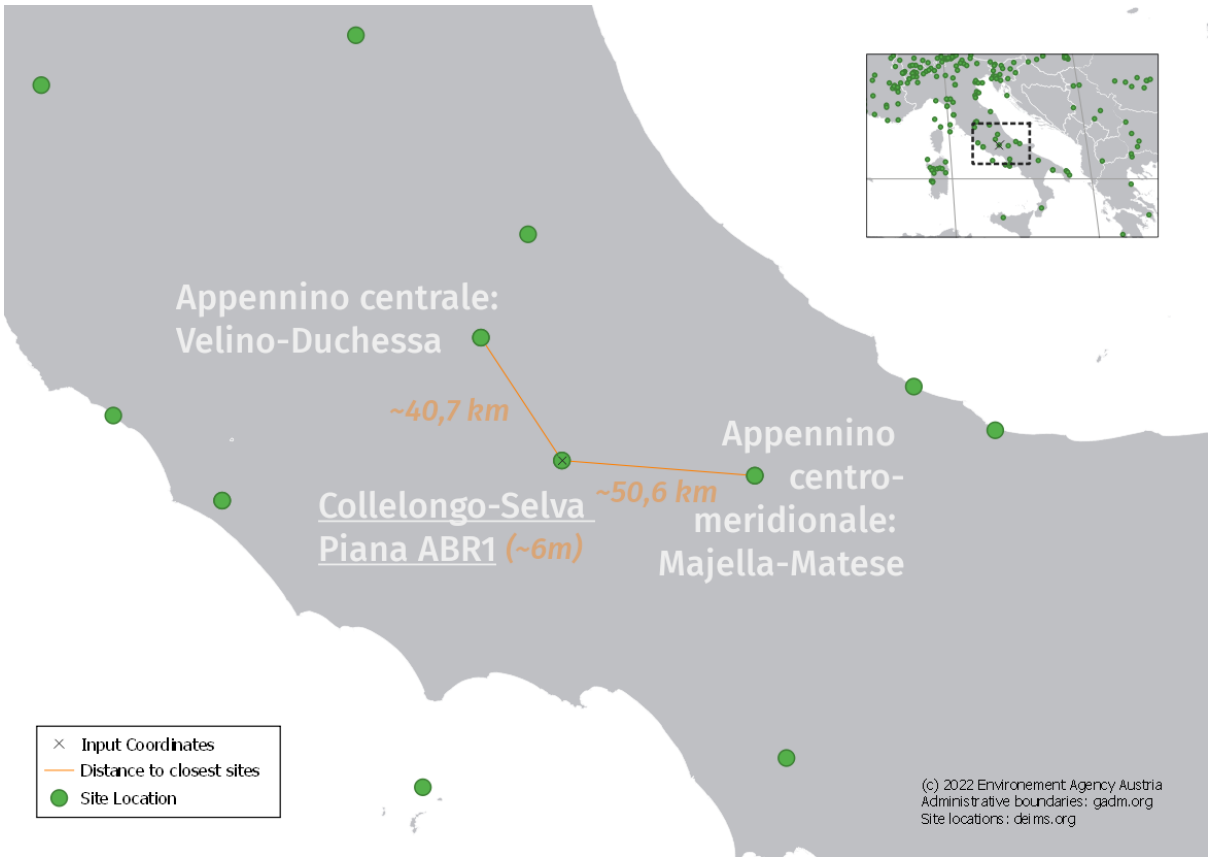

Figure 10: Cartographic representation of the distance calculation for FLUXNET station IT-Col (Lat: 41.84936, Lon: 13.58814)

Table 15: Mapping of FLUXNET sites to eLTER sites documented on DEIMS-SDR (filter <1.5km)

FLUXNET site	lat	lon	closest site on deims-sdr within search radius (5km)	dist in km
AT-Neu	47.11667	4.6599	Stubai (combination of Neustift meadows and Kaserstattalm) - Austria	1.328
BE-Bra	51.30761	4.51984	Brasschaat - De Inslag - Belgium	0.023
BE-Lon	50.55162	4.74623	Lonzée - Belgium	0.008
BE-Vie	50.30493	5.99812	Vielsalm Terrestrial Observatory - Belgium	0.016
CH-Dav	46.81533	9.85591	Davos Seehornwald - Switzerland	0.053
CH-Lae	47.47833	8.36439	Laegeren - Switzerland	0.003
CZ-BK1	49.50208	18.53688	Bily Kriz - Czechia	0.002
CZ-BK2	49.49443	18.54285	Bily Kriz - Czechia	0.956

FLUXNET site	lat	lon	closest site on deims-sdr within search radius (5km)	dist in km
CZ-wet	49.02465	14.77035	Třeboň wet meadows - Czechia	0.007
DE-RuR	50.62191	6.30413	TERENO - Rollesbroich - Germany	0.361
DE-RuS	50.86591	6.44714	TERENO - Selhausen - Germany	0.018
DE-Seh	50.87062	6.44965	TERENO - Selhausen - Germany	0.55
FI-Hyy	61.84741	24.29477	Hyytiälä SMEAR II LTER - Finland	0.012
IT-Col	41.84936	13.58814	Collelongo-Selva Piana ABR1 - Italy	0.006
IT-Ren	46.58686	11.43369	Renon BOL1 - Italy	0.01
IT-Tor	45.84444	7.57806	Torgnon grassland Tellinod (IT19 Aosta Valley) - Italy	0.516

(source data: https://FLUXNET.org/sites/site-list-and-pages/)

The current function uses the centroid coordinates provided by DEIMS-SDR. It can be applied to any input source providing a name and coordinates as lat/lon (WGS84). Further work could include e.g. the use of site boundaries and locations of the eLTER sites enhancing the current capabilities.

To simplify the process for the prototype implementation of the workflow we focused on the following sites showcasing the workflow:

- Stubai (Austria, https://deims.org/324f92a3-5940-4790-9738-5aa21992511c)
- Rollesbroich (Germany, https://deims.org/356417de-5a3c-429d-82c1-08a4e924ab3b)
- Hyytiälä SMEAR II LTER (Finland, https://deims.org/663dac80-211d-4c19-a356-04ee0da0f0eb)
- Torgnon Larch forest Tronchaney (IT19 Aosta Valley) (Italy, https://deims.org/4312983f-c36a-4b46-b10a-a9dea2172849)

Currently this was done based on the mapping table created prior to the workflow. In the later phase of the implementation the list of sites will be extended to all available sources in the input dataset.

Selection of variables

To simplify the process, for the prototype a number of variables have been selected with daily resolution (DD) which could be selected by the users. This is a subset of the full variable list provided by FLUXNET[20] which will be implemented in a later version of the prototype. As normally more than one variable is of interest for the user, a multi-select function for the variables of interest have been implemented.

Table 16: Table X: Selected list of FLUXNET variables for the prototype

Variable	Description
TA_F_MDS (incl. _QC)	Air temperature, gap filled using MDS method
P_F (incl. _QC)	Precipitation consolidated from P and P_ERA
WS_F (incl. _QC)	Wind speed, consolidated from WS and WS_ERA
SWC_F_MDS_1 (incl. _QC)	Soil water content, gap filled with MDS (numeric index "#" increases with the depth, 1 is shallowest)

[20] see https://fluxnet.org/data/fluxnet2015-dataset/fullset-data-product/

Variable	Description
SW_IN_F_MDS (incl. _QC)	Shortwave radiation, incoming, gap filled using MDS (negative values set to zero, e.g., negative values from instrumentation noise)
VPD_F_MDS (incl. _QC)	Vapour Pressure Deficit, gap filled using MDS
LE_F_MDS (incl. _QC)	Latent heat flux, gap filled using MDS method
NEE_VUT_REF (incl. _QC)	Net Ecosystem Exchange, using Variable Ustar Threshold (VUT) for each year, reference selected on the basis of the model efficiency (MEF). The MEF analysis is repeated for each time aggregation
NEE_VUT_USTAR50 (incl. _QC)	Net Ecosystem Exchange, using Variable Ustar Threshold (VUT) for each year, from 50 percentile of USTAR threshold
RECO_NT_VUT_REF	Ecosystem Respiration, from Nighttime partitioning method, reference selected from RECO versions using model efficiency (MEF). The MEF analysis is repeated for each time aggregation
RECO_NT_VUT_USTAR50	Ecosystem Respiration, from Nighttime partitioning method, based on NEE_VUT_USTAR50
GPP_NT_VUT_REF	Gross Primary Production, from Nighttime partitioning method, reference selected from GPP versions using model efficiency (MEF). The MEF analysis is repeated for each time aggregation
GPP_NT_VUT_USTAR50	Gross Primary Production, from Nighttime partitioning method, based on NEE_VUT_USTAR50

(Source: https://FLUXNET.org/data/FLUXNET2015-dataset/fullset-data-product/)

Extraction and transformation workflow

R scripts have been developed implementing the defined workflow steps and a R-Shiny Interface has been built to allow the user interaction. Users can select (a) the data source, (b) variables to be extracted, and (c) select the eLTER site of interest (based on a spatial mapping of the FLUXNET/ICOS sites to the respective eLTER site). Figure X shows a screenshot of the data extraction and transformation interface. The application can be accessed through the DataLabs at https://elter-cookieproto.datalabs.ceh.ac.uk/. A user account is needed to access the application at the moment.

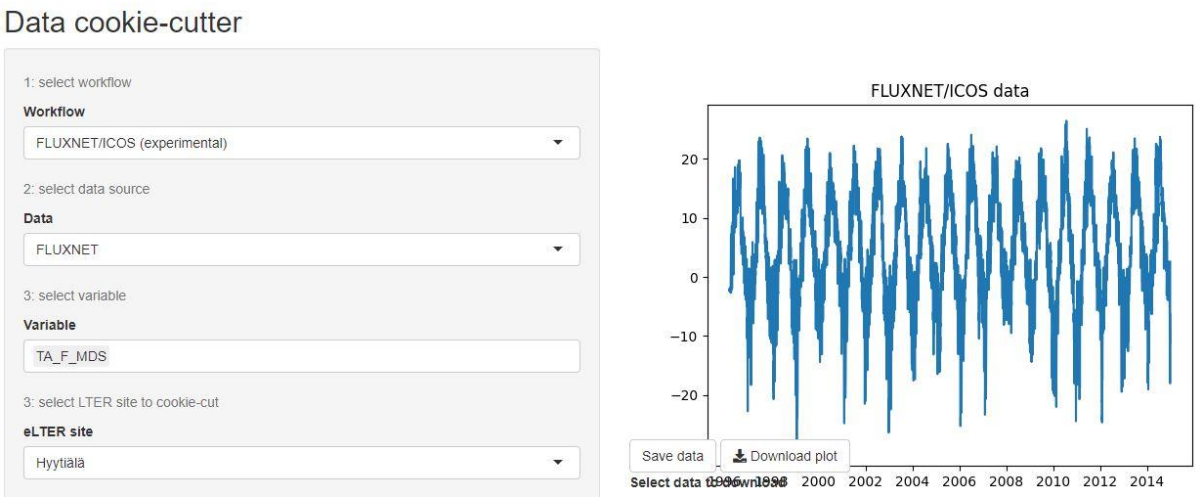

Figure 11: Screenshot of DataLabs for data retrieval and harmonisation

The workflow results a zip-archive containing

(a) the transformed data based on the specifications of the eLTER Data Reporting Format as csv file (<data>.csv)
(b) the reference for the variables provided based on the specification of FLUXNET (reference.csv)
(c) the methods and units for the variables based on the specification of FLUXNET (method.csv)

(d) a basic provenance documentation containing the version and link of the extraction workflow, the user inputs selected in the extraction workflow, the reference to the input dataset, and the link to the data licence of the input dataset (metadata.csv)

Figure 12 shows an example of the transformed input dataset using the eLTER Data Reporting Format (version 1.2). Figure 13 shows an example of the metadata file containing the provenance information.

	A	B	C	D	E	F	G	H	I
1	SITE_CODE	STATION_CODE	VARIABLE	LEVEL	TIME	VALUE	UNIT	FLAGQUA	FLAGSTA
2	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	TA_F_MDS	NA	2002-01-01	-8.85	deg C	1	NA
3	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	SW_IN_F_MDS	NA	2002-01-01	22.505	W m-2	1	NA
4	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	VPD_F_MDS	NA	2002-01-01	0.63	hPa	1	NA
5	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	P_F	NA	2002-01-01	0	mm	1	NA
6	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	WS_F	NA	2002-01-01	0.609	m s-1	0.9375	NA
7	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	SWC_F_MDS_1	NA	2002-01-01	NA	%	NA	NA
8	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	LE_F_MDS	NA	2002-01-01	0.408684	W m-2	0.9375	NA
9	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	NEE_VUT_REF	NA	2002-01-01	-1.38934	µmolCO2 m-2 s-1	0.9375	NA
10	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	NEE_VUT_USTAR50	NA	2002-01-01	-1.38934	µmolCO2 m-2 s-1	0.9375	NA
11	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	RECO_NT_VUT_REF	NA	2002-01-01	0	µmolCO2 m-2 s-1	TRUE	NA
12	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	RECO_NT_VUT_USTAR50	NA	2002-01-01	0	µmolCO2 m-2 s-1	TRUE	NA
13	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	GPP_NT_VUT_REF	NA	2002-01-01	1.38934	µmolCO2 m-2 s-1	TRUE	NA
14	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	GPP_NT_VUT_USTAR50	NA	2002-01-01	1.38934	µmolCO2 m-2 s-1	TRUE	NA
15	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	TA_F_MDS	NA	2002-01-02	-8.699	deg C	1	NA
16	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	SW_IN_F_MDS	NA	2002-01-02	20.913	W m-2	1	NA
17	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	VPD_F_MDS	NA	2002-01-02	0.662	hPa	1	NA
18	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	P_F	NA	2002-01-02	0	mm	1	NA
19	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	WS_F	NA	2002-01-02	0.457	m s-1	0.916667	NA
20	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	SWC_F_MDS_1	NA	2002-01-02	NA	%	NA	NA
21	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	LE_F_MDS	NA	2002-01-02	0.764895	W m-2	0.9375	NA
22	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	NEE_VUT_REF	NA	2002-01-02	-1.35749	µmolCO2 m-2 s-1	0.9375	NA
23	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	NEE_VUT_USTAR50	NA	2002-01-02	-1.35749	µmolCO2 m-2 s-1	0.9375	NA
24	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	RECO_NT_VUT_REF	NA	2002-01-02	0	µmolCO2 m-2 s-1	TRUE	NA
25	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	RECO_NT_VUT_USTAR50	NA	2002-01-02	0	µmolCO2 m-2 s-1	TRUE	NA
26	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	GPP_NT_VUT_REF	NA	2002-01-02	1.35749	µmolCO2 m-2 s-1	TRUE	NA
27	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	GPP_NT_VUT_USTAR50	NA	2002-01-02	1.35749	µmolCO2 m-2 s-1	TRUE	NA
28	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	TA_F_MDS	NA	2002-01-03	-9.459	deg C	1	NA
29	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	SW_IN_F_MDS	NA	2002-01-03	18.539	W m-2	1	NA
30	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	VPD_F_MDS	NA	2002-01-03	0.512	hPa	1	NA
31	https://deims.org/324f92a3-5940-4790-9738-5aa21992511c	AT-Neu	P_F	NA	2002-01-03	0	mm	1	NA

Figure 12: Harmonised FLUXNET dataset following the eLTER Reporting Format specification.

```
METADATA
Workflow version: 0.1
Workflow source: https://github.com/eLTER-RI/spatial-data-processor
Source dataset: FLUXNET2015
Selected site: AT-Neu, https://deims.org/324f92a3-5940-4790-9738-5aa21992511c
Selected variables: ['TA_F_MDS', 'P_F', 'WS_F']
Output filename: AT-Neu-fluxnet-2022-03-28-16-16-37.csv

ACKNOWLEDGEMENTS
Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina
Poindexter, et al. 2020. “The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data.”
Scientific Data 7 (1): 225. https://doi.org/10.1038/s41597-020-0534-3.

Link to data policy: https://fluxnet.org/data/data-policy/
```

Figure 13. Example for provenance metadata (metadata.txt)

In addition, basic provenance metadata are provided including the reference to the original data and data licence. The file metata.txt contains the following information:

- version of the cookie-cutter version
- link to the source code of the workflow on github
- source dataset selected
- selected stations/sites and mapping to the respective eLTER site (DEIMS.ID)
- selected variables

- output filename
- acknowledgement and citation of the input data source (including DOI if available)
- link to the data policy and data licence (if available)

This information is currently stored for the prototype workflow but should in future be provided through the sign-post records for the data source from the Digital Asset Registry (DAR).

Furthermore, according to the specification of the eLTER Data Reporting Format the descriptions of the variables (variables.csv) and methods (method.csv) provided by FLUXNET are attached to the zipped output. Currently the export of the station listing (station.csv) is not supported but site locations can be accessed through the DEIMS-SDR REST-API for the respective eLTER site.

References to the source code:

- https://elter-cookieproto.datalabs.ceh.ac.uk/
- https://github.com/eLTER-RI/spatial-data-processor/tree/41-workflow

Outlook:

The current version of the gas-flux demonstrator focuses on a limited list of sites and variables and further evaluation of end-users is needed. This is done in consultation with SC contributors further improving the functionality. Furthermore, recent developments within ICOS provided a python library[21] (ICOS, 2020) still needs to be assessed and tested.

5.3 Demonstrator Climate Diagram

Christoph Wohner

Providing information on eLTER sites including ecological characteristic is one of the main functions of DEIMS-SDR (Wohner et al., 2019). This includes the climatic characteristics, which currently are addressed by annual mean air temperature and annual sum of precipitation. Currently no information is provided on the length of the meteorological measurements, which were used to calculate these climatic characteristics for the site. To enhance these capabilities, existing pan-European climate data, such as the E-OBS daily gridded meteorological dataset (Copernicus Climate Change Service, 2020), can be used to derive average values for air temperature, precipitation and other meteorological parameters for defined reference periods. The derived values can be compared with legacy data, to assess the accuracy of the latter. Information could be used to complement information provided by DEIMS-SDR in a site selection or analysis processes or used to update information on the eLTER sites in DEIMS-SDR if long-term meteorological measurements are missing. By this, the function provides input to the Information Clusters enhancing the description of the eLTER sites.

A basic use case was defined as the basis for the prototype. Here a series of steps can be defined from a user point of view. The main success story defines the sequence of steps which allow the user to achieve the goal defined by the use case.

[21] See https://github.com/ICOS-Carbon-Portal/pylib

Goal of the use case: Retrieve long-term monthly average climate values as service for a given eLTER site for a given reference period (e.g. climate normal reference period).

Main Success Scenario

1. Central user logs in to DataLabs and executes the workflow
2. Define the list of sites relevant
3. Calculate the long-term monthly average for air temperature and precipitation for the normal reference period 1991-2020
4. Download information from the workflow for local use

A python script was written that cuts out the relevant thematic and temporal layer from the netCDF file and intersects it with the point coordinates of a given site. The resulting time series is then used for calculating monthly average values for the normal reference period 1991 - 2020. A visualisation for this data was built in the context of the H2020 e-shape project. An example is given in Figure 14.

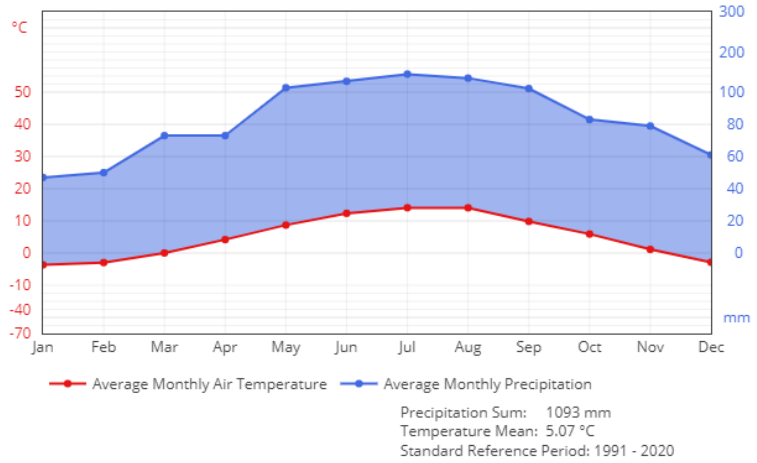

Figure 14: Climate chart of Gesäuse National Park (Austria) based on E-OBS (normal reference period 1991-2020)

The script can be accessed on the DataLabs. A user account is needed.

References to the source code:

- https://datalab.datalabs.ceh.ac.uk/resource/elter/tempprecvalues/tree/

Outlook:

This information could be used to enhance information provided through the DEIMS-SDR REST-API e.g. in an extended site selection process. Nevertheless, this approach can be applied to extract summarised characteristics for other variables like elevation characteristics, snow cover or phenology. Future extension are planned and are discussed in the current report. In principle, such a workflow can also be applied to other data such as elevation models, land cover classifications or, as already carried in earlier projects, for climate projection data (Rennie et al., 2021). For this, desired data products need to be defined and suitable input data provided that covers all sites in the desired geographical area of interest. Generated data products could then be deposited in persistent data repositories and linked to site records using the eLTER Information System.

5.4 Demonstrator Gridded Observations and model data for specific sites

Antonello Provenzale, Alice Baronetti & Angelica Parisi

Gridded observation and model data can be adapted to the needs of specific sites, using EURO-CORDEX simulations with bias correction and temperature downscaling. During the Ecopotential project[22] this procedure was applied to the ensemble of Protected Areas included in the project, many of which are eLTER sites. Here we report the main steps of the procedure, extracted from the Deliverable 8.1 of the ECOPOTENTIAL project (Palazzi & Terzago, 2017) which are applied in the eLTER PLUS context. The work builds on existing experiences and workflows and evaluates procedures developed by Ecopotential and analysis to which extent the same procedure can be applied to other target areas on demand.

E-OBS gridded data

E-OBS is a standard European land-only daily high-resolution gridded dataset for precipitation and minimum, maximum, and mean surface air temperature, starting in 1950 and regularly updated. The dataset builds on daily observations at more than 2,000 single-site locations available through the European Climate Assessment and Data set portal (ECA&D; http://eca.knmi.nl/; Haylock et al., 2008). E-OBS was produced by interpolating the monthly precipitation totals and monthly mean temperature using three-dimensional thin-plate splines, then interpolating the daily anomalies using indicator and universal kriging for precipitation and kriging with an external drift for temperature and, third, combining the monthly and daily estimates (Haylock et al., 2008). E-OBS is provided for an area extending from 25N-75N and from 40W-75E. The data files are in the compressed NetCDF format, thus fully compatible with the data format of climate model outputs, and at various spatial resolutions. The finest resolution is 0.25° on a lat-lon regular grid, and it is the one that is used here. The information on precipitation and temperature is then extracted from the original gridded data by using the shapefiles of the areas of interest.

Regional Climate Projections – bias correction procedure

Most climate model outputs, regional or global, are affected by temperature and precipitation biases. This means that the average climatological temperature and precipitation fields provided by the models in a given period have different means or intensity scales with respect to the observed data. To compare future climatic states to current ones, then, a bias correction procedure is often applied. This procedure requires to use observed data to correct offsets or scales of model outputs on a period during which both model results and observed data are available.

In the ECOPOTENTIAL procedure, a simple bias correction scheme based on the adjustment of the mean value was adopted, either by adding a temporally constant offset (for the temperature correction) or by using a constant multiplicative correction factor (for precipitation) for the simulated data. This method was chosen for its easy applicability and because it has the property of preserving the annual cycle and the linear trend in time of the original model variable, which should not be destroyed.

To bias-correct regional climate model projections, the EOBS data were used (of course, if more recent European-scale datasets become available, the procedure can be repeated with the new datasets). For the climate models, in ECOPOTENTIAL it was opted to use the family of simulations produced by the EURO-CORDEX initiative and we followed the same procedure here.

Bias-corrected RCM outputs

In such procedure, the model values are corrected using an offset or multiplicative correction factor pixel by pixel and constant in time that corrects for differences in the long-term climatology between

[22] See www.ecopotential-project.eu

the simulated and observed average in the common period chosen to compare model outputs and data. The bias correction fields are calculated at the spatial grid of the reference dataset (E-OBS, 0.25° lon-lat resolution), which does not have the same spatial resolution as the model grid (0.11° lon-lat). In order to overcome this scale mismatch the following procedure was adopted:

- the RCA4 model temperature and precipitation fields were conservatively remapped (Jones, 1999) at the same grid as E-OBS (0.25°);
- the correction factor was calculated at the E-OBS grid and then remapped at the original RCA4 grid using the nearest neighbour interpolation;
- the correction factor field, now defined at the same spatial grid as the model, was applied to the model fields to get the bias corrected field at the original model resolution.

This procedure guarantees that the bias corrected model outputs have the same climatology of E-OBS at the spatial resolution of the gridded data.

A subsequent procedure of temperature downscaling can then be applied, using the adiabatic lapse rate. Detailed information on these procedures can be found at Palazzi & Terzego (2017).

Data storage and archive properties

Currently, datasets are available through the public repository http://data.dta.cnr.it/ecopotential/, organising the data in folders named according to the areas to which they refer.

For each data request, metadata are provided through a "README" file located within each area folder. By clicking on the README.html file the user can access information on:

- the original CORDEX-RCM datasets, including the references to the CORDEX model run, the data archive, the RCA4 model and its outputs;
- the RCA4 simulations used for the RCM simulations
- the post-processing that was adopted, namely the cropping of the original CORDEX data over the domain of interest, the possible re-projections onto a different geographical reference system, the temporal aggregations, the bias correction, and the spatial downscaling when applied.
- the characteristics of the outputs, eventually re-projected/bias corrected/downscaled, including the list of variables provided, units, geographical reference system, spatial and temporal resolution.

In addition to the work done in Ecopotential, in eLTER PLUS the whole procedure was cleaned and made usable for areas of interest extracted with boundaries provided via the DEIMS REST-API. It can now be applied to any other sites for which model runs and E-OBS gridded data are available.

Reference to the source code:

- The procedure currently was enhanced and run locally. An evaluation of the implementation at the DataLabs have to be done as next step.

5.5 Demonstrator MODIS Land Surface temperature (Gran Paradiso National Park, Italy)

Antonello Provenzale, Alice Baronetti & Angelica Parisi

In order to provide temperature data enabling linking to *in-situ* measurements at the sites a demonstrator was set-up using the Collection-6 (C6) MODIS Terra Land Surface Temperature (LST) products. This is linked to activities carried out by task 4.3 in the eLTER PLUS project but tried to

provide added value to the retrieval of *in-situ* data by adding additional layers. The procedure was prototyped for the eLTER site Gran Paradiso National Park (Italy) but can be applied to other selected sites, once successfully evaluated in the eLTER PLUS project context.

The eight-day MODIS Terra LST product MOD1A2 at 1 km of spatial resolution were downloaded from the NASA/USGS Land Resources Distributed Active Archive Centre (LP DAAC) (https://lpdaac.usgs.gov/) for the 2001-2020 period. An eight-day compositing period was chosen because twice of such period is the exact ground track repeat period of the Terra platform. Figure 15 shows an example of the downloaded MODIS Terra daily footprint that includes the Gran Paradiso National Park site. The MODIS Terra MOD1A2 is archived in Hierarchical Data Format - Earth Observing System (HDF-EOS) format files.

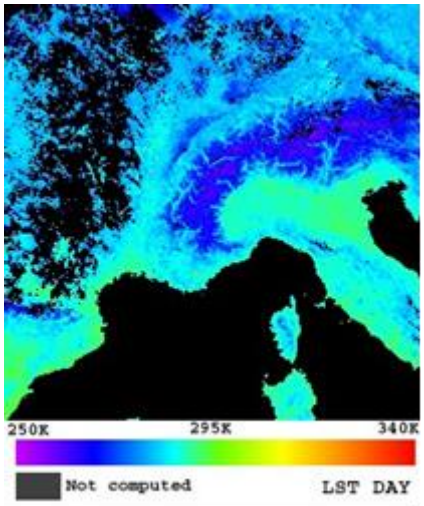

Figure 15: MODIS Terra footprint, an example of the LST MODIS1A2 product [degree Kelvin]

The quality indicators (QC_Day) and (QC_Night) layers are stored in a bit-encoded manner and in table 1 the bit flags for the QC layers are shown (Table 17). The purpose of these indicators is to give, for each pixel, information on LST_Day_1km and LST_Night_1km quality. In fact, they tell if LST results were nominal, abnormal, or if other defined conditions were encountered for a pixel

Table 17: Bit flags defined by QC Day and Night indicators for the MOD1A2 product (Source: User manual)

bits	Long Name	Key
1 & 0	Mandatory QA flags	00=LST produced, good quality, not necessary to examine more detailed QA 01=LST produced, other quality, recommend examination of more detailed QA 10=LST not produced due to cloud effects 11=LST not produced primarily due to reasons other than cloud
3 & 2	Data quality flag	00=good data quality 01=other quality data 10=TBD 11=TBD
5 & 4	Emis Error flag	00=average emissivity error <= 0.01 01=average emissivity error <= 0.02 10=average emissivity error <= 0.04 11=average emissivity error > 0.04
7 & 6	LST LST Error flag	00=average LST error <= 1K 01=average LST error <= 2K 10=average LST error <= 3K 11=average LST error > 3K

In these regards, all the MOD1A2 layers available for the 2001-2020 period were downloaded from the NASA Earth Data Search tool (https://lpdaac.usgs.gov/tools/earthdata-search/). The LST layers (Day and Night) with the respective QC indicators were extracted from the HDF files and were saved as Tagged Image File Format (TIFF). The QC information was then applied to determine the usefulness of the LST data. All the low-quality pixels, or that include no data, or that were covered by clouds, were removed from all the LST layers. Then the Daytime and Night-time LST values were converted from Kelvin to degrees Celsius. Finally, the Day and Night LST datasets were cropped to fit the Gran Paradiso National Park boundaries and the WGS84 Coordinate Reference System was set. Boundaries were downloaded from the DEIMS-SDR (Dynamic Ecological Information Management System - Site and dataset registry) database, where all the eLTER PLUS sites are stored. In Figure 16 an example of the raw and the processed Day LST for January 1st, 2001 is shown. The raw Day LST footprint extracted from the MODIS1A2 product is shown in the left panel. On the right panel the Day LST quality controlled and cropped on the Gran Paradiso National Park boundaries. The provided methodology can easily extended to other areas.

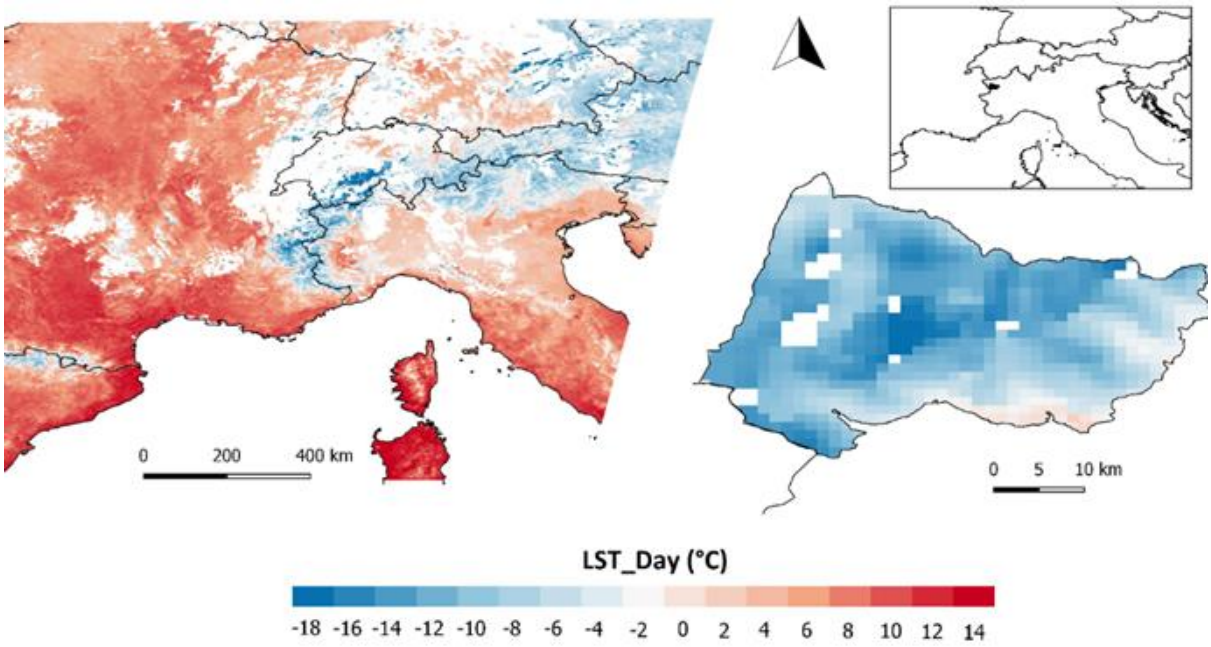

Figure 16: Day LST MOD1A2 product for January 1st, 2001. The generated eight-day Daily and Night time Land Surface Temperature datasets for the 2001-2020 period are made up of 920 layers each, with a spatial resolution of 1 km.

To provide user-friendly dataset of Day and Night LST, the LST layers were stored into NetCDF4 files, called "LSTDay_GranParadiso_2001_2020.nc" and "LSTNight_GranParadiso_2001_2020.nc". NetCDF4 is a binary and header-based format and data is stored in multidimensional arrays. The LST dataset has three dimensions: latitude, longitude and time and each file contains only one data variable (Day or Night LST, Figure 17). The dataset is in the NetCDF4 format, where the variables latitude, longitude, time and Day LST (LST_MAX) are stored. The provided methodology is easily replicable for all the other eLTER PLUS sites

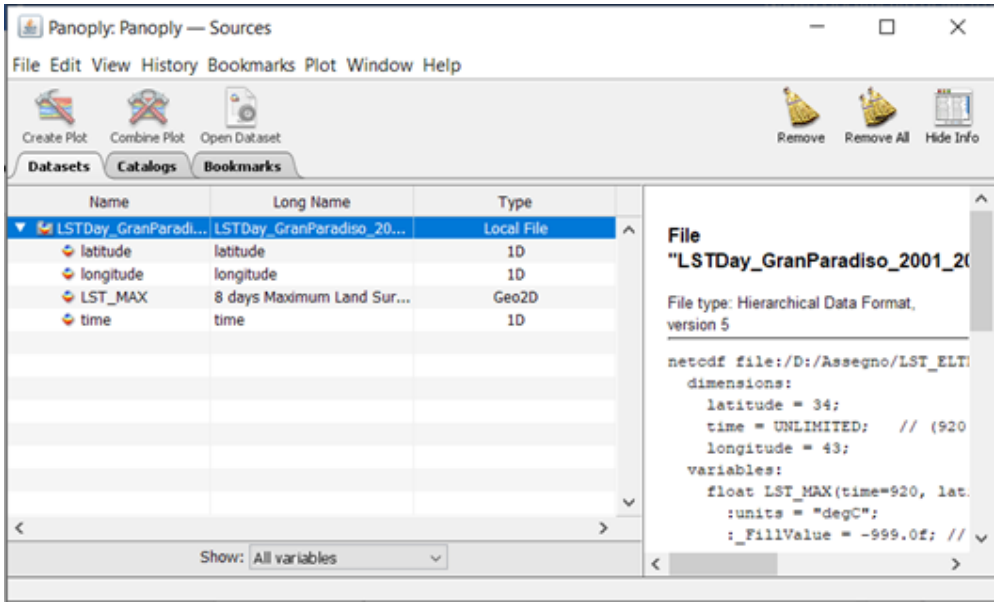

Figure 17: Structure of the LSTDay_GranParadiso_2001_2020.nc file.

The Day and Night datasets are easily readable offline with the Panoply Data Viewer tool (https://www.giss.nasa.gov/tools/panoply/). Panoply is a tool that allows to plot geo-gridded arrays stored in NetCDF4 format. In the figure below, Day LST for January 17, 2001 is shown (Figure 18).

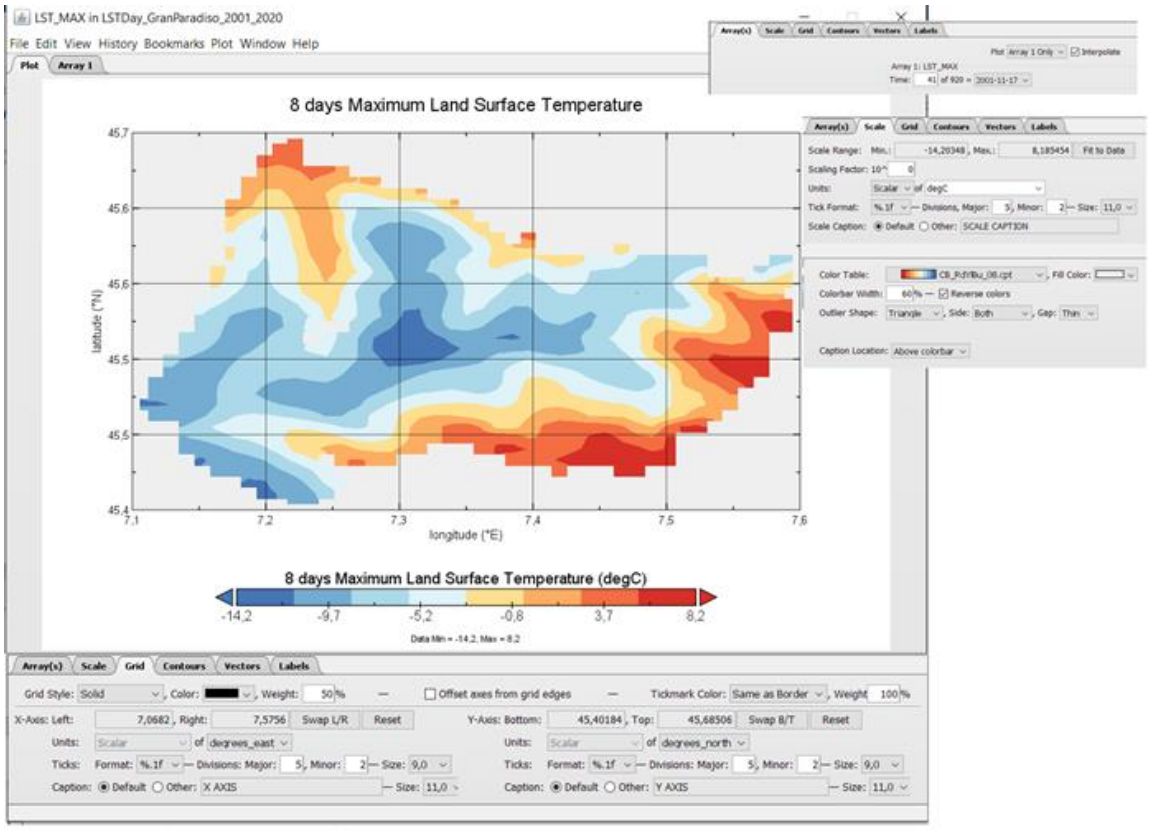

Figure 18: Example for Day LST (MODIS) spatial distribution for January 17, 2001.

The two datasets will be available also at the THREDDS Data Server (TDS) for the online visualisation. This is a web server that provides data access for scientific datasets, using OPeNDAP, OGC WMS and WCS and HTTP. The server is available at the link: https://www.unidata.ucar.edu/software/tds/current/. Integration with the eLTER DIP still needs to be evaluated and is a task in the design of the eLTER Information System.

6 Discussion and outlook

Johannes Peterseil

The work done in the context of task 4.1 aimed to provide (a) an overview on relevant *in-situ* legacy and third-party data sources as well as (b) exemplarily demonstrate access and harmonisation of *in-situ* data to test and support the concept of the eLTER Information Clusters as defined in the introduction. We mainly focused on the eLTER RI as well as on the research community as main stakeholders. This is in line with the project scope but could be extended to other stakeholders in future. Based on our findings Information Clusters in the context of eLTER provide information on different level of scale, which are described in the following.

Catalogue or relevant data sources – The first layer of the Information Clusters is the description of the relevant *in-situ* data sources which should be based on information provided in Annex 3. This could be either on the level of data portals providing a range of relevant datasets or on the level of single datasets of interest. Metadata needs either be collected via harvesting mechanisms linking to open metadata endpoints following specific metadata standards or by manual entry as sign-post records. Within the eLTER Information System the DAR will be the catalogue to collect and share this information. The work done is a first evaluation with regard to the current requirements defined by the eLTER SO as well as the project data needs. Further work is needed to extend this catalogue as the eLTER RI evolves.

Tools and services – The Information Clusters not only consists of metadata and datasets, but moreover on a set of functions and tools to retrieve and harmonise information for further processes. This needs to be collected in a service catalogue shared for the eLTER RI. The current work done within this task but also as outlined by Halada et al. (2022) contributes to this. Tools like ReLTER, deimsPy or the FLUXNET/ICOS cookie-cutter provide functions and interfaces for end-users to interact with a range of data sources relevant for the project.

Extracted Time series datasets – An important layer is access to *in-situ* data, which can be linked to single eLTER sites to run data analysis pipelines. Still, a subject of discussion is whether data should be calculated on-demand or pre-processed and stored for each eLTER Site. This mainly needs to be discussed in the frame of the governance for the eLTER CDN. In this respect, work done by Rennie et al. (2021) who extracted EURO-CORDEX data for eLTER Sites within the eLTER project can be seen as a first contribution to the idea of the Information Clusters using DEIMS-SDR to share the information. Similar work done by Palazzi & Terzago (2017) can also be seen in this respect. The ReLTER or FLUXNET/ICOS demonstrator can be seen as an example for on-demand retrieval.

Extracted Aggregated datasets – In order to extend the capabilities for site documentation and discovery existing datasets on European scale can be used. Information like climate characteristics (see chapter 5.3) could be used in this respect. Aggregated information is used and linked to the area covered by the eLTER Site and can be used for site selection or analysis. This might be extended to land-cover, elevation or similar characteristics.

In this respect, the generic workflow for building Information Clusters reflects these levels. It is an input for designing the eLTER Information System in general and the capabilities of the eLTER CDN specifically (Koivula et al., 2022).

Nevertheless, a number of issues needs to be taken into account when building the Information Clusters for eLTER. This ranges from site linking, thematic scoping and extent to the question of governance and provenance.

One ring to find them …[23]

When focusing on data from third-party sources (e.g. monitoring networks or RIs) an issue, which needs to be tackled in the eLTER context, is to assign place-based *in-situ* observations to a respective eLTER Site of relevance. Different naming conventions or identifiers are used, thus making an unambiguous mapping difficult. To overcome that we issue the deims.id, a unique persistent and resolvable identifier (PID) for all eLTER Sites documented in DEIMS-SDR. This allows cross-RI identification of research sites (Wohner et al., 2019). The deims.id is used in the mapping tables and is also shared via the REST-API. Currently, evaluation of DataCite DOIs is ongoing whether it could be used for referencing the basic site records in future and enable cross citations. Within DEIMS-SDR this '*co-location aspect*' is managed by the affiliation of an eLTER Site to monitoring networks or research infrastructure including the respective local identifier in this catalogue. Unfortunately, this information sometimes is not complete or up-to-date and needs continuous updates by the Site and Platform Coordinators. We tried to showcase a solution when mapping the FLUXNET stations to the eLTER Sites in the surrounding landscape. Depending on the spatial accuracy of the input data on the station location (e.g. rounded geographic coordinates), a mapping might be difficult or leads to ambiguous results. Therefore, a manual interaction is needed to assess the quality of the mapping and select the appropriate matching site. Further information, e.g. on the ecosystems and habitats where the observation took place, would improve the mapping and reducing the ambiguity. This information could be either taken from the input station catalogue or from European datasets, like CORINE Land Cover. Functions, like the *getSitesWithinRadius()* of deimsPy can support this workflow and might be used also for quality control processes within DEIMS-SDR.

Especially when dealing with legacy data stored in generic repositories like Dryad or Zenodo, often a proper reference to the eLTER Site is missing or just done by the site name. In addition, datasets exists which have been documented and published prior to the implementation of the deims.id or in catalogues not supporting the deims.id, so a reference to data and publications resulting from eLTER activities sometimes is difficult. In this respect, the exchange with e.g. Pangaea as an important repository in the Earth Science domain on this issue should be initiated as part of the data governance of the eLTER network and RI.

Recommendation 1: Include eLTER Site reference by the deims.id in the metadata and signpost metadata records on the eLTER DAR to enable an unambiguous identification of the eLTER Site.

Need Focus …[24]

The work showed that a broad variety of information relevant for the eLTER context exist. We provided a first screening of data sources on different scales, which is far from complete. Especially European and national scale datasets and data sources linked to eLTER priorities might provide valuable input for workflows and enhance site information. We used the eLTER SOs to assess the thematic scope and relevance of the identified data sources but we also identified a number of data sources contributing to additional variables not listed in the SOs, which could be of interest in future and should therefore be included in the Information Clusters.

In addition, the question on how to deal with potential overlaps and conflicts in data provided by similar data sources needs to be tackled. In the current work, we were not able to address this issue thoroughly as the focus was on the identification of relevant data sources and showcasing the retrieval. A single station could e.g. provide data for the same period through FLUXNET as well as ICOS.

[23] Tolkien, Lord of the Rings

[24] 2019, song by TeeZandos

Both apply their own data quality routines and data flagging which potentially leads to slightly varying datasets. Therefore, a prioritisation to select the most relevant data source is needed. In the current project, we solved this via consultations with the SC contributors to select a showcase of interest and relevance. However, this needs to be done on a broader basis which could not be supported by the current task.

Recommendation 2: Implement a governance process for eLTER to identify the focus of data sources to be included in the Information Clusters.

My priorities …[25]

An aspect that needs to be discussed in the frame of the eLTER Information Clusters is whether data from existing sources should be pre-calculated for all or relevant sites or calculated on demand for direct use. This addresses aspects relevant for the design of the eLTER Information System architecture. One is the calculation time and processing requirements, second the storage requirements, and third the accessibility of the data source. Whereas simple clipping mechanisms, e.g. for CORINE Land Cover, do not require massive computing resources and could be done on-demand (see e.g. Halada et al., 2022) others might be more demanding in terms of calculation time and computing power (see e.g. Rennie et al., 2021, Palazzi & Terzago, 2017).

On the other side data, which are required for central purposes (e.g. generation of eLTER data products) and are not easily accessible, could be pre-calculated and stored within eLTER. This requires negotiated data use and exchange with the data provider. This would also be true, e.g. if data from an eLTER Site are reported to one monitoring network or research infrastructure (e.g. ICP Forest) which would be used as SO data pipeline. For these pre-calculated workflows, if used to provide eLTER Standard Data Products, would be the option of choice. So for the time being no definitive answer can be given, as the definition process of the eLTER Standard Data Products is only as its beginning in eLTER PLUS. These activities need to be aligned with the further work on the eLTER SOs in order to enable this focus related to on-demand versus pre-calculated. Nevertheless, the work done provided first insights elemental aspects and potential challenges. Some of them relate to data access and documentation.

Recommendation 3: Prioritisation of data sources and processing mechanisms (on-demand versus pre-calculated) needs to be defined in the eLTER governance process, also in relation to the eLTER Standard Data Products.

Recommendation 4: For each data source, a list of prioritised tools and services (e.g. ReLTER) should be provided as part of the Information Clusters.

Words don't come easy …[26]

The application of a common data model and format for the integration and retrieval of data via the eLTER Information Clusters is an important aspect. This applies to the structural and syntactic interoperability. The publication of data papers from the eLTER network (e.g. Zilioli et al., 2019, Rennie et al., 2020) is one of the options and shows the effort which needs to be taken to harmonise and integrate long-term data. In the current work, we used the eLTER Data Reporting Format that applies

[25] 2007, song by Prodigy

[26] 1982, song by F. R. David

to many observation data types collected in the frame of eLTER, but not for all. The eLTER Data Reporting Format (Peterseil & Geiger, 2020) was refined based on the initial eLTER proposal and is an extension of the specifications and recommendations of the ICP Integrated Monitoring. The aim is to provide an easy integration of observation data from different sources based on a common core data structure. Additional data specifications, e.g. Darwin Core, FLUXNET, ICOS are currently discussed to ensure interoperability across networks and RIs.

For the biodiversity and gas-flux (daily aggregations) demonstrator the use of the eLTER Data Reporting specification worked out well as we mainly dealt with observation data. For climate data model outputs the eLTER Data Reporting Format could not be applied, as it does not support continuous value surfaces as represented in netCDF format.

Recommendation 5: Ensure structural and syntactic interoperability by applying a common data format, e.g. provided by the eLTER Data Reporting specification.
Recommendation 6: Checking of the common data format (including existence of provenance) should be done when ingesting data to the eLTER CDN.

This work is also linked to the eLTER SO activity driven by WP3 and the definition of the eLTER Standard Data Products driven by WP10. The semantic harmonisation as well as quality control has not been addressed by the current task as it reflects the transition from level 0 datasets in the eLTER context to level 1 data.

Stairway to heaven … [27]

For the assessed data sources provided in Annex 3 basic metadata have been provided allowing a basic description for the assessment. As these data sources are potential input to the eLTER Information Clusters, a documentation within the Digital Asset Registry (DAR) should be provided. In the current report as well as in Halada et al. (2022) an overview on relevant metadata information is given, which needs to be mapped to the eLTER Metadata Community Profile (Oggioni & Kliment, 2011) and further developments done on the dataset metadata model in DEIMS-SDR. This work needs to be done to ensure the usability and sustainability of the implementation of the eLTER Information Clusters that focuses (a) on the level of the data sources (e.g. data portal) and (b) to specific datasets.

Recommendation 7: Provide core metadata elements to describe relevant legacy and third-party data sources as well as respective tools and services linked to them in the eLTER DAR.
Recommendation 8: Enable the discovery of metadata on relevant data sources, linked datasets and tools through the eLTER DIP by the means of metadata.

Hells Bells …[28]

Finally yet importantly, the issue of data licensing, legal interoperability and data re-sharing needs to be considered. Especially when focusing on third-party *in-situ* data specific data licences and data policies are or might be attached to the data, which hinder or limit data use and re-sharing in the context of the eLTER Information Clusters and data pipelines. This is also relevant, when specific data access procedures are implemented (e.g. ICP Forest) which limit the data use to the specific defined

[27] 1971, song by Led Zeppelin

[28] 1980, song by ACDC

context. We identified several cases, which needs to be addressed in future. This list needs to be seen as very preliminary and further research is needed also in relation to the eLTER data governance:

- data are open access (e.g. CC0, CC-BY) and DOI is provided - in this case the appropriate metadata and provenance needs to be provided to ensure credits for the data provider. Re-sharing should not be an issue if the provenance and credit is provided.
- data are open access (e.g. CC0, CC-BY) and no DOI is provided, but for the clipped product a DOI is issued by the eLTER CDN - in this case appropriate metadata and provenance needs to be provided, but based on a URL or literature reference if available. Re-sharing should not be an issue if the provenance and credit is provided.
- data have restricted access (e.g. CC-BY-NC) - derived data might only be used for specific purposes and only be used internally to ensure re-usability and reproducibility of the results (e.g. gap-filling, validation).

Recommendation 9: For core datasets with regard to the production of the eLTER Standard Data Products open data licenses (e.g. CC-BY 4.0 International) needs to be negotiated with the data providers.

Sweet dreams (Are made of this) …[29]

Interaction with the design and the development of the eLTER Information System and the eLTER CDN (Kuivola et al., 2022) is needed in this respect. The current catalogue and list of data sources needs to be seen as a living document which reflects the current status of the requirements and the eLTER SOs.

[29] 1983, song by Eurythmics

7 Conclusion

The aim of task 4.1 was to develop and test workflows for access and harmonisation of relevant *in-situ* data sources on global, continental and national scale. With respect to the implementation of the eLTER RI we focused on data requirements defined both by the Research Challenges addressed in the eLTER PLUS project as well as the needs for supporting the implementation of data flows defined by the eLTER SO. We identified 176 legacy and third party data sources, which could be assigned to the respective eLTER SOs providing a good coverage for the different variables of the Ecological Integrity components. The authors recognize that there is a broad range of data sources available on a European scale, which could not be assessed within the resources of this work package e.g. thematic data repositories like Pangaea and site legacy data. However, the approach tested in this task provides a clear strategy to enable further integration of other data sources in the future. In addition, mobilisation of eLTER legacy data is still an issue but improved significantly in the eLTER PLUS project.

Metadata collected on relevant data sources needs to be incorporated into the Digital Asset Registry (DAR). This is still an open task which needs to be conducted in the coming year. This information will form the base layer for the information clusters which is complemented by services (e.g. ReLTER) providing on-demand access and data retrieval for given eLTER sites and platforms.

A prioritisation of variables is needed for pre-calculated datasets being of common interest across all eLTER Sites and Platforms. This information could not be extracted from the eLTER SOs but needs additional discussion on the governance level of eLTER and the emerging eLTER RI. Furthermore, the eLTER SOs are currently in a refinement phase. So further inputs need to be taken into account when information is available.

We showcased examples for the retrieval and harmonisation of selected data sources of different technological readiness levels (TRL). Whereas ReLTER and DEIMSpy are on TRL 6-7, the extended cookie-cutter version for ICOS/FLUXNET would be TRL 5 with further extensions needed to automate and integrate the workflow components. It could be shown that the selected workflows are operational and are useful for the eLTER PLUS users. This was done in a co-design process including the respective Research Challenge leads and Science Case contributors in the design and implementation phase on a regular basis. However, the decision needs to be made if we focus on on-demand services or pre-calculated datasets in the context of the eLTER Information System and workflows.

The results of the work done in task 4.1 will be provided as input to the design and architecture of the extended eLTER Information System led by WP11 and the further definition of workflows towards the eLTER Standard Data Products led by WP10.

Acknowledgements and project reference

The work was performed within the project eLTER PLUS (Grant agreement No. 871128) funded under the Horizon 2020 programme. The research challenges and science cases contributed to the definition of the requirements for the data sources and WP11 supported the implementation of the demonstrators. The work done was aligned with the requirements defined by the eLTER PPP project.

8 References

Acri, F.; Bastianini, M.; Bernardi Aubry, F.; Camatti, E.; Boldrin, A.; Bergami, C. et al. (2020): A long-term (1965-2015) ecological marine database from the LTER-Italy Northern Adriatic Sea site: plankton and oceanographic observations. In: Earth Syst. Sci. Data 12 (1), S. 215–230. DOI: 10.5194/essd-12-215-2020.

Bauer, P., Dueben, P. D., Hoefler, T., Quintino, T., Schulthess, T. C., & Wedi, N. P. (2021a). The digital revolution of Earth-system science. Nature Computational Science, 1(2), 104-113.

Bauer, P., Stevens, B., & Hazeleger, W. (2021b). A digital twin of Earth for the green transition. Nature Climate Change, 11(2), 80-83.

Copernicus Climate Change Service (2020). E-OBS daily gridded meteorological data for Europe from 1950 to present derived from *in-situ* observations [Data set]. ECMWF. https://doi.org/10.24381/CDS.151D3EC6

Dirnböck, T., Peterseil, J., Forsius, M., Holmberg, M., Kühn, I., Futter, M., Vereecken, H., Méndez, P.F. (2020) Report about data requirements (CS1-4) to be used in VA and WP10. Deliverable D9.5, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 12pp.

Dorren, L.K.A., Berger, F., Imeson, A.C., Maier, B., Rey, F. (2004) Integrity, stability and management of protection forests in the European Alps. - Forest Ecology and Management 195: 165–176. doi:10.1016/j.foreco.2004.02.057

Drought 2018 Team, and ICOS Ecosystem Thematic Centre (2020). "Drought-2018 Ecosystem Eddy Covariance Flux Product for 52 Stations in FLUXNET-Archive Format." Collection of FLUXNET product ZIP archives, March. https://doi.org/10.18160/YVR0-4898.

eLTER PLUS (2019) European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS (eLTER PLUS), Part B, Description of Action. Internal document.

Fowler, Martin (2004) UML Distilled: A Brief Guide to the Standard Object Modeling Language. Boston: Addison-Wesley

Giorgi, F., Jones, C., & Asrar, G. R. (2009). Addressing climate information needs at the regional level: the CORDEX framework. World Meteorological Organization (WMO) Bulletin, 58(3), 175.

Haase, P., Tonkin, J.D., Stoll, S., Burkhard, B., Frenzel, M., Geijzendorffer, I.R., Häuser, Ch., 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 Total Environment 613-614: 1376-1384 https://doi.org/10.1016/j.scitotenv.2017.08.111

Halada, L., Dick, J., Bolton, W., Gašparovičová, P., Hilbert, H., Baránková, Z., Gemmelová, L., Kozelová, I., Kenderessy, P., Rusňák, T. (2022) Workflows for retrieval and harmonisation of data from official statistics. Deliverable D4.2, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 129pp.

Haylock, M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, J. Geophys. Res., 113, D20119, doi:10.1029/2008JD010201.

ICOS (2020) ICOS Carbon Portal Python Package. [https://github.com/ICOS-Carbon-Portal/pylib; last accessed 2022-04-11]

Koivula, H., Wohner, C., Rennie, S., Watkins, J., Brown, M., Bolton, W., Peterseil, J., Oggioni, A. (2022) Concept and extension of central data node capabilities. Deliverable D11.2, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128.

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, 626:1439-1462, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2017.12.001.

Mollenhauer, H., Kasner, M., Haase, P., Peterseil, J., Wohner, Ch., 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 Total Environ. 624: 968-978.

Nativi, S., Mazzetti, P., & Craglia, M. (2021). Digital ecosystems for developing digital twins of the earth: The destination earth case. Remote Sensing, 13(11), 2119.

Oggioni, A., Kliment, T. (2011). EnvEurope (LTER-Europe) Metadata Specification for Dataset Level. Deliverable PDA1.1.4, Life EnvEurope Project, Grant agreement No. Life Environment Projekt LIFE08/ENV/IT/000339. 88pp.

Oggioni, A., Silver, M., Ranghetti, L., Tagliolato, P. (2021). oggioniale/ReLTER: ReLTER v1.0.0 (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.5576813)

Palazzi, E., Terzago, S. (2017) High resolution (1-10km) climate, land use and ocean change scenarios. Deliverable 8.1, EU Horizon 2020 Project, Grant Agreement No. 641762. 68pp. [Link: https://www.ecopotential-project.eu/images/ecopotential/documents/D8.1.pdf, last accessed 2022-04-11]

Pastorello, Gilberto, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, et al. (2020). "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." Scientific Data 7 (1): 225. https://doi.org/10.1038/s41597-020-0534-3.

Peterseil, J., Geiger, S. (2020) Field specification for data reporting. Technical Document TechDoc.01. EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. [https://doi.org/10.5281/zenodo.6373410]

Peterseil, J., Parland-von Essen, J., Lukkarinen, A., Pursula, A., Oggioni, A., Zilioli, M., Watkins, J., Rennie, S., Brown, M., Magagna, B., Goldfarb, D., Dirnböck, T., Haase, P., Baatz, R., Dick, J., Orenstein, D. (2020) Report on Data and ICT needs from Research Challenges (RCs) to be used in Virtual Access (VA). Deliverable D10.2, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 99pp.

PyPI (2022) deimsPy. [https://pypi.org/project/deims/; last accessed: 2022-04-11]

Rennie, S., Andrews, Ch.; Atkinson, S.; Beaumont, D.; Benham, S.; Bowmaker, V. et al. (2020): The UK Environmental Change Network datasets – integrated and co-located data for long-term environmental research (1993–2015). In: Earth Syst. Sci. Data 12 (1), S. 87–107. DOI: 10.5194/essd-12-87-2020.

Rennie, S., Goergen, K.; Wohner, Ch.; Apweiler, S.; Peterseil, J.; Watkins, J. (2021): A climate service for ecologists: sharing pre-processed EURO-CORDEX regional climate scenario data using the eLTER Information System. In: Earth Syst. Sci. Data 13 (2), S. 631–644. DOI: 10.5194/essd-13-631-2021.

rOpenSci, Anderson, B., Chamberlain, S., DeCicco, L., Gustavsen, J., Krystalli, A., Lepore, M., Mullen, L., Ram, K., Ross, N., Salmon, M., Vidoni, M., Riederer, E., Sparks, A., Hollister, J. (2021). rOpenSci Packages: Development, Maintenance, and Peer Review (0.7.0). Zenodo. https://doi.org/10.5281/zenodo.5704478

Schentz, H., Peterseil, J. & Bertrand, N. (2013) EnvThes- interlinked thesaurus for long term ecological research, monitoring, and experiments. Proceedings EnviroInfo 2013: Environmental Informatics and Renewable Energies. Shaker Verlag, Aachen.

Schentz, H., Peterseil, J., Magagna, B. & Mirtl, M. (2011) Semantics in Ecosystem Research and Monitoring. Proceedings EnviroInfo 2011: Innovations in Sharing Environmental Observation and Information. Shaker Verlag, Aachen.

Skiba, U., Vicente, J., Lomba, A., Diaz-Pines, E., Nikolaidis, N., Gaube, V., Haase, P., Dirnböck, T., Peterseil, J. (2020) Report on data requirements (CS1-4) from VA and WP10. Deliverable D8.5, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 12pp.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M. , Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, Ph.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, Ch.T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J.G., Groth, P., Goble, K., Grethe, J. S., Heringa, J., 't Hoen, P. A.C, Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., ons, A., Packer, A. L., Persson, B., Rocca-Serra, Ph., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, Th., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016): The FAIR Guiding Principles for scientific data. management and stewardship. Scientific Data 3 [2016/03/15/online; http://dx.doi.org/10.1038/sdata.2016.18]

Wohner, C., Peterseil, J., Genazzio, M. A., Guru, S., Hugo, W., & Klug, H. (2020). Towards interoperable research site documentation – Recommendations for information models and data provision. In Ecological Informatics (Vol. 60, p. 101158). Elsevier BV. https://doi.org/10.1016/j.ecoinf.2020.101158

Wohner, C., Peterseil, J., Klug. H. (in prep) Designing and implementing a data model for describing environmental monitoring and research sites. Ecological Informatics (in review 03/2022)

Wohner, C., Peterseil, J., Poursanidis, D., Kliment, T., Wilson, M., Mirtl, M., & Chrysoulakis, N. (2019). DEIMS-SDR – A web portal to document research sites and their associated data. In Ecological Informatics (Vol. 51, pp. 15–24). Elsevier BV. https://doi.org/10.1016/j.ecoinf.2019.01.005

Wyborn, C., Louder, E., Harrison, J. Montambault, J. , Montana, J., Ryan, M., Bednarek, A., Nesshöver, C., Pullin, A., Reed, M., Dellecker, E., Kramer, J., Boyd, J., Dellecker, A., Hutton J. (2018). Understanding the Impacts of Research Synthesis, Environmental Science & Policy (Vol. 86, pp. 72-84), ISSN 1462-9011, https://doi.org/10.1016/j.envsci.2018.04.013 .

Zacharias, S., Anttila, S., Bäck, J., Böttcher, K., Mallast, U., Mirtl, M., Schaub, M., Trostiuk, V. (2021) Discussion Paper on eLTER Standard Observations (eLTER SOs). Deliverable D3.1, EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128. 55pp.

Zilioli M, Oggioni A, Tagliolato P, Pugnetti A, Carrara P. (2019) Feeding Essential Biodiversity Variables (EBVs): actual and potential contributions from LTER-Italy. Nature Conservation 34: 477–503. https://doi.org/10.3897/natureconservation.34.30735

Online resources

Destination Earth - https://digital-strategy.ec.europa.eu/en/policies/destination-earth

EUDAT B2SHARE is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store, publish and share research data in a FAIR manner. B2SHARE facilitates research data storage, guarantees long-term persistence of data and allows data, results or ideas to be shared worldwide. B2SHARE supports community domains with metadata extensions, access rules and publishing workflows. https://sp.eudat.eu/catalog/resources/709c053d-3b3f-4c54-8ef4-6efeea387816

9 Annexes

9.1 List of SOs addressed by data sources

See separate annex document – Annex 1

9.2 Thematic mapping of data sources

See separate annex document – Annex 2

9.3 List of data sources

See separate annex document – Annex 3

9.4 Thematic mapping of data sources

See separate annex document – Annex 4

9.5 Example of national data sources

See separate annex document – Annex 5

List of figures

List of tables

Annex 1 - List of SOs addressed by data sources

In total we could identify 96 of 173 eLTER SO in the selected data sources. The list of SOs addressed by selected data sources (based on Zacharias et al., 2021). In column P the associated priority is reported (A = very high B = further discussion required as defined in D3.1 SO. O=other relevant variables, not included in the SO list as explained in report text)

Ecosystem integrity component	Compartment	Parameter	P
Abiotic characteristic	Climate	Air temperature	A
Abiotic characteristic	Climate	Precipitation	A
Abiotic characteristic	Climate	Relative air humidity	A
Abiotic characteristic	Climate	Wind speed / Wind direction	A
Abiotic characteristic	Lake	Vertical profiles of water temperature, pH, EC, turbidity	A
Abiotic characteristic	Soil	Belowground biomass	O
Abiotic characteristic	Soil	Soil base saturation	O
Abiotic characteristic	Soil	Soil cation exchange capacity	O
Abiotic characteristic	Soil	Soil inventory	A
Abiotic characteristic	Soil	Soil organic C content (per horizon)	A
Abiotic characteristic	Soil	Soil pH (in H2O/KCl/CaCl2)	A
Abiotic characteristic	Soil	Soil total N content (per horizon)	A
Abiotic characteristic	Soil	Soil total P content (per horizon)	A
Abiotic characteristic	Soil	Various	A
Abiotic characteristic	Soil	Soil temperature	A
Abiotic characteristic	Streams/Rivers	pH, EC, water temperature	A
Abiotic characteristic	Streams/Rivers/Lake	Climate	O
Abiotic characteristic	Streams/Rivers/Lake	Various	O
Abiotic characteristic	Terrestrial/Aquatic	Climate	O
Biotic heterogeneity	Lake	Algal community (quantitative)	B
Biotic heterogeneity	Lake	Fish community (quantitative)	B
Biotic heterogeneity	Lake	Macrophyte community (quantitative)	B
Biotic heterogeneity	Lake	Zooplankton (quantitative)	B
Biotic heterogeneity	Other	Archaea, Bacteria, Fungi, Chromista	O
Biotic heterogeneity	Soil	Various	O
Biotic heterogeneity	Streams/River	Fish community (quantitative)	B
Biotic heterogeneity	Streams/River	Macroinvertebrate community (quantitative)	B
Biotic heterogeneity	Streams/Rivers/Lake	Algal community (quantitative)	B
Biotic heterogeneity	Streams/Rivers/Lake	Fish community (quantitative)	B
Biotic heterogeneity	Streams/Rivers/Lake	Macroinvertebrate community (quantitative	B
Biotic heterogeneity	Streams/Rivers/Lake	Macrophyte community (quantitative)	B
Biotic heterogeneity	Streams/Rivers/Lake	Various	O
Biotic heterogeneity	Streams/Rivers/Lake	Zooplankton (quantitative)	B
Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using acoustic recording	A
Biotic heterogeneity	Terrestrial	Flying insects	A

Ecosystem integrity component	Compartment	Parameter	P
Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	A
Biotic heterogeneity	Terrestrial	Mammals	B
Biotic heterogeneity	Terrestrial	Plant phenology	A
Biotic heterogeneity	Terrestrial	Various	O
Biotic heterogeneity	Terrestrial/Aquatic	Various	O
Energy budget	Biomass	Aboveground biomass	A
Energy budget	Biomass	Aboveground vegetation growth	O
Energy budget	Biomass	Gross primary production	B
Energy budget	Biomass	Leaf area Index (LAI)	A
Energy budget	Biomass	Net primary production (EC-Station)	A
Energy budget	Biomass	Net primary production (dendrometer)	B
Energy budget	Biomass	Transpiration	B
Energy budget	Net Ecosystem Exchange (NEE)	H2O concentration	B
Energy budget	Net Ecosystem Exchange (NEE)	H2O flux	B
Energy budget	Radiation Budget	Ground heat flux	B
Energy budget	Radiation Budget	Latent heat flux	B
Energy budget	Radiation Budget	Sensible heat flux	B
Matter budget	Atmospheric deposition	Bulk NH4-N, NO3-N, Ntot, P, K deposition in precipitation	B
Matter budget	Atmospheric deposition	Bulk NH4-N, NO3-N, Ntot, P, K deposition in canopy throughfall (forests)	B
Matter budget	Atmospheric deposition	Bulk pH, anion, cation deposition in canopy throughfall (forests)	B
Matter budget	Atmospheric deposition	Bulk pH, anion, cation deposition in precipitation	B
Matter budget	Atmospheric deposition	stemflow NH4-N, NO3-N, Ntot, P, K, pH, cation, anion deposition in stemflow (forests)	B
Matter budget	Atmospheric deposition	Bulk pH, anion, cation deposition in precipitation	B
Matter budget	Atmospheric deposition	Dry deposition of N-components	B
Matter budget	Atmospheric deposition	Various	O
Matter budget	Biomass	Aboveground litterfall (forests)	B
Matter budget	Biomass	Leaf C, N, K, P, Ca, Mg, Mn content	B
Matter budget	Groundwater	DOM composition	B
Matter budget	Groundwater	Major ion concentrations: Cl, SO4, Br, Na, K, Mg, Ca, B	B
Matter budget	Groundwater	Nutrient concentration: TP, SRP, TDN, NO3, NO2, NH4, DOC, DIC	B
Matter budget	Groundwater	Radioactive Isotopes (14C, T/3He, T)	B
Matter budget	Lake	In-situ vertical profiles and inflow concentrations of TP, SRP, NO3, DOC, SAC254	A
Matter budget	Lake	In-situ vertical profiles and inflow concentrations TOC, POC, TN, NO2, NH4, SRSi, DIC	B
Matter budget	Lake	Vertical profiles of chl a, pigments	A
Matter budget	Lake	Vertical profiles of dissolved oxygen	A

Ecosystem integrity component	Compartment	Parameter	P
Matter budget	Lake	Vertical profiles of major ion concentrations: Cl, SO4, Na, K, Mg, Ca	B
Matter budget	Net Ecosystem Exchange	CH4 concentration	B
Matter budget	Net Ecosystem Exchange	CH4 flux	B
Matter budget	Net Ecosystem Exchange	CO2 concentration	B
Matter budget	Net Ecosystem Exchange	CO2 flux	B
Matter budget	Net Ecosystem Exchange	CO2 flux (Gas flux chamber)	B
Matter budget	Net Ecosystem Exchange	N2O flux	B
Matter budget	Net Ecosystem Exchange	N2O, NO, NOx flux (Gas flux chamber)	B
Matter budget	Radiation Budget	PAR	A
Matter budget	Soil water	Conductivity	B
Matter budget	Soil water	DOC concentration	B
Matter budget	Soil water	NH4-N, NO3-N, DON concentration	B
Matter budget	Soil water	P concentration	B
Matter budget	Soil water	pH value	B
Matter budget	Soil water	Anion concentrations	B
Matter budget	Soil water	Cation concentrations	B
Matter budget	Soil water	DOC leaching	B
Matter budget	Soil water	NH4-N, NO3-N, DON leaching	B
Matter budget	Streams/River	TP, SRP, NO3, DOC, SAC254	A
Matter budget	Streams/Rivers	Turbidity	A
Matter budget	Streams/Rivers	Cl, SO4, Na, K, Mg, Ca	B
Matter budget	Streams/Rivers	In-situ vertical profiles and inflow concentrations TOC, POC, TN, NO2, NH4, SRSi, DIC	B
Matter budget	Streams/Rivers	TP, SRP, NO3, DOC, SAC254	A
Matter budget	Streams/Rivers/Lake	Various	O
Matter budget	Terrestrial	Cl, SO4, Na, K, Mg, Ca	O
Matter budget	Terrestrial	Various	O
Other	Marine	Various	O
Other	Soil	Microbial activity (e.g. ATP conc.)	O
Other	Streams/Rivers/Lake	Various	O
Other	Terrestrial/Aquatic	Various	O
Socio-Ecology	Agriculture and Forestry	Agricultural products	A
Socio-Ecology	Agriculture and Forestry	Fertiliser input (N, P, K fertilisation, liming, pesticides)	B
Socio-Ecology	Agriculture and Forestry	Grazing timing, intensity; livestock (livestock units)	B
Socio-Ecology	Agriculture and Forestry	Harvest (cropland, grassland, forest) (t/ha)	A
Socio-Ecology	Land use and land cover change	Land cover (CORINE)	A
Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)	A
Socio-Ecology	Land use and land cover change	Land use (historic)	A
Socio-Ecology	Land use and land cover change	Various	O
Water balance	Groundwater	Groundwater level	A

Ecosystem integrity component	Compartment	Parameter	P
Water balance	Lake	Inflow/outflow	A
Water balance	Lake	Ice cover	A
Water balance	Lake	Water level	A
Water balance	Soil water	Soil water content	A
Water balance	Streams/River	Discharge	A
Water balance	Streams/Rivers/Lake	Various	O
Water balance	Terrestrial	Ice cover	O
Water balance	Terrestrial	Snow cover	A
Water balance	Terrestrial	Snow density	B
Water balance	Terrestrial/Aquatic	Discharge	O
Water balance	Terrestrial/Aquatic	Various	O

Annex 2 – Thematic mapping of data sources to eLTER SOs

Each of the data sources has been assigned to one or many eLTER SOs (Zacharias et al., 2021). The table shows the information on the level of the Ecological Integrity Component as condensed overview. 'Other' is referring to other variables, which has been identified as important, but being not addressed by the eLTER SOs.

ID	Data source	Abiotic characteristic	Biotic heterogeneity	Energy budget	Matter budget	Water balance	Socio-Ecology	Other	Sum of EI components adr.
1	ICP forests	x	x	x	x				4
2	ICP vegetation				x				1
3	ICP Integrated Monitoring	x	x		x			x	4
4	ICP Modelling and Mapping				x				1
5	"INTERACT International Network for Terrestrial Research and Monitoring in the Arctic"	x	x		x	x			4
6	"GEM Greenland ecosystem monitoring"	x	x		x	x		x	5
7	AMAP-Arctic monitoring and Assessment Programme working Group				x				1
8	EMEP - European Monitoring and Evaluation Programme				x				1
9	ACTRIS				x				1
10	EBAS				x				1
11	ICOS - Integrated Carbon Observation System				x				1
12	GLORIA	x	x						2
13	forestREplot		x						1
14	Global soil partnership	x							1
15	SoilBON		x						1
16	GCP Global Carbon Project				x				1
17	TeaComposition							x	1
18	LUCAS 2009 TOPSOIL data (ESDAC)	x							1
19	LUCAS 2015 TOPSOIL data (ESDAC)	x							1
20	ESDAC European Soil Data Centre	x							1
21	ESDAC						x		1
22	Soil Geographical Database of Eurasia	x							1
23	PASERA						x		1
24	BDAT	x							1
25	BASOL	x							1
26	DONESOL	x							1
27	Art.11 of habitat directive (FFH)		x						1

ID	Data source	Abiotic characteristic	Biotic heterogeneity	Energy budget	Matter budget	Water balance	Socio-Ecology	Other	Sum of EI components adr.
28	EUNIS		x						1
29	GBIF		x						1
30	EGDI European Geological Data Infrastructure	x							1
31	OpenOBS		x						1
32	FungalRoot	x	x						2
33	Global Ants Database		x						1
34	RMQS	x	x						2
35	CRYOBS-CLIM					x			1
36	GLACIOCLIM	x				x			2
37	The Farmland Bird Index		x						1
38	EuroBirdPortal		x						1
39	EBBC Atlas		x						1
40	TOURBIERES	x							1
41	ALTI	x							1
42	BDD link to Global Mammal Assessment		x						1
43	BDD link to Global Amphibian Assessment		x						1
44	MoveBank		x						1
45	DRILOBase		x						1
46	SoilTemp	x							1
47	International Soil Moisture Network					x			1
48	ESA CCI BIOMASS v3.0			x					1
49	ESA CCI LAKES v1.1					x			1
50	ESA CCI LAND COVER v2.0.7						x		1
51	ESA CCI SNOW					x			1
52	ESA CCI SM v6.1					x			1
53	Copernicus Global Land Services SSM					x			1
54	Copernicus Global Land Services SWI					x			1
56	CES OSO						x		1
57	GLASS			x					1
58	"SMOS Soil Moisture and Ocean Salinity"					x			1
59	MODIS			x					1
60	Copernicus Global Land Services LAI300			x					1
61	Copernicus Global Land Services LAI 1km			x					1
62	GLIMS					x			1
63	European Fluxes Database Cluster				x				1
64	BDCHARM 50	x							1
65	BDIFF						x		1

ID	Data source	Abiotic characteristic	Biotic heterogeneity	Energy budget	Matter budget	Water balance	Socio-Ecology	Other	Sum of EI components adr.
66	Open Government Data							x	1
67	INSPIRE	x	x				x		3
68	THEIA/OZCAR							x	1
69	Pangea		x						1
70	Zeonodo		x					x	2
71	FigShare							x	1
72	VIGIE-NATURE		x						1
73	Knowledge Base "Status"		x						1
74	COPERNICUS HRL Layers						x		1
75	CLC+ work form MC						x		1
76	CHELSA v2.1	x							1
77	WorldClim v2.1	x							1
78	E-OBS	x							1
79	CHPclim	x							1
80	GPCC	x							1
81	TRMM	x							1
82	ERA-Interim	x							1
83	ERA5	x							1
84	DRIAS	x				x			2
85	AERIS	x							1
86	HyMeX	x							1
87	RADOME	x							1
91	NIVOSE	x							1
95	ATMO-AASQA				x				1
96	ICP Waters		x		x				2
97	Danubius-RI					x			1
98	GLEON -Global Lake Ecological Observatory Network							x	1
99	MarineStrategyFrameworkDirective				x				1
100	Water Framework Directive	x	x		x	x			4
101	TeaComposition H2O							x	1
102	SANDRE		x						1
103	Waste water Treatment Plant (Urban Wastewater)					x			1
104	BRGM	x				x			2
106	WATERBASE -Water quality ICM				x				1
107	WATERBASE -Biology		x						1
108	CCM (Catchment Characterisation and Modelling)					x			1
109	BANQUE HYDRO					x			1

ID	Data source	Abiotic characteristic	Biotic heterogeneity	Energy budget	Matter budget	Water balance	Socio-Ecology	Other	Sum of EI components adr.
110	ADES					x			1
111	DATA EAU FRANCE-Fish		x						1
112	DATA EAU FRANCE-Contaminants				x				1
113	NAIADES	x	x		x				3
114	CYBERCAROTHEQUE						x		1
115	ODATIS							x	1
116	Sextant		x						1
117	UWWTD						x		1
118	BioFresh		x		x	x			3
119	COPERNICUS	x				x	x		3
124	ESA-CCI+	x				x			2
125	AQUACOSM		x		x				2
126	TERENO	x				x			2
128	Droughtnet	x	x				x		3
129	TreeDivNet		x						1
130	NutNet	x	x	x			x		4
131	ANAEE-France							x	1
132	OLA	x				x			2
133	F-ORE-T				x				1
134	ACBB		x						1
135	PRO	x							1
136	EJP SOIL	x							1
137	H+					x			1
138	CensusHub							x	1
139	INSEE						x		1
140	SIDDT						x		1
141	AGRESTE						x		1
142	CORINE LAND COVER						x		1
143	RPG (Parcel Registry Graphic)						x		1
144	ZNIEFF		x						1
145	REMONTER LE TEMPS						x		1
146	CDDA		x						1
147	European Social Survey						x		1
148	DINAMIS						x		1
149	SRTM	x							1
150	LTER France Data Node							x	1
151	Google Earth Engine	x							1

ID	Data source	Abiotic characteristic	Biotic heterogeneity	Energy budget	Matter budget	Water balance	Socio-Ecology	Other	Sum of EI components adr.
152	PANGAEA	x	x			x		x	4
153	BD Forêt V2			x					1
154	SapFluxNet Databse			x		x			2
155	EFAS snow equivalent					x			1
156	NitroEurope IP				x				1
157	INI-International Nitrogen Initiative				x				1
158	Global Partnership on Nutrient Management							x	1
159	SITES		x					x	2
160	"SRDB v4.0 Global Database of Soil Respiration"				x				1
161	Critical Zone Observatories				x				1
162	ECN	x	x		x	x			4
163	FLUXNET				x				1
164	CORINE LANDCOVER (CLMS)						x		1
165	Ecosystem types of Europe		x						1
166	CDDA (National designated areas)		x						1
167	Land Use Harmonisation 2						x		1
168	EuroCordex	x							1
169	ESGF	x							1
170	PhenoCam		x						1
171	GBIF Api		x						1
172	BioTime		x						1
173	ECAD dataset (e-obs gridded dataset)	x							1
174	ECAD daily data	x							1
175	OBIS API		x						1
176	iNaturalist		x						1

Annex 3 List of data sources

The following data sources have been identified within task 4.1 of the eLTER PLUS project. Entrance 'NA' is indicating when information could not be retrieved in the runtime of the task implementation. This information will be updated in a later stage of the implementation of the Information Clusters. Contact information provided in this table is taken from the websites.

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
1	ICP forests	Terrestrial	http://icp-forests.net/	Forest monitoring network: Level I and Level II under UNECE LRTAP convention. Level1: annual assesment of crown condition (defoliation, discolouration, demages visible on the threes). Assessment of soil condition (soil chemistry), Foliar survey (foliar chemistry) Level II: investigate the causes/effects of different stress factors on forest ecosystems	european	-	en	data available upon registration (request)	data access is given via user account for the data portal. Data can be downloaded as .csv files	http://icp-forests.net/page/plots-data	A complete set of forms (tables) includes: a reduced plot file; data files; data accompanying reports (optional); a laboratory QA file called LQA (in case of surveys with data from laboratory analyses)	http://icp-forests.net/page/changes-database	Till Kirchner (Thünen Institut) till.kirchner@thuenen.de
2	ICP vegetation	Terrestrial	https://icpvegetation.ceh.ac.uk/	International research programme under UNECE LRTAP convention. In-situ measurments of impacts of ozone pollution on vegetation (since 1996) & atmospheric deposition of heavy metals (mk kg-1), nitrogen (% mass), persistent organic pollutants to vegtation	european	-	en	all data are copyrighted to ICP Vegetation	Data is for free. Formats: .xls; .pdf	https://icpvegetation.ceh.ac.uk/data-and-maps/data	NA	https://icpvegetation.ceh.ac.uk/biblio	https://icpvegetation.ceh.ac.uk/contact
3	ICP Integrated Monitoring	Terrestrial	https://www.syke.fi/en-US/Research__Development/Nature/Monitoring/Integrated_Monitoring	Multi-disciplinary research programme under UNECE LRTAP convention. In-situ monitoring of effects of air pollutants and climate change on different ecosystem compartmants at the catchment level (10-1000 ha) with mandatory and optionals monitoring variables	european	-	en	SYKE applies Creative Commons By 4.0 International license for open datasets. data available upon request through the Programme Centre of ICP IM (Pernilla.Ronnback@slu.se)	IM data are freely available and can be extracted in desirable format	https://www.syke.fi/en-US/Open_information/Spatial_datasets	SYKE's INSPIRE content mainly through Atom feeds as GML files; LAPIO; WFS direct access	https://www.syke.fi/en-US/Research__Development/Nature/Monitoring/Integrated_Monitoring/Publications_from_ICP_IM	Pernilla.Ronnback@slu.se
4	ICP Modelling and Mapping	Terrestrial	https://www.unece-wge.org/?page_id=23	International program center under UNECE LRTAP convention. National data submitted by national focla centers & critical loads calculated based on European background data base Air pollution effects, risks and trends	european	-	en	data use agreement	data avilable in .csv-format and can be accessed through CCE (cce@uba.de)	https://www.umweltbundesamt.de/en/cce-data-models	Call for Data - National data submitted by National Focal Centres (NFCs); European background database (EU-DB) calculated by the CCE	https://www.umweltbundesamt.de/en/document/call-for-data-2015-2017	Co-ordination Centre for Effects of ICP Modelling & Mapping: cce@uba.de
5	INTERACT	Terrestrial	https://eu-interact.org/	International Network for Terrestrial Research and Monitoring in the Arctic Circumarctic network. Glaciology, permafrost, climate, ecology, biodiversity and biogeochemical cycling	global	-	en	data available upon request. Restriction are given only in INSPIRE directive and the Directive 2003/4/EC	web access	https://dataportal.eu-interact.org/dataset	NA	NA	margareta.johansson@nateko.lu.se

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
6	GEM Greenland ecosystem monitoring	Terrestrial	https://g-e-m.dk/	Integrated monitoring programme. In-situ monitoring of climate warming on arctic ecosystems in Greenland. It includes 5 programs providing data on meteorology & hydrology, geomorphology, flux monitoring (CO2, CH4, H2O), soil properties, snow properties, glaciar surface mass balance, monitoring of flora, arthropod, birds, mammals, and freshwater biotic and abiotic dynamics as well as physical, chemical and biological data from coastal zone	national	Greenland	en	data avilable upon registration	open web access, .xls format	https://data.g-e-m.dk/	NA	NA	g-e-m@au.dk
7	AMAP-Arctic monitoring and Assessment Programme working Group	Terrestrial	https://www.amap.no/	Working group of Arctic council. Atmospheric contaminats data Marine contaminant data radioactivity data freshwater & terrestrial contaminants datasets from arctic and subarctic areas Data are stored in diverse databases	european	-	en	AMAP Data Policy	Data are stored in thematic data centers: a) atmospheric contaminants data (http://www.nilu.no/) b) marine contaminants data (http://www.ices.dk/) c) radioactivity data (https://www.regjeringen.no/en/dep) d)freshwater and terrestrial contaminants datasets in SynCon database	https://ebas.nilu.no/data-access/ https://data.ices.dk/	NA	NA	amap@amap.no
8	EMEP - European Monitoring and Evaluation Programme	Terrestrial	https://www.ceip.at/webdab-emission-database http://ebas.nilu.no/ https://www.emep.int/mscw/mscw_moddata.html	Co-operative programme under UNECE LRTAP for monitoring and evaluation of the long-range transmission of air pollutants in Europe. Collect emission data, measurments of air and precipitation quality, modelling of atmospheric transport and deposition of air pllutions	european	-	en	https://data.norge.no/nlod/en/2.0 https://creativecommons.org/licenses/by/4.0/	NetCDF. Data are freely available through the web portal. Calculated gridded data is available in NetCDF format and country source receptor tables on CSV format	https://www.emep.int/mscw/mscw_moddata.html	gridded data on NetCDF format and country source receptor tables on CSV format	NA	emep.mscw@met.no
9	ACTRIS	Terrestrial	https://www.actris.eu/	Pan-European aerosol, clouds and trace gases research infrastructure. Detecting changes and trends in aerosols, clouds, and trace gases through in-situ and remote sensing measurements	european	-	en	https://actris.nilu.no/Data/Policy/ https://actris.nilu.no/Content/Documents/DataPolicy.pdf	Available at no cost	https://actris.nilu.no/	NA	NA	ebas@nilu.no
10	EBAS	Terrestrial	http://ebas.nilu.no/	EBAS is a database infrastructure developed and operated by NILU – Norwegian Institute for Air Research. It is designed to document, quality assure, secure long-term storage and provide users for access to atmospheric composition data generated by international and national frameworks and research projects. Atmospheric composition data	global	-	en	most open data, restricted data requires the user to approve the associated data policy	netCDF	http://ebas-data.nilu.no/	INSPIRE, ISO, CF Convention and GEO Air Quality Community of Practice (GEO AQ CoP), harmonized with current and future reporting to the EU Exchange of Information (EoI)	https://ebas.nilu.no/thredds/	kt@nilu.no (Kjetil Tørseth)
11	ICOS - Integrated Carbon Observation System	Terrestrial	https://www.icos-cp.eu/	European-wide greenhouse gas research infrastructure. In-situ greenhouse gas data from atmosphere, ecosystems, oceans & meteorological parameters	european	-	en	ICOS CC4BY licence	Json; csv; xml; TSV or Turtle	https://data.icos-cp.eu/portal/#%7B%22filterCategories%22:%7B%22project%22:%5B%22icos%22%5D,%22level%22:%5B1,2%5D,%22stationclass%22:%5B%22ICOS%22%5D%7D%7D	INSPIRE	SPARQL endpoint https://meta.icos-cp.eu/sparqlclient/?type=CSV	Contact Carbon Portal: info@icos-cp.eu (Alex Vermeulen)

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
12	GLORIA	Terrestrial	https://www.gloria.ac.at/scope/aims	Long-term observation network. Mapping of vascular plants, bryophytes, lichens in alpine ecosystems	global	-	en	data available upon request & consent of data owner	as Access-DB & MariaDB and can be exported in csv. format	NA	NA	NA	harald.pauli@oeaw.ac.at
13	forestREplot	Terrestrial	http://www.forestreplot.ugent.be/about.html	European and North American database on forest vegetation resurveys in temperate, decideous forests. Forest vegetation resurveys	global	-	en	data available within the network upon request and evaluation https://forestreplot.ugent.be/forestREplot_Content-DataPolicy.pdf	excel tables	NA	NA	NA	Kris Verheyen, Genth University, Belgium Kris.Verheyen@UGent.be
14	Global soil partnership	Terrestrial	http://www.fao.org/global-soil-partnership/en/	Collect data on soil organic carbon , soil pollution, soil erosion, soil biodiversity, salinity	global	-	en	http://www.fao.org/global-soil-partnership/insii	web services:Global soil information system, soil spatial data infrastructure	NA	NA	NA	GSP-Secretariat@fao.org
15	SoilBON	Terrestrial	https://geobon.org/bons/thematic-bon/soil-bon/	Global soil biodiversity observation network. Monitoring of temporal dynamics of soil biodiversity and function globally	global	-	en	NA	NA	NA	NA	NA	soilbon.europe-bounces@lists.uni-leipzig.de
16	GCP Global Carbon Project	Terrestrial	https://www.globalcarbonproject.org/index.htm	Global Research Project of Future Earth. Global budgets of carbondioxide, methane and nitous oxide	global	-	en	use of data is conditional	excel, ppt, pdf	https://www.globalcarbonproject.org/carbonbudget/archive.htm	NA	NA	info@globalcarbonproject.org
17	TeaComposition	Terrestrial	https://www.teacomposition.org/	Remaining mass of standardized litter type over time is collected. Mass loss, soil data, climate	global	-	en	free access	open access upon request. .csv.format	https://www.teacomposition.org/data-management/	NA	NA	ika.djukic@umweltbundesamt.at
18	LUCAS 2009 TOPSOIL data (ESDAC)	Terrestrial	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	Soil samples have been analysed for the percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	european	-	en	free access but required registration	Open access upon request. Excel file (for geographical coordinates & soil properties), RDATA and .csv (for multispectral absorbance data), shapefile	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	NA	NA	ec-esdac@ec.europa.eu
19	LUCAS 2015 TOPSOIL data (ESDAC)	Terrestrial	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	JRC ESDAC / Land Use/Cover Area frame statistical Survey. Soil samples have been analysed for the percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	european	-	en	free access but required registration	Open access upon request. Excel file (for geographical coordinates & soil properties), RDATA and .csv (for multispectral absorbance data), shapefile	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	NA	NA	ec-esdac@ec.europa.eu

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
20	ESDAC European Soil Data Centre	Terrestrial	https://esdac.jrc.ec.europa.eu/resource-type/datasets	JRC-European Soil Data Center. Soil point data, soil threats data, soil functions data	european	-	en	creative Commons Attribution 4.0 International (CC BY 4.0) licence	Open access upon request.	https://esdac.jrc.ec.europa.eu/resource-type/datasets-list	NA	NA	NA
21	ESDAC	Terrestrial	https://esdac.jrc.ec.europa.eu/resource-type/datasets	JRC ESDAC / Land Use/Cover Area frame statistical Survey.	european	-	en	free access but required registration	raster, csv...	https://esdac.jrc.ec.europa.eu/resource-type/datasets-list	NA	NA	marie-noelle.pons@univ-lorraine.fr
22	Soil Geographical Database of Eurasia	Terrestrial	https://esdac.jrc.ec.europa.eu/esbn/EUSIS.html	Joint research project: main soil types is based on the terminology of the F.A.O. Soil type	european	-	en	NA	NA	NA	NA	NA	marc.van-liedekerke@jrc.ec.europa.eu
23	PASERA	Terrestrial	https://www.isric.org/projects/pan-european-soil-erosion-risk-assessment-pesera	Within Pan-European Soil Erosion Risk Assessment Project the soil erosion was quantified at the regional scale. Modelled potential monthly soil erosion	european	-	en	NA	NA	NA	NA	NA	stephan.mantel@wur.nl
24	BDAT	Terrestrial	https://www.gissol.fr/donnees/tableaux-de-donnees/donnees-de-la-bdat-3028	Sample collectd by farmers	national	France	fr	free access; licence CC-BY-NC-SA	csv	https://www.gissol.fr/donnees/tableaux-de-donnees/donnees-de-la-bdat-3028	NA	NA	infosol@inrae.fr
25	BASOL	Terrestrial	https://basol.developpement-durable.gouv.fr/tableaux/home.htm#nature%20polluants	Provided by local environmental agency. sample of polluted soil	national	France	fr	free access to the mapping; quid for extraction ?	NA	NA	NA	NA	marie-noelle.pons@univ-lorraine.fr
26	DONESOL	Terrestrial	https://dw3.gissol.fr/login; http://www.gissol.fr/donnees/webservices	GIS SOL (Group of Scientific Interest for the SOIL) and Soil Inventory, Management and Conservation Program (IGCS). Auger drilling or soil pits, soil profile, horizon	national	France	fr	free access but required registration	NA	https://dw3.gissol.fr/fichiers/dictionnaire_donesol_igcs_3-11_01-03-2020.pdf	NA	NA	infosol@inrae.fr
27	Art.11 of habitat directive (FFH)	Terrestrial	https://ec.europa.eu/environment/nature/knowledge/rep_habitats/index_en.htm	Results of the reporting of species status and trends according to art.11 of habitat and species directive; data content and quality needs to the checked. Habitat types & species of community interest	european	-	en	free access https://ec.europa.eu/environment/nature/knowledge/rep_habitats/index_en.htm	html, doc, pdf, xls	https://cdr.eionet.europa.eu/help/habitats_art17	NA	NA	NA
28	EUNIS	Terrestrial	https://eunis.eea.europa.eu/	European Nature Information System. Species, habitat types and protected sites (Natura 2000) across Europe	european	-	en	free access	RDF/XML	https://eunis.eea.europa.eu/index.jsp	NA	NA	NA

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
29	GBIF	Terrestrial	https://www.gbif.org/dataset/search	the Global Biodiversity Information Facility, an international network and research infrastructure. Occurence dataset	global	-	en	free access	NA	https://www.gbif.org/country/FR/summary	NA	NA	yvan.le-bras@mnhn.fr
30	EGDI European Geological Data Infrastructure	Terrestrial	https://www.europe-geology.eu/about-egdi/	EuroGeoSurveys' European Geological Data Infrastructure. Onshore Geology, Marine Geology, Mineral Resources, Geohazards, Energy, Soil, Groundwater	european	-	en	NA	NA	http://www.europe-geology.eu/	https://www.europe-geology.eu/metadata/	https://egdi.geology.cz	contact@europe-geology.eu egdi.metadata(at)geology.cz
31	OpenOBS	Terrestrial	https://openobs.mnhn.fr/	Mixt Unit of Services dedicated to biodiversity observation data (UMS Patrinat; SINP). Occurence dataset	national	France	fr	free access upon registration	various	https://inpn.mnhn.fr/espece/extraction-sinp/preambule (https://inpn.mnhn.fr/accueil/actualites/sommaire)	NA	NA	laurent.poncet@ofb.gouv.fr; yvan.le-bras@mnhn.fr
32	Fungal Root	Terrestrial	https://www.gbif.org/dataset/744edc21-8dd2-474e-8a0b-b8c3d56a3c2d	Database contains global data on the type and intensity of mycorrhizal colonization of vascular plants, provides also information on soil chemical data	global	-	en	free access upon registration at: https://www.gbif.org/	NA	https://www.gbif.org/dataset/744edc21-8dd2-474e-8a0b-b8c3d56a3c2d	NA	https://files.plutof.ut.ee/public/orig/50/F2/50F278A0EFB3400395CC8E65D69527C7DA8D8F7415F6E421C316F56B858A6265.zip	n.a.soudzilovskaia@cml.leidenuniv.nl
33	Global Ants Database	Terrestrial	http://globalants.org/	https://onlinelibrary.wiley.com/doi/pdf/10.1111/icad.12211 Data on ant species richness, abundance, composition and functional traits at the community level	global	-	en	http://globalants.org/static/intellectual-property-guidelines-for-the-global-ants-database.pdf	open access; data avilable upon request via web portal	NA	NA	NA	Heloise Gibb:h.gibb@latrobe.edu.au
34	RMQS	Terrestrial	https://www.gissol.fr/le-gis/programmes/rmqs-34	GIS SOL (Group of Scientific Interest for the SOIL) and Soil Inventory, Management and Conservation Program (IGCS). Survey on 2170 sites	national	France	fr	free access but registration needed	NA	NA	NA	NA	NA
35	CRYOBS-CLIM	Terrestrial	http://data.cryobsclim.fr	eLTER Sites - OZCAR. Sample/observation	global	-	en	free access but required registration	NA	https://cryobsclim.osug.fr/	NA	NA	delphine.six@univ-grenoble-alpes.fr
36	GLACIOCLIM	Terrestrial	https://glacioclim.osug.fr/-Acces-aux-donnees-79-	eLTER Sites - OZCAR. Metherological and glaciological data from alps, Andes, Antarctica, Himalayas; sample/observation	global	-	en	free access	NA	https://cryobsclim.osug.fr/	NA	NA	delphine.six@univ-grenoble-alpes.fr

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
37	The Farmland Bird Index	Terrestrial	https://pecbms.info/trends-and-indicators/indicators/	Across Europe, the Farmland Bird Index is used to monitor the effects of farm managment on the population trands of the farmland bird species. The Farmland Bird Index is used by the European Union's member states to evaluate measures implemented under Rural Development Programme. Along specified route at 10-20 counter points (with 300-400m bee-line distance between the counter points) two times a year the presence of farmland birds is recorded.	european	-	en	freely accessible https://pecbms.info/use-of-the-results/data-access-policy/	web access Excel sheets	https://pecbms.info/trends-and-indicators/indicators/	NA	NA	klvanova@birdlife.cz
38	EuroBirdPortal	Terrestrial	https://www.eurobirdportal.org/ebp/en/#home/HIRRUS/r52weeks/CUCCAN/r52weeks/	European Bird Census Council (EBCC). Birds observation	european	-	en	freely accessible https://eurobirdportal.org/ebp/en/help/	web access	https://www.ebba2.info/contribute-with-your-data/links-to-data-portals/	NA	NA	anella@ornitologia.org
39	EBBC Atlas	Terrestrial	https://www.ebcc.info/?ID=34	European Ornithological Atlas project initiated in 1971. Observation	european	-	en	https://www.ebba2.info/results/data-access-policy/	NA	https://www.gbif.org/dataset/c779b049-28f3-4daf-bbf4-0a40830819b6	NA	NA	Eaton@rspb.org.uk
40	TOURBIERES	Terrestrial	https://www.sno-tourbieres.cnrs.fr/	eLTER Sites - OZCAR. Sample/observation of peatlands, 4 main focus: Meteorology and soil physics, greenhouse gas fluxes, hydrology and hydrochemistry, biodiversity	national	France	en/fr	free access upon registration	NA	https://data-snot.cnrs.fr/data-access/; https://data-snot.cnrs.fr/snot/login.jsf	NA	NA	sebastien.gogo@univ-orleans.fr
41	ALTI	Terrestrial	https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#les-modeles-num%C3%A9riques-3d	National organism focus on geographic information (IGN). Numerical model 3D	national	France	fr	free access only for 75m, then for scientific partnership	NA	https://geoservices.ign.fr/documentation/services/utilisation-web/bibliotheque-dacces	NA	NA	europe-international@ign.fr
42	BDD link to Global Mammal Assessment	Terrestrial	https://globalmammal.org/activities/data-sets/	Prgm. Modeling Global Mammal data sets	global	-	en	contact carlo.rondinini@uniroma1.it, for colloboration	data available upon request the GMA team	NA	NA	NA	julien.renaud@univ-grenoble-alpes.fr
43	BDD link to Global Amphibian Assessment	Terrestrial	https://amphibiaweb.org/data/access.html	Prgm. Modeling synthesizing and sharing information about amphibians aims to establish a knowledge-base for all amphibians	global	-	en	free access	19 MB xml file, dynamics flux	https://amphibiaweb.org/cgi/amphib_query?rel-isocc=like&orderbyaw=Order&where-isocc=Austria	NA	NA	julien.renaud@univ-grenoble-alpes.fr
44	Movebank	Terrestrial	https://www.movebank.org/cms/movebank-content/about-movebank	GPS records manage, share, analyze and archive animal movement data	global	-	en	partly free access, for details a permission is needed	NA	https://www.movebank.org/cms/webapp?gwt_fragment=page=search_map	HTTP/CSV or JSON/JavaScript	https://github.com/movebank/movebank-api-doc	simon.chamaille@cefe.cnrs.fr

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
45	DRILO Base	Terrestrial	http://geo.drilobase.org/	Earthworms Database	global	-	en	NA	NA	http://drilobase.org/	NA	NA	NA
46	SoilTemp	Terrestrial	https://soiltemp.weebly.com/	Global initiative. In-situ microclimate measurments of surface soil temperature (up to 10cm below the surface).	global	-	en	creative Commons Attribution 4.0 International (CC BY 4.0) licence	.tiff, data are in preprint, there might be significant modifications in any updated version	https://soiltemp.weebly.com/data-products.html	NA	NA	jonas.lembrechts@uantwerpen.be
47	International Soil Moisture Network	Terrestrial	https://ismn.geo.tuwien.ac.at/en/	Global in-situ soil moisture database. Soil moisture from different networks.	global	-	en	free access upon registration	NA	https://www.geo.tuwien.ac.at/insitu/data_viewer/	NA	https://ismn.geo.tuwien.ac.at/en/data/advanced-download/	ismn@geo.tuwien.ac.at (Wouter Dorigo)
48	ESA CCI BIOMASS v3.0	Terrestrial	https://climate.esa.int/en/projects/biomass/	Global maps of above-ground biomass for four epochs (mid 1990s, 2010, 2017 and 2018).	global	-	en	https://artefacts.ceda.ac.uk/licences/specific_licences/esacci_biomass_terms_and_conditions.pdf	netCDF and geotiff, public access	https://data.ceda.ac.uk/neodc/esacci/biomass/data/agb/maps/v3.0/	INSPIRE	NA	Scientific Leader: Shaun Quegan, s.quegan@sheffield.ac.uk Project Manager: Richard Lucas, richard.lucas@aber.ac.uk Assistant Project Manager: Heather Kay, hek4@aber.ac.uk ESA Technical Officer: Frank Martin Seifert, Frank.Martin.Seifert@esa.int
49	ESA CCI LAKES v1.1	Aquatic	https://climate.esa.int/en/projects/lakes/	Lake Water Level, Lake Water Extent, Lake Surface Water temperature, Lake Ice Cover and Thickness, and Lake Surface Reflectance from satellite observations	global	-	en	https://artefacts.ceda.ac.uk/licences/specific_licences/esacci_lakes_terms_and_conditions.pdf	netCDF, public access	https://data.ceda.ac.uk/neodc/esacci/lakes/data/lake_products/L3S/v1.1/	Climate and Forecast (CF) and INSPIRE	NA	Science Leader: Jean François Crétaux, jean-francois.cretaux@legos.obs-mip.fr Science Leader: Stefan Simis, stsi@pml.ac.uk Project Manager: Bruno Coulon, bcoulon@gro Project Officer: Beatriz Calmettes, bcalmettes@groupcls.com

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
50	ESA CCI LAND COVER v2.0.7	Terrestrial	https://climate.esa.int/en/projects/land-cover/	Land cover maps Land surface seasonality products all from satellite observations Global map of open water body Surface reflectance products	global	-	en	http://licences.ceda.ac.uk/image/data_access_condition/esacci_landcover_terms_and_conditions.pdf	netCDF and geotiff, public access	https://data.ceda.ac.uk/neodc/esacci/land_cover/data/land_cover_maps/v2.0.7 http://maps.elie.ucl.ac.be/CCI/viewer/index.php	Climate and Forecast (CF) and INSPIRE	NA	Scientific Leader: Pierre Defourny, Pierre.Defourny@uclouvain.be Scientific and Project Manager: Céline Lamarche, celine.lamarche@uclouvain.be ESA Technical Officer: Olivier Arino, Olivier.Arino@esa.int Technical Questions: Technical Questions, contact@esa-landcover-cci.org
51	ESA CCI SNOW	Terrestrial	https://climate.esa.int/en/projects/snow/	Daily global Snow Cover Fraction - snow on ground (SCFG) in percentage (%) per pixel Snow water equivalent (SWE)	global	-	en	open access	netCDF, free and open access	https://data.ceda.ac.uk/neodc/esacci/snow/data/scfg/AVHRR_MERGED/v1.0; https://catalogue.ceda.ac.uk/uuid/93cf539bc3004cc8b98006e69078d86b	Climate and Forecast (CF) and INSPIRE	NA	Scientific Leader: Thomas Nagler, thomas.nagler@enveo.at Project Manager: Gabriele Schwaizer, gabriele.schwaizer@enveo.at ESA Technical Officer: Anna Maria Trofaier, Anna.Maria.Trofaier@esa.int
52	ESA CCI SM v6.1	Terrestrial	https://www.esa-soilmoisture-cci.org/index.php?q=node	Surface soil moisture (~0-5cm of depth) from satellite observations	global	-	en	free access upon registration	data is provided as global, daily netCDF files	https://climate.esa.int/en/projects/soil-moisture/data/	Climate and Forecast (CF) version 1.7	NA	cci_sm_contact@eodc.eu (E-mails sent to this address reaches both data managers and key project scientists)
53	Copernicus Global Land Services SSM	Terrestrial	https://land.copernicus.eu/global/products/ssm	Online resource. Surface Soil Moisture (%) from satellite observations	european	-	en	open access	multi-band netCDF4 files	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Time=NORMAL,NORMAL,-1,,,-1,,	Climate and Forecast (CF) conventions (v1.6)	NA	helpdeskticket@vgt.vito.be

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
54	Copernicus Global Land Services SWI	Terrestrial	https://land.copernicus.eu/global/products/swi	Soil Water Index (up to 5cm)	european	-	en	open access	multi-band netCDF4 files	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=514690;Collection=1000281;	Climate and Forecast (CF) conventions (v1.6)	NA	helpdeskticket@vgt.vito.be
55	Copernicus Global Land Services SWI	Terrestrial	https://land.copernicus.eu/global/products/swi	Online resource. Soil Water Index (up to 5cm)	global	-	en	open access	multi-band netCDF4 files	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=514690;Collection=735734	Climate and Forecast (CF) conventions (v1.6)	NA	helpdeskticket@vgt.vito.be
56	CES OSO	Terrestrial	https://www.theia-land.fr/en/ceslist/land-cover-sec/	CESBIO Laboratory. Interpretation of satellite images for land use	national	France	en, fr	free access	NA	https://labo.obs-mip.fr/multitemp/fully-automatic-land-cover-map-generation-at-country-scale-over-france/	NA	NA	vincent.thierion@inrae.fr
57	GLASS	Terrestrial	http://www.glass.umd.edu/Overview.html	Long-term Global Land Surface Satellite Dataset. Remote sensing	global		en	free access	hdf /netcdf	https://doi.org/10.1080/17538947.2013.805262	NA	NA	geodata@igsnrr.ac.cn
58	SMOS Soil Moisture and Ocean Salinity	Terrestrial	https://smos-diss.eo.esa.int/oads/access/	soil moisture, vegetation optical depth and other ancillary data derived during the processing (surface temperature, roughness parameter, dielectric constant, brightness temperature at the top of the atmosphere and at the surface) with their corresponding uncertainties. Remote sensing	global	-	en	free access	Earth Explorer and NetCDF	https://earth.esa.int/eogateway/catalog/smos-nrt-data-products	NA	NA	smos.dissemination.support@esa.int
59	MODIS	Terrestrial	http://www.ntsg.umt.edu/project/modis/mod17.php	Remote sensing	global	-	en	free access	netcdf	https://modis-land.gsfc.nasa.gov/npp.html#	NA	NA	john.kimball@mso.umt.edu
60	Copernicus Global Land Services LAI300	Terrestrial	https://land.copernicus.eu/global/products/lai	Remote sensing	global	-	en	free access	multi-band NetCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=512260;Collection=1000062;Time=NORMAL,NORMAL,-1,,,-1,,	Climate and Forecast (CF) conventions v1.6 and INSPIRE	NA	helpdeskticket@vgt.vito.be
61	Copernicus Global Land Services LAI 1km	Terrestrial	https://land.copernicus.eu/global/products/lai	Remote sensing	global	-	en	free access	multi-band NetCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=512260;Collection=1000083;Time=NORMAL,NORMAL,-1,,,-1,,	Climate and Forecast (CF) conventions v1.6 and INSPIRE	NA	helpdeskticket@vgt.vito.be

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
62	GLIMS	Terrestrial	http://glims.colorado.edu/glacierdata/	National Snow and Ice Data Center; GLIMS for Geospatial Glacier Database. Remote sensing	global	-	en	free access	ESRI Shapefile, MapInfo Table Format, GML (Geography Markup Language), KML (Keyhole Markup Language for viewing in Google Earth), GMT (Generic Mapping Tools)	https://www.glims.org/download/	NA	NA	http://www.glims.org/maps/contact_info.html
63	European Fluxes Database Cluster	Terrestrial	http://www.europe-fluxdata.eu/	European database of greenhouse gas fluxes. In-situ greenhouse gas fluxes between ecosystems and the atmosphere, supporting meteorological, soil and vegetation data	european	-	en	data available upon registration; site-specific data policy, mostly open	access via web portal, export as csv	http://www.europe-fluxdata.eu/home/log-in/	NA	NA	database@unitus.it
64	BDCHARM 50	Terrestrial	http://infoterre.brgm.fr/page/geoservices-ogc	BRGM. Maps	national	France	fr	free access partially	layer, WMS, WFC, KML	http://infoterre.brgm.fr/page/geoservices-ogc	NA	NA	NA
65	BDIFF	Terrestrial	https://bdiff.agriculture.gouv.fr/	Observation of forest fire events from various sources (6 categories)	national	France	fr	required agreement	NA	NA	NA	NA	NA
66	Open Government Data	Terrestrial	https://www.europeandataportal.eu/en	official portal for European data The portal provides access to open data from international, EU, national, regional, local and geo data portals. It replaces the EU Open Data Portal and the European Data Portal. Sections: Searching Data, Providing Data, Using data	european	-	en	https://data.europa.eu/en/legal-notice	NA	NA	NA	NA	https://data.europa.eu/en/feedback/form
67	INSPIRE	Terrestrial	https://inspire-geoportal.ec.europa.eu/	Data on air quality, Water framework directive, habitates, flora, fauna, birds. Observation	european	-	en	https://inspire.ec.europa.eu/privacy-policy/59294	NA	NA	NA	https://catalog.inspire.geoportail.lu/geonetwork/srv/ger/catalog.search#/home	jrc-inspire-support@ec.europa.eu
68	THEIA/OZCAR	Terrestrial	https://in-situ.theia-land.fr/	Data Center for Earth's Surface Datasets. NA	national	France	en	free access Metadata	various	NA	NA	NA	isabelle.braud@inrae.fr
69	Pangea	Terrestrial	https://www.pangaea.de/	Species. Georeferenced data	global	-	en	open source libraries ; https://wiki.pangaea.de/wiki/Data_policy	NA	NA	NA	https://pypi.org/project/pangaeapy/	https://www.pangaea.de/contact/

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
70	Zenodo	Terrestrial	https://www.eui.eu/Research/Library/ResearchDataServices	Species. Multi-disciplinary data	european	-	en	files may be deposited under closed, open, or embargoed access;https://about.zenodo.org/terms/	supports all file formats	https://www.zenodo.org/communities/	Metadata is licensed under CC0, except for email addresses. All metadata is exported via OAI-PMH and can be harvested	OAI-PMH: https://zenodo.org/oai2d REST API: https://zenodo.org/api/records/ Documentation: https://developers.zenodo.org	https://zenodo.org/support
71	FigShare	Terrestrial	https://figshare.com/	Species. Multi-disciplinary data; e.g eath and evionmental science figshare is a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner	global	-	en	https://help.figshare.com/article/preservation-and-continuity-of-access-policy	Public pages on Figshare are compliant with the following accessibility standards:European accessibility standard EN, 301 549 WCAG 2.1 AA, Section 508	https://figshare.com/ https://api.figshare.com/v2	DublinCore + user extensions	https://api.figshare.com/v2	info@figshare.com
72	VIGIE-NATURE	Terrestrial	NA	UMS PatriNat. Observation from citizen science project	national	France	fr/en	request form https://framaforms.org/demande-de-donnees-issues-du-programme-vigie-nature-1524124340	NA	http://www.vigienature.fr/fr/acces-donnees-3611	NA	NA	dozieres@mnhn.fr
73	Knowledge Base "Status"	Terrestrial	https://inpn.mnhn.fr/telechargement/referentielEspece/bdc-statuts-especes	UMS PatriNat. From ~150 regulations in FR, information on habiates, species	national	France	en	free access	NA	https://inpn.mnhn.fr/programme/base-de-connaissance-statuts/presentation	NA	NA	https://inpn.mnhn.fr/contact/contacteznous
74	COPERNICUS HRL Layers	Terrestrial	https://land.copernicus.eu/pan-european/high-resolution-layers	High resolution land cover layer for Europe under COPERNICUS work programme. Land cover date for land cover elements: Imperviousness, forest, grassland, wettness/water and small woody features in 3-year time teps 2009-2012-2015-2018 to be continued; with 20m and since 2018 with 10m resolution	european	-	en	COPERNICUS free and open policy	GRID-data, free download from Website	https://land.copernicus.eu/pan-european/high-resolution-layers - pages for the respective datasets including links to WMS/WFS services	ISO19115/INSPIRE	metadata provided for single datasets but no common catalogue tob e harvested	https://land.copernicus.eu/contact
75	CLC+ work form MC	Terrestrial	https://land.copernicus.eu/eagle	Standardization and harmonization from national LANDUSE data using EAGLE data model and upscaling it to european reference grid. Work in progress, to be pulbished 2021	european	-	en	COPERNICUS free and open policy	GRID data, tbd.	NA	NA	NA	https://land.copernicus.eu/contact
76	CHELSA v2.1	Meterology	https://chelsa-climate.org/downloads/	Online resource. Globla air temperature and precipitation data; especially precipitation patterns are more accurate.	Global	-	en	free access	.sdat, sgrd, .prj…	https://envicloud.wsl.ch/#/?prefix=chelsa%2Fchelsa_V2%2FGLOBAL%2F	proprietary	NA	julien.renaud@univ-grenoble-alpes.fr
77	World Clim v2.1	Meterology	https://worldclim.org/	Online resource. Historical climate data (1970-2000)	global	-	en	free access	GeoTiff	https://worldclim.org/data/worldclim21.html	NA	NA	info@worldclim.org

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
78	E-OBS	Meterology	https://www.ecad.eu/download/ensembles/download.php	Online resource. Daily gridded weather data	european	-	en	free access upon registration	netCDF	https://www.ecad.eu/download/ensembles/ensembles.php	CF Convention	NA	NA
79	CHPclim	Meterology	https://www.chc.ucsb.edu/data/chpclim	High-quality monthly rainfall climatology (in situ measurements + satellite)	global	-	en	open access	netCDF or GeoTiff	http://data.chc.ucsb.edu/products/CHPclim/	NA	NA	pete@geog.ucsb.edu (Pete Peterson)
80	GPCC	Meterology	https://www.dwd.de/EN/ourservices/gpcc/gpcc.html	Gridded gauge-analysis products derived from quality controlled station data	global	-	en	open access	netCDF or ascii	https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre	https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre	NA	gpcc@dwd.de
81	TRMM	Meterology	https://climatedataguide.ucar.edu/climate-data/trmm-tropical-rainfall-measuring-mission	Tropical and subtropical precipitation	global	-	en	free access upon registration	netCDF, HDF	https://gpm.nasa.gov/data/directory	https://climatedataguide.ucar.edu/climate-data/trmm-tropical-rainfall-measuring-mission	NA	https://ncar.ucar.edu/who-we-are/contact-us https://www.ucar.edu/who-we-are/contact-us
82	ERA-Interim	Meterology	https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim	ECMWF is the European Centre for Medium-Range Weather Forecasts. It has been superseded by the ERA5 reanalysis. research institute and a 24/7 operational service, producing global numerical weather predictions and other data for our Member and Co-operating States and the broader community. The Centre has one of the largest supercomputer facilities and meteorological data archives in the world. Other strategic activities include delivering advanced training and assisting the WMO in implementing its programmes.	global	-	en	https://www.ecmwf.int/en/privacy	Forecast Access upon registration	https://www.ecmwf.int/en/forecasts/accessing-forecasts	NA	NA	https://www.ecmwf.int/en/about/contact-us
83	ERA5	Meterology	https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5	Estimation of atmospheric, land and oceanic climate variables	global	-	en	free access; datasets are held in the ECMWF data archive (MARS)	netCDF, GRIB2, GRIB1	https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset	NA	NA	NA
84	DRIAS	Meterology	http://drias-climat.fr/accompagnement/section/40	Geoportal. Atmospheric expriments (ex. EUROCORDEX climate projections) and impact experiments (ex. ADAMONT)	national	France	en	NA	NA	NA	NA	NA	NA
85	AERIS	Meterology	https://en.aeris-data.fr/catalogue/	Data Center for Atmosphere Datasets. NA	national	France	en	free access Metadata	various	NA	NA	NA	NA

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
86	HyMeX	Meterology	https://mistrals.sedoo.fr/HyMeX/	Prgm. Hydrological cycle in the Mediterranean EXperiment, geophysical in situ observations, satellite products, model outputs, questionnaire data. Searchable by parameters, instruments or events.	global	-	en	https://mistrals.sedoo.fr/HyMeX/Data-Policy/	free access but registration needed	https://mistrals.sedoo.fr/HyMeX/Browse-Catalogue/	various	NA	databasecontact@hymex.org
87	RADOME	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance). Basic parameters (Air T, humidity, wind speed and direction, precipitation) and additional parameters according to the instrumentation (soil T, visibility, radiation)	national	France	fr	required agreement	ASCII or BUFR for hourly data, ASCII for data 6 min	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
88	RADOME	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance). Basic parameters (Air T, humidity, wind speed and direction, precipitation) and additional parameters according to the instrumentation (soil T, visibility, radiation)	national	France	fr	required agreement	ASCII or BUFR for hourly data, ASCII for data 6 min	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
89	RADOME	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance). Basic parameters (Air T, humidity, wind speed and direction, precipitation) and additional parameters according to the instrumentation (soil T, visibility, radiation)	national	France	fr	required agreement	ASCII or BUFR for hourly data, ASCII for data 6 min	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
90	RADOME	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance). Basic parameters (Air T, humidity, wind speed and direction, precipitation) and additional parameters according to the instrumentation (soil T, visibility, radiation)	national	France	fr	required agreement	ASCII or BUFR for hourly data, ASCII for data 6 min	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
91	NIVOSE	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance) in high altitude. Samples	national	France	fr	required agreement	ASCII	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
92	NIVOSE	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance) in high altitude. Samples	national	France	fr	required agreement	ASCII	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
93	NIVOSE	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance) in high altitude. Samples	national	France	fr	required agreement	ASCII	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA
94	NIVOSE	Meterology	https://donneespubliques.meteofrance.fr/	National organism focus on meteorological data (MeteoFrance) in high altitude. Samples	national	France	fr	required agreement	ASCII	http://www.meteofrance.fr/prevoir-le-temps/observer-le-temps/moyens/les-stations-au-sol	NA	NA	NA

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
95	ATMO-AASQA	Meterology	https://atmo-france.org/les-donnees/	Federation of air pollutant associations (18 in total AASQA; Atmo France). Station measurement data, emission data	national	France	fr	free access	csv	https://atmo-france.org/wp-content/uploads/2020/04/notice_6avril2020.pdf	NA	NA	data@atmo-france.org
96	ICP Waters	Aquatic	http://www.icp-waters.no/	International cooperative programme. In-situ monitoring of effect of atmospheric pollution on surface water (lake, rivers) since 1985	european	-	en	NA	data avilable in .csv-format. There is no open access, data providers need to contact national agency website (http://www.icp-waters.no/useful-links/)	Data – ICP Waters (icp-waters.no)	NA	NA	kari.austnes@niva.no
97	Danubius-RI	Aquatic	https://www.danubius-ri.eu/index.html	Pan-European research infrastructure on intradisciplinary research on river-delta-sea systems.	european	-	en	open access	Variable	https://gis.geoecomar.ro/danubius/dataportal/general_data_list.php?orderby=did	None	NA	NA
98	GLEON - Global Lake Ecological Observatory Network	Aquatic	https://gleon.org/data	The Global Lake Ecological Observatory Network conducts innovative science by sharing and interpreting high-resolution sensor data to understand, predict and communicate the role and response of lakes in a changing global environment.	global	-	en	https://gleon.org/sites/default/files/pdf/data/2009_October_15_GLEON_data_access_policy.pdf	GLEON data may be accessed via the EDI search, the DataONE search or the Google data set search. High frequency sensor data from GLEON buoys may also be submitted and accessed in the CUAHSI data system	https://hiscentral.cuahsi.org/pub_network.aspx?n=3555 https://hiscentral.cuahsi.org/pub_network.aspx?n=3552 https://hiscentral.cuahsi.org/pub_network.aspx?n=3553	NA	NA	NA
99	MarineStrategyFrameworkDirective	Aquatic	https://ec.europa.eu/environment/marine/eu-coast-and-marine-policy/marine-strategy-framework-directive/index_en.htm	The aim of the European Union's ambitious Marine Strategy Framework Directive is to protect more effectively the marine environment across Europe.	european	-	en	https://ec.europa.eu/info/privacy-policy_en	NA	https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1495097018132&uri=CELEX:32017D0848	NA	NA	https://ec.europa.eu/info/departments/environment_en#contact
100	Water Framework Directive	Aquatic	https://ec.europa.eu/environment/water/water-framework/info/intro_en.htm	Water protection is therefore one of the priorities of the Commission. European Water Policy should get polluted waters clean again, and ensure clean waters are kept clean. The following will provide an overview on development, present state and future of European Water Policy.	european	-	en	https://ec.europa.eu/info/privacy-policy_en	NA	https://water.europa.eu/	NA	NA	https://ec.europa.eu/info/departments/environment_en#contact
101	TeaComposition H2O	Aquatic	https://www.bluecarbonlab.org/teacomposition-h2o/	Standardized litter material was incubated around the world in coastal wetlands (mangrove, tidal marshes, and seagrass), freshwater wetlands (bogs, fens, riverine, lacustrine) and aquatic ecosystems (seaweeds, streams, ponds) over period of 3 years and the mass loss has been detected over time. Litter mass loss and microbial community composition	global	-	en	data avilable upon request	.csv fromat	https://www.bluecarbonlab.org/teacomposition-h2o/	NA	NA	s.trevathantackett@deakin.edu.au
102	SANDRE	Aquatic	http://www.sandre.eaufrance.fr/atlas/srv/fre/catalog.search#/home	Geoportal for French Data on Freshwater and Coastal Ecosystems. Referentials	national	France	fr	free access Metadata	various	NA	NA	NA	NA

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
103	Waste water Treatment Plant (Urban Waste water)	Aquatic	http://assainissement.developpement-durable.gouv.fr/	French public data. Measurement station	national	France	fr	free access	csv, ods	NA	NA	NA	laurence.lestel@sorbonne-universite.fr (ZAS)
104	BRGM	Aquatic	https://infoterre.brgm.fr/page/cartes-geologiques	Geology maps.	national	France	fr/en	free access	NA	NA	NA	NA	NA
105	BRGM	Aquatic	https://www.brgm.fr/projet/referentiel-hydrogeologique-francais-bdlisa	BDLISA BRGM (aquifers).	national	France	fr/en	free access	NA	NA	NA	NA	NA
106	WATERBASE-Water quality ICM	Aquatic	https://www.eea.europa.eu/data-and-maps/data/waterbase-water-quality-icm-1	European public data. Sample	european	-	en	free access	csv (very large), SQLite in 2 parts	https://discodata.eea.europa.eu/	NA	NA	marie-noelle.pons@univ-lorraine.fr
107	WATERBASE-Biology	Aquatic	https://www.eea.europa.eu/data-and-maps/data/waterbase-biology	European public data. Sample Waterbase is the generic name given to the EEA's databases on the status and quality of Europe's rivers, lakes, groundwater bodies and transitional, coastal and marine waters, on the quantity of Europe's water resources, and on the emissions to surface waters from point and diffuse sources of pollution	european	-	en	free access	csv, SQLite	NA	NA	NA	marie-noelle.pons@univ-lorraine.fr
108	CCM (Catchment Characterisation and Modelling)	Aquatic	https://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23	JRC. Hydrographic network (modeling)	european	-	en	free access upon registration	ESRI File Geodatabase	https://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=50	NA	NA	https://edo.jrc.ec.europa.eu/edov2/php/index.php?id=1038
109	BANQUE HYDRO	Aquatic	http://www.hydro.eaufrance.fr/indexd.php?connect=1	Provided by local environmental agency. Station, measurement	national	France	fr	free access upon registration	csv, txt	NA	NA	NA	NA
110	ADES	Aquatic	https://ades.eaufrance.fr/	National organism focus on groundwater data (BRGM). Sensor and human based, analyses in lab	national	France	fr	free access upon registration	NA	https://ades.eaufrance.fr/Recherche	NA	NA	a.winckel@brgm.f

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
111	DATA EAU FRANCE-Fish	Aquatic	http://www.data.eaufrance.fr/	French public data. Electric fishing for national survey	national	France	fr	free access	NA	NA	NA	NA	NA
112	DATA EAU FRANCE-Contaminants	Aquatic	http://www.data.eaufrance.fr/	French public data. Contaminants in sediments, fish	national	France	fr	free access	NA	NA	NA	NA	NA
113	NAIADES	Aquatic	http://www.naiades.eaufrance.fr/france-entiere#/	French public data. Sample for freshwater biodiversity, measurements ofr hydromorphology, lab analyses for PC, sensor for T)	national	France	fr	free access	NA	http://naiades.eaufrance.fr/gestion-url	NA	NA	NA
114	CYBERCAROTHEQUE	Aquatic	https://cybercarotheque.fr	Consortium of national research organisms. Sample, and high-resolution multi-proxy (sedimentological, palynological, and geochemical) analysis	national	France	fr	access under conditions, contact cecile.pignol@univ-smb.fr	NA	NA	NA	NA	cecile.pignol@univ-smb.fr
115	ODATIS	Aquatic	https://www.odatis-ocean.fr/en/data-and-services/data-access/direct-access-to-the-data-catalogue#/search?from=1&to=30	French public data. In-situ observations as well as remote sensing data on marine biogeochemistry, geology, meteorology, phyiscal oceanography are collected.	national	France	en	publicly avilable. For some data restictive licenses might be needed	NA	https://www.odatis-ocean.fr/en/data-and-services/data-access/direct-access-to-the-data-catalogue#/search?from=1&to=30	INSPIRE	NA	https://www.odatis-ocean.fr/en/contact
116	Sextant	Aquatic	https://sextant.ifremer.fr/eng	National organism focus on coastal/marine data (IFREMER). Data on marine ecosystems	national	France	en	required agreement	NA	https://sextant.ifremer.fr/eng/Data	NA	https://sextant.ifremer.fr/eng/Services/Use-the-Sextant-API	sextant@ifremer.fr
117	UWWTD	Aquatic	https://uwwtd.eu/	Size, compliance of Urban WWTP	european	-	en	free access	depends upon the type of data (xlsx, xml, csv, shp, kml)	NA	NA	NA	NA
118	BioFresh	Aquatic	http://www.freshwaterplatform.eu/	An internet platform bringing together information and data on freshwater biodiversity. Freshwater Macro-invertebrates Laboratory quality control samples Freshwater Diatoms Soil Solution Chemistry Invertebrates: Moths Atmospheric Chemistry: NO2 Freshwater Flow Freshwater Macro-invertebrates Invertebrates: Spittle	european	-	en	NA	NA	http://www.freshwaterplatform.eu/	http://data.freshwaterbiodiversity.eu/metadb/about_metadata/	NA	http://www.freshwaterplatform.eu/index.php/contact.html

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
119	COPERNICUS	Aquatic	https://land.copernicus.eu/global/themes/water	Water COPERNICUS. Lake surface water temperature (LSWT) in K	global	-	en	open access	netCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=1100100;Collection=1100102;Time=NORMAL,NORMAL,-1,,,-1,,	metadata according to the Climate and Forecast (CF) conventions (v1.6)	NA	Accountable contact: European Commission Directorate – General Joint Research Centre, Email address: copernicuslandproducts@jrc.ec.europa.eu Scientific contact: BC and PML, Email address: lwq_info@brockmann-consult.de Production and distribution contact: BC and PML, Email address: lwq_info@brockmann-consult.de
120	COPERNICUS	Aquatic	https://land.copernicus.eu/global/themes/water	Water COPERNICUS. Lake turbidity, trophic state index, lake surface reflectances	global	-	en	open access	netCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=1100500;Collection=1100502;Time=NORMAL,NORMAL,-1,,,-1,,	metadata according to the Climate and Forecast (CF) conventions (v1.6)	NA	Accountable contact: European Commission Directorate – General Joint Research Centre, Email address: copernicuslandproducts@jrc.ec.europa.eu Scientific contact: BC and PML, Email address: lwq_info@brockmann-consult.de Production and distribution contact: BC and PML, Email address: lwq_info@brockmann-consult.de

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
121	COPERNICUS	Aquatic	https://land.copernicus.eu/global/themes/water	Water COPERNICUS. Lake turbidity, trophic state index, lake surface reflectances	global	-	en	open access	netCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=1100500;Collection=1100701;Time=NORMAL,NORMAL,-1,,,-1,,	metadata according to the Climate and Forecast (CF) conventions (v1.6)	NA	Accountable contact: European Commission Directorate – General Joint Research Centre, Email address: copernicuslandproducts@jrc.ec.europa.eu Scientific contact: BC and PML, Email address: lwq_info@brockmann-consult.de Production and distribution contact: BC and PML, Email address: lwq_info@brockmann-consult.de
122	COPERNICUS	Aquatic	https://land.copernicus.eu/global/themes/water	Water COPERNICUS. Min and max extent of water bodies	global	-	en	open access	netCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=514888;Collection=1000152;Time=NORMAL,NORMAL,-1,,,-1,,	metadata according to the Climate and Forecast (CF) conventions (v1.6)	NA	Accountable contact: European Commission Directorate – General Joint Research Centre, Email address: copernicuslandproducts@jrc.ec.europa.eu Scientific contact: BC and PML, Email address: lwq_info@brockmann-consult.de Production and distribution contact: BC and PML, Email address: lwq_info@brockmann-consult.de

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
123	COPERNICUS	Aquatic	https://land.copernicus.eu/global/themes/water	Water COPERNICUS. Rivers and lakes water level	global	-	en	open access	netCDF4	https://land.copernicus.vgt.vito.be/PDF/portal/Application.html#Browse;Root=1000120;Collection=1000122;Time=NORMAL,NORMAL,-1,,,-1,,	metadata according to the Climate and Forecast (CF) conventions (v1.6)	NA	Accountable contact: European Commission Directorate – General Joint Research Centre, Email address: copernicuslandproducts@jrc.ec.europa.eu Scientific contact: BC and PML, Email address: lwq_info@brockmann-consult.de Production and distribution contact: BC and PML, Email address: lwq_info@brockmann-consult.de
124	ESA-CCI+	Aquatic	netCDF and geotiff, public access. It is possible to extract the CCI Land Cover products at specific locations. The output is a CSV.	ESA Lakes Climate Change Initiative. Data relevant to Atmospheric Sciences and Earth Observation fields; Lake water level, lake water extent, Lake surface water temperature, lake Ice cover,Lake water Leaving Reflectance	global	-	en	open access upon registration:https://help.ceda.ac.uk/article/4660-depositing-data-at-ceda-a-step-by-step-guide	https://help.ceda.ac.uk/article/98-accessing-data wide range of file formats:https://help.ceda.ac.uk/article/104-file-formats	https://data.ceda.ac.uk/neodc/esacci/lakes/data/lake_products/L3S/v1.0/	NA	NA	https://help.ceda.ac.uk/
125	AQUACOSM	Experiments	https://www.aquacosm.eu/	EU network of mesocosms facilities for research on marine and freshwater ecosystems. Pyhtoplantkon, Zooplanktion, Microbial Planktion, Periphyton, Water Chemistry,	european	-	en	open access 6 months after completion of the publishable dataset	NA	https://www.aquacosm.eu/data/	https://www.aquacosm.eu/geonetwork/srv/eng/catalog.search#/search?resultType=details&sortBy=relevance&_cat=datasets&from=1&to=20	NA	https://www.aquacosm.eu/contact/
126	TERENO	Experiments	https://www.tereno.net/	Soil temperature and soil moisture at different depths (5, 20 and 50 cm). Plus precipitation, evapotranspiration, runoff and periodically also different chamber measurements were made to access soil or plant gas exchange.	national	Germany	en	https://ddp.tereno.net/ddp/docs/TERENO-Data%20Policy%20V2.0.pdf	netCDF, open access via OPENDAP or HTTPServer	https://service.tereno.net/thredds/terenoCatalog.html	Dublincore, ISO19115	NA	support-tereno@tereno.net
127	TERENO	Experiments	https://www.tereno.net/	Soil temperature and soil moisture at different depths (5, 20 and 50 cm). Plus precipitation, evapotranspiration, runoff and periodically also different chamber measurements were made to access soil or plant gas exchange.	national	Germany	en	Open access https://ddp.tereno.net/ddp/docs/TERENO-Data%20Policy%20V2.0.pdf	netCDF, open access via OPENDAP or HTTPServer	https://service.tereno.net/thredds/terenoCatalog.html	NA	NA	support-tereno@tereno.net

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
128	Droughtnet	Experiments	https://drought-net.colostate.edu/	Coordinated global experimental network. Site and vegetation data, weather data	global	-	en	https://drought-net.colostate.edu/guidelines-participation-die	NA	NA	NA	NA	https://drought-net.colostate.edu/contact
129	TreeDivNet	Experiments	http://www.treedivnet.ugent.be/	Global experimental network. Global data on the effect of tree species diversity on the ecosystem functioning	global	-	en	data avilable upon request	data avilable in .xls-format; no central data base on data but on experiments (http://www.treedivnet.ugent.be/measurements.html)	NA	NA	NA	https://treedivnet.ugent.be/contact.html
130	NutNet	Experiments	https://nutnet.org/	Coordinated global experimental network.	global	-	en	data avilable upon request	.csv /EML format	NA	NA	NA	wilf0020@umn.edu
131	ANAEE - France	Experiments	http://w3.avignon.inra.fr/geonetwork_anaee/srv/eng/catalog.search#/home	European RI. NA	national	France	en	free access Metadata	various	NA	NA	NA	christian.pichot@inrae.fr
132	OLA	Experiments	https://si-ola.inra.fr/si_lacs/login.jsf	ANAEE Database. High altitude lake	national	France	en	free access upon registration	NA	https://www6.inrae.fr/soere-ola/ (FR)	NA	NA	ghislaine.monet@inrae.fr
133	F-ORE-T	Experiments	http://www.gip-ecofor.org/f-ore-t/	ANAEE Database. Flow data from EC towers (recently accumulation chambers), meteorological data and soil climate data	national	France	en/fr	free access upon registration	NA	https://si-foret-ecofor.fr/si_foret-prod/login.jsf	NA	NA	damien.maurice@inrae.fr
134	ACBB	Experiments	https://si-acbb.inra.fr/acbb-web-1912.01/login.jsf;jsessionid=78D5DFD19E2F0692EA74E544FC251ED6.worker1	ANAEE Database. Temporary grassland	national	France	en	free access upon registration	NA	http://www.soere-acbb.com/caracteristiques (fr)	NA	NA	NA
135	PRO	Experiments	distributed	ANAEE Database. Environmental impacts of Organic Waste Products recycling on field crops at long time scale	national	France	en	NA	NA	https://www6.inrae.fr/qualiagro_eng/Nos-partenaires/The-SOERE-PRO-network	NA	NA	NA
136	EJP SOIL	Experiments	https://ejpsoil.eu/	EJP SOIL is a European Joint Programme Cofund on Agricultural Soil Management contributing to key societal challenges including climate change, water and future food security. The objectives are to develop knowledge, tools and an integrated research community to foster climate-smart sustainable agricultural soil management	european	-	en	https://international.au.dk/about/profile/privacy-policy/	https://www.was.digst.dk/au-dk	NA	NA	NA	ejpsoilcoord@inrae.fr

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
137	H+	Experiments	http://hplus.ore.fr/en/	eLTER Sites - OZCAR. Sample/observation	national	France	en	free access upon registration	NA	http://hplus.ore.fr/base-de-donnees-fr	NA	NA	annick.battais@univ-rennes1.fr
138	Census Hub	Socio-Ecology	https://ec.europa.eu/CensusHub2/query.do?step=selectHyperCube&qhc=false	Eurostat is the statistical office of the European Union. the European Statistical System (ESS) was built up gradually with the objective of providing comparable statistics at EU level. Our mission is to provide high quality statistics and data on Europe.	european	-	en	https://ec.europa.eu/eurostat/about/policies/accessibility	https://ec.europa.eu/eurostat/about/policies/accessibility	NA	NA	NA	https://ec.europa.eu/eurostat/help/support
139	INSEE	Socio-Ecology	https://www.insee.fr/fr/statistiques/3677855	French public data. Survey	national	France	fr	free access	xls	NA	NA	NA	NA
140	SIDDT	Socio-Ecology	psql9.grenoble.cemagref.fr/module_statistiques/mod_extraction_bdd/liste_tables.php	French public data. Indicators	national	France	fr	free access	NA	https://siddt.irstea.fr/;psql9.grenoble.cemagref.fr/module_statistiques/fichiers_aide/DOC-inventaire_donnees-siddt.pdf	NA	NA	frederic.bray@inrae.fr
141	AGRESTE	Socio-Ecology		French public data. Data from individuals link to farm/crop management	national	France	fr	free access for statistical data; requires approval for scientific program and an access to individual practices	NA	https://agreste.agriculture.gouv.fr/agreste-web/servicon/I.1/listeTypeServicon/	NA	NA	claude.napoleone@inrae.fr
142	CORINE LAND COVER	Socio-Ecology	https://www.data.gouv.fr/fr/datasets/corine-land-cover-occupation-des-sols-en-france/#_	Open platform for French public data. Visual interpretation of satellite images for land use	national	France	fr	free access	NA	NA	NA	NA	NA
143	RPG (Parcel Registry Graphic)	Socio-Ecology	https://www.data.gouv.fr/fr/datasets/registre-parcellaire-graphique-rpg-contours-des-parcelles-et-ilots-culturaux-et-leur-groupe-de-cultures-majoritaire/	National organism focused on geographic information (IGNF). Crops types declared by farmers to receive Common Agricultural Policy aids	national	France	fr	free access (open Licence 1.0)	shapefile	NA	NA	NA	NA
144	ZNIEFF	Socio-Ecology		National organism (MNHN; SINP). Observation, expertise	national	France	fr/en	partly free access	wfs, wms	https://inpn.mnhn.fr/telechargement/documentation/znieff	NA	NA	flepareur@mnhn.fr

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
145	REMONTERLE TEMPS	Socio-Ecology	https://remonterletemps.ign.fr/	National organism focus on geographic information (IGN). Scan of historical land cover (>1950)	national	France	fr	free access	NA	https://remonterletemps.ign.fr/	NA	NA	NA
146	CDDA	Socio-Ecology	https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-16	European Common Database on Designated Areas. GIS	european	-	en	free access	NA	https://www.eionet.europa.eu/	NA	https://www.eea.europa.eu/data-and-maps/RSS	NA
147	European Social Survey	Socio-Ecology	https://www.europeansocialsurvey.org/	Annual social survey - core questions (every year) + rotating module (one-time questions). Media, politics, wellbeing, identity, socio-demographic profile, values	european	-	en	free access	SPSS/Stata/SAS	https://www.europeansocialsurvey.org/data/	NA	NA	ess@city.ac.uk
148	DINAMIS	Diverse	https://dinamis.data-terra.org/	Platform for the acquisition and dissemination of Earth observation space data; optical data with very high spatial resolution. Remote sensing	national	France	fr	NA	NA	http://www.theia-land.art-geodev.fr/infrastructures-partenaires/metacatalogue-dinamis/	NA	NA	NA
149	SRTM	Diverse	http://srtm.csi.cgiar.org/srtmdata/	Consortium for Spatial Information (CGIAR-CSI).	global	-	en	free access	Geo TIFF, Esri ASCII	https://cgiarcsi.community/		NA	a.jarvis@cgiar.org
150	LTER France Data Node	Diverse	http://meta.data-za.org/geonetwork/srv/fre/catalog.search#/home	RI RZA. NA	national	France	fr/en	free access Metadata	various	NA	NA	NA	virginie.girard@univ-grenoble-alpes.fr
151	Google Earth Engine	Diverse	https://earthengine.google.com/	Analysis system and library of remote sensing data. Multiple, primarily remote sensing but also GIS data layers. EROS (USGS/NASA) Landsat catalog, numerous MODIS datasets, Sentinel-1 data, NAIP data, precipitation data, sea surface temperature data, CHIRPS climate data, and elevation data	global	-	en	free access upon registration	various	https://developers.google.com/earth-engine/datasets/catalog	NA	https://developers.google.com/earth-engine/guides	NA
152	PANGAEA	Diverse	https://www.pangaea.de/	Georeferenced data related to earth science fields. Our services are generally open for archiving, publishing, and re-usage of data. The World Data Center PANGAEA is member of the World Data System.	global	-	en	https://www.pangaea.de/about/privacypolicy.php	Open access	NA	Dublin Core, DIF, ISO 19115, XML, Java	NA	https://www.pangaea.de/contact/
153	BD Forêt V2	Terrestrial	https://professionnels.ign.fr/bdforet	National organism (IGN). Geographical reference system for the description of forest species	national	France	fr	free access (licence Etalab 2.0)	shapefile	https://professionnels.ign.fr/documentation/donnees/vecteur/bdforet	Unkown	https://professionnels.ign.fr/sites/default/files/2021-07/IGNF_BDFORETr_2-0.html	virginie.girard@univ-grenoble-alpes.fr

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
154	SapFluxNet Database	Terrestrial	https://zenodo.org/record/3971689#.YGxkOugzbcs	SapFluxNet contains globally harmonized data on whole-plant transpiration from sap flow measurments. Globally distributed datasets with sap flow time series for 2714 plants, mostly trees, of 174 species	european	-	en	freely available from the Zenodo repository	sapfluxnetr R package, designed to access, visualise and process SAPFLUXNET data is available from CRAN	http://sapfluxnet.creaf.cat/shiny/sfn_progress_dashboard/	NA	NA	r.poyatos@creaf.uab.cat
155	EFAS snow equivalent	Terrestrial	https://www.efas.eu/en/data-access	This data set contains for the whole of Europe two layers on snow water equivalent. One based on satellite information and the other one a hydrological model output.	european	-	en	https://www.efas.eu/en/terms-and-conditions	NA	https://confluence.ecmwf.int/display/COPSRV/European+Flood+Awareness+System	NA	https://confluence.ecmwf.int/display/COPSRV/European+Flood+Awareness+System	peter.salamon@ec.europa.eu
156	NitroEurope IP	Terrestrial	http://www.nitroeurope.ceh.ac.uk/ https://www.peer.eu/projects/peer-flagship-projects/nitroeurope	NitroEurope database contains data on N fluxes, net GHG exchanges, C:N pool.	european	-	en	free access upon registration	(http://www.nitroeurope.ceh.ac.uk/sites/nitroeurope.ceh.ac.uk/files/neu_data/Component7/web/NEU_Data_Policy.pdf)	NA	NA	NA	Dr. Mark Sutton Telephone: +44 (0)131 4454343 Email: ms@ceh.ac.uk
157	INI-International Nitrogen Initiative	Terrestrial	https://initrogen.org/	NA	global	-	en	NA	NA	https://www.inms.international/ https://futureearth.org/	NA	https://initrogen.org/rss.xml	raghuram98@hotmail.com; europe@initrogen.org (Dr Kevin Hicks)
158	Global Partnership on Nutrient Management	Terrestrial	https://www.unep.org/science-data	UNEP employs seven interlinked subprogrammes for action: Climate Action, Chemicals and Pollutions Action, Nature Action, Science Policy, Environmental Governance, Finance and Economic Transformations and Digital Transformations. The UN Environment Programme offers more than 15,000 items, from real-time data tools and platforms to key reports, publications, fact sheets, interactives and more.	global	-	en	https://www.unep.org/privacy	NA	https://wesr.unep.org/ https://www.unep.org/resources/frontiers-2022-noise-blazes-and-mismatches https://www.unep.org/data-resources	NA	NA	NA
159	SITES	Terrestrial	https://www.fieldsites.se/sv-SE	SITES is a national infrastructure for ecosystem research that facilitates long-term, world-class field based ecosystem research by offering a unique infrastructure and expertise to the Swedish and international research community.	national	Sweden	en	https://data.fieldsites.se/licence	Open ressource	https://www.fieldsites.se/en-GB/sites-thematic-programs-32634372	NA	https://data.fieldsites.se/portal/	holger.villwock@slu.se info@fieldsites.se
160	SRDB v4.0 Global Database of Soil Respiration	Terrestrial	https://daac.ornl.gov/SOILS/guides/SRDB_V4.html	The database provides soil respiration data from 1458 published studies with measurements taken between 1961 and 2016. Annual soil respiration (gC/m2/yr), Mean seasonal soil respiration (umol/m2/s) Annual or seasonal partitioning of soil respiration sources Q10 and associated temperature range R10 (soil respiration at 10 degrees C)	global	-	en	open access upon registration through the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC)	.csv format	https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1578	NA	NA	uso@daac.ornl.gov
161	Critical Zone Observatories	Terrestrial	https://czo-archive.criticalzone.org/national/	Diverse datasets across disciplines and spatial and temporal scales. Stream Water Chemistry - Cations, Anions, Metals (1982-2015) Climate, Flux Tower, Streamflow / Discharge - CUAHSI WDC web services (1968-2015) Soil Gas - CO2 and O2 (2014-2017) LiDAR, Land Cover, GIS/Map Data - OpenTopography (2010-2017) Air Temperature, Flux Tower, Meteorology - NADP and NOAA or other weather stations (2017)	national	US	en	https://czo-archive.criticalzone.org/national/data/czo-data-policies/	diverse formats, including CSV, Excel, and JPG. Open access	https://www.hydroshare.org/community/1/	NA	NA	bill@eps.berkeley.edu

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
				Flux Tower - AmeriFlux Network data (2007-2018) Streamflow / Discharge - USGS and USDA Data Resources (1985-2017) Water Budget Data Products - Precip, Throughfall,...									
162	ECN	Terrestrial / Aquatic	http://data.ecn.ac.uk/ https://catalogue.ceh.ac.uk/eidc/documents	In-situ monitoring datasets from the Environmental Change Network (ECN) in the United Kindom. The ECN Data Centre manages data for five UK integrated environmental monitoring networks: The UK Environmental Change Network, The UK Environmental Change Biodiversity Network, The Upland Waters Monitoring Network, The UK Lake Ecological Observatory Network, Habitats Monitoring Network. Together these networks cover 137 sites where a variety of physical, chemical and biological measurements are made, using a variety of data collection methods - from field sampling to automated telemetry. The datasets covers a broad range of observational data from long term monitoring sites. Most of the sites are also registered for eLTER. Freshwater Chemistry Freshwater Macro-invertebrates Laboratory quality control samples Freshwater Diatoms Soil Solution Chemistry Invertebrates: Moths Atmospheric Chemistry: NO2 Freshwater Flow Freshwater Macro-invertebrates Invertebrates: Spittle Bug Nymphs Freshwater Macrophytes Freshwater Zooplankton Surface Water: Chemistry Freshwater Phytoplankton Freshwater Chemistry Invertebrates: Ground Predators Vertebrates - Mammals: Rabbits Invertebrates: Butterflies Vertebrates - Birds: Moorland Birds Vertebrates - Birds: Breeding Birds Survey Invertebrates: Tipulids Vegetation - Semi-natural: Coarse-grain monitoring Vegetation - Managed: Permanent pasture Vegetation - Semi-natural: Baseline Invertebrates: Spittle bug Adults Precipitation Chemistry Vertebrates - Birds: Common Birds Census Soil Characterisation and Change: Coarse-grain monitoring Meteorology: Manual Met Office recording Vertebrates - Amphibians: Frogspawn Vegetation - Semi-natural: Woodland monitoring Meteorology: Automatic Weather Station recording Vegetation - Managed: Hedgerows and field margins Soil Survey and Classification: Baseline Vegetation - Semi-natural: Fine-grain monitoring Surface Water: Discharge Vertebrates - Mammals: Bats Soil Characterisation and Change: Fine-grain monitoring Vegetation - Managed: Cereals Freshwater Automatic Measurements	national	United Kingdom	en	http://data.ecn.ac.uk/datapolicy.asp	csv	http://data.ecn.ac.uk/request_form.asp	NA	NA	srennie@ceh.ac.uk

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
163	FLUXNET	Terrestrial	https://fluxnet.org/	Briefly, at each tower site, the eddy covariance method is applied to quantify the fluxes of scalars (e.g., CO2, CH4, water vapor) and energy (e.g., sensible, latent heat) between the biosphere and atmosphere. In addition, continuous measurements of ancillary physical variables (e.g., air temperature, precipitation, radiation) are acquired from a large number of sensors at high temporal resolution. The half-hourly or hourly fluxes are calculated and quality-controlled by the local tower teams. The data are then transferred to the Regional Networks and the FLUXNET Data Portal. In this process, the Regional Network and FLUXNET teams standardize the data format, perform uniform data quality checks, and produce value-added products using highly vetted gap-filling and flux partitioning software developed by the European team and the AmeriFlux team (see data processing). The processed and standardized data are then archived and prepared for querying, distributing, and downloading. A list of data services available via FLUXNET are here. Greenhous gas flux data	global	-	en	https://fluxnet.org/data/data-policy/	csv	https://fluxnet.org/data/data-services-and-products/	NA	NA	fluxdata-support@fluxdata.org
164	CORINE LANDCOVER (CLMS)	Terrestrial	https://land.copernicus.eu/pan-european/corine-land-cover	The CORINE Land Cover (CLC) inventory was initiated in 1985 (reference year 1990). Updates have been produced in 2000, 2006, 2012, and 2018. It consists of an inventory of land cover in 44 classes. CLC uses a Minimum Mapping Unit (MMU) of 25 hectares (ha) for areal phenomena and a minimum width of 100 m for linear phenomena. The time series are complemented by change layers, which highlight changes in land cover with an MMU of 5 ha. Different MMUs mean that the change layer has higher resolution than the status layer. Due to differences in MMUs the difference between two status layers will not equal to the corresponding CLC-Changes layer. If you are interested in CLC-Changes between two neighbour surveys always use the CLC-Change layer. Land cover	european	-	en	open access	gdb, shp	https://land.copernicus.eu/pan-european/corine-land-cover	INSPIRE	NA	EEA
165	Ecosystem types of Europe	Terrestrial	https://www.eea.europa.eu/data-and-maps/data/ecosystem-types-of-europe-1	The dataset combines the Copernicus land service portfolio and marine bathymetry and seabed information with the non-spatial EUNIS habitat classification for a better biological characterization of ecosystems across Europe. As such it represents probabilities of EUNIS habitat presence for each MAES ecosystem type. Ecosystem types	european	-	en	open access	gdb, shp	https://www.eea.europa.eu/data-and-maps/data/ecosystem-types-of-europe-1	INSPIRE	NA	EEA
166	CDDA (National designated areas)	Terrestrial	https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-15	The European inventory of nationally designated protected areas holds information about designated areas and their designation types, which directly or indirectly create protected areas. This is version 18 and covers data reported until March 2020. Protected areas	european	-	en	open access	gdb, shp, csv	https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-15	INSPIRE	NA	EEA
167	Land Use Harmonisation 2	Terrestrial	https://luh.umd.edu/	UH2 v2e Release (02/22/19): LUH2 v2e extends three harmonized future land-use forcing datasets (from LUH2 v2f) for the years 2100-2300. All land-use states and management variables are held constant at year 2100 values during these extension years, and as a result almost all transitions between land-use states are set to zero, with the exception of crop rotations, shifting cultivation, and wood harvest (which uses year 2099 national wood harvest demands for all years from 2100 to 2299). LUH2 v2e is now available for SSP1 RCP2.6, SSP5 RCP3.4OS, and SSP5 RCP8.5. Land use	global	-	en	open access	gdb, shp, csv	https://luh.umd.edu/index.shtml	not found	NA	Louise Chini Assistant Research Professor (301) 405-4050 lchini@umd.edu George Hurtt Professor (301) 405-8541 gchurtt@umd.edu Steve Frolking

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
													Research Professor (603) 862-0244 steve.frolking@unh.edu Ritvik Sahajpal Assistant Research Professor (301) 832-4118 ritvik@umd.edu
168	EuroCordex	Meterology	https://cordex.org/?option=com_content&view=featured&Itemid=476 and https://cordex.org/domains/cordex-region-euro-cordex/	EURO-CORDEX is the European branch of the international CORDEX initiative, which is a program sponsored by the World Climate Research Program (WRCP) to organize an internationally coordinated framework to produce improved regional climate change projections for all land regions world-wide. The CORDEX-results will serve as input for climate change impact and adaptation studies within the timeline of the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) and beyond. Climate variables (https://is-enes-data.github.io/CORDEX_variables_requirement_table.pdf)	european	-	en	open access	netDCF	https://euro-cordex.net/	not found	NA	gerics-it@hzg.de
169	ESGF	Meterology	https://esgf-data.dkrz.de/projects/esgf-dkrz/	Climate models DKRZ, the German Climate Computing Center, provides these tools and the associated services which are needed to investigate the processes in the climate system: Computer power, data management, and guidance to use these tools efficiently.	global	-	en	https://esgf-data.dkrz.de/projects/esgf-dkrz/privacy	NA	NA	Built with Sphinx	NA	ESGF esgf-user@lists.llnl.gov
170	PhenoCam	Terrestrial	https://phenocam.sr.unh.edu/webcam/	Vegetation phenology (seasonal changes) by recording time-lapse images of a fixed scene	global	-	en	open access upon registration https://phenocam.sr.unh.edu/webcam/fairuse_statement/	.jpg, .csv,	https://phenocam.sr.unh.edu/webcam/accounts/login/?next=/webcam/network/download/	proprietary	https://phenocam.sr.unh.edu/api/	andrew.richardson@nau.edu; phenocam@nau.edu
171	GBIF Api	Diverse	https://www.gbif.org/	Species occurence GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.	global	-	en	https://www.gbif.org/terms/privacy-policy	NA	NA	https://www.gbif.org/publishing-data	https://www.gbif.org/developer/summary	helpdesk@gbif.org
172	BioTime	Diverse	https://biotime.st-andrews.ac.uk/	BioTIME is a comprehensive collection of assemblage time-series in which the abundances of the species that comprise ecological communities have been monitored over a number of years. BioTIME data span the globe and encompass land and seas; they also include freshwater systems. The current version of BioTIME contains over 12 million records, features almost 50 thousand species, covers over 600 thousand distinct geographic locations and is representative of over 20 biomes, occurring over 6 different climatic zones. BioTIME follows the guiding principles of FAIR data (Findable, Accessible, Interoperable, Re-usable). Species observations	global	-	en	https://biotime.st-andrews.ac.uk/contribute.php	csv (data and metadata)	https://biotime.st-andrews.ac.uk/aboutUs.php	proprietary	none	biotimeproj@st-andrews.ac.uk

ID	Data source	Modalities	Web link	Description	Scales of data collection	Country	Language	Data policy	Data format & access	Web link (access)	Metadata Standard	Catalogue endpoint (API, CSW)	Contact
173	ECAD dataset (e-obs gridded dataset)	Meterology	https://www.ecad.eu/dailydata/index.php	E-OBS comes as an ensemble dataset and is available on a 0.1 and 0.25 degree regular grid for the elements daily mean temperature TG, daily minimum temperature TN, daily maximum temperature TX, daily precipitation sum RR, daily mean sea level pressure PP, daily mean wind speed FG and global radiation QQ. They cover the area: 25N-71.5N x 25W-45E. The data files are in NetCDF-4 format. The Global 30 Arc-Second Elevation Data Set (GTOPO30), a global raster Digital Elevation Model (DEM) with a horizontal grid spacing of 30 arc seconds (approximately 1 kilometer) developed by USGS is used for the elevation file as well. Daily observations at meteorological stations	european	-	en	https://www.ecad.eu/documents/ECAD_datapolicy.pdf	netCDF-4, registration for download needed.	https://www.ecad.eu/download/ensembles/download.php	proprietary	NA	eca@knmi.nl
174	ECAD daily data	Meterology	https://www.ecad.eu/dailydata/index.php	The ECA dataset contains series of daily observations at meteorological stations throughout Europe and the Mediterranean. Part of the dataset is freely available for non-commercial research and education: see our data policy for more details. To download this part of the data select one of the options below. The non-downloadable series may be available from the data provider directly, as well as additional data. Please direct your inquiries to obtain these data to the ECA&D Project Team. Note that a gridded version with daily temperature, precipitation and pressure fields is also available, as well as a predefined set of aggregated indices data. Daily observations at meteorological stations	european	-	en	https://knmi-ecad-assets-prd.s3.amazonaws.com/documents/ECAD_datapolicy.pdf	csv (zipped), direct download	https://www.ecad.eu/dailydata/index.php	proprietary	NA	eca@knmi.nl
175	OBIS API	Aquatic	https://obis.org/	OBIS aims to to be the most comprehensive gateway to the world's ocean biodiversity and biogeographic data and information required to address pressing coastal and world ocean concerns. It builds and maintains a global alliance that collaborates with scientific communities to facilitate free and open access to, and application of, biodiversity and biogeographic data and information on marine life. Species occurence	global	-	en	https://obis.org/manual/policy/	Darwin Core and Darwin Core Archive and dataset structure (https://obis.org/manual/contribute/)	https://api.obis.org/ https://obis.org/data/access/ https://mapper.obis.org/	EML 2.0	NA	https://obis.org/contact/
176	iNaturalist	Diverse	https://www.inaturalist.org/observations	Species occurence	global	-	en	https://www.inaturalist.org/pages/how+can+i+use+it	JSON/JSONP and PNG	https://www.inaturalist.org/users/sign_in	proprietary	https://api.inaturalist.org/v1/docs/	help@inaturalist.org

Annex 4 Assignment of data sources to eLTER SO

The following table represent the mapping of the provided data by the data sources to the eLTER SO (Zacharias et al., 2021). Not for all varibles in the input data source, a mapping was possible. The codes 'NA' and 'various' reflect this. If no appropriate EI concept could be found and the variable has been identified as important, e.g. based on the requirements from the science cases in the project, the mapping is indicated by 'other'.

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
1	ICP FORESTS	Level1: annual assesment of crown condition (defoliation, discolouration, demages visible on the threes). Assessment of soil condition (soil chemistry), Foliar survey (foliar chemistry) Level II: investigate the causes/effects of different stress factors on forest ecosystems	Crown condition	Matter budget	Atmospheric deposition	NA			annually	
			Foliar chemistry	Matter budget	Biomass	Leaf C, N, K, P, Ca, Mg, Mn content			every 2 years	
			Soil chemistry	Abiotic characteristic	Soil	Soil inventory; Soil organic C content (per horizon); Soil total N content (per horizon); Soil total P content (per horizon); Soil pH (in H2O/KCl/CaCl2); Soil cation exchange capacity; Soil base saturation			every 10 years	
			Tree growth	Energy budget	Biomass	Net primary production (dendrometer); Gross primary production; Aboveground vegetation growth; Aboveground biomass			every 5 years	
			Ground vegetation	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)			every 5 years	
			Stand structure incl. Deadwood	Energy budget	Biomass	Aboveground biomass			test phase	
			Epiphytic lichens	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)			test phase	
			Soil solution chemistry	Matter budget	Soil water	Soil water - pH value; Soil water - Conductivity; Soil water - NH4-N, NO3-N, DON concentration;Soil water - DOC concentration; Soil water - P concentration; Soil water - Cation concentrations; Soil water - Anion concentrations; Soil water - NH4-N, NO3-N, DON leaching; Soil water - DOC leaching			continously	
			Atmospheric deposition	Matter budget	Atmospheric deposition	Atmospheric Deposition - Bulk NH4-N, NO3-N, Ntot, P, K deposition in precipitation; Atmospheric Deposition -Bulk pH, anion, cation deposition in precipitation; Atmospheric Deposition -Bulk NH4-N, NO3-N, Ntot, P, K deposition in canopy throughfall (forests); Atmospheric Deposition -Bulk pH, anion, cation deposition in canopy throughfall (forests);Atmospheric Deposition -Stemflow NH4-N, NO3-N, Ntot, P, K, pH, cation, anion deposition in stemflow (forests);Atmospheric Deposition -Dry deposition of N-components			continously	
			Ambient air quality	Matter budget	Atmospheric deposition	NA			continously	
			Meteorology	Abiotic characteristic	Climate	Precipitation, Air temperature			continously	
			Pheneology	Biotic heterogeneity	Terrestrial	Plant phenology			several times per year	
			Litterfall	Matter budget	Biomass	Aboveground litterfall (forests)			continously	
2	ICP VEGETATION	in-situ measurments of impacts of ozone pollution on vegetation (since 1996) & atmospheric deposition of heavy metals (mk kg-1), nitrogen (% mass), persistent organic pollutants to vegtation		Matter budget	Atmospheric deposition	Atmospheric Deposition - Bulk NH4-N, NO3-N, Ntot, P, K deposition in precipitation; Atmospheric Deposition -Bulk pH, anion, cation deposition in precipitation; Atmospheric Deposition -Bulk NH4-N, NO3-N, Ntot, P, K deposition in canopy throughfall (forests); Atmospheric Deposition -Bulk pH, anion, cation deposition in canopy throughfall (forests);Atmospheric Deposition -Stemflow NH4-N, NO3-N, Ntot, P, K, pH, cation, anion deposition in stemflow (forests);Atmospheric Deposition -Dry deposition of N-components			every 5 years	
3	ICP Integrated Monitoring	in-situ monitoring of effects of air pollutants and climate change on different ecosystem compartmants at the catchment level (10-	Meteorology	Abiotic characteristic	Climate	Precipitation, Air temperature			continously/daily	
			Air chemistry	Matter budget	Atmospheric deposition	Dry deposition of N-components			continously/daily/weekly	

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
		1000 ha) with mandatory and optionals monitoring variables	Precipitation chemistry	Matter budget	Atmospheric deposition	Atmospheric Deposition - Bulk NH4-N, NO3-N, Ntot, P, K deposition in precipitation; Atmospheric Deposition -Bulk pH, anion, cation deposition in precipitation			daily/weekly/monthly	
			Throughfall	Matter budget	Atmospheric deposition	Atmospheric Deposition -Bulk NH4-N, NO3-N, Ntot, P, K deposition in canopy throughfall (forests);Atmospheric Deposition -Bulk pH, anion, cation deposition in canopy throughfall (forests)			weekly/monthly	
			Soil chemistry	Abiotic characteristic	Soil	Soil inventory; Soil organic C content (per horizon); Soil total N content (per horizon); Soil total P content (per horizon); Soil pH (in H2O/KCl/CaCl2); Soil cation exchange capacity; Soil base saturation;			every 5 years	
			Soil water chemistry	Matter budget	Soil water	Soil water - pH value; Soil water – Conductivity; Soil water – Percolation; Soil water - NH4-N, NO3-N; DON concentration; Soil water - DOC concentration; Soil water - P concentration; Soil water - Cation concentrations; Soil water - Anion concentrations; Soil water - NH4-N, NO3-N, DON leaching; Soil water - DOC leaching			weekly	
			Runoff water chemistry	Matter budget	Streams/River	TP, SRP, NO3, DOC, SAC 254			daily/weekly/monthly	
			Foliage chemistry	Matter budget	Biomass	Leaf C, N, K, P, Ca, Mg, Mn content			yearly	
			Vegetation	Biotic heterogeneity	Terrestrial	Habitat structure, vegetation/plant phenology based on satellite remote sensing (European scale)			1-5 year	
			Trunk epiphyhtes	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)			1-5 year	
			Metal chemistry of mosses	Matter budget	Atmospheric deposition	NA			every 5 years	
			Stemflow	Matter budget	Atmospheric deposition	Atmospheric Deposition -Stemflow NH4-N, NO3-N, Ntot, P, K, pH, cation, anion deposition in stemflow (forests)			weekly/monthly	
			Groundwater chemistry	Matter budget	Groundwater	Groundwater - Nutrient concentration: TP, SRP, TDN, NO3, NO2, NH4, DOC, DIC; Groundwater - Major ion concentrations: Cl, SO4, Br, Na, K, Mg, Ca, B; Groundwater - DOM composition			2 monthly	
			Lake water chemistry	Matter budget	Lake	Lake - Vertical profiles of major ion concentrations: Cl, SO4, Na, K, Mg, Ca			2-6 monthly	
			Hydrobiology of streams	Biotic heterogeneity	Streams/River	NA			6 monthly	
			Hydrobiology of lakes	Biotic heterogeneity	Lake	NA			monthly/2m (in spring/autmn)	
			Forest damage	Other	Terrestrial	NA			yearly	
			Tree bioelements and tree indication	Other	Terrestrial	NA			every 5 years	
			Vegetation structure and species cover	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)			every 10-20 years	
			Aerial green algae	Other	Terrestrial	NA			yearly	
			Microbial decomposition	Other	Terrestrial	Microbial activity (e.g. ATP conc.)			yearly	
			Toxicity assessment	Other	Terrestrial	NA				
			Inventory of birds	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using voice recording			every 3-5 years	
			Phneology	Biotic heterogeneity	Terrestrial	Plant phenology			daily/weekly	
4	ICP Modelling and Mapping	national data submitted by national focla centers & critical loads calculated based on European background data base Air pollution effects, risks and trends	Critical loads, Critical levels (0.5x0.25 in the Long/Lat System)	Matter budget	Atmospheric deposition	Dry deposition of N-components			every 3-5 years	
5	INTERACT International Network for Terrestrial Research and Monitoring in the Arctic	glaciology, permafrost, climate, ecology, biodiversity and biogeochemical cycling	Multiple - no standard.	Water balance	Terrestrial	Snow cover, Snow density			Various- no standard.	Various- no standard.
				Abiotic characteristic	Terrestrial	Climate				
				Biotic heterogeneity	Terrestrial	Various				
				Matter budget	Terrestrial	Various				
				Water balance	Terrestrial	Ice cover				

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
				Water balance	Terrestrial	Snow cover				
				Water balance	Terrestrial	Snow denisty				
6	GEM Greenland ecosystem monitoring	in-situ monitoring of climate warming on arctic ecosystems in Greenland. It includes 5 programs providing data on meteorology & hydrology, geomorphology, flux monitoring (CO_2, CH_4, H_2O), soil properties, snow properties, glaciar surface mass balance, monitoring of flora, arthropod, birds, mammals, and freshwater biotic and abiotic dynamics as well as physical, chemical and biological data from coastal zone		Abiotic characteristic	Terrestrial/Aquatic	Climate			yearly	3 stations (Disko, Nuuk, Zackenberg)
				Matter budget	Net Ecosystem Exchange	CH_4 flux, CO_2 flux			yearly	3 stations (Disko, Nuuk, Zackenberg)
				Abiotic characteristic	Soil	Various			yearly	3 stations (Disko, Nuuk, Zackenberg)
				Water balance	Terrestrial/Aquatic	Various			yearly	3 stations (Disko, Nuuk, Zackenberg)
				Biotic heterogeneity	Terrestrial/Aquatic	Various			yearly	3 stations (Disko, Nuuk, Zackenberg)
				Other	Terrestrial/Aquatic	Various			yearly	3 stations (Disko, Nuuk, Zackenberg)
7	AMAP-Arctic monitoring and Assessment Programme working Group	Atmospheric contaminats data Marine contaminant data radioactivity data freshwater & terrestrial contaminants datasets from arctic and subarctic areas Data are stored in diverse databases		Matter budget	Groundwater	Groundwater - Radioactive Isotopes (14C, T/3He, T)				
8	EMEP - European Monitoring and Evaluation Programme	collect emission data, measurments of air and precipitation quality, modelling of atmospheric transport and deposition of air pllutions	depostion of sulphur, nitrogen, heavy metals, POPs	Matter budget	Atmospheric deposition	Atmospheric deposition			year-month-day-hour EMEP 0.1º x 0.1º longitude-latitude grid	
9	ACTRIS	detecting changes and trends in aerosols, clouds, and trace gases through in-situ and remote sensing measurments	135 different atmospheric variables	Matter budget	Atmospheric deposition	Various				
10	EBAS	atmospheric composition data	atmospheric composition measurements from fixed sites or mobile platforms, 71 different countries, 1060 stations, 608 component types, 23 matrix types, 94 instrument types	Matter budget	Atmospheric deposition	Various				
11	ICOS - Integrated Carbon Observation System	in-situ greenhouse gas data from atmosphere, ecosystems, oceans & meteorological parameters	CO_2, CH_4	Matter budget	Net Ecosystem Exchange	CO_2 flux; CH_4 flux	2008	present		140 measurement stations
12	GLORIA	mapping of vascular plants, bryophytes, lichens in alpine ecosystems	species richness, plant species composition, abudance/cover, soil temperature, snow cover period	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)			5-10 years	Plot and mountain sumit
13	forestREplot	forest vegetation resurveys	plant abundance	Biotic heterogeneity	Terrestrial	Habitat structure, vegetation/plant phenology based on satellite remote sensing (European scale)			various	typically small regions with repeated plot records
14	Global soil partnership	collect data on soil organic carbon , soil pollution, soil erosion, soil biodiversity, salinity		Abiotic characteristic	Soil	Soil organic C content (per horizon);				
15	SoilBON	monitoring of temporal dynamics of soil biodiversity and function globally		Biotic heterogeneity	Soil	Various	2022	ongoing	annual	15x15m2
16	GCP Global Carbon Project	global budgets of carbondioxide, methane and nitous oxide	CO_2, CH_4, N_2O	Matter budget	Net Ecosystem Exchange	CO_2 flux; CH_4 flux; N_2O flux	2001	present	annual	global

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
17	TeaComposition	mass loss, soil data, climate	litter mass loss	Other	Soil	Microbial activity (e.g. ATP conc.)	2016	ongoing	annualy	1m2
18	LUCAS 2009 TOPSOIL data (ESDAC)	Soil samples have been analysed for the percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	Abiotic characteristic	Soil	Soil cation exchange capacity; Soil pH; Soil organic C content; Soil total N content; Soil total P content	2009	2012	1 sample/3 years cycle	2 km grid, 19.967 geo-referenced samples distributed in 25 EU countries; 0.5 kg of topsoil (0-20 cm)
19	LUCAS 2015 TOPSOIL data (ESDAC)	Soil samples have been analysed for the percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	percentage of coarse fragments, particle size distribution (% clay,silt and sand content), pH (in CaCl2 and H2O), organic carbon (g/kg), carbonate content (g/kg), phosphorous content (mg/kg), total nitrogen content (g/kg), extractable potassium content (mg/kg) , cation exchange capacity (cmol(+)/kg) and multispectral properties	Abiotic characteristic	Soil	Soil cation exchange capacity; Soil pH; Soil organic C content; Soil total N content; Soil total P content	2015	2018	1 sample/3 years cycle	2 km grid, 21.859 geo-referenced samples distributed in 28 countries; 0.5 kg of topsoil (0-20 cm)
20	ESDAC European Soil Data Centre	Soil point data, soil threats data, soil functions data		Abiotic characteristic	Soil	Soil cation exchange capacity; Soil pH; Soil organic C content; Soil total N content; Soil total P content; Belowgroud biomass			variuos	various
21	ESDAC			Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)				
22	Soil Geographical Database of Eurasia	Soil type	soil type according FAO nomenclatur	Abiotic characteristic	Soil	soil inventory				1:1,000,000
23	PASERA	modelled potential monthly soil erosion	soil erosion	Socio-Ecology	Land use and land cover change	NA			2000-2003	1km grid
24	BDAT	sample collectd by farmers	soil analyses linked to fertilisation (31 PC parameters as carbon, PH...)	Abiotic characteristic	Soil	Soil organic C, N, P content, Soil pH (in H2O/KCl/CaCl2), Soil cation exchange capacity, Fertilizer input (N, P, K fertilisation, liming, pesticides)			? (from 1990 to 2014)	
25	BASOL	sample of polluted soil	chemical parameters (arsenic, baryum, cadnium, cobalt, etc.)	Abiotic characteristic	Soil	Soil inventory				
26	DONESOL	auger drilling or soil pits, soil profile, horizon	type of soil unit, strata, map unit of soil	Abiotic characteristic	Soil	Soil inventory			variable	variable
27	Art.11 of habitat directive (FFH)	habitat types & species of community interest	Status and trends	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	1994	2018	period of 6 years	EU members (inside and outside Natura 2000 sites
28	EUNIS	species, habitat types and protected sites (Natura 2000) across Europe	Find species, habitat types and protected sites across Europe	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)				EU members/Europe
29	GBIF	occurence dataset	species, taxon	Biotic heterogeneity	Terrestrial	Mammals; vascular plants, Archaea, Bacteria, Fungi, Chromista				
30	EGDI European Geological Data Infrastructure	Onshore Geology, Marine Geology, Mineral Resources, Geohazards, Energy, Soil, Groundwater	geological features	Abiotic characteristic		Soil inventory				pan-european and national scale
31	OpenOBS	occurence dataset	species, habitat, taxon	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichen mosses/vegetation based on UAV remote sensing (local)				France, distributed
32	FungalRoot	Database contains global data on the type and intensity of mycorrhizal colonization of vascular plants, provides also information on soil chemical data	Mycorrhiza, soil properties	Biotic heterogeneity	Terrestrial	Various				

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
32	FungalRoot	Database contains global data on the type and intensity of mycorrhizal colonization of vascular plants, provides also information on soil chemical data	Mycorrhiza, soil properties	Abiotic characteristic	Soil	Soil inventory				
33	Global Ants Database	Data on ant species richness, abundance, composition and func-tional traits at the community level	ant diversity and traits	Biotic heterogeneity	Terrestrial	Various				
34	RMQS	survey on 2170 sites	C, soil microbial richness and diversity, Cd, Co, Cr,Cu, Mo, Ni, Pb, Tl, As, Hg, PAH, PCB, pesticides, dioxines, furanes	Biotic heterogeneity	Soil	Various			10-15 years	survey on 2170 16kmx16km grid
34	RMQS	survey on 2170 sites	C, soil microbial richness and diversity, Cd, Co, Cr,Cu, Mo, Ni, Pb, Tl, As, Hg, PAH, PCB, pesticides, dioxines, furanes	Abiotic characteristic	Soil	Various				
35	CRYOBS-CLIM	sample/observation	Ice, Meteo, Precipitation, Radiation, Snow	Water balance	Terrestrial	Ice cover; Precipitation; Snow cover;				
36	GLACIOCLIM	Metherological and glaciological data from alps, Andes, Antarctica, Himalayas; sample/observation	glaciology (Snow density)	Abiotic characteristic	Climate	Precipitation; Air temperature;				
36	GLACIOCLIM	Metherological and glaciological data from alps, Andes, Antarctica, Himalayas; sample/observation	glaciology (Snow density)	Water balance	Terrestrial	snow density				
37	The Farmland Bird Index	Along specified route at 10-20 counter points (with 300-400m bee-line distance between the counter points) two times a year the presence of farmland birds is recorded.	Populations of common birds of farmed land	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using voice recording	1980	2019	annualy	1km line assessmen
38	EuroBirdPortal	birds observation		Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using voice recording	2010	present	month - one year	30x30km
39	EBBC Atlas	observation	breeding birds location	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using voice recording	1971	1997		Europe
40	TOURBIERES	sample/observation of peatlands, 4 main focus: Meteorology and soil physics, greenhouse gas fluxes, hydrology and hydrochemistry, biodiversity	36 variables : atmo pressure, precipiation, wind direction, wind speed, air/soil temperature, air/soil moisture, soil heat flux, GW flux direction, COS/CH4 flux, pH, conductivity, discharge, plant population, green area index, etc.	Abiotic characteristic	Soil	Soil temperature				
41	ALTI	numerical model 3D	MNT, altitude	Abiotic characteristic	Terrestrial	NA			once	France (75x75m, 25m, 5m, 1m)
42	BDD link to Global Mammal Assessment	modeling	distribution based on habitat suitability, distribution based on scenarios of land cover and climate change, Values of Extinction Risk reduction Opportunity for mammals, Generation length for mammals, Retrospective extinction risk assessments for carnivores and ungulates	Biotic heterogeneity	Terrestrial	Mammals				
43	BDD link to Global Amphibian Assessment	modeling	distribution	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using acoustic recording				
44	MoveBank	gps records	specie movement	Biotic heterogeneity	Terrestrial	Birds; mammals;	2007	present		geo-referenced points
45	DRILOBase		earthworms	Biotic heterogeneity	Terrestrial	NA	2013	present		

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
46	SoilTemp	in-situ microclimate measurments of surface soil temperature (up to 10cm below the surface).	Soil temperature [C]	Abiotic characteristic	Soil	Soil temperature	2020	ongoing	hourly	global
47	International Soil Moisture Network	soil moisture from different networks	Soil moisture	Water balance	Soil	Soil water content			different temporal sampling intervals (daily average, hourly, several times per hour)	point
48	ESA CCI BIOMASS v3.0	global maps of above-ground biomass for four epochs (mid 1990s, 2010, 2017 and 2018)	Above-ground biomass from satellite observations	Energy budget	Biomass	Aboveground biomass	1990	2018	-	1 km
49	ESA CCI LAKES v1.1	Lake Water Level, Lake Water Extent, Lake Surface Water temperature, Lake Ice Cover and Thickness, and Lake Surface Reflectance from satellite observations	Lake Water Level, Lake Water Extent, Lake Surface Water temperature, Lake Ice Cover and Thickness, and Lake Surface Reflectance	Water balance	Lake	Water level	Sep 1992	Dec 2019	daily	per lake
50	ESA CCI LAND COVER v2.0.7	Land cover maps Land surface seasonality products all from satellite observations Global map of open water body Surface reflectance products	Land cover maps Land surface seasonality products all from satellite observations Global map of open water body Surface reflectance products	Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)	1992	2015	annual	300 m
51	ESA CCI SNOW	Daily global Snow Cover Fraction - snow on ground (SCFG) in percentage (%) per pixel Snow water equivalent (SWE)	Snow cover fraction Snow water equivalent	Water balance	Terrestrial	Snow cover	1982	present	daily	between 0.01° and 0.05°
52	ESA CCI SM v6.1	Surface soil moisture (~0-5cm of depth) from satellite observations	Surface soil moisture (~0-5cm of depth). Percentage of saturation [%] units for the ACTIVE product, and volumetric [m3m-3] units for the PASSIVE and COMBINED products	Water balance	Soil	Soil water content	November 1978 (PASSIVE and COMBINED product) January 1991 (ACTIVE product)	2020	daily	0.25°
53	Copernicus Global Land Services SSM	Surface Soil Moisture (%) from satellite observations	Surface soil moisture (%) from satellite observations	Water balance	Soil	Soil water content	2015	present	daily	1°/112, WGS 1984
54	Copernicus Global Land Services SWI	Soil Water Index (up to 5cm)	Soil Water Index [%]	Water balance	Soil	Soil water content	2015	present	daily	1 km
54	Copernicus Global Land Services SWI	Soil Water Index (up to 5cm)	Soil Water Index [%]	Water balance	Soil	Soil water content	2007	present	daily	Global, 0.1°, WGS 1984
56	CES OSO	interpretation of satellite images for land use	land use	Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)		present		10 and 20 m
57	GLASS	remote sensing	LAI / GPP	Energy budget	Biomass	Leaf area Index (LAI); Gross primary production	1981	2018	weekly	0.05°
58	SMOS Soil Moisture and Ocean Salinity	remote sensing	Soil moisture	Water balance	Soil	Soil water content	2010	present		30-50 km
59	MODIS	remote sensing	GPP, NPP	Energy budget	Biomass	Gross primary production; Net primary production	2000	present	daily, annual	500 m
60	Copernicus Global Land Services LAI300	remote sensing	LAI	Energy budget	Biomass	Leaf area Index (LAI)	2014	present	10-daily	300 m
61	Copernicus Global Land Services LAI 1km	remote sensing	LAI	Energy budget	Biomass	Leaf area Index (LAI)	1999	2020		1 km
62	GLIMS	remote sensing	glaciology	Water balance	Lake	Ice Cover	1850	2018	irregular	5-30 m
63	European Fluxes Database Cluster	In-situ greenhouse gas fluxes between ecosystems and the atmosphere, supporting meteorological, soil and vegetation data	CO2 flux, H2O flux, CH4 flux (Eddy covariance towers)	Matter budget	Net Ecosystem Exchange	Net Ecosystem Exchange CO2 flux	1996	present	different time resolutions, from halfhourly to yearly	plot scale

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
64	BDCHARM 50	maps	geological features, natural and industrial risks, Materials Observatory and marine Aggregates	Abiotic characteristic		Soil inventory				
65	BDIFF	observation of forest fire events from various sources (6 categories)		Socio-Ecology	Land use and land cover change	Various				
66	Open Government Data			Other	Terrestrial/Aquatic	various				
67	INSPIRE	observation	e.g. habitates and biotopes	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)				
67	INSPIRE	observation	e.g.atmopsheric conditions	Abiotic characteristic	Climate	Precipitation; Air temperature; Relative air humidity				
67	INSPIRE	observation	e.g. land use	Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)				
67	INSPIRE	observation	e.g. Soil	Abiotic characteristic	Soil	Soil inventory				
68	THEIA/OZCAR	NA	NA	Other	Terrestrial/Aquatic	NA			NA	NA
69	Pangea	georeferenced data		Biotic heterogeneity	Terrestrial	NA				
70	Zeonodo	multi-disciplinary		Biotic heterogeneity	Terrestrial	Various				
71	FigShare	multi-disciplinary data; e.g earth and evionmental science		Other	Terrestrial/Aquatic	Various				
72	VIGIE-NATURE	observation from citizen science project	occurence of Urban plant; Photomonitoring of pollinating insects; Butterfly; Snail ; Bumblebees ; Common birds; Winter Bird ; Dragonfly; Bat;Vascular Plants; Agricultural Biodiversity	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation				
72	VIGIE-NATURE	observation from citizen science project	occurence of Urban plant; Photomonitoring of pollinating insects; Butterfly; Snail ; Bumblebees ; Common birds; Winter Bird ; Dragonfly; Bat;Vascular Plants; Agricultural Biodiversity	Biotic heterogeneity	Terrestrial	Flying insects				
72	VIGIE-NATURE	observation from citizen science project	occurence of Urban plant; Photomonitoring of pollinating insects; Butterfly; Snail ; Bumblebees ; Common birds; Winter Bird ; Dragonfly; Bat;Vascular Plants; Agricultural Biodiversity	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers)				
73	Knowledge Base "Status"	from ~150 regulations in FR, information on habiates, species	status	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)				
74	COPERNICUS HRL Layers	land cover date for land cover elements: Imperviousness, forest, grassland, wettness/water and small woody features in 3-year time teps 2009-2012-2015-2018 to be continued; with 20m and since 2018 with 10m resolution	degree of soil sealing (percentage value 0-100)	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m
74	COPERNICUS HRL Layers		degree of crown cover density (percentage vlaue 0-100)	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m
74	COPERNICUS HRL Layers		dominant tree type (coniferous, decidous)	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m
74	COPERNICUS HRL Layers		grassland	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
74	COPERNICUS HRL Layers		water and temporary wet areas	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m
74	COPERNICUS HRL Layers		small woody features	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			every 3 years	10m (since 2018), before 20m
75	CLC+ work form MC	work in progress, to be pulbished 2021	land use classes according to EAGLE data model	Socio-Ecology	Land use and land cover change	Land cover (CORINE)			tbd.	variable, reference grid: 100*100m
76	CHELSA v2.1	globla air temperature and precipitation data; especially precipitation patterns are more accurate.	air temperature [C], precipitation (mm), Clouds, Radiation, DEM, land cover	Abiotic characteristic	Climate	Air temperature; Precipitation			daily, monthly means	30 arc-second ~1km
76	CHELSA v2.1			Abiotic characteristic	Climate	Precipitation				
77	WorldClim v2.1	Historical climate data (1970-2000)	air temperature [°C]	Abiotic characteristic	Climate	Air temperature	1970	2000	monthly	four spatial resolutions, between 30 seconds (~1 km2) to 10 minutes (~340 km2)
			precipitation [mm]	Abiotic characteristic	Climate	Precipitation				
			solar radiation [kJ m-2 day-1]	Abiotic characteristic	climate	NA				
			wind speed [ms-1]	Abiotic characteristic	climate	Wind speed / Wind direction				
			water vapor pressure [kPa]	Abiotic characteristic	climate	Relative air humidity				
		Historical monthly weather data (1960-2018)	average minimum temperature (°C), average maximum temperature (°C)	Abiotic characteristic	climate	Air temperature	1960	2018	monthly	2.5 minutes (~21 km2)
			total precipitation (mm)	Abiotic characteristic	Climate	Precipitation				
78	E-OBS	daily gridded weather data	daily mean temperature , daily minimum temperature, daily maximum temperature, daily precipitation sum, daily averaged sea level pressure and daily mean global radiation.	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1950	present	daily	0.1 and 0.25 degree grids
79	CHPclim	High-quality monthly rainfall climatology (in situ measurements + satellite)	precipitation	Abiotic characteristic	Climate	Precipitation			monthly	0.05 degree
80	GPCC	gridded gauge-analysis products derived from quality controlled station data	precipitation	Abiotic characteristic	Climate	Precipitation	1891	2018	Climatology, Daily, Monthly	0.5°, 1.0°, 2.5°
81	TRMM	tropical and subtropical precipitation	precipitation	Abiotic characteristic	Climate	Precipitation	1998	2019	Sub-daily, Daily, Monthly	0.25° , ~40S - 40N and ~50S - 50N
82	ERA-Interim	It has been superseded by the ERA5 reanalysis.	climate	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1979	2019		80 km
83	ERA5	estimation of atmospheric, land and oceanic climate variables	climate	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1979	present	hourly	30 km
84	DRIAS	atmospheric expriments (ex. EUROCORDEX climate projections) and impact experiments (ex. ADAMONT)	atmospheric variables (temperature, precipitation, number of frost days, maximum temperature record) and snowfall impact indices: snow depth, snow water equivalent, number of days with snow depth > thresholds, etc.	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity				
84	DRIAS	atmospheric expriments (ex. EUROCORDEX climate projections) and impact experiments (ex. ADAMONT)	atmospheric variables (temperature, precipitation, number of frost days, maximum temperature record) and snowfall impact indices: snow depth, snow water equivalent,	Water balance	Terrestrial	snow cower, snow density				

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
			number of days with snow depth > thresholds, etc.							
85	AERIS	NA	NA	Abiotic characteristic	climate	NA			NA	NA
86	HyMeX	HYdrological cycle in the Mediterranean EXperiment, geophysical in situ observations, satellite products, model outputs, questionnaire data. Searchable by parameters, instruments or events.	https://mistrals.sedoo.fr/HyMeX/Parameter-search/	Abiotic characteristic	climate	Relative air humidity	various	various	various	various
87	RADOME	Basic parameters (Air T, humidity, wind speed and direction, precipitation) and additional parameters according to the instrumentation (soil T, visibility, radiation)	Precipitatition (Kg/m2)	Abiotic characteristic	climate	Precipitation	-	present	hourly and/or intra-hourly time steps (6 min)	approximately 30 km (RADOME network 550 stations)
			Wind speed (m/s) & direction (degree)	Abiotic characteristic	climate	Wind speed / Wind direction	-	present	hourly and/or intra-hourly time steps (6 min)	approximately 30 km (RADOME network 550 stations)
			Air temperature (K)	Abiotic characteristic	Climate	Air temperature	-	present	hourly and/or intra-hourly time steps (6 min)	approximately 30 km (RADOME network 550 stations)
			Air relative humidity (%)	Abiotic characteristic	Climate	Relative air humidity	-	present	hourly and/or intra-hourly time steps (6 min)	approximately 30 km (RADOME network 550 stations)
91	NIVOSE	samples	Snow depth	Abiotic characteristic	Climate	Snow cover	-	present	daily	Pyrénées, Alps, Corse, 28 stations
			temperature	Abiotic characteristic	Climate	Air temperature	-	present	daily	Pyrénées, Alps, Corse, 28 stations
			humidity	Abiotic characteristic	Climate	Relative air humidity	-	present	daily	Pyrénées, Alps, Corse, 28 stations
			wind	Abiotic characteristic	Climate	Wind speed / Wind direction	-	present	daily	Pyrénées, Alps, Corse, 28 stations
95	ATMO-AASQA	station measurement data, emission data	air pollutants (major pollutants: fine particules (PM10, PM2,5), nitrogen oxide (NOx), sulfur dioxide (SO2), carbon monoxide (CO), ozone (O3), benzene (C6H6), lead (Pb), cadmium (Cd), arsenic (As), nickel (Ni), volatile organic compounds (COVM), ammonia (NH3))	Matter budget	Atmospheric deposition	Atmospheric Deposition -Dry deposition of N-components			hours, months, years - 5-year series depth	France (thousand of station)
96	ICP Waters	in-situ monitoring of effect of atmospheric pollution on surface water (lake, rivers) since 1985	water chemistry	Matter budget	Streams/Rivers/Lake	Lake - Vertical profiles of chl a, pigments; Lake - Vertical profiles of dissolved oxygen; Lake - In-situ vertical profiles and inflow concentrations of TP, SRP, NO3, DOC, SAC 254; Lake - In-situ vertical profiles and inflow concentrations TOC, POC, TN, NO2, NH4, SRSi, DIC; Rivers – Turbidity; Rivers -TP, SRP, NO3, DOC, SAC 254; Rivers -TOC, POC, TDN, NO2, NH4, SRSi, DIC; Rivers -Cl, SO4, Na, K, Mg, Ca				
			Invertebrates (emphemeroptera, plecoptera, trichoptera, turbellaria,	Biotic heterogeneity	Streams/Rivers/Lake	Macroinvertebrate community (quantitative) - River				

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
			mollusca, Hirudinea, Daphnia, Malacostrace)							
			fish communities	Biotic heterogeneity	Streams/Rivers/Lake	Fish community (quantitative) –River				
			trace metals in surface water	Matter budget	Streams/Rivers/Lake	NA				
			trace metals in sediments	Matter budget	Streams/Rivers/Lake	NA				
			POP in water	Matter budget	Streams/Rivers/Lake	NA				
			POP in sediments	Matter budget	Streams/Rivers/Lake	NA				
			trace metals and POP in fish	Matter budget	Streams/Rivers/Lake	NA				
97	Danubius-RI			Water balance	Streams/River	Discharge	2009			
98	GLEON -Global Lake Ecological Observatory Network			Other	Lake	NA				
99	MarineStrategyFrameworkDirective			Matter budget	Atmospheric deposition	NA				
100	Water Framework Directive	Provides data from the 1st and 2nd River Basin Management Plans reported by EU Members States, Norway and the United Kingdom according to article 13 of the Water Framework Directive (WFD). The database includes information about surface water bodies (number and size, water body category, ecological status or potential, chemical status, significant pressures and impacts, and exemptions) and about groundwater bodies (number and size, quantitative status, chemical status, significant pressures and impacts, and exemptions).		Abiotic characteristic	Streams/Rivers/Lake	Various				
				Biotic heterogeneity	Streams/Rivers/Lake	Various				
				Matter budget	Streams/Rivers/Lake	Various				
				Water balance	Streams/Rivers/Lake	Various				
101	TeaComposition H2O	Litter mass loss and microbial community composition	Litter mass loss and microbial community composition	Other	Streams/Rivers/Lake	Various	2017	ongoing	annaual	plot
102	SANDRE	referentials	administrative limits, works, fish zonation...	Biotic heterogeneity	Streams/River	Fish community (quantitative)			NA	NA
103	Waste water Treatment Plant (Urban Wastewater)	measurement station	annual waste water discharge, Sludge production, Maximum input load...	Water balance	Streams/River	Discharge			yearly since 2009	France, ~ 22000 stations
104	BRGM			Abiotic characteristic	Soil	Soil inventory			NA	France
				Water balance	Groundwater	Groundwater level			NA	France
106	WATERBASE -Water quality ICM	sample	chemistry	Matter budget		Cl, SO4, Na, K, Mg, Ca				Europe
107	WATERBASE -Biology	sample	biology	Biotic heterogeneity	Streams/Rivers/Lake	Fish community (quantitative) - Lake; Fish community (quantitative) –River				Europe
108	CCM (Catchment Characterisation and Modelling)	Hydrographic network (modeling)		Water balance	Streams/River	Discharge				
109	BANQUE HYDRO	station, measurement	discharge [m3/s], water depth [m]	Water balance		Discharge			variable (from hourly to daily or monthly)	France, 5000 hydrological stations (3200 active hydro. stations)
110	ADES	sensor and human based, analyses in lab	water depth [m], chemical parameters [micro-g/L]	Water balance	Groundwater	Groundwater level			?	France, 5049 stations, >79000 sample points with piezo, >75000 sample points for quality water

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
111	DATA EAU FRANCE-Fish	electric fishing for national survey	fish population and fish caracteristic	Biotic heterogeneity	Streams/Rivers/Lake	Fish community (quantitative)			annual	Results provided by region (x22)
112	DATA EAU FRANCE-Contaminants	contaminants in sediments, fish	contaminants (PCB-DL, PCDD/F, PCBi, Hg)	Matter budget	Streams/Rivers/Lake	NA			annual (2008, 2009 and 2010)	France,
113	NAIADES	sample for freshwater biodiversity, measurements ofr hydromorphology, lab analyses for PC, sensor for T)	hydrobiology (taxon), hydromoprhology (data per transect), water temperature, physico-chemical (data per sample)	Biotic heterogeneity	Streams/Rivers/Lake	various			annual	France
			hydrobiology (taxon), hydromoprhology (data per transect), water temperature, physico-chemical (data per sample)	Abiotic characteristic	Streams/Rivers	water temperature			annual	France
			hydrobiology (taxon), hydromoprhology (data per transect), water temperature, physico-chemical (data per sample)	Matter budget	Streams/Rivers	NA				
114	CYBERCAROTHEQUE	sample, and high-resolution multi-proxy (sedimentological, palynological, and geochemical) analysis	marine, lake (and soon continental and glacial) sedimentary records	Socio-Ecology	Land use and land cover change	Land use (historic)				Global; 1906 sedimentary records from 243 missions
115	ODATIS	In-situ observations as well as remote sensing data on marine biogeochemistry, geology, meteorology, phyiscal oceanography are collected.	biochemistry, biology, physics, geology (temperature, salinity, pressure, turbidity, CDOM, Chlorophyll, pH, dissolved oxygen, TSS, metals, nutrient salts, pigments, current, sea level, wave, weather, biology (plankton, toxin, benthic habitat...), coastline, bathymetry)	Other	Marine	various				
116	Sextant	data on marine ecosystems		Biotic heterogeneity	Marine	NA	1983	ongoing		
117	UWWTD	size, compliance of Urban WWTP		Socio-Ecology	Resource Use	NA				Europe
118	BioFresh	Freshwater Macro-invertebrates Laboratory quality control samples Freshwater Diatoms Soil Solution Chemistry Invertebrates: Moths Atmospheric Chemistry: NO2 Freshwater Flow Freshwater Macro-invertebrates Invertebrates: Spittle	biology info about fish, macrophytes, macrobenthos and zooplankton	Biotic heterogeneity	Streams/Rivers/Lake	Macroinvertebrate community (quantitative) - River, Macrophyte community (quantitative) - Lake, Zooplankton (quantitative) - Lake, Fish community (quantitative) - Lake,				Europe
			biology info about fish, macrophytes, macrobenthos and zooplankton	Matter budget	Soil water; Soil	Soil water - pH value, Soil water – Conductivity, Soil water - NH4-N, NO3-N, DON, DOC, P, cation and anion concentrations;				
			biology info about fish, macrophytes, macrobenthos and zooplankton	Water balance	Streams/Rivers/Lake	Discharge				
119	COPERNICUS	Lake surface water temperature (LSWT) in K	Lake surface water temperature (LSWT)	Abiotic characteristic	Climate	Relative air humidity	2002	present	10-daily	1 km
		Lake turbidity, trophic state index, lake surface reflectances	Lake turbidity, trophic state index, lake surface reflectances	Abiotic characteristic	Lake	Vertical profiles of water temperature, pH, EC, turbidity	2002	present	10-daily	300 m, 1 km
		Lake turbidity, trophic state index, lake surface reflectances	Lake turbidity, trophic state index, lake surface reflectances	Abiotic characteristic	Lake	Vertical profiles of water temperature, pH, EC, turbidity	2019	2020	10-daily	100 m
		Min and max extent of water bodies	Water bodies extent	Socio-Ecology	Land use and land cover change	Land use change (CORINE)	2014	present	Monthly	100 m, 300 m
		Rivers and lakes water level	Water level	Water balance	Lake	Water level	1992 (lakes), 2002 (rivers)	present	10-daily	per lake and river

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
124	ESA-CCI+	Data relevant to Atmospheric Sciences and Earth Observation fields; Lake water level, lake water extent, Lake surface water temperature, lake Ice cover,Lake water Leaving Reflectance	Turbidity, Water extent uncertainty, concentration of chlorophyll-a, lake ice cover, lake ice cover uncertainty, lake surface skin temperature, quality levels, water surface height above geoid, water surface height uncertainty	Water balance	Lake	Ice cover				
			Turbidity, Water extent uncertainty, concentration of chlorophyll-a, lake ice cover, lake ice cover uncertainty, lake surface skin temperature, quality levels, water surface height above geoid, water surface height uncertainty	Water balance	Lake	Water level; Inflow/outflow				
			Turbidity, Water extent uncertainty, concentration of chlorophyll-a, lake ice cover, lake ice cover uncertainty, lake surface skin temperature, quality levels, water surface height above geoid, water surface height uncertainty	Abiotic characteristic	Lake	Vertical profiles of water temperature, pH, EC, turbidity				
125	AQUACOSM	Pyhtoplantkon, Zooplanktion, Microbial Planktion, Periphyton, Water Chemistry,	dissolved inorganic nutrients,particulate organic matter, microplastic effect on food web	Matter budget	Lake	Lake - In-situ vertical profiles and inflow concentrations of TP, SRP, NO3, DOC, SAC 254				
			phytoplancton, zooplancton, prokaryotic community composition	Biotic heterogeneity	Lake	Zooplankton (quantitative), Algal community (quantitative)				
126	TERENO	Soil temperature and soil moisture at different depths (5, 20 and 50 cm). Plus precipitation, evapotranspiration, runoff and periodically also different chamber measurements were made to access soil or plant gas exchange.	Soil temperature (°C)	Abiotic characteristic	Soil	Soil temperature	2010	present	Monthly, annual	4 observatories in Germany
			Soil water content (%)	Water balance	Soil	Soil water content	2010	present	Monthly, annual	4 observatories in Germany
128	Droughtnet	Site and vegetation data, weather data		Abiotic characteristic	climate	Precipitation				
				Abiotic characteristic	climate	Air temperature				
				Abiotic characteristic	climate	Soil organic C content				
				Abiotic characteristic	climate	Soil total N content				
				Abiotic characteristic	soil	Soil pH (in H2O/KCl/CaCl2)				
				Socio-Ecology	Land use and land cover change	Land use				
				Biotic heterogeneity	Terrestrial	Habitat structure, vegetation/plant phenology				
129	TreeDivNet	Global data on the effect of tree species diversity on the ecosystem functioning	species richness, functional diversity, genetic diversity, phylogenetic diversity, phylogenetic evenes, insect herbivory, tea decomposition	Biotic heterogeneity	Terrestrial	Habitat structure, vegetation/plant phenology based on satellite remote sensing (European scale)				
130	NutNet	data on management, treatment, deposition etc. in approx.. 130 grassland sites on global scale.		Socio-Ecology	Land use and land cover change	Land cover (Orthophotos); Land use (historic)				
				Energy budget	Radiation Budget	PAR				
				Energy budget	Biomass	Aboveground biomass				
				Biotic heterogeneity	Terrestrial	Plant phenology				
				Abiotic characteristic	Soil	Soil inventory				
				Socio-Ecology	Agriculture and Forestry	Fertilizer input (N, P, K fertilisation, liming, pesticides)				
131	ANAEE-France	NA	NA	Other	Terrestrial/Aquatic	NA			NA	NA

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
132	OLA	high altitude lake		Water balance	Lake	Discharge; Water level				4 large lakes : lacs d'Annecy, du Bourget, Léman and Pavin; 20 small lake in Alps
				Abiotic characteristic	Lake	Vertical profiles of water temperature, pH, EC, turbidity				
133	F-ORE-T	flow data from EC towers (recently accumulation chambers), meteorological data and soil climate data		Matter budget	Net Ecosystem Exchange	Net Ecosystem Exchange CO2 flux (Gas flux chamber)	1992	present	various	10 sites, recofor network = 102 permanent plots (~2 hectares), cataneat network (atmo deposition) = 27 permanent plots
134	ACBB	temporary grassland		Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)				Lusignan (Poitou-Charentes)
135	PRO	environmental impacts of Organic Waste Products recycling on field crops at long time scale		Abiotic characteristic	Soil/climate	NA				5 sites (QualiAgro, Ile-de-France; EFELE, Ile-et-Villaine; La Bouzule, Lorraine; Couhins, Aquitaine; Colmar, Alsace), ~40plots of 450m2
136	EJP SOIL			Abiotic characteristic	Soil/climate	NA				
137	H+	sample/observation	groundwater level and quality	Water balance		Groundwater level				
138	CensusHub			Other	Terrestrial/Aquatic	NA				
139	INSEE	survey	working people, modes of transport to and from work, housing, jobs, families, nationality, education, households, population	Socio-Ecology	Population	NA			from 1876 to today, annual	France, municipality
140	SIDDT	indicators	(1) working people, modes of transport to and from work, housing, jobs, families, nationality, education, households, population (2) Permanent Equipment Base (BPE) (3) tourist accommodation capacity (4) Tax homes	Socio-Ecology	Population	NA			(1) 1990-2016, annual (2) BPE 2007-2018 (3) 1999-2019 (4) 1996-2011	France, municipality
140	SIDDT		agricultural practices (Livestock, crops, heads of holdings, otex, agricultural areas, etc.).	Socio-Ecology	Agriculture and Forestry	Grazing timing, intensity; livestock (livestock units)			1988-2000-2010	France, municipality and canton
141	AGRESTE	data from individuals link to farm/crop management	land use, farming pratices	Socio-Ecology	Agriculture and Forestry	Grazing timing, intensity; livestock (livestock units)				
142	CORINE LAND COVER	visual interpretation of satellite images for land use	land use	Socio-Ecology	Land use and land cover change	Land cover (CORINE)				
143	RPG (Parcel Registry Graphic)	Crops types declared by farmers to receive Common Agricultural Policy aids	Agricultural land use by plot	Socio-Ecology	Agriculture and Forestry	Harvest (cropland, grassland, forest) (t/ha); Agricultural products			yearly since 2013	Agricultural plot
144	ZNIEFF	observation, expertise	habitat characteristics, threats, issues	Biotic heterogeneity	Terrestrial	NA			since 30 years	

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
145	REMONTER LE TEMPS	scan of historical land cover (>1950)	land use	Socio-Ecology	Land use and land cover change	Land cover (Orthophotos)			variable according to the site	
146	CDDA	GIS	limits of natural parks	Biotic heterogeneity	Terrestrial	NA				
147	European Social Survey	Media, politics, wellbeing, identity, socio-demographic profile, values	Multiple quantitive and qualitative data sets	Socio-Ecology	Population	NA			Biannually since 2002	
148	DINAMIS	remote sensing		Socio-Ecology	Land use and land cover change	Land cover (CORINE)				
149	SRTM		MNT, altitude	Abiotic characteristic	Terrestrial	NA				(30m, 250m)
150	LTER France Data Node	NA	NA	Other	Terrestrial/Aquatic	NA			NA	NA
151	Google Earth Engine	multiple, primarily remote sensing but also GIS data layers. EROS (USGS/NASA) Landsat catalog, numerous MODIS datasets, Sentinel-1 data, NAIP data, precipitation data, sea surface temperature data, CHIRPS climate data, and elevation data	Climate and weather (surface temperature, climate, atmospheric, weather), satellite imagery (Landsat, Sentinel, MODIS, High-Resolution Imagery), geophysical data	Abiotic characteristic	Terrestrial	various	1984	present		
152	PANGAEA			Other	Terrestrial/Aquatic	NA				
153	BD Forêt V2	Geographical reference system for the description of forest species	forest and natural plant formations using a soil cover approach (cover density of the stand, its composition and the species) for elements over 5,000 m2	Energy budget	Biomass	NA	2007	2018	NA 'actual data from 2016	0.5 ha
154	SapFluxNet Databse	202 globally distributed datasets with sap flow time series for 2714 plants, mostly trees, of 174 species	Sapflow data, soil water content	Energy budget	Biomass	NA			sub-daily time series of sap flow and hydrometeorological drivers for one or more growing seasons. The datasets encompass the period between 1995 and 2018, with 50 % of the datasets being at least 3 years long.	
			Sapflow data, soil water content	Water balance	Soil	Soil water content				
155	EFAS snow equivalent		Snow equivalent	Water balance	Terrestrial	Snow cover				horizontal resolution is 5km
156	NitroEurope IP			Matter budget		CO2 flux (Gas flux chamber); CH4 flux (Gas flux chamber); N2O, NO, NOx flux (Gas flux chamber)	2006	2011		
157	INI-International Nitrogen Initiative			Matter budget		CO2 flux (Gas flux chamber); CH4 flux (Gas flux chamber); N2O, NO, NOx flux (Gas flux chamber)	2003			
158	Global Partnership on Nutrient Management			Other	Terrestrial/Aquatic	NA				
159	SITES	A multitude of long-term data collection is conducted within SITES Thematic Programms: 1) SITES Spectral: phenological observations 2) SITES Water: measurements of hydrological, physical, chemical, and biological parameters within lakes and streams 3) SITES AquaNet: replicated large-scale enclosure experiments	NA	Biotic heterogeneity	Terrestrial/Aquatic	Plant phenology				
			NA	other	Terrestrial/Aquatic	various				

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
160	SRDB v4.0 Global Database of Soil Respiration	Annual soil respiration (gC/m2/yr), Mean seasonal soil respiration (umol/m2/s) Annual or seasonal partitioning of soil respiration sources Q10 and associated temperature range R10 (soil respiration at 10 degrees C)	Soil respiration Q10	Matter budget	Net Ecosystem Exchange	Net Ecosystem Exchange CO2 flux	1961	2016	seasonal and annual fluxes	Point-based measurements
161	Critical Zone Observatories	Stream Water Chemistry - Cations, Anions, Metals (1982-2015) Climate, Flux Tower, Streamflow / Discharge - CUAHSI WDC web services (1968-2015) Soil Gas - CO2 and O2 (2014-2017) LiDAR, Land Cover, GIS/Map Data - OpenTopography (2010-2017) Air Temperature, Flux Tower, Meteorology - NADP and NOAA or other weather stations (2017) Flux Tower - AmeriFlux Network data (2007-2018) Streamflow / Discharge - USGS and USDA Data Resources (1985-2017) Water Budget Data Products - Precip, Throughfall,...		Matter budget	Net Ecosystem Exchange	Net Ecosystem Exchange CO2 flux (Gas flux chamber)	2014	2017		
162	ECN	Freshwater Chemistry Freshwater Macro-invertebrates Laboratory quality control samples Freshwater Diatoms Soil Solution Chemistry Invertebrates: Moths Atmospheric Chemistry: NO2 Freshwater Flow Freshwater Macro-invertebrates Invertebrates: Spittle Bug Nymphs Freshwater Macrophytes Freshwater Zooplankton Surface Water: Chemistry Freshwater Phytoplankton Freshwater Chemistry Invertebrates: Ground Predators Vertebrates - Mammals: Rabbits Invertebrates: Butterflies Vertebrates - Birds: Moorland Birds Vertebrates - Birds: Breeding Birds Survey Invertebrates: Tipulids Vegetation - Semi-natural: Coarse-grain monitoring Vegetation - Managed: Permanent pasture Vegetation - Semi-natural: Baseline Invertebrates: Spittle bug Adults Precipitation Chemistry Vertebrates - Birds: Common Birds Census Soil Characterisation and Change: Coarse-grain monitoring Meteorology: Manual Met Office recording Vertebrates - Amphibians: Frogspawn Vegetation - Semi-natural: Woodland monitoring Meteorology: Automatic Weather Station recording Vegetation - Managed: Hedgerows and field	Meteorology: Automatic Weather Station recording Meteorology: Manual Met Office recording	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1991	data products available until 2015	continous measurements based on measurement protocols, varying temporal resolution (continous, weekly, annual)	site based measurements
			Atmospheric Chemistry: NO2	Matter budget	Atmospheric deposition	NA	1993			
			Freshwater Automatic Measurements Freshwater Chemistry Freshwater Chemistry Laboratory quality control samples	Matter budget	Streams/Rivers/Lake	NA	1992			
			Freshwater Flow	Water balance		Discharge	1993			
			Precipitation Chemistry	Matter budget	Atmospheric deposition	Bulk NH4-N, NO3-N, Ntot, P, K deposition in precipitation; Bulk pH, anion, cation deposition in precipitation	1992			
			Soil Characterisation and Change: Coarse-grain monitoring Soil Characterisation and Change: Fine-grain monitoring Soil Survey and Classification: Baseline	Abiotic characteristic		Soil inventory	1992			
			Soil Solution Chemistry	Matter budget	Soil water	pH value; Conductivity; NH4-N, NO3-N, DON concentration; DOC concentration; P concentration; Cation concentrations; Anion concentrations	1992			
			Surface Water: Chemistry	Matter budget	Streams/Rivers/Lake	various	1992			
			Surface Water: Discharge	Water balance		Discharge	1993			
			Vegetation - Managed: Cereals Vegetation - Managed: Hedgerows and field margins	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	1991			

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
		margins Soil Survey and Classification: Baseline Vegetation - Semi-natural: Fine-grain monitoring Surface Water: Discharge Vertebrates - Mammals: Bats Soil Characterisation and Change: Fine-grain monitoring Vegetation - Managed: Cereals Freshwater Automatic Measurements	Vegetation - Managed: Permanent pasture Vegetation - Semi-natural: Baseline Vegetation - Semi-natural: Coarse-grain monitoring Vegetation - Semi-natural: Fine-grain monitoring Vegetation - Semi-natural: Woodland monitoring							
			Vertebrates - Amphibians: Frogspawn Vertebrates - Birds: Breeding Birds Survey Vertebrates - Birds: Common Birds Census Vertebrates - Birds: Moorland Birds Vertebrates - Mammals: Bats Vertebrates - Mammals: Rabbits	Biotic heterogeneity	Terrestrial	Birds, bats, frogs, some insects (e.g., grasshoppers) using acoustic recording	(1971) / 1993			
			Freshwater Macro-invertebrates Freshwater Macro-invertebrates Freshwater Macrophytes Freshwater Phytoplankton Freshwater Zooplankton Freshwater Diatoms	Biotic heterogeneity	Streams/Rivers/Lake	Macroinvertebrate community (quantitative); Macrophyte community (quantitative); Zooplankton (quantitative); Algal community (quantitative)	1992 (?)			
			Invertebrates: Butterflies Invertebrates: Ground Predators Invertebrates: Moths Invertebrates: Spittle bug Adults Invertebrates: Spittle Bug Nymphs Invertebrates: Tipulids	Biotic heterogeneity	Terrestrial	Flying insects	1992			
163	FLUXNET	Greenhous gas flux data	greenhouse gas flux	Matter budget	Atmospheric deposition	CO2 flux; CO2 concentration; CH4 flux; CH4 concentration	1996	2015	continous measurement	point measurement (flux tower)
164	CORINE LANDCOVER (CLMS)	Land cover	land cover	Socio-Ecology	Land use and land cover change	Land cover (CORINE)	1990	2018	6 years	25ha MMU / 100m
165	Ecosystem types of Europe	Ecosystem types	Habitat	Biotic heterogeneity	Terrestrial	Habitat Structure, vegetation/plant phenology based on satellite remote sensing (European extent)	2012	2012		25ha MMU / 100m
166	CDDA (National designated areas)	Protected areas	Protected areas	Biotic heterogeneity	Terrestrial	NA	2020			detailled delineation
167	Land Use Harmonisation 2	Land use	Land use	Socio-Ecology	Land use and land cover change	Land use (historic)	850		yearly	0.25 x 0.25 resolution
168	EuroCordex	Climate variables (https://is-enes-data.github.io/CORDEX_variables_requirement_table.pdf)	Climate	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity			yearly	0.11 degree (EUR-11, ~12.5km)
169	ESGF	Climate models		Abiotic characteristic	Climate	Air temperature; Precipitation				
170	PhenoCam	vegetation phenology (seasonal changes) by recording time-lapse images of a fixed scene	Plant phenology	Biotic heterogeneity	Terrestrial	Plant phenology	2000	present	high-frequency (typically, 30 minute)	point
171	GBIF Api	species occurence		Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	2001		observation	global
172	BioTime	species observations	Species	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	1874		observation	Plot / plot location geographic with two digit precision

ID	Data source	Collected data	Variable description	SO Component	SO Compartement	SO Variable(s)	Start year	End year	Temporal resolution	Spatial resolution
173	ECAD dataset (e-obs gridded dataset)	daily observations at meteorological stations	precipitation, temperature, sea level pressure, relative humidity, wind speed and global radiation in Europe based on ECA&D information	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1950	2020	daily	0.25∘ × 0.25∘ grid
174	ECAD daily data	daily observations at meteorological stations	Daily maximum temperature TX Daily minimum temperature TN Daily mean temperature TG Daily precipitation amount RR Daily mean sea level pressure PP Daily cloud cover CC Daily humidity HU Daily snow depth SD Daily sunshine duration SS Global radiation QQ Daily mean wind speed FG Daily maximum wind gust FX Daily wind direction DD	Abiotic characteristic	Climate	Air temperature; Precipitation; Relative air humidity	1950	2020	daily	per station
175	OBIS API	species occurence		Biotic heterogeneity	Marine	NA				
176	iNaturalist	species occurence	Nature observations, species identification	Biotic heterogeneity	Terrestrial	Habitat structure, vascular plants, lichens, mosses/vegetation based on UAV remote sensing (local)	2008	present	-	-

Annex 5 – Exemplary list of national scale data sources

National scale *in-situ* data sources provide an additional layer relevant for building Information Clusters in the context of eLTER. Based on experience and expertise within the task team we did a in-depth screening of relevant sources based on the example of France. This is not possible for all countries contributing to the eLTER network due to resource constraints.

Data source	***Link and description***	***Modalities***	***Country***	***Language***
DINAMIS	*https://dinamis.data-terra.org/* *The National Institutional Device for Mutualized Supply of Satellite Imagery (DINAMIS) is a platform for the acquisition and dissemination of Earth observation space data dedicated to French and foreign institutional users under conditions. Its objective is to facilitate access to and develop the public use of space data for :* *- Research and development: fundamental and applied, seed and demonstration phase,* *- Territorial planning and development,* *- Environmental monitoring and management,* *- Innovation and creation of added value.*	*All*	*France*	*fr*
LTER France Data Node	*http://meta.data-za.org/geonetwork/srv/fre/catalog.search#/home* *The Zones Ateliers (ZA) focus on a functional unit (a river and its watershed, landscapes - agricultural or urban - and biodiversity, from Antarctica to sub-Saharan Africa, or the coastline, or the environments characterised by chronic radiation of natural or enhanced natural origin) and develop a specific scientific approach based on observations and experiments on workshops sites, to conduct multidisciplinary research in the long term. A ZA is therefore, most often, a network of workshop sites. The RZA has a portal and a geocatalogue (Geonetwork 3.4.2), which is currently being updated. This catalogue harvests the ZAs catalogues, if they exist.*	*All*	*France*	*fr/en*
ADES	*https://ades.eaufrance.fr/* *ADES is the national portal of Access to Groundwater Data for metropolitan France and overseas departments. It gathers on a public website quantitative and qualitative data related to groundwater.*	*Aquatic*	*France*	*fr*
BRGM	*https://www.brgm.fr/projet/referentiel-hydrogeologique-francais-bdlisa* *The hydrogeological repository BDLISA is a national tool for locating groundwater data. BDLISA aims to provide a map of hydrogeological entities for the whole territory of France and its overseas territories.*	*Aquatic*	*France*	*fr/en*
CYBERCAROTHEQUE	*https://cybercarotheque.fr* *The objective of this portal is to present the metadata associated with marine and lake sediment cores (and soon continental and glacial cores) stored in French laboratories. To date, 2366 cores from 277 missions are already integrated, from existing inventories of laboratories.*	*Aquatic*	*France*	*fr*
HydroPortail	*https://www.hydro.eaufrance.fr/* *The Hydroportal is the reference site for access to to hydrometric and hydrological data in France. These data (flowrate and water height) are available for* *all the measuring stations existing or having existed on the French French rivers, including overseas. This represents approximately 5,000 stations, among which more than 3,000 are active* *and regularly feed the national database.*	*Aquatic*	*France*	*fr*

Data source	***Link and description***	***Modalities***	***Country***	***Language***
NAIADES	*http://www.naiades.eaufrance.fr/france-entiere#/* *Naïades is the national interface for accessing assessment data (status and pressures/quality) of continental surface waters (rivers and lakes). It allows users to access data on chemical concentrations, species inventories and associated assessment indicators, physico-chemical parameters and hydromorphology.*	*Aquatic*	*France*	*fr*
ODATIS	*https://www.odatis-ocean.fr/en/data-and-services/data-access/direct-access-to-the-data-catalogue#/search?from=1&to=30* *ODATIS (Ocean Data and Services Cluster) helps to describe, quantify and understand the ocean as a whole: ocean dynamics and thermodynamics, evolution of its physico-chemical properties, biochemical cycles, marine ecosystems functions, ocean evolution and ocean-climate link from the past evolution (paleo-oceanography). It also covers coastal-specific topics (including estuaries and lagoons): morpho-dynamic evolution of the shoreline, coastline and sea-level, pollutions and eutrophication, coastal ecosystem changes and evolution.*	*Aquatic*	*France*	*en*
SANDRE	*http://www.sandre.eaufrance.fr/atlas/srv/fre/catalog.search#/home* *The Sandre's mission is to establish and make available the WIS water data repository. This repository, composed of technical specifications and lists of free codes, describes the methods of exchange of water data on the French scale. From an IT point of view, the Sandre guarantees the interoperability of water-related information systems *.*	*Aquatic*	*France*	*fr*
Sextant	*https://sextant.ifremer.fr/eng* *Sextant's objective is to document, disseminate and promote a catalogue of data relating to the marine environment. Aimed at Ifremer's laboratories (https://wwz.ifremer.fr/) and partners, as well as national and European players working in the marine and coastal field, Sextant provides tools that promote and facilitate the archiving, consultation and availability of these geographic data.*	*Aquatic*	*France*	*en*
Waste water Treatment Plant (Urban Wastewater)	*http://assainissement.developpement-durable.gouv.fr/* *Database on the 22,331 wastewater treatments operated in France in 2020.*	*Aquatic*	*France*	*fr*
PNDB	*https://data.pndb.fr/* *The National Biodiversity Data Center is aiming at for scientists producing, managing and analysing biodiversity data*	*Biodiversity*	*France*	*fr*
ACBB	*https://www.soere-acbb.com/* *Long-term observation and experimentation system for environmental research - agro-ecosystem, biogeochemical cycle and biodiversity. The database includes biological, chemical, physical and agroecosystem management practices data.*	*Experiments*	*France*	*en*
ANAEE-France	*http://w3.avignon.inra.fr/geonetwork_anaee/srv/eng/catalog.search#/home* *Understanding the dynamics and sensitivity of ecosystems to global change is a fundamental research challenge. In this perspective, we need to manipulate ecosystems to characterise their dynamic properties, the complexity of the interactions at work and develop a fundamental corpus necessary for the implementation of adaptation*	*Experiments*	*France*	*en*

Data source	*Link and description*	*Modalities*	*Country*	*Language*
	and restoration measures. *The AnaEE-France infrastructure (Analysis and Experimentation on Ecosystems), supported by CNRS, INRAE and Grenoble Alpes University, meets these objectives by providing a coherent set of services dedicated to experimentation for the study of terrestrial and aquatic ecosystems. With centralised access to experimental platforms, analysis instruments and modelling platforms, AnaEE-France facilitates the use of experimental platforms, the emergence of innovative projects and the reuse of data.*			
F-ORE-T	*https://si-foret-ecofor.fr/* *The observatory F-ORE-T aims to understand the functioning of these ecosystems by analysing, in particular, the stocks and flows of carbon, water and mineral elements and to evaluate their response to changes, whether slow or rapid, natural or anthropogenic (climate, forestry, land use change). The management of several types of data has been developed, flow data from flux towers, meteorological data and soil climate data. The feeding of the SI with current and historical data is in progress.*	*Experiments*	*France*	*en/fr*
H+	*http://hplus.ore.fr/en/* *The first goal of the H^+ observatory is to maintain and coordinate a network of experimental sites capable of providing data - including chronicles or data on long term experiments - relevant to the understanding of the water cycle and of the motion of solute elements in aquifers. The H^+ Database includes all the data acquired on the H^+ network.*	*Experiments*	*France*	*en*
OLA	*https://si-ola.inra.fr/si_lacs/login.jsf* *The Observatory's database includes all the data related to the monitoring of the lakes whatever the theme (phytoplankton, zooplankton, fish, physicochemistry, etc.). The data in presence are on the one hand data obtained in situ as for example the probe data giving profiles on all the depth of the lake (pH, T°, turbidity, transparency, fluorescence, etc) and on the other hand data resulting from counts or analyses, obtained in laboratory from the samples. The observation system is centred on long-term monitoring, mainly of the large peri-alpine lakes that are Lake Geneva, Lake Annecy, Lake Bourget and Lake Aiguebelette, but also concerns small high altitude lakes (sentinel lakes).*	*Experiments*	*France*	*en*
PRO	*https://www6.inrae.fr/qualiagro_eng/Nos-partenaires/The-SOERE-PRO-network* *The QualiAgro programme pertains to the SOERE PRO network working on environmental impacts of Organic Waste Products recycling on field crops at long time scale. This network groups together 5 French field experiments from different soil-climate contexts lead by the EGC INRAE Unit of Grignon.*	*Experiments*	*France*	*en*
AERIS	*https://en.aeris-data.fr/catalogue/* *The French thematic consortium on atmosphere federates at the national level data management activities and scientific expertise in the atmosphere.* *Whether it is data from long-term observation sites that record the causes and signs of climate change, results from large-scale field or airborne campaigns to better understand the chemistry and physics of*	*Meteorology*	*France*	*en*

Data source	***Link and description***	***Modalities***	***Country***	***Language***
	our atmosphere, measurements from satellites that continuously observe the Earth and probe its atmosphere, or even results from laboratory experiments, all atmospheric research data are intended to be stored, distributed and combined by the AERIS cluster			
ATMO-AASQA	*https://atmo-france.org/les-donnees/* *In France, air monitoring is mandatory and regulated. Since the law on Air of December 30 1996, known as the LAURE law, the Environment Code entrusts the implementation of air quality monitoring to the AASQAs. The national network of AASQAs is currently made up of 18 approved associations present in each administrative region of France and its overseas territories, as well as an equivalent non-approved association.* *administrative region of France and overseas as well as an equivalent non-accredited association* *in New Caledonia.*	*Meteorology*	*France*	*fr*
DRIAS	*http://drias-climat.fr/accompagnement/section/40* *Drias[les futurs du climat] aims to make available regionalized climate projections made in French climate modelling laboratories (IPSL, CERFACS, CNRM). The climate information is delivered in different graphical or numerical forms.*	*Meteorology*	*France*	*en*
NIVOSE	*https://donneespubliques.meteofrance.fr/* *Nivôse is the name of the network of automatic weather stations in high mountains, created by Météo-France. The Nivôses stations were created to allow meteorologists and more widely the public to access to real time weather data concerning difficult to access mountainous areas.*	*Meteorology*	*France*	*fr*
RADOME	*https://donneespubliques.meteofrance.fr/* *Real time data measured at automatic stations of the French network with hourly and/or sub-hourly time steps (6mn). Basic parameters (temperature, humidity, wind direction and force, precipitation) and additional parameters according to instrumentation (ground temperature, visibility, ground condition, insolation, global radiation). Declined either for a single station, or for the stations of the RADOME network (about 550 stations), or for the whole extended network (about 1150 stations including RADOME stations). An annual subscription is required.*	*Meteorology*	*France*	*fr*
AGRESTE	*https://agreste.agriculture.gouv.fr/agreste-web/* *Results for the whole of France including DOM* *Crop production: areas, yields, harvested or marketed production* *Animal production: Livestock present on the farms, finished animals produced. (production, average weight, weight produced), production and use of milk on the farm.* *Distribution of the territory in (annual agricultural statistics): arable land, permanent crops, utilized agricultural area, total area.*	*Socio-Ecology*	*France*	*fr*
CORINE LAND COVER	*https://www.data.gouv.fr/fr/datasets/corine-land-cover-occupation-des-sols-en-france/#_*	*Socio-Ecology*	*France*	*fr*
INSEE	*https://www.insee.fr/fr/statistiques/3677855*	*Socio-Ecology*	*France*	*fr*
REMONTER LE TEMPS	*https://remonterletemps.ign.fr/* *Going back in time allows to observe the evolution of the territory:*	*Socio-Ecology*	*France*	*fr*

Data source	*Link and description*	*Modalities*	*Country*	*Language*
	urbanisation (extension of urban areas, new cities, large industrial developments...), modification of natural spaces (coastal areas...), evolution of communication routes (road network...). *Ti is possible to consult online historical geographical data (old maps, aerial photographs), and compare them with current maps and to download historical aerial photographs (since 1919).* *More than three million historical aerial photographs and old maps are available for download on the French territory.*			
RPG (Graphical Parcel Registry)	*https://geoservices.ign.fr/rpg#telechargement* *The graphical parcel register is a geographical database used as a reference for the instruction of the Common Agricultural Policy (CAP) aids.* *The anonymized version distributed here as part of the public service of reference data availability contains the graphic data of the parcels (basic land unit of the farmers' declaration) with their main crop. These data have been produced since 2007.* *The anonymous data of the RPG are vintage and contain parcels corresponding to those declared for the campaign N in their known situation and stopped by the administration, in general on January 1st of the year N+1. These data cover the entire French territory, including Mayotte and Saint-Martin, but excluding Saint-Barthélemy.*	*Socio-Ecology*	*France*	*fr*
SIDDT	*https://siddt.inrae.fr/* *Information System Dedicated to Territories. This portal puts at your disposal the "territory" database of the LESSEM research unit (Laboratory of Ecosystems and Societies in Mountains) of INRAE Grenoble: large national statistical files, heading "BD Communale", GIS data, section "Geographic information".*	*Socio-Ecology*	*France*	*fr*
ZNIEFF	*https://inpn.mnhn.fr/programme/inventaire-znieff/presentation#* *Launched in 1982, the inventory of Natural Areas of Ecological, Faunistic and Floristic Interest (ZNIEFF) aims to identify and describe, throughout the national territory, sectors of greater ecological interest sheltering heritage biodiversity with a view to creating a base of knowledge but also a tool to assist decision-making (protection of space, land use planning). There are two types of ZNIEFF :* *ZNIEFF Type I : ecologically homogeneous areas, defined by the presence of rare, remarkable or characteristic species, species associations or habitats of the regional natural heritage. These are the most remarkable areas of the territory;* *ZNIEFF type II: areas that include functional natural and landscape ensembles, with a high degree of cohesion and richer than the surrounding environments.* *The ZNIEFF inventory concerns the whole French territory: metropolitan France and overseas territories, continental and marine environments. The ZNIEFF database downloadable in a Zip file in text format corresponds to the export of all the information recorded in the ZNIEFF forms and updated twice a year on the INPN.*	*Socio-Ecology*	*France*	*fr/en*
BD Forêt V2	*https://geoservices.ign.fr/bdforet* *The BD Forêt is the reference database for forest areas and semi-natural environments. It describes, on a departmental basis, the forest and natural vegetation formations by a land cover approach translating a description of the stand's cover density and composition for the elements of more than 5000 m2. The V2 version, not yet available for all the departments, will be updated every 3 years.*	*Terrestrial*	*France*	*fr*

Data source	***Link and description***	***Modalities***	***Country***	***Language***
BDAT	*https://www.gissol.fr/tag/bdat* *Knowledge of the variability of surface horizons of cultivated soils, with about 250,000 soil analyses performed each year, mainly at the request of farmers for a better management of fertilisation.*	*Terrestrial*	*France*	*fr*
BDCHARM 50	*http://infoterre.brgm.fr/formulaire/telechargement-cartes-geologiques-departementales-150-000-bd-charm-50* *The Bd Charm-50 is the georeferenced database of the harmonised and vectorized 1/50 000 geological maps. With its 6 layers of detailed information (shape .shp format), the harmonised geological map provides continuous information.*	*Terrestrial*	*France*	*fr*
BDIFF	*https://bdiff.agriculture.gouv.fr/* *The database on forest fires (BDIFF) is an internet application in charge of centralising all the data on forest fires on the French territory since 2006 and of making all this information available to the public and to the State services.*	*Terrestrial*	*France*	*fr*
CES OSO	*https://www.theia-land.fr/en/ceslist/land-cover-sec/* *Production of land cover maps based on satellites images (17 classes for 2016 and 2017) and 23 classes for 2018, 2019 and 2020), with a spatial resolution of 10 m (raster) and 20 m (vector) and an updating frequency of one year*	*Terrestrial*	*France*	*en, fr*
DONESOL	*https://dw3.gissol.fr/login;* *http://www.gissol.fr/donnees/webservices* *National database structuring and gathering point and surface data from soil surveys*	*Terrestrial*	*France*	*fr*
GEORISQUES	*https://www.georisques.gouv.fr/risques/sites-et-sols-pollues/donnees#/* *Dataset on the pollution of soils, including old industrial sites*	*Terrestrial*	*France*	*fr*
Knowledge Base "Statuts"	*https://inpn.mnhn.fr/telechargement/referentielEspece/bdc-statuts-especes* *In terms of biodiversity management, the first step is to gather knowledge on the status and evolution of biodiversity. Each year, a multitude of data on the conservation status of species is produced by many actors in the territories, using different methods and formats, making exchanges and analyses complex on a global scale. It is becoming necessary to produce summaries that can be used by the greatest number of people.* *The "Statuts" knowledge base is part of this approach.*	*Terrestrial*	*France*	*en*
OpenOBS	*https://openobs.mnhn.fr/* *OpenObs allows users to view and download observation data on species available in the National Inventory of Natural Heritage, the national platform of the SINP (Information System of the Natural Heritage Inventory).*	*Terrestrial*	*France*	*fr*
RGE ALTI	*https://geoservices.ign.fr/rgealti* *Gridded digital terrain model (DTM) which describes the relief of the French territory on a large scale. A refinement program was initiated in 2009: its objective is to create a digital terrain model (DTM) with a 1 m resolution for the whole of France.*	*Terrestrial*	*France*	*fr*
RMQS	*https://www.gissol.fr/le-gis/programmes/rmqs-34* *A long-term soil monitoring tool to monitor the evolution of soils*	*Terrestrial*	*France*	*fr*

Data source	***Link and description***	***Modalities***	***Country***	***Language***
	under the effect of major natural factors and human activities (uses, land development, agricultural practices, sludge spreading, atmospheric deposition, accidental pollution, etc.). 2240 sites are sampled every 15 years.			
SNO-Tourbières	*https://www.sno-tourbieres.cnrs.fr/* *Access to the data (meteorology, energy and GES fluxes, hydrologie, hydrochemistry, biodiversity) collected on the four peatland sites (Bernadouze, La Guette, Frasne and Landemarais)*	*Terrestrial*	*France*	*en/fr*
THEIA/OZCAR	*https://www.theia-land.fr/en/homepage-en/* *Theia is pursuing four main objectives :* *1. Promoting and facilitating the use of space data, for science and public actors, in terms of imagery, added-value products as well as in-situ data;* *2. Developing added -value products and services for the science communities and national public actors;* *3. Developing networks of competences* *4. Supporting French realisations at European and international level.*	*Terrestrial*	*France*	*en*
VIGIE-NATURE	*https://www.vigienature.fr/fr* *Vigie-Nature is a participatory science program open to all those who are curious about nature, from beginners to the most experienced. Based on simple and rigorous protocols, it offers everyone the opportunity to contribute to research by discovering the biodiversity.*	*Terrestrial*	*France*	*fr*
GEM Greenland ecosystem monitoring	*https://g-e-m.dk/*	*Terrestrial*	*Greenland*	*en*
Information Grid (InGrid)	*https://www.ingrid-oss.eu/latest/about/applications.html* *Information grid (InGrid) is applied for research, visualisation, and download of Geodata at national, regional, and international level*	*Terrestrial*	*Germany*	*de*
UBA-DE	*https://www.umweltbundesamt.de/daten* *Environment Agency Germany data portal provides access to a wide range of ecosystem data.*	*Terrestrial*	*Germany*	*de*