ENVRI^{plus} DELIVERABLE

D19.7
FINAL DATA MANAGEMENT PLAN

WORK PACKAGE 19 – PROJECT MANAGEMENT

LEADING BENEFICIARY: UNIVERSITY OF HELSINKI

Author(s):	Beneficiary/Institution
Ari Asmi	University of Helsinki
Jean-Daniel Paris	CEA
Zhiming Zhao	University of Amsterdam
Anders Tjulin	EISCAT association
Helen Glaves	NERC
Antti Pursula	CSC

Accepted by: Werner Kutsch

Deliverable type: [REPORT]

Dissemination level: PUBLIC

Deliverable due date: 31.3.2018/M36

Actual Date of Submission: 29.7.2019/M46

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654182

FINAL DATA MANAGEMENT PLAN

ABSTRACT

This document presents the updated ENVRIplus data management plan (DMP), developed by the project Data Management Team in their workshops and email conversation. This is the second version of the DMP and will be updated annually. The document describes shortly the project, scope of DMP, default policies and then explains the current plan on data set collection level.

DATA MANAGEMENT TEAM

The DMP team met first time in during ENVRIplus kickoff meeting in Helsinki, Finland in May 2015. The initial meeting concluded with discussion on the overall procedure towards proper data management within the ENVRIplus cluster. The draft version of the document was then developed during the first 6 months of the project, using emails and teleconferences within the DMT, and with the relevant project participants.

The overall structure of this document is based on the DMP templates available in the UK Digital Curation Center (http://www.dcc.ac.uk/resources/data-management-plans).

ENVRI PLUS AS A PROJECT

The objective of ENVRI PLUS is to provide common solutions to shared challenges for European Environmental and Earth System Research Infrastructures (RIs) in their efforts to deliver new services for science and society. The main actions of this work are (with their relevant data types and sources as identified in the beginning of the project):

1. Improve the RI's abilities to observe the Earth System, particularly in developing and testing new sensor technologies, harmonizing observation methodologies and developing methods to overcome common problems associated with distributed remote observation networks;
    a. measurements from new sensors under tests or intercomparison of sensors and sampling systems – these are rarely needed for long term preservation, but important during project period;
    b. Data collection of field measurements from RI scientists;
2. Generate common solutions for shared information technology and data related challenges of the environmental RIs in data and service discovery and use, workflow management and documentation, data identification and citations methodologies, data provenance, and user characterization and interaction;
    a. Software products for data management and applications:
    b. RI questionnaires and interview results on existing and future needs for IT solutions;
3. Develop harmonized policies for access (physical and virtual) for the environmental RIs, including access services for the multidisciplinary users;
    a. RI questionnaires and interview results on existing and future needs for access, including experiences on new TNA structures, details of the applications, etc.
    b. (potentially) data sets provided by TNA activities
4. Investigate the interactions between RIs and society: Find common approaches and methodologies how to assess the RIs' ability to answer the economical and societal challenges, develop ethics guidelines for RIs and investigate the possibility to enhance the use Citizen Science approaches in RI products and services;
    a. Questionnaires and interviews of RIs
    b. Analysis results of societal interaction methodologies

c. Citizen science participant data, and produced data sets

5. Ensure the cross-fertilisation and knowledge transfer of new technologies, best practices, approaches and policies of the RIs by generating training material for RI personnel to use the new observational, technological and computational tools and facilitate inter-RI knowledge transfer via a staff exchange program;
    a. Experience (questionnaires) from personnel exchange
    b. Training materials
    c. Personnel exchange personal information
6. Create RI communication and cooperation framework to coordinate activities of the environmental RIs towards common strategic development, improved user interaction and interdisciplinary cross-RI products and services.
    a. Reports of meetings, their discussions and participants
    b. Dissemination material

Thus, the overall data products generated are heterogeneous, and mostly for internal consortium use, and expected storage needs are minor in volume. Relatively lot of "soft" research information is collected in contrast to direct observational "hard" research data.

Related policies

The RIs have their own associated data management policies, either existing or in development. Similarly, the Beneficiaries might have their own overreaching policies, although the effect of these will be controlled by the relationship defined in the Consortium Agreement document of the ENVRIplus. This DMP is descriptive to only the data directly generated within the project ('foreground' in the CA).

Scope of the DMP

The DMT had several discussions on the scope of the DMP. After these discussion it was decided the that the ENVRIplus project DMT will have following scope definitions:

- It covers all research data sets produced directly in ENVRIplus project activities mentioned in the Description of Action. It specifically does not cover the auxiliary data sets produced by the Research Infrastructures using the ENVRIplus products (e.g. using common harmonized metadata system, etc.) as these are subject to the data management plans of the individual Research Infrastructures.
- Even though the project partners or non-partner personnel taking a part of the Trans National Activities (TNA) are strongly suggested to use the overall DMP structure presented here, the data sets produced by non-ENVRIplus partners in their TNA activities are outside the scope of this DMP and the Consortium Agreement. However, clauses regarding DMP used were included in the TNA agreements, and most of the TNA partners submitted their data according to the requirements listed here.
- This DMP covers observational data, results from questionnaires, interviews and developed software. For the software, only relevant parts of the DMP template need to be covered.
- The overall level of the DMP is more on the "collection" level than individual data sets. Similar data sets from similar source are considered a collection, if the policies attached to them are similar or the data sets can be considered to form a single whole.

DEFAULT ENVRIPLUS DATA POLICIES

This section presents the default data policies, which can be then referred in individual data collection policy results later in the document.

Metadata and documentation

ENVRIplus contains a work package on harmonization of metadata for Earth System observation RIs. This work is naturally intended for long-term harmonization, and thus most of the data produced during the project will not immediately benefit for this work. However, the RIs taking part of the project have their own metadata schemes, and the overall **ENVRIplus data policy is to use these schemes and documentation methods as much as possible**. Use of suitable international standards (e.g. INSPIRE directive) are strongly recommended.

For non-observational data or data produced by non-ENVRIplus partners, other methods are needed. Major part of the project is concerned on human or policy data collected form the participating RIs. These data sets are not part of the typical RI collected data sets and thus require more attention on metadata and documentation, which need to be defined case-by-case basis, with overall requirement of ensuring data usability for re-use.

Similarly, the citizen science data products and non-ENVRIplus partners data from Pilot Transnational Access are not necessarily covered by typical RI metadata standard procedures. The overall ENVRIplus data policy in these cases requires inclusions of sufficient and suitable metadata for these datasets. The WP leaders of relevant WPs are required to provide information on these datasets on the next iteration of the DMP.

Ethical and privacy issues

Standard (RI) observational data sets in ENVRIplus are not expected to have ethical or privacy concerns. This should be evaluated on case-to-case basis by the WP leader in question.

However, the questionnaire data and Citizen Science collected datasets can have significant privacy concerns. For data sets which have these concerns, the ENVRIplus general policy is to follow the procedure:

1. Collection of any private data (including e.g. opinions, names, positions, etc) is to be avoided if not useful for the purpose of the study.
2. If such information is needed, the target of the study must be informed before the data collection about the
    a. Collected information,
    b. Why it is needed,
    c. How the targets were selected,
    d. Who has access to the data,
    e. Of any anonymization scheme (if any) is used,
    f. How the data will be analyzed,
    g. How the target and either agree on disagree on the terms,
    h. How the access to the data will be controlled and,
    i. How long and where the data will be stored.
3. The data retention must be carefully considered, and if the raw data sets do not need to be stored, they must be destroyed efficiently after use.
4. Data storage must be adequate to the level of sensitivity of the data.

5. Access control of the data must be adequate and clearly defined, including access policies in the long-term storage.

All of these procedures of course need a formal ethical review for doing the study. The template and procedure above should be used as the basis of the ethical approval requests.

The procedure for questionnaires was approved by the University of Helsinki Ethical review board. This is to make sure that each questionnaire in the project will not need to be separately evaluated in participating institutions. Each ENVRIplus beneficiary making a questionnaire inside the project will instead follow the accepted procedure.

IPR issues

Ownership of the data sets produced within the project partners are defined in the Consortium Agreement. Ownership of non-partner data sets produced within the project envelope (e.g. Citizen Science and TNA partners) must be clearly defined in the Terms of Collaboration to be defined with these partners.

Ownership of the TNA partners data was in the end left to the TNA partners. This also means that the datasets produced in their activities are outside of the ENVRIplus data products, and of this document.

Storage and backup during project time

The individual partners (and RIs) are responsible for the project time storage and backup of the collected data sets. However, the ENVRIplus has collaboration with the EUDAT2020 H2020 project, which provides B2DROP service for data set storage (max size 2Gb per file), sharing and backup. The DMT recommends partners to use this facility.

Access management

For many datasets produced, the storage and access management can be done using the RIs own repositories. Access to other ENVRIplus partners to these data sets (not others) must be provided as defined in the Consortium Agreement. For the data stored in EUDAT2020 services, the access management will be provided by their Access Management modules. Access must be provided to the Commission officials and their appointed reviewers. Access to sensitive data must be adequately controlled.

Retention and preservation

Decision on retention of the data sets should be done carefully. In the case of sensitive data, the retention must consider the absolute need of storing such information, and effective anonymization procedures must be followed. If sensitive data is to be destroyed, such actions must be done effectively and with clear responsible person determined before data collection.

The retention of the general data sets done on the decision of the WP leader in question, however this decision can be changed if the data originator or other project participant complains the situation to the ENVRIplus Executive Board. Project Executive Board has final say on the retention issues. In the data retention decision the following aspects are to be considered: 1) Re-usability of the data (including metadata), 2) needed resources for long term storage (size, access), 3) expected storage period, 4) possibility of external data storage using non-project related repositories.

Long term preservation plan

Data selected to preservation must include long-term preservation plan. Data can be stored in generally accepted long-term preservation system, with minimum of 10 years of guaranteed storage. Such preservation systems must also provide acceptable sustainability plan for migration to other storage systems. The DMT suggests use of EUDAT services (B2SHARE, B2SAFE), or Zenodo for such storage.

Sharing policy

Unless required by Consortium Agreement or sensitive data policies, all ENVRIplus data products must be made openly accessible and findable (i.e. via metadata harvesting access) in reasonable time after data collection. Default time for this is 6 (six) months data end of data collection. This is controlled by the relevant WP leaders. Exceptions for this rule can be made on for good reasons by Project Executive Board on request.

Restrictions on sharing

Only restrictions on sharing data sets are connected to sensitive information, or temporary grace period right after data collection. Otherwise all ENVRIplus results are open to use for any purpose, although all data sets must be appropriately licensed using very open licensing policy (e.g. Creative Commons 4.0BY or similar).

WP5 raw interview data is not shared outside of the participating institution. Processed (not personal) datasets from them are documented in Deliverable D5.1. All deliverables are available in the project website.

Responsible person

Each data set produced must have an appointed responsible person. The data providing institution must provide a way to maintain the responsibility even in the case of personnel change – i.e. a clear procedure must be detailed on how the responsibility can be transferred and how this can be controlled.

Resources used

Data set storage, curation and maintenance costs are valid ENVRIplus costs. The long term resources needed for storage must be considered.

DATA SETS DESCRIPTION AND AVAILABILITY

During the last month of the project, a project-wide survey of generated datasets, reports, and software was launched. The survey results were then analysed and confirmed with the WP leaders of the project. This section can be compared with the D19.3, which included the plans for managing the project data earlier.

Work Package:	3	**Version:**		
Title:	Oceanographic data measured during the WaveGlider demo mission			
Authors:	Christoph Waldmann, Robert Huber			
Description:	During the WaveGlider demo mission which will take place June 2019 at the coastal area of Gran Canaria, near surface temperature, salinity, oxygen and current data will be emasured. All data will be published at the PANGAEA data repository.			
Format	PANGAEA ASCII			
Metadata	Directory Interchange Format (DIF), ISO19139, Dublin Core, DataCite			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	http://www.pangaea.de			
Persistent identifiers	DOI, ORCID			
Open access	Yes	**Licence**	CC4.0BY	
Access limitations:	none			
Vocabularies, standards	See metadata standards given above			
Quality control	First level QC done by data set authors, second level data curation done by PANGAEA curators			
Long term availability	PANGAEA is a certified long term data archive (DSA, WDC, WMO)			

Work Package:	4	**Version:**	1.0	
Title:	Methane cruise dataset			
Authors:	Ruffine, Paris, Grilli, Italiano, Schumacher et al.			
Description:	Data collected during the Black Sea cruise to investigate the joint deployment of RIs on the transverse topic of Methane from seafloor to the atmosphere. Useful for its results and possibly reanalysis.			
Format	heterogeneous dataset			
Metadata	Description in associated deliverable. Preparation of metadata ongoing			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	>=1Gb but <1Tb			
Stored in public repository	Yes			
Location of storage:	EUDAT b2drop, b2share			
Persistent identifiers	to be created when dataset complete			
Open access	Yes	**Licence**	CC4.0BY	
Access limitations:	embargo until July 2019			
Vocabularies, standards				
Quality control	respective applicable domain/RI QC procedure			
Long term availability	EUDAT			

Work Package:	7	**Version:**		
Title:	Data Analytics Solution			
Authors:	Leonardo Candela, Gianpaolo Coro, Pasquale Pagano, Giancarlo Panichi			
Description:	This component is part of a large software system called gCube. It implements the solution for data processing and analytics developed in the context of the ENVRIplus project. The distinguishing features of the proposed solution are (a) to be suitable for serving the needs of scientists involved in ENVRI RIs, (b) to be open and extensible both with respect to the algorithms and methods it enables and the computing platforms it relies on to execute those algorithms and methods, (c) to be open-science-friendly, i.e. it is capable of incorporating every algorithm and method integrated into the data processing framework as well as any computation resulting from the exploitation of integrated algorithms into a research objectcatering for citation, reproducibility, repeatability and provenance.			
Format	Software artifacts mainly written in Java			
Metadata	Every software artifact is endowed with descriptive files including LICENCE and README			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	>=1Gb but <1Tb			
Stored in public repository	Yes			
Location of storage:	The various versions of the artifact (consisting in several modules) have been stored in Zenodo (special collection for gCube) https://zenodo.org/communities/gcube-system?page=1&size=20			
Persistent identifiers	Every artifact has its own DOI.			
Open access	Yes	**Licence**	The software is available in Open Access with EUPL licence.	
Access limitations:				
Vocabularies, standards				
Quality control	The software is produced and released according to documented policies described here https://wiki.gcube-system.org/gcube/Software_Integration_and_Distribution:_Procedures			
Long term availability	This is guaranteed by Zenodo.			

Work Package:	8	**Version:**		
Title:	Prototype Catalog B2FIND Format			
Authors:	Claudia Martens, Erwann Quimbert			
Description:	A prototype catalog with converted records from several ENV RIs in EUDAT B2FIND (CKAN-based) format			
Format	bit encoded data			
Metadata	The data is metadata (catalog entries describing datasets)			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	No, explain below			
Location of storage:	http://eudat7-ingest.dkrz.de/dataset?groups=envriplus			
Persistent identifiers				
Open access	No	**Licence**		
Access limitations:	prototype for feasibility (not available for general use)			
Vocabularies, standards	EUDAT CKANbased metadata Schema			
Quality control				
Long term availability	No sustainability plan and funding commitment			

Work Package:	8	**Version:**		
Title:	CERIF metadata catalog			
Authors:	Keith Jeffery, Daniele Bailo, Erwann uimbrt			
Description:	A catalog in CERIF format of EPOS assets			
Format	CERIF (EU recommendation to Member States) as bitcode data			
Metadata	CERIF itself: the catalog is metadata records describing assets			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):			n/a	
Stored in public repository			No, explain below	
Location of storage:			The original CERIF data in the EPOS catalog is FAIR. The copy at IFREMER for prototyping is not publicly available	
Persistent identifiers				
Open access	No	**Licence**		
Access limitations:	Prototype only			
Vocabularies, standards	CERIF			
Quality control				
Long term availability	The original CERIF records in EPOS are FAIR and have a sustainability plan. The copy at IFREMER for ENVRI prototyping has no sustainability / funding.			

Work Package:	9	**Version:**		
Title:	ANIMATE project data integration into 52North and ERDDAP software			
Authors:	National Oceanography Centre, Southampton			
Description:	Historic open data from the ANIMATE project were used: https://www.bodc.ac.uk/resources/inventories/edmed/report/281 / This are historic open data as a demonstrator for interoperbility of web services			
Format	OGS Sensor observation service, NOAA ERDDAP dataset			
Metadata	OGC SensorML asscіated with data			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):			n/a	
Stored in public repository			Yes	
Location of storage:			52North endpoint http://linkedsystems.uk/52n-sos-webapp/index ERDDAP endpoint https://linkedsystems.uk/erddap/tabledap/ENVRIplus_b122nnnn.html	
Persistent identifiers	52North endpoint http://linkedsystems.uk/52n-sos-webapp/index ERDDAP endpoint https://linkedsystems.uk/erddap/tabledap/ENVRIplus_b122nnnn.html			
Open access	Yes	**Licence**	CC4.0	
Access limitations:				
Vocabularies, standards	NERC vocabulary service vocabularies, Marine SWE profile SensorML template			
Quality control	Standard BODC ingestion https://www.bodc.ac.uk/submit_data/what_do_we_do_with_your_data/data_processing_steps/			
Long term availability	BODC are a long term IODE accredited data archive centre			

Work Package:	9	**Version:**		
Title:	ENVRI+ Theme2 Science Demonstrators and Service Portfolio			
Authors:	WP9 agile groups			
Description:	In order to validate Theme2 services with realistic community use cases			
Format	Wiki, Youtube Videos, Software code			
Metadata	ENVRI+ Theme2 Science Demonstrators and Service Portfolio			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	https://confluence.egi.eu/display/EC/Science+Demonstrators			
Persistent identifiers				
Open access	Yes	**Licence**	Depends on service, cc4.0by as default	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	They are in ENVRI community repository, depending on community policy how long to support			

Work Package:	9	**Version:**	0.9.0	
Title:	New Particle Formation Event Analysis Virtual Research Environment			
Authors:	Markus Stocker			
Description:	The VRE supports new particle formation event analysis on interoperable e-Infrastructures. It provides access to Jupyter notebooks to classify events and process information about them. It integrates the SMEAR Research Infrastructure (provider of primary data) and uses EGI and D4Science services to support primary data interpretation and the cataloging of data derived in analysis. Access to the VRE requires a D4Science account.			
Format	Python			
Metadata	N/A			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	https://github.com/markusstocker/pynpf-d4science https://marketplace.eosc-portal.eu/services/new-particle-formation-event-analysis https://services.d4science.org/web/particleformation/			
Persistent identifiers	http://doi.org/10.5281/zenodo.1438314			
Open access	Yes	**Licence**	MIT	
Access limitations:				
Vocabularies, standards	Linking Open Descriptions of Events (LODE) ontology; OWL Time; GeoNames			
Quality control				
Long term availability	Published on Zenodo			

Work Package:	11	**Version:**	1.0	
Title:	ENVRIplus_Stromboli_Data_September_2017_June_2018			
Authors:	Ben Esse			
Description:	Data sets from fieldwork on Stromboli for the UPDASH ENVRIplus project, looking at detecting volcanic ash using depolarisation of sunlight			
Format	Plain text and images			
Metadata	Basic data included as file headers for spectra and in image file name for imagery			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	>=1Gb but <1Tb			
Stored in public repository	Yes			
Location of storage:	www.zenodo.org			
Persistent identifiers	http://doi.org/10.5281/zenodo.3250968			
Open access	Yes	**Licence**	Cc4.0by	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	Uploaded to online repository and backed up locally			

Work Package:	11	**Version:**	1.0	
Title:	Data resulting from the ENVRIplus TNA campaign RELECT at SMEAR II HYYTIÄLÄ multi-disciplinary RI platform			
Authors:	Susana Barbosa			
Description:	Atmospheric electric field, ambient gamma radiation and soil gas radon concentration monitored at the SMEAR II - Hyytiälä RI from June to November 2017. Data collected to improve understanding on the inter-relations between atmospheric electricity and natural radioactivity.			
Format	plain text datafiles			
Metadata	Electric field, gamma radiation and soil gas radon concentration data from ENVRIplus TNA campaign RELECT at SMEAR II HYYTIÄLÄ multi-disciplinary RI platform			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	https://rdm.inesctec.pt/dataset/cs-2017-010 https://rdm.inesctec.pt/dataset/cs-2017-009 https://rdm.inesctec.pt/dataset/cs-2017-011			
Persistent identifiers	https://doi.org/10.25747/JXKE-2J39 https://doi.org/10.25747/SSCT-JD93 https://doi.org/10.25747/JBVT-0940			
Open access	Yes	**Licence**	Creative Commons Attribution Share-Alike	
Access limitations:	No limitations to access			
Vocabularies, standards				
Quality control	Check for missing values (and NA flag inserted for missing times), inspection of the time series for obvious outliers			
Long term availability	Data stored in institutional repository			

Work Package:	11	**Version:**	1	
Title:	Branch fluxes and mole fractions of COS and CO2, LRU, stomatal conductance, internal conductance, eddy-covariance fluxes of COS and CO2, GPP, and meteorological information			
Authors:	Linda M. J. Kooijmans, Wu Sun, Juho Aalto, Kukka-Maaria Erkkilä, Kadmiel Maseyke, Ulrike Seibtb Timo Vesala, Ivan Mammarella, and Huilin Chen			
Description:	The dataset was collected to characterize the use of COS as a tracer for GPP.			
Format	plain text, see https://zenodo.org/record/1211481#.XB4Lb9IzbIU			
Metadata	see https://zenodo.org/record/1211481#.XB4Lb9IzbIU			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):			n/a	
Stored in public repository			Yes	
Location of storage:			Zenodo	
Persistent identifiers	10.5281/zenodo.1211481			
Open access	Yes	**Licence**	CC4.0	
Access limitations:				
Vocabularies, standards				
Quality control	Internal			
Long term availability	Zenodo storage			

Work Package:	11	**Version:**		
Title:	Ash fragmentation at Mount Etna and implications of different particle shape on ash dispersal in the atmosphere, ETNASH			
Authors:	M Polacci, D Andronico, A Cristaldi, M de Michieli Vitturi			
Description:	Ash is a common manifestation of explosive basaltic volcanism. At Mount Etna ash emissions accompany different eruptive styles, from mild to moderate Strombolian explosions to high energy fire-fountain activity. Based on visual observations of the eruptive activity and textural and compositional features of ash samples, it was found that the characteristics of coarse ash particles at Mount Etna vary systematically with the eruptive style. For example, ash emitted at the peak of high energy activity is more vesicular, less crystallised and less compositionally evolved than that erupted during lower energy explosive activity or at the end of a long-lasting explosive eruption, and it contains fewer or no lithic material (Taddeucci et al. 2004JVGR, 137, Andronico et al. 2008JVGR, 173). Abundant ash venting has characterised activity at Etna since 1998 (Andronico et al. 2013JGR, 118, 2014BV, 76), deeply affecting peoples everyday life and the overall economy in Eastern Sicily. Previous research on ash characteristics and the link between ash and eruption behaviour has improved our general knowledge on the dynamics of ash emissions at Mount Etna. However, no systematic description on mechanisms of ash fragmentation at Mount Etna has ever been provided. In addition, different ash characteristics translate into different transport patterns when the particles are released in the atmosphere, with implications for the environment and ground/air transportation that go far beyond the local economy and climate. The present proposal addresses this topic head-on.			
Format	PDF			
Metadata	Methods, experimental setup, preliminary results			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	No, but will be			
Location of storage:	The data will be published in a paper in Frontiers in Earth Science. At present the paper is under revision.			
Persistent identifiers	There will be a doi when the paper is published			
Open access	No	**Licence**	They will be when the paper is published	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	The data will be stored electronically in the journal Frontiers in Earth Science			

Work Package:	11	**Version:**	1	
Title:	IBAIRN time series			
Authors:	John Crowley			
Description:	Time series of trace gas measurements made by the Max-Planck-Institut for Chemistry (Mainz, Germany) during the IBAIRN campaign (Sept. 2016).			
Format	plain text			
Metadata	Chemical identifiers, concentration units, responsible PI			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):			n/a	
Stored in public repository			Yes	
Location of storage:			Zenodo	
Persistent identifiers	10.5281/zenodo.3254828			
Open access	Yes	**Licence**	CC4.0	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	Zenodo			

Work Package:	11	**Version:**	1.0	
Title:	Impact of land-use changes on soil health and greenhouse gases emissions			
Authors:	Nataliya Bilyera, Evgenia Blagodatskaya, Cornelia Rumpel, Yakov Kuzyakov, Abad Chabbi			
Description:	Data were collected for study of land-use changes on soil parameters			
Format	PDF			
Metadata	Descriptions of the methods used for soil analysis is included			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	http://doi.org/10.5281/zenodo.3255319			
Persistent identifiers	http://doi.org/10.5281/zenodo.3255319			
Open access	Yes	**Licence**	Cc4.0	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	It is located in zenodo repository			

Work Package:	11	**Version:**	1.0	
Title:	MACRORE_rawDataset			
Authors:	Irene Olivé, Emilio García-Robledo, João Silva, Pascale Cuet, Patrick Frouin			
Description:	Origin: Reunion Island Reef lagoon Purpose: Physico-chemical description of the natural variability in the lagoon system and estimation of productivity of seagrasses			
Format	csv			
Metadata	Usual codes and abbreviations are used for description of parameters (day, time, temperature, species, station, salinity, oxygen, pH, alkalinity, nitrate, nitrite, silicate, phosphate). Abbreviations or symbols from international system are used. Units are indicated in the file			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):			n/a	
Stored in public repository			Yes	
Location of storage:			Zenodo repository (www.zenodo.org)	
Persistent identifiers	DOI: 10.5281/zenodo.3256079			
Open access	No	**Licence**	Access for consulting is open but not use is allowed without consent of the authors. The reason is because this is still an incomplete dataset not totally curated	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	Database is incomplete. Once the database is completed it will be updated and an eventual DOI from a publication will be issued.			

Work Package:	13	**Version:**	1.0	
Title:	Questionnaire Ethics Dataset			
Authors:	Silvia Peppoloni, Giuseppe Di Capua, Florian Haslinger			
Description:	The dataset contains the answers to the questionnaire, that was addressed to participants in the ENVRI PLUS project (researchers and technicians, involved or not as Task, WP and Theme leaders). Its goal was to investigate to what extent each of them is aware of and takes into consideration ethical issues in relation to his/her scientific activities. Moreover, the answers allowed to understand how participating organizations in ENVRIplus are dealing with ethical and social aspects related to their institutional tasks. The dataset is an useful input to RI managers, to improve their institutional activity through the implementation of structures and offices dedicated to ethical and social issues, the improvement of the work environment, and other aspects that could increase the quality of the relationships with their stakeholders and society in general sense.			
Format	Microsoft Excel (.xlsx) file			
Metadata	Data are grouped in sheets in a functional way, useful for data elaboration			
Includes personal info	Yes (names of the persons)			
Personal information handling (if relevant):	Limited access			
Size (if imporant for sustainablity):	n/a			
Stored in public repository	No, explain below			
Location of storage:	At the present the dataset is stored on the pc of one of the authors,			
Persistent identifiers	At the present the dataset doesn't have any persistent identifier, but it could have it, if it was functional for similar or additional activities on the same subject.			
Open access	No	**Licence**	not open access	
Access limitations:	Personal data			
Vocabularies, standards	-			
Quality control	In order to control the quality and reliability of the answers, some crossed checks on some specific questions were applied, in particular related to participants belonging to the same RI.			
Long term availability	A long term access would be useful. But this would require to transfer the dataset on the ENVRIplus website and removal of personal information			

Work Package:	15	**Version:**		
Title:	Marine Science contents for multimedia game			
Authors:	Marina Locritani, Laura Beranzoli			
Description:	The documents were been produced to realize the "Marine Science" topic of the ENVRIgame.			
Format	PDF			
Metadata	key words: marine science, game 3b. Scientific_Game_ScienzeMarine.pdf: it describes the contents for ENVRIgame "Marine Science" topic Marine Science_INGV.pdf: it is an information document to support the ENVRIgame "Marine Science" topic			
Includes personal info				
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	https://zenodo.org/record/3296951#.XSXDtCMS9T4			
Persistent identifiers	https://zenodo.org/badge/DOI/10.5281/zenodo.3296951.svg			
Open access	Yes	**Licence**	CC4.0	
Access limitations:				
Vocabularies, standards				
Quality control	Internal			
Long term availability	zenodo			

Work Package:	15	**Version:**		
Title:	Dataset of questionnaires for teachers developed in the frame of ENVRIPLUS project			
Authors:	Giuliana D'Addezio, Marina Locritani			
Description:	The data was collected to improve the content of the ENVRIPLUS e-Learning Platform for multimedia education of secondary school level teachers and students.			
Format	.xlsx .docs			
Metadata	module_google.docx it is the used questionnaire Questionnaire_ENVRIplus_all.xlsx it is the results of collected questionnaires			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	https://zenodo.org/record/3296468#.XSXBlugza70			
Persistent identifiers	https://zenodo.org/badge/DOI/10.5281/zenodo.3296468.svg			
Open access	Yes	**Licence**		
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	Data is downloadable from Zenodo			

Work Package:	15	**Version:**	1.0	
Title:	Practical Introduction to the ENVRI RM			
Authors:	Abraham Nieva de la Hidalga			
Description:	As part of WP15, Task 15.1 we developed a series of learning materials which were published on the ENVRI learning platform and used as an introduction to the use of the ENVRI RM			
Format	SCORM 1.2			
Metadata	A set of 9 lessons introducing the ENVRI RM and its use with a use case scenario.			
Includes personal info	no			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	2			
Stored in public repository	Yes			
Location of storage:	Zenodo			
Persistent identifiers	https://doi.org/10.5281/zenodo.3269798			
Open access	Yes	**Licence**	Cc4.0	
Access limitations:	Packages can be installed in learning management environments such as moodle. They can also be imported into Xerte for editing/updating			
Vocabularies, standards	ENVRI RM Terminology UML4ODP RM-ODP			
Quality control	Review by peers Evaluation by user groups Feedback on use			
Long term availability	According to zenodo policy, items are retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.			

Work Package:	15	**Version:**	1.0	
Title:	Evaluation questionnaire scores from two Management Training workshops			
Authors:	Jacco Konijn, Enrico Guarini			
Description:	The dataset (in Excel) represents the outcomes of the evaluation forms, collected after the WP 15.3 RI management training workshops in Zandvoort (May 2018) and Milan (February 2019).			
Format	XLS			
Metadata	No relevant			
Includes personal info	No			
Personal information handling (if relevant):				
Size (if imporant for sustainablity):	n/a			
Stored in public repository	Yes			
Location of storage:	www.zenodo.org			
Persistent identifiers	10.5281/zenodo.3327902 https://doi.org/10.5281/zenodo.3327902			
Open access	Yes	**Licence**	CC Attribution 4.0 International	
Access limitations:				
Vocabularies, standards				
Quality control				
Long term availability	The relevance for future use is extremely limited, if not zero. Stored in Zenodo for reporting reasons.			

Work Package:	17	**Version:**		
Title:	Documents related to BEERi (Board of European Environmental Research Infrastructures) activity.			
Authors:	Ulpu Leijala, BEERi members.			
Description:	Meeting minutes, presentations, background material, consultations, etc.			
Format	pdf, word.			
Metadata	Own templates.			
Includes personal info	Yes (names)			
Personal information handling (if relevant):	Confidential data			
Size (if imporant for sustainablity):				
Stored in public repository	No, explain below			
Location of storage:	Material is confidential, access is only given to the BEERi members.			
Persistent identifiers	Documents are not meant to be linked to other purposes.			
Open access	No	**Licence**	Confidentiality issues.	
Access limitations:	Only for BEERi members.			
Vocabularies, standards				
Quality control	BEERi chair has reviewed the notes.			
Long term availability	Long-term location for the BEERi material is under consideration.			

CONCLUSIONS

IMPACT ON PROJECT

This is the initial version of the Data Management Plan. It is clear that this document needs to be further developed, detailed and corrected during project period. However, it presents the overall DM principles in the ENVRIplus and most likely data types collected.

IMPACT ON STAKEHOLDERS

Many of the Data Management Plan actions are crucial for RIs, as this document will clearly present that the produced documents, software and key datasets are available to all personnel.

A document of ENVRI^{plus} project - www.envri.eu/envriplus

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654182

APPENDICES

ANNEX I DESCRIPTION OF PROJECT INTERNAL PERSONNEL INFORMATION STORAGE

DESCRIPTION OF THE FILE
Personal Data Act (523/1999) section 10

D ate of drafting: 29.10.2015

Use an enclosure if necessary.

1. Controller	Name Ari Asmi
	Contact information (address, tel. etc...) P.O.Box 48, 00014 UHEL, Finland, +358407709729, ari.asmi@helsinki.fi
2. The person in charge and/or contact person	Name Magdalena Brus
	Contact information (address, tel. etc...) P.O.Box 48, 00014 UHEL, Finland, +358504154762, magdalena.brus@helsinki.fi
3. Name of the register (should describe the content)	Contact information of ENVRIplus project participants
4. The purpose for processing the personal data / the purpose for the use of a register (If the processing of personal data is outsourced [external service providers are used for the processing], a mention about it can be included to this point)	Maintaining communication and organisational information for the ENVRIplus H2020 project. This is meant as a register of the project participants for the internal use of the project. The Grant Agreement of the project (between the University of Helsinki and European Commission) specifically mentions "In particular, the Coordinator shall be responsible for: .... keeping the address list of Members and other contact persons updated and available"

[Unofficial translation]

5. Content of the register (For instance: name, address and telephone number of the data subject)	Name, professional title, address, telephone number, email address, institution they work in, (potentially) associated Research Infrastructure, involvement on specific tasks in the project.
6. Regular sources of information * (Which data is received, by whom and on what ground. For instance: consent or provision of a law)	The information is collected from the project participants, specifically from the Primary Investigators of project Beneficiaries.
7. Regular destinations of disclosed data and whether the data is transferred to countries outside the the European Union or the European Economic Area	Data is available to the Project Beneficiaries, and people working directly (project participants) within the project. Data is also available for the participants outside of the European Union (Switzerland, Norway), but the participating organizations have declared that they will follow the ethical regulations of the Horizon 2020. Similarly, the data can be shared to the European Commission services, if needed in their reporting or auditing purposes.
8. The principles how the data file/register is secured.	A. Manual register (place of storage and the methods of protection) Copy of the register is stored in encrypted drive of the personal work computer of M. Brus. B. Data register/ADP register (principles for the right to use a register, monitoring of the use and actual protection of hardware) Data is stored in the Activecollab project collaboration site, which is password protected. The project partners have a right to use the register. The site access is controlled by the Controller (Ari Asmi) and the Contact Person (Magdalena Brus). They also monitor the site access.

* Section 10 of the Personal Data Act does not require to make a note of this information, but it's recommendable to do so, regarding the right of access of the data subjects. Furthermore, it also describes the content of the register.

ANNEX II

Removed in this version, as ANNEX IV is a newer version

ANNEX III

QUESTIONNAIRE ETHICAL REVIEW TEMPLATE

This is the Ethical review document prepared by Malcolm Atkinson and Cristina Alexandru (U. Edinburgh) to their ethical board. This can be used as basis for similar documents in other parts of the project.

ANNEX IV

Guidelines for ENVRIplus Questionnaires

Before making any questionnaires, consider the following requirements:

1. Collection of any private data (including e.g. opinions, names, positions, etc) **is to be avoided** if not useful for the purpose of the study. Make sure that all information you collect are **strictly required** for the actions described in the Description of Action document of the Grant Agreement with the European Commission.
2. If such information is needed, the target of the study must be informed before the data collection about the
    a. Collected information,
    b. Why it is needed,
    c. How the targets were selected,
    d. Who has access to the data,
    e. Of any anonymization scheme (if any) is used,
    f. How the data will be analyzed,
    g. How the target and either agree on disagree on the terms,
    h. How the access to the data will be controlled and,
    i. How long and where the data will be stored.

    This is best done by including a header information on your questionnaire. COPY/PASTE the text under the line on the next page your questionnaire and fill in the information in the BOLDED parts.

3. The data retention must be carefully considered, and if the raw data sets (answers) do not need to be stored, they must be destroyed after use. We require that all personal information is destroyed latest on the project end, preferably immediately after the conclusions of the study are finalized.
4. You must indicate a responsible person for the questionnaire who is responsible that the material is adequately stored and handled. He or she is also responsible on the access control, data storage and on destruction of personal information.
5. Data storage must be adequate to the level of sensitivity of the data. Use of separate protected areas in the common ActiveCollab virtual platform can be arranged - please contact the project office if needed. Also protected services such as EUDAT B2DROP can be used in this context, if the access is carefully considered.
6. Access control of the data must be adequate and clearly defined, including access policies in the long-term storage (if needed).
7. You MUST fill any registration information legally required by your country of operations regarding the storage of personal information. This is YOUR responsibility.

IMPORTANT: Including the header information is only valid if you are NOT collecting any sensitive information. For sensitive information, you MUST also include the consent form agreement.

YOU MUST INFORM THE ENVRIPLUS PROJECT OFFICE (envriplus-coordination@helsinki.fi) ON ANY QUESTIONNAIRES YOU ARE CONDUCTING AND PROVIDE INFORMATION ON HOW THIS TEMPLATE IS USED.

INFORMATION ON THIS QUESTIONNAIRE

This questionnaire is meant to collect your professional knowledge related to the ENVRIplus research infrastructures which you are connected to. As we can also collect your name, position and professional position and potentially other personal information, it is important that you understand the reason and procedure of this questionnaire.

Information on ENVRIplus project: (http://www.envriplus.eu)

ENVRIPLUS is a cluster project, funded by the European Commission Horizon 2020 programme, on collaboration of research infrastructures (RIs) in Environmental and Earth System sciences, built around ESFRI roadmap and associating leading e-infrastructures and Integrating Activities together with technical specialist partners. ENVRIPLUS is driven by 3 overarching goals: 1) favouring cross-fertilization between infrastructures, 2) implementing innovative concepts and devices across RIs, and 3) facilitating research and innovation in the field of environment to an increasing number of users outside the RIs. ENVRIPLUS organizes its activities along a main strategic plan where sharing multidisciplinary expertise will be most effective. It aims to improve Earth observation monitoring systems and strategies, including actions towards harmonization and innovation, to generate common solutions to many shared information technology and data related challenges, to harmonize policies for access and provide strategies for knowledge transfer amongst RIs. ENVRIPLUS develops guidelines to enhance transdisciplinary use of data and data products supported by applied use cases involving RIs from different domains. ENVRIPLUS coordinates actions to improve communication and cooperation, addressing Environmental RIs at all levels, from management to end-users, implementing RI staff exchange programs, generating material for RI personnel, and proposing common strategic developments and actions for enhancing services to users and evaluating the socioeconomic impacts. ENVRIPLUS is expected to facilitate structuration and improve quality of services offered both within single RIs and at pan RI level. It promotes efficient and multidisciplinary research offering new opportunities to users, new tools to RI managers and new communication strategies for environmental RI communities. The produced solutions, services and other project results are made available to all environmental RI initiatives, thus contributing to the development of a consistent European RI ecosystem

Responsible person for this questionnaire: [**Name**], [**email**], [**Institution**]

You can always ask for further information from the responsible person above, or from the ENVRIplus project office: envriplus-coordination@helsinki.fi

The questionnaire aims at providing [**SPECIFY INFORMATION TYPE, e.g. technical details on energy requirements of observations**] information on [**SPECIFY REASON**] regarding European public Research Infrastructures (RI). Answering the questionnaire is voluntary and you can stop answering at any moment. The questionnaire will be done using a [**SPECIFY, e.g. web-based form**], and will take approximately [**XXX**] minutes to answer. You have been selected to answer the questionnaire as your professional capacity as the representative of the RI you are working with.

[SELECT ONE:

No personal data is requested, although you have an opportunity to optionally leave your contact information for further information, if needed.

OR

Your personal information: name, contact information, organization, and position in your organization are stored for analysis purposes.**]**

All data will be stored securely on [**FILL IN STORAGE LOCATION, e.g. Secure Servers in University of Helsinki, Department of Physics**] and will only be used within the framework of the ENVRIplus project. Access to the answers is restricted to the responsible person and the data anyzers selected by him/her. The answers will be analyzed offline. The questionnaire technical results and conclusions deducted from the results can be published within the ENVRIplus project deliverables, reports and documentation, however no personal information will be published in any form. All questionnaire answers will be deleted latest at the end of the ENVRIplus project. If your contact information is stored with your answers, you can also request to be informed on the reports and documents generated from the information collected in this questionnaire.

CONSENT FORM

in addition to the information form above if sensitive information, e.g. age, sex, opinions, political or religious opinions, medical information or similar, is strictly needed for the questionnaire. Please contact the project office if you are unsure of this. This can be done as a virtual form. If some points are not filled YES, you must make sure that there is no possibility to continue.

CONSENT FORM

Title of Project: ENVRIplus

Questionnaire title: [**title**]

Responsible person for this questionnaire: [**Name**], [**email**], [**Institution**]

Date: [fill or auto-fill]

This questionnaire can include potentially sensitive information on you. Please confirm the following

1. I confirm that I have read and understand the information for the above project and have had the opportunity to ask questions.

[yes]

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without my legal rights being affected.

[yes]

3. I agree for the answers to be recorded, and for each recording to be kept until associated report is completed. No identification information will published in any form and the personal information will be destroyed immediately after the questionnaire results are analyzed.

[yes]

ANNEX V

Consideration on publishing the Deliverables in ZENODO service.

Some of the ENVRIplus partners asked if the project public deliverables could be published in Zenodo (or similar) service, giving them DOIs. The general agreement to publish (public) deliverables in ZENODO or a similar service was approved by the Commission officer. However, this document needs still to be discussed and approved in the ENVRIplus Executive Board. At this stage, this represents the issues to be discussed on this stage.

1) Who does the submission? The submission process is a little involved, and has also a responsibility issue to take the relevant contributors (see below) into account. Even references can be included (but are not absolutely necessary). If coordination does this, some might be missed, and is an extra requirement, which would in some cases better fit to the main author (see 2).

DMT recommendation: The submission is done primarily by the main author. However, if needed the Project Office can do this as well. A mention and the DOI of the submission must then be submitted to the Project Office and to the Project internal Activecollab site.

2) Who are the authors? Even though we do have author list, the Zenodo metadata also includes term “contributors”. We need a policy on difference between these.

DMT recommendation: The authors of each deliverable should be all people who have significantly contributed to the creation of the document. For minor contributions, the metadata field "contributors" can be used. Responsibility of naming these persons in with the main author of the deliverable, although the deliverable (internal) review is responsible to check these issues.

3) Even though the GA and CA strictly require us to make the deliverables openly available, it is not so clear will this also include publishing (as much as Zenodo is publishing, see below). This is not so much a legal issue than a common decision in the consortium. Of course if all involved are clearly asked their permission, this is fine. We should aim for consortium-level agreement, but…

DMT recommendation: The Commission officer already agreed on publication also in Zenodo, so we suggest that we will have a General Assembly agreement on our internal policies.

4).. IF Zenodo is publication some might not be willing to do this. This is because at least traditionally parts of Deliverables have been used as parts of peer reviewed journal publications. This has been quite common practice in some fields, where the first report in the Deliverable is used as the basis of a more mature report in a journal. If Zenodo is publication, this might lead to issues when the journal asks if it is published earlier in some form. Even though our deliverables ARE peer reviewed (in a fashion), the Web of Science unfortunately does not see it so (a fact which I do not like, but what can one do).

DMT recommendation: For each individual case, we need to make sure we have agreement of all partners. This requires still further discussion within the Consortium.

5) We need a formal decision on using Zenodo, and not some other service (e.g. Figshare). We should have all the deliverables in a same place. It could be good to have the DOA in the same place as well, so we can use it as a “related publication”.

DMT recommendation: Single service (e.g. Zenodo) should be chosen. Key part is that it should be free, as reliable as possible, and preferably located within the European Union. The decision on which service (e.g. Zenodo, Figshare, B2SHARE) should be well explained.

6) In DOA we specify CC4.0BY, so this must be used in publishing

DMT recommendation: General Assembly should be informed also on license when making the decision.

7) The ENVRIPLUS project must be properly included in the metadata (field Funding)

DMT recommendation: Always include the Project number and funder in the metadata.

8) Keywords should be at least somewhat similar in similar deliverables

DMT recommendation: Project office should prepare list of potential keywords, which can then be supplemented by partners. ActiveCollab site is a good tool for this.

9) Which version is included? Now we have already several versions of the D5.1… DOIs currently do not properly handle versioning and child/parent/sibling relations, thus the old version must then always be remembered to be include in the "related publication" fields.

DMT recommendation: At least the version sent to the Commission must be then published. Further developments MUST use the Related Publication Field.

10) Zenodo offers possibility to "reserve" a DOI. This way it could be used to form a workflow a) upload a document b) include metadata c) get a DOI d) put proper citation information INTO the deliverable (e.g. first page) including the DOI e) change the file on the bottom of the Zenodo metadata input form. A little involved system, but proper citation method should be somehow included.

DMT recommendation: This procedure should be made a part of the General Assembly decision.